Prevention Science (2021) 22:986–1000
https://doi.org/10.1007/s11121-021-01257-0

Can Policy Promote Adoption or Outcomes of Evidence-based Prevention Programming?: a Case Illustration of Positive Behavioral Interventions and Supports

Catherine P. Bradshaw[1,2] · Elise T. Pas[2] · Rashelle J. Musci[2] · Joseph M. Kush[1] · Ji Hoon Ryoo[3]

Accepted: 7 May 2021 / Published online: 8 June 2021

Abstract
This study examined the impact of a state policy requiring that any school with a habitual truancy rate of 8% or higher to be trained in Tier 1 school-wide Positive Behavioral Interventions and Supports (SW-PBIS). A regression discontinuity (RD) design was used to examine how the schools' mandate status related to SW-PBIS training as well as student suspensions, truancy, and achievement in 410 public middle and high schools, of which 261 were affected by the mandate. We further examined the growth trajectories (i.e., improvement) of implementation fidelity over time using growth mixture modeling (GMM). Contrary to the intent of the policy to improve student outcomes, the RD results suggested that the mandate did not significantly impact reading and math achievement, truancy rates, or SW-PBIS training in 2010–2011 through 2013–2014. Mandated schools had higher suspension rates in 2010–2011 through 2013–2014 than the non-mandated schools; however, these differences in the suspension rates appear to have persisted from years prior to the mandate. Descriptive analyses suggested that mandated schools had statistically significantly higher rates of training, and the GMM analyses on the fidelity data indicated that mandated schools were significantly more likely to be in an improving implementation growth trajectory over time. Taken together, results suggested that the policy showed some promise for improving SW-PBIS training and fidelity over time, but it had little to no impact on student outcomes.

Keywords Regression discontinuity design · State policy · SW-PBIS

Although decades of educational research highlight the importance of adopting and scaling up evidence-based prevention models to improve behavioral and academic outcomes, there has been relatively limited uptake of some of the most effective prevention programs and frameworks in schools across the USA (Fagan et al., 2019; Glasgow et al., 2012; Spoth et al., 2013). A recent trend is the use of policy to promote scale-up of evidence-based programs, either through incentive or in reaction to mandate (Fagan et al., 2019; Sheras & Bradshaw, 2016). Yet, there has been limited research on how policy impacts the implementation of prevention programs or the outcomes achieved. The current paper focused on this recent policy trend by considering the effects of a state educational policy that required schools with high rates of student truancy (i.e., chronic absenteeism) to implement an evidence-based prevention framework called School-Wide Positive Behavioral Interventions and Supports (SW-PBIS; Sugai & Horner, 2006). This multi-tiered framework aims to create refined systems and procedures for preventing and responding to student behavior in all classroom and non-classroom contexts.

Specifically, this study examined the impact of a state policy which mandated the implementation of Tier 1, SW-PBIS in schools with truancy rates of 8% or higher. Using a regression discontinuity design (Thistlewaite & Campbell, 1960; Trochim, 2001), we estimated the effect of the state policy on student academic and behavioral outcomes and SW-PBIS training status (i.e., whether or not a school had been trained in SW-PBIS) for schools above and below this 8% truancy threshold. We also examined the changes in implementation

✉ Catherine P. Bradshaw
cbradsha@jhsph.edu

1 School of Education and Human Development, University of Virginia, Charlottesville, USA

2 Bloomberg School of Public Health, Johns Hopkins University, Baltimore, USA

3 Department of Education, Yonsei University, Seoul, South Korea

fidelity of the SW-PBIS framework in schools and explored whether specific growth patterns of fidelity were more common among those schools affected by the policy, as compared to those that were not. The findings are considered with regard to the broader movement to leverage policy to promote the dissemination and scale-up of evidence-based prevention programming (Fagan et al., 2019), and the impacts of these policies on both implementation fidelity and the outcomes achieved.

School-wide Positive Behavioral Interventions and Supports

Positive Behavioral Interventions and Supports (PBIS) is a multi-tiered framework (Sugai & Horner, 2002, 2006) for setting-level implementation of tiered prevention programing that aims to systematically and consistently prevent student behavior problems and promote a positive school environment. The full three-tiered PBIS framework incorporates universal or school-wide supports (Tier 1 or SW-PBIS), which are accompanied by two tiers of more intensive interventions and supports for students at risk for and/or already displaying behavioral issues to complement the universal school-wide components (Sugai & Horner, 2006). SW-PBIS operates through the articulation of positive behavioral expectations, the creation of systems to support positive behavior, and training in data-based decision-making. It provides a positive, proactive, public health preventive orientation toward addressing the behavioral and academic concerns experienced by schools (Sugai & Horner, 2006). To date, most states implementing the PBIS framework have focused on Tier 1 (SW-PBIS) implementation (Kittelman et al., 2019). Tier 1 SW-PBIS is the focus of the current policy evaluation.

In terms of anticipated effects of SW-PBIS, the impacts are theorized to operate through state and district infrastructure, organizational change at the school level, and behavioral changes in the staff, all of which translate to improved student outcomes. The state and district infrastructure provides training and on-going technical assistance to ensure that schools can successfully implement SW-PBIS. From an organizational perspective, the enhanced communication, collaborative decision-making, and consistent management practices are theorized to improve the climate and organization of the school (Bradshaw et al., 2008a, b). In turn, these changes are theorized to lead to improved classroom management and teacher efficacy. Improved classroom context is expected to result in decreased student problem behaviors (Bradshaw et al., 2012a, b; Waasdorp et al., 2012) and improved academic achievement and engagement (Barrett et al., 2008; Horner et al., 2009; Madigan et al., 2016; Sugai & Horner, 2006). Specifically, reductions in discipline problems, disruptions, and office disciplinary referrals are expected to result in more time spent in the classroom and on academics, thereby translating into improved academic performance. For additional detail on the theory of change, see Sugai and Horner (2006) and Barrett et al. (2008).

Prior SW-PBIS Research

Prior effectiveness research provides evidence that many of these theorized effects do occur when SW-PBIS is implemented with high fidelity at both the elementary and secondary levels (e.g., Mercer et al., 2017; Madigan et al., 2016; Horner et al., 2009; also see Horner et al., 2010; Lee & Gage, 2020). For example, randomized studies conducted in elementary schools, testing the universal, SW-PBIS model, have shown it to reduce student office discipline referrals and suspensions and improve school climate and student achievement (see Bradshaw et al., 2008a, b; Bradshaw et al., 2009a, b; Bradshaw et al., 2010; Horner et al., 2009; Madigan et al., 2016). A randomized effectiveness study of Tier 1 SW-PBIS in elementary schools also showed that students were rated by their teachers as having fewer behavioral problems (e.g., aggressive behavior, concentration problems, bullying, rejection) than students in non-PBIS schools (Bradshaw et al., 2012; Waasdorp et al., 2012). In addition, significant effects have been observed on teacher ratings of school climate (Bradshaw et al., 2008a, b). Although the randomized trial research on SW-PBIS has largely been conducted in elementary schools, quasi-experiments have also been conducted in secondary schools. When taken to scale, quasi-experimental studies have demonstrated that SW-PBIS improves out-of-school suspensions for elementary and secondary schools (i.e., a main effect for all school types; Gage et al., 2019), with some promising effects also on academic performance both at the elementary and secondary levels (e.g., Lee & Gage, 2020; Madigan et al., 2016; Pas et al., 2019). Notably, the Pas et al. (2019) study examined outcomes separately for secondary schools and found improvements in suspensions, truancy, and math and reading achievement in the secondary schools implementing SW-PBIS. At the time of this policy, there was no published research evidence to support its effects on truancy (i.e., the focus of research was on exclusionary discipline and academic achievement); Pas and colleagues (2019), did however, report declines in truancy in secondary schools implementing SW-PBIS in 2007–2008.

Leveraging Policy to Promote Scale-up of SW-PBIS

Given the promise of SW-PBIS and the relatively limited cost associated with the model (Lindstrom Johnson et al., 2020), it is perhaps not surprising that it is currently one of the most widely disseminated evidence-based programs in schools; it is estimated by the Office of Special Education Programs (i.e., OSEP, 2019) funded National PBIS Technical Assistance Center (www.pbis.org) that nearly 26,000 schools across the USA have been trained in Tier 1 SW-PBIS. Increasing concerns about school safety also have likely contributed to its broad dissemination over the past two decades. For example, in Maryland, following increasing national concerns about school safety resulting from the wave of school shootings in the 1990s, the Maryland State Department of Education began a state-wide scale-up of SW-PBIS. In 1999, Positive Behavioral Interventions and Supports (PBIS) Maryland was created with the aim of providing training, support, and evaluation of the PBIS model and related prevention programming throughout the state in order to improve conditions for learning in Maryland schools (Bradshaw et al., 2012a, b), with a specific focus on Tier 1 (SW-PBIS). Consistent with the National PBIS Technical Assistance Center's model for training and implementation of SW-PBIS (www.pbis.org), schools in Maryland volunteered to adopt SW-PBIS and attended a state-wide 2-day training event that covered the foundational elements of the implementation of the SW-PBIS framework. Schools implementing the model in Maryland were expected to meet a number of conditions, including assessing for buy-in from 80% of school staff, the formation of a five- to six-person school-level PBIS team (including an administrator and a team leader), providing a 3-year implementation commitment, and the identification of an internal behavior support coach (e.g., school psychologist) to give on-going support regarding implementation. Each year, school-level PBIS teams were also expected to attend 1- or 2-day summer booster trainings, provided by districts or regionally. These teams were expected to train the other staff and students within their school. Implementation status has been tracked by the PBIS Maryland Partnership, through the submission of bi-annual implementation data on the Implementation Phases Inventory (Bradshaw, 2009a, b). Further, schools submitted data on the School-wide Evaluation Tool (Horner et al., 2004) for recognition of exemplary implementation.

Following more than a decade of dissemination of the framework in the state, the Maryland State legislature passed *State Code §7–304.1 PBIS Program* indicating that any school with a habitually truant rate (i.e., students with 20 or more unexcused absences) at or exceeding 8% in the 2008–2009 school year was mandated to be trained in and implement SW-PBIS in the summer of 2010, or to implement "an alternative behavior modification program developed in collaboration with the Department" (PBIS Program, 2008). The Maryland State Department of Education (MSDE) reviewed Maryland public schools' truancy data to identify schools exceeding the threshold and contacted the district superintendents with school names. Schools identified as meeting or exceeding the 8% truancy rate and *not yet trained* in SW-PBIS were expected to receive training for the upcoming school year (see Supplemental File 1). Due to Maryland's extensive infrastructure for SW-PBIS training and the advantages of PBIS (e.g., flexibility), the assumption was that mandated schools would not seek to implement an alternative program. To our knowledge, Maryland is the only state to pass a mandate based on truancy; however, other states (e.g., Florida, Illinois), as well as several federal legislators, have proposed bills related to PBIS that have not passed (e.g., H.R.3165, the *Positive Behavior for Safe and Effective Schools Act* in 2011). Similarly, there have been some state bullying-specific policies (e.g., New Jersey) that have required implementation of a bullying prevention program (for a study of the impact of the bullying policies, see Hatzenbuehler et al., 2015). In summary, this particular truancy-focused policy is unique and thus provides a potentially instructive opportunity to examine the impacts of the policy on both outcomes and implementation of this prevention framework.

Overview of the Current Study

The primary purpose of the current study was to examine whether schools affected by the mandate (as indicated by meeting or exceeding the truancy threshold of 8% in the 2008–2009 school year, and thus appearing on a list sent to their superintendent indicating that they were mandated to implement SW-PBIS in 2010; see Supplemental File 1) showed improvements in student truancy, suspensions, and academic achievement over the course of the subsequent 4 years (i.e., the 2010–2011 through 2013–2014 school years). We focused on these specific student outcomes because they were the main outcomes of interest to policymakers, and are largely consistent with the theory of change process associated with SW-PBIS described above and elsewhere (see Sugai & Horner, 2006). Importantly, we did not expect such changes to occur immediately, as the effects of policies often require a few years to reach full implementation, much less affect the intended outcomes (Sheras & Bradshaw, 2016). Secondarily, we examined actual uptake of SW-PBIS, which was operationalized by receipt of training in SW-PBIS, as well as fidelity of implementation. We used school data from 410 secondary schools and focused on middle and high schools, as they were most likely to be affected by the mandate given its focus on truancy and the general tendency for secondary schools to experience higher rates of truancy relative to elementary schools (USDOE, 2019). Using these data, we aimed to address the following primary research question: *Are mandated schools (i.e., those that reach or exceed the 8% threshold) experiencing the intended improvements in truancy, suspensions, and academic achievement?* Given this clear assignment threshold on the truancy variable (i.e., 8%), we utilized a regression discontinuity

(RD) design (Thistlewaite & Campbell, 1960; Trochim, 2001) to analyze the effect of the mandate on student outcomes. In light of prior research on SW-PBIS and the underlying theory of change process (e.g., Bradshaw et al., 2008a, b; Bradshaw, 2009), we hypothesized that schools affected by the mandate would achieve improvements in both behavioral and academic outcomes.

We further leveraged state-collected data on SW-PBIS training status (i.e., whether or not a school was trained) to examine our second research question: *To what extent did the mandate improve SW-PBIS training rates?* Using an additional regression discontinuity (RD) analysis, we assessed whether the mandate led to increased training rates among mandated schools in SW-PBIS. Since both mandated and non-mandated schools may have been trained in SW-PBIS prior to or following the mandate, we also conducted descriptive analyses of training status, year-by-year, over multiple years to determine whether and when mandated status was associated higher training rates. This approach allowed for a more nuanced exploration of issues related to training and took into consideration issues related to timing, as the effect of the mandate to access training may not have been immediate. Although not directly measured in this study, it is possible that readiness and buy-in may have been different for schools mandated to implement SW-PBIS compared to those that may have volunteered; this in turn could affect training status, as well as implementation fidelity, among the mandated schools relative to non-mandated schools (Bradshaw & Pas, 2011; Sheras & Bradshaw, 2016).

Again using state-collected data on implementation fidelity, we examined our third research question: *Was the mandate associated with changes in implementation fidelity of SW-PBIS over time?* Given that the mandated schools already trained in SW-PBIS were expected to *improve* implementation of SW-PBIS, we hypothesized that there would be an increase or growth in fidelity of SW-PBIS implementation across the subsequent 4 years. Therefore, we conducted growth mixture modeling to test this third research question. In summary, this particular policy is of interest not only with regard to impacts on student behavioral outcomes, like suspension, truancy, and academics, but also with regard to training in and implementation of SW-PBIS. As such, the findings from this study may inform other policy efforts related to the scale-up of evidence-based programs in schools.

Method

Sample

Within the state of Maryland, there are 24 districts or local education agencies (i.e., 23 counties and 1 city), all of which participate in the Maryland SW-PBIS Initiative. The focus of this study is on secondary schools including traditional middle schools serving grades 6–8, traditional high schools serving grades 9–12, and combined middle and high schools serving grades 6–12. Sample data consisted of all 410 secondary schools in the state including 212 middle (51.7%), 179 high (43.7%), and 19 combined (4.6%) schools. In the spring of 2010, the first year that mandates were issued, 261 schools (63.7%) were identified as being above the truancy threshold, based on their 2008–2009 truancy data at or exceeding 8% (see Supplemental File 1). Of those mandated, 104 were middle schools (39.8% of all schools), 142 were high schools (54.4%), and 15 were combined schools (5.7%); 127 schools (48.7% of all mandated schools) had been trained in SW-PBIS in or before the summer 2009. Therefore, the remaining 134 schools were on the mandated list sent to the district superintendents and were expected to receive SW-PBIS training during the summer of 2010 and implement in 2010–2011.

Measures

School-Level Outcomes The school outcome data were provided by the Maryland State Department of Education for the 2008–2009 through 2013–2014 school years. These included the suspension rates (i.e., total suspension events divided by total school enrollment times 100), truancy rates (i.e., percent of students in the school missing 20 or more days of school across a given school year), and the percent of students within each school that were proficient on the Maryland School Assessment (MSA) grades 6–8 tests of reading and math and proficient in English 2 (i.e., typically taken in 10th grade) and Algebra on the High School Assessment (HSA) for each year. School proficiency rates for high schools were calculated by summing the percent of students who scored in the proficient and advanced ranges for the English 2 and Algebra exams in the given year. For middle schools, the percent proficient and advanced in each assessed grade (i.e., 6–8) was averaged across grades to calculate the school average proficiency rate. The suspension rate reflects the number of suspension events divided by the school enrollment. The baseline data were included for the 2008–2009 school year (see Table 1). We examined outcomes in school years 2010–2011 through 2013–2014.

School-Level Demographic Characteristics We leveraged baseline data for school-level demographics provided by MSDE from 2008 to 2009 (i.e., first year of study), including student mobility (i.e., number of students who entered the school, plus the number who withdrew from the school, divided by total student enrollment), the percent of students who were White, school enrollment, and percent of students receiving free and reduced-priced meals; these variables served as controls in the outcome analyses, based on prior research (Bradshaw & Pas, 2011; Pas & Bradshaw, 2012; Stuart et al., 2015).

Implementation Status and Fidelity Training data have been tracked annually by the PBIS Maryland partnership, whereby 1 = trained and 0 = not trained (see Bradshaw et al., 2012a, b). SW-PBIS implementation status following the offering of SW-PBIS training in the summer of 2010, the expected training year, through 2014 was of interest. Training rates in 2009 were also included in the models to provide a pre-mandate context for training rates in the state. SW-PBIS trained schools in Maryland have also submitted fidelity data to the PBIS Maryland Partnership annually via the Implementation Phases Inventory (IPI; Bradshaw et al., 2009a, b); therefore, IPI data are only available for SW-PBIS trained schools, but not for untrained schools. The IPI assesses the presence of 44 key elements of SW-PBIS following a "stages of change" theoretical model. The four stages assessed are: *preparation* (Cronbach's alpha from Bradshaw et al. (2009) [α] = 0.65, e.g., "PBIS team has been established"), *initiation* (α = 0.80, e.g., "A strategy for collecting discipline data has been developed"), *implementation* (α = 0.90, e.g., "Discipline data are summarized and reported to staff"), and *maintenance* (α = 0.91, e.g., "A set of materials has been developed to sustain PBIS"). The IPI was completed by the PBIS coach who indicated, for each item, whether the extent to feature is *not in place* (0), *partially in place* (1), or *fully in place* (2). Schools received a percentage of implemented elements for each stage; the stage scores were averaged to compute a total score, which was utilized here. Prior research on the psychometric properties of the IPI found it to have adequate internal consistency (α = 0.94) and reliability (test–retest correlation of 0.80; Bradshaw et al., 2009a, b); the IPI has also been shown to have predictive validity with regard to suspension rates (Pas & Bradshaw, 2012).

Data Analysis

Research Question 1 Are Mandated Schools (i.e., those that reach or exceed the 8% threshold) Experiencing the Intended Improvements in Truancy, Suspensions, and Academic Achievement?

We utilized a regression discontinuity (RD) design (Ryoo & Pullen, 2017; Thistlewaite & Campbell, 1960; Trochim, 2001) to examine student outcomes using the RDDtools (Stigler & Quast, 2015), RDDrobust (Calonico et al., 2016), and RDD (Dimmery, 2016) packages in the R software (R Core Team, 2018). In conducting the RD analyses, we were able to determine whether the truancy cutpoint issued during the spring of the 2009–2010 school year had a significant impact on each of four student outcomes: (1) truancy, (2) suspensions, (3) math achievement, and (4) reading achievement in 2010–2011 through

Table 1 Baseline sample descriptive data

	All		Mandated		Non-mandated		*t* test	
	M	*SD*	*M*	*SD*	*M*	*SD*	*d*	*p*
Enrollment	1005.79	518.86	1060.42	553.93	910.09	436.40	0.29	.003
% Receiving special education	10.69	5.06	11.53	5.70	9.22	3.21	0.47	< .001
% Receiving free and reduced meals	32.43	20.74	38.66	19.31	21.52	18.59	.090	< .001
% Receiving limited English proficiency	2.89	3.51	2.98	3.62	2.72	3.30	0.08	.446
% Mobility	18.35	14.52	22.54	15.11	11.01	9.81	0.86	< .001
Student- teacher ratio	19.29	3.32	19.58	3.42	18.79	3.09	0.08	.441
% Asian	5.34	6.41	3.72	4.19	8.19	8.36	− 0.74	< .001
% Hispanic	7.80	10.96	8.40	11.61	6.76	9.65	0.15	.125
% African American	35.27	30.59	43.15	31.53	21.44	23.19	0.01	.936
% White	49.98	32.72	43.29	33.02	61.69	28.73	− 0.58	< .001
Suspension rate	20.01	17.40	26.00	18.58	9.53	7.50	1.01	< .001
Math proficiency	79.28	16.61	75.41	17.61	86.05	12.03	− 0.67	< .001
Reading proficiency	84.81	12.04	81.27	12.96	91.03	6.68	− 0.88	< .001
Truancy Rate	13.42	11.58	18.10	12.19	5.23	1.80	1.31	< .001
Average IPI overall scores	83.66	15.08	81.53	15.72	87.00	13.46	− 0.09	.022
Average years of training after summer 2009	3.22	2.28	3.40	2.17	2.89	2.43	0.23	.035

All data from the 2008–2009 school year

IPI Implementation Phases Inventory, *d* = Cohen's *d*

2013–2014. Within the model, c represents the cutpoint of assignment variable X. Treatment receipt is denoted by a binary variable D, such that $D = 1$ if $X \geq c$, while $D = 0$ if $X < c$. The outcome variable Y is modeled as a linear function:

$$Y_i = a_1 + a_2 D_i + b_1 \cdot \left(X_i - c\right) + b_2 D_i \cdot \left(X_i - c\right) + \epsilon_i, \quad (1)$$

where a_1 represents the estimated outcome for those just below the cutpoint, a_2 represents the estimated treatment effect for those just equal or above the cutpoint, b_1 represents the estimated slope for those just below the cutpoint , b_2 represents the estimated difference in slopes for those just above the cutpoint (as compared to those just below the cutpoint), and ϵ_i represents a residual error term (Stigler & Quast, 2015). Within the model presentation, a_2 represents the estimated treatment effect, or the difference in outcomes between those just above and just below the 8% cutpoint.

Valid interpretations of RD analyses rely on three main assumptions: (1) continuity assumption, (2) exogeneity assumption, and (3) functional form assumption (Jacob et al., 2012; Lee & Lemieux, 2010; Ryoo & Pullen, 2017). The continuity assumption requires the average potential outcome to be a continuous function of the cutpoint, conditional on treatment status. This assumption was explored by visually examining a scatterplot of the outcome variable and the assignment variable. Further, under this assumption, any discontinuity in the treatment effect is expected to occur at the specified cutpoint, and not at other cutpoints. This was explored by estimating treatment effects at various other cutpoints (i.e., "placebo tests"; Thoemmes et al., 2017). The exogeneity assumption requires that units to the left of the cutpoint are equivalent in expectation to units on the right. This assumption was formally tested using McCrary's (2008) density test. To help bolster this assumption, the association between covariates and the assignment variable should be smooth around the cutpoint. That is, there should be no inherent discontinuities between those above and below the cutpoint on any variables beside the treatment indicator. Finally, the functional form assumption requires the correct functional form of the relationship between the assignment variable and the outcome. Testing this assumption consisted of fitting models with different forms of the assignment variable, including linear, quadratic, and interactions with the treatment. To determine the model of best fit, models were compared using the Akaike Information Criterion (AIC; Jacob et al., 2012). The AIC takes into consideration the trade-off between bias and variance in the model, such that the lowest AIC reflects better model fit.

While we present the RD results utilizing the entire sample of schools, we recognize that modern RD analysis often presents results from local analyses, wherein an optimal bandwidth around the cutpoint is selected, such that the sample is balanced on key characteristics across the cutpoint. Narrower bandwidths both reduce the statistical power, but also reduce the external validity of the present analyses. However, full sample analyses increase power and generalizability but decrease internal validity, to the extent to which the samples above and below the cutpoint are not balanced on key variables. Sensitivity analyses on the optimal bandwidth were performed for all student outcomes and are summarized in the "Results" section (for details see Imbens & Kalyanaraman, 2009). We calculated power for the RD using the "rddapp" package in R, basing it upon our sample of 410 schools and the summary statistics provided in Tables 1 and 2. The power for suspensions, truancy, and math achievement were 1.0, and for reading achievement, the power was 0.988. Thus, the RD design has sufficient power with 410 schools.

Research Question 2 To What Extent Did the Mandate Improve SW-PBIS Training Rates?

As with research question 1, we again utilized a RD design to examine student outcomes using the RDDtools, RDDrobust, and RDD (Dimmery, 2016) packages in the R software. In conducting the RD analyses, we were able to determine whether the truancy cutpoint issued during the spring of the 2009–2010 school year had a significant impact on training status across the subsequent 4 years. To further address the timing of training, we cross-tabulated the mandate status with SW-PBIS training and implementations status (i.e., whether schools were or were not mandated or trained and implementing) and conducted t tests to determine whether there was a statistically significant association between mandate status and SW-PBIS training rates in 2008 and through 2014.

Research Question 3 Was the Mandate Associated with Changes in Implementation Fidelity of SW-PBIS Over Time?

As in research question 2, we conducted t tests to determine whether there was a statistically significant association between mandate status and SW-PBIS fidelity (i.e., average IPI scores) in 2008 and through 2014. We further examined fidelity growth trajectories (i.e., changes in fidelity over time) based on the IPI data during the year in which the mandate was determined (i.e., 2008–2009), the year during which the mandate was issued (i.e., 2009–2010), and the 4 years following the mandate (i.e., 2010–2011 through 2013–2014). To do so, we fit a series of growth mixture models to determine (a) the patterns of change in fidelity scores over time, (b) whether the intercepts (i.e., 2009 IPI scores) and slopes (i.e., change over time) differed for schools that were affected by the mandate or were not affected by the mandate, and (c) whether mandated schools were more or less likely

Table 2 Parameter estimates of regression discontinuity models

Outcomes	Intercept	p	a_2	p	b_1	p	b_2	p
2010–2011 outcomes								
Math	**82.10**	**<.01**	−0.38	.87	**−1.45**	**.02**	0.88	.17
Reading	**88.05**	**<.01**	0.30	.84	**−1.10**	**.01**	0.39	.35
Suspensions	**11.53**	**<.01**	**9.45**	**<.01**	0.90	.16	−0.58	.36
Truancy	**8.05**	**<.01**	0.35	.65	0.51	**.01**	0.36	.08
PBIS training	0.32	.10	0.10	.65	0.05	.38	−0.05	.37
2011–2012 outcomes								
Math	**84.52**	**<.01**	−0.73	.75	−1.16	.06	0.50	.42
Reading	**88.00**	**<.01**	0.04	.98	−0.87	.06	0.12	.79
Suspensions	**10.33**	**<.01**	**7.03**	**<.01**	0.70	.20	−0.35	.53
Truancy	**8.31**	**<.01**	0.21	.81	**0.54**	**.02**	0.29	.21
PBIS training	**0.40**	**.04**	0.14	0.53	0.07	.25	−0.07	.23
2012–2013 outcomes								
Math	**81.70**	**<.01**	−1.15	.65	−1.35	**.05**	0.79	.25
Reading	**89.00**	**<.01**	−0.55	.74	−0.73	.10	0.01	.98
Suspensions	**8.95**	<.01	**6.94**	**<.01**	0.72	.13	−0.59	.22
Truancy	**8.77**	**<.01**	0.29	.76	**0.60**	**.02**	0.19	.45
PBIS training	**0.42**	**.03**	0.21	0.34	0.07	.25	−0.07	.22
2013–2014 outcomes								
Math	**75.30**	**<.01**	−0.83	.80	−1.56	.07	1.13	.20
Reading	**86.34**	**<.01**	−0.64	.75	−0.77	.14	0.10	.86
Suspensions	**7.74**	**<.01**	**6.22**	**<.01**	0.56	.22	−0.50	.28
Truancy	**8.30**	**<.01**	−0.46	.65	0.51	.06	0.30	.03
PBIS training	**0.46**	**.02**	0.19	.39	0.08	.19	−0.07	.21

All models controlled for 2008–2009 baseline data for the targeted outcome and school demographic variables; student outcomes models also controlled for 2010 PBIS training status. Bolded indicates a significant effect at $p < .05$

to demonstrate heterogeneity in specific growth patterns. The growth mixture models (GMMs) were fit in M*plus* Version 8 (Muthén & Muthén, 2002–2018) to model patterns of growth over time in SW-PBIS fidelity scores on the IPI. To explore different growth patterns and initial discrepancies, we fit the linear growth models for each homogeneous group. Model fitting was conducted iteratively, by adding one growth class at a time and assessing whether the addition of a class achieved better model fit as demonstrated through decreasing values of the following three fit indices: Akaike Information Criteria (AIC), Bayesian Information Criterion (BIC), and sample size adjusted BIC. Further, a statistically significant Vuong-LMR likelihood ratio test (Muthén & Muthén, 2002–2018) was utilized to indicate improved fit. We also considered entropy scores closest 1.00 and latent class probabilities greater than 0.70 (Ramaswamy et al., 1993) as additional fit indicators; meaningful class sizes, along with conceptual and theoretical relevance, were also considered in final model selection (Nylund et al., 2007).

After enumerating the number of classes, we examined whether mandated status was a significant predictor of the intercept and slope for the classes as well as a predictor of class membership. In the exploratory analysis, we began with the unconditional GMM and then added covariates in the GMM (Ryoo et al., 2018). Covariates of interest, added in the final model, were the same as in the RD analysis and included 2008–2009 data regarding school size (enrollment), the percent of students who were mobile during the 2008–2009 school year and received free and reduced-price meals, and the percent of White students.

Results

Research Question 1: Are Mandated Schools Experiencing the Intended Improvements in Truancy, Suspensions, and Academic Achievement?

Checking of RD Assumptions To investigate the continuity assumption, scatterplots of the outcome variable and the assignment variable were visually examined, in which there appeared to be a smooth, continuous relationship on both sides of the cutpoint. Further, 95% confidence intervals from placebo tests indicated there were no other cutpoints at which discontinuities existed. Results from McCrary's density test indicated there was no discontinuity of density of observations on either side of the cutpoint ($z = -1.85$, $p = 0.06$). Further, scatterplots of the assignment variable and multiple covariates as outcomes indicated no discontinuities or differences in densities between those to the left and right of the cutpoint. Finally, the functional form was examined, in which a linear, quadratic, and cubic assignment variable, as well as an assignment and treatment interaction variable, were modeled. Results from AIC model comparisons indicated no models fit the data better than the original linear model. As such, the linear model remained the model used throughout analyses.

Regression Discontinuity Models Models with and without baseline demographic covariates were fit to the data. The specific baseline variables were selected based on prior research demonstrating that these variables are associated with being trained in SW-PBIS (Bradshaw & Pas, 2011; Pas & Bradshaw, 2012; Pas et al., 2019; Ryoo et al., 2018; Stuart et al., 2015) and thus are potentially confounding variables of intervention status. While point estimates were similar between models with and without covariates, the baseline demographics were ultimately included to both improve precision and reduce sample bias (Lee, 2008). In examining the a_2 estimates and p values, the truancy cutpoint (i.e., of 8%) was significantly associated with the suspension rate in every year, such that mandated schools had significantly higher suspensions than those not mandated. The suspension rates appeared to decline over time across the full sample of schools (i.e., both mandated and non-mandated); however, the total suspension rate for mandated schools remained statistically significantly higher than for non-mandated schools across all years. Although the RD analysis achieved overall statistically significant balance on baseline suspension rates, particularly within the sensitivity analysis using the optimal bandwidth, it is worth noting that the gap between the two groups in the unadjusted average suspension rates in years prior to the mandate was rather sizable (i.e., 8–10%). These unadjusted differences appear to have shrunk with time, but within the RD analyses were statistically significantly higher among the mandated schools. On the other hand, the treatment effect estimates were non-significant for the other three outcomes in each year, indicating there were no significant differences in reading and math achievement as well as truancy rates between those just above and just below the truancy cutpoint.

The b_1 estimates represent the slopes for each outcome, and the p value indicates whether the slope was significantly different than 0. These findings range across years, whereby in 2010–2011, slopes for each outcome except suspensions were significantly different from zero. Thus, schools with higher rates of truancy in 2008–2009 had lower rates of reading and math achievement in 2010–2011 and higher rates of truancy in 2010–2011, as compared to those schools with lower rates of truancy in 2008–2009. For 2011–2012 outcomes, only the slope of truancy was significantly different than 0, indicating that schools with higher rates of truancy in 2008–2009 also had higher rates of truancy in 2011–2012 as compared to those schools with lower rates of truancy in 2008–2009. For the 2012–2013 outcomes, the slopes of math achievement and truancy were significantly different than 0, indicating that schools with higher rates of truancy in 2008–2009 had lower rates of math achievement in 2010–2011 and higher rates of truancy in 2012–2013. The truancy rates in 2008–2009 were not related to outcomes in 2013–2014, as indicated by the non-significant p values for the b_1 estimates. Finally, b_2 estimates and p values indicated that there were no significant differences in the slopes between schools at or above and below 8% truancy for any of these student outcomes, in any year. This means that the relation between truancy in 2008–2009 and the 2010–2011, 2011–2012, 2012–2013, and 2013–2014 outcomes did not differ as a function of whether schools had less than 8% habitually truant students as compared to those > 8% habitually truant. In other words, the analyses did not suggest that the association between early truancy and later outcomes varied between the schools above and below the mandate threshold.

Sensitivity Analyses Utilizing the optimal bandwidth for the 2010–2011 outcomes dropped schools on the tail ends of the data and retained schools that were the closest to the cutpoint and thus most similar on key variables. Models utilizing the optimal bandwidth retained, on average, 84% of the sample of schools ($n = 346$). These analyses demonstrated the same findings as the models conducted with the entire sample and presented above. Therefore, the findings were robust to changes in the sample (i.e., full versus a trimmed sample). While we believe that there is substantive support to allow the slopes to vary across the cutpoint (Lee & Lemieux, 2010), we conducted additional sensitivity analyses to further illustrate this point. Specifically, we compared the fit between two models, one constraining the slope to be the same across the cut point and one allowing the slope to vary. While there was a slight improvement in the AIC in the constrained model, it was not significant enough to warrant constraining the slope in the analytic model (AIC in constrained model = 500.09; AIC in unconstrained model = 500.53). Prior work has suggested that constraining the slope to be the same across the cutpoint would result in an artificial amplification of the magnitude of the treatment group slope and dampens the magnitude of the control group slope, implying a significant impact of the treatment in question when there is no true impact (Jacob et al., 2012). Therefore, we maintained a more conservative approach, allowing the slopes to vary across the cutpoint.

Research Question 2: To What Extent Did the Mandate Improve SW-PBIS Training Rates?

The same RD analyses were conducted for the binary SW-PBIS training status variable (i.e., 0 = not trained, 1 = trained; see Table 2). As described above, all assumptions were tested and met. The a_2 estimates were non-significant for training status in each year, indicating there were no significant differences in training status across years comparing mandated to non-mandated schools. The b_1 estimates, or slopes, also had non-significant p values, indicating that the slope was not significantly different than 0. Finally, b_2 estimates and p values indicated that there were no significant differences in the slopes between mandated and non-mandated for training status in any year. This means that the association between the mandate and subsequent training status did not vary as a function of whether schools were above or below the mandate threshold. We also conducted descriptive analyses to provide greater detail and context for the training rates throughout the state during this time frame. Of the 261 secondary schools that were mandated based on their 2008–2009 truancy data, 127 had been trained in or before summer 2009 (i.e., 48.7% of all trained), leaving 134 to be trained. Among the 149 non-mandated schools, 134 schools had been trained in or before 2009 (i.e., 51.3% of all trained), leaving only 15 to be trained. After the summer

Table 3 SW-PBIS training and implementation fidelity descriptive statistics

	Non-Mandated (n = 149)		Mandated (n = 261)		t test	
Trained by:	n	%	n	%	d	p
2009	81	54.36	127	48.66	−0.11	.268
2010	82	55.03	154	59.00	0.08	.435
2011	85	57.05	171	65.52	0.18	.093
2012	87	58.39	180	68.97*	0.22	.034
2013	88	59.06	187	71.65*	0.27	.011
2014	89	59.73	196	75.10*	0.34	.002
Overall IPI Score in:	M	SD	M	SD	d	p
2009	87.00	13.46	81.53	15.72	−0.37	.107
2010	89.06	9.25	81.94	17.12	−0.49	< .001
2011	91.65	8.46	79.51	19.91	−0.71	< .001
2012	91.04	9.92	79.00	19.41	−0.71	< .001
2013	89.84	12.45	84.11	16.37	−0.38	.027
2014	87.91	14.45	82.43	18.47	−0.32	.068
IPI > = 80% in:	n	%	n	%	d	p
2009	65	78.46	102	63.73	−0.32	.037
2010	74	82.43	122	66.39	−0.36	.010
2011	69	91.30	139	61.15	−0.70	< .001
2012	73	87.67	149	55.03	−0.72	< .001
2013	69	82.61	132	73.48	−0.22	.130
2014	74	79.73	149	68.46	−0.25	.065

In total, 167 out of 208 trained schools (80.3%) provided IPI data in 2009 (i.e., 102 mandated and 65 non-mandated), 196 out of 236 trained schools (83.1%) provided IPI data in 2010 (122 mandated and 74 non-mandated), 208 out of 256 trained schools (81.3%) provided IPI data in 2011 (139 mandated and 69 non-mandated), 222 out of 267 trained schools (83.1%) provided IPI data in 2012 (149 mandated and 73 non-mandated), 201 out of 275 trained schools (73.1%) provided IPI data in 2013 (132 mandated and 69 non-mandated), and 223 out of 285 trained schools (78.2%) provided IPI data in 2014 (149 mandated and 74 non-mandated). There were only 29 schools in total that were trained in SW-PBIS but never submitted IPI data, 19 of which were mandated schools

d = Cohen's d

*$p < .05$ for independent t tests comparing mandated to non-mandated during each year

2011 training, 65.5% ($n = 171$) of all mandated and 57.0% ($n = 85$) of all non-mandated schools were trained in SW-PBIS. A similar pattern emerged through 2014 (see Supplemental File 1 for annual training information). Although the number of mandated schools trained each year appears substantially higher than the non-mandated secondary schools, t tests reveal that the total proportion of trained schools that were mandated versus non-mandated schools did not differ significantly in 2009–2011 (see Table 3). However, by 2012 and through 2014, there was a statistically significantly higher proportion of mandated schools trained than non-mandated schools (see Table 3). By 2014, three-quarters of all trained secondary schools (i.e., 290 of the total 410 schools) in the state had been mandated to do so.

Research Question 3: Was the Mandate Associated with Changes in Implementation Fidelity of SW-PBIS Over Time?

With regard to implementation fidelity, descriptive analyses were conducted among the trained schools that submitted data (see the note with Table 3 for details on IPI data provided). Overall, the fidelity levels, as indicated by mean IPI scores, were statistically significantly higher in 2010–2013 in non-mandated schools, but were no longer higher in 2014. The unconditional GMM model results indicated that the best fitting model was a 3-class model (see Table 4), wherein there was a growth trajectory composed of 19.2% schools whose fidelity scores began well below adequate but incrementally improved over time (i.e., estimated intercept of 56.12% fidelity; slope of 6.73% per year) to reach 80% on average; a growth trajectory composed of 11.5% of schools whose scores declined over time (i.e., estimated intercept of 77.89% fidelity; slope of − 7.51% per year), and the majority of schools (69.3%) where fidelity scores were high and stable over time (i.e., estimated intercept of 88.69% fidelity; slope of 0.42% per year; see Supplemental File 2). Next, the intercept, slope, and class membership were regressed on the mandate status (1 = mandated, 0 = not mandated). Results indicated that mandated schools had a significantly lower intercept (estimate = − 3.94, $SE = 1.37$, $p = 0.004$) but not slope (estimate = − 0.25, $SE = 0.33$, $p = 0.45$). The intercept and slope were not significantly associated with one another. With regard to class membership, mandated schools were 6.5 times more likely to be in the improving growth trajectory class (estimate = 1.87, $SE = 0.59$, $p = 0.00$, odds ratio [OR] = 6.50) than in the high-stable scores class. The mandated schools were no more likely to be in the declining growth class than they were to be in the high-stable class. GMMs that examined fidelity and adjusted for baseline enrollment, the percent of students receiving FARMs, the mobility rate, and the percent of White students in the school indicated slightly different results. There were no significant associations between the mandated status and the intercept and slope estimates. On the other hand, there was a significant association between the intercept and enrollment (estimate = − 0.10, $SE = 0.002$, $p = 0.00$), the percent of students who were mobile (estimate = − 0.24, $SE = 0.11$, $p = 0.03$), and the percent of White students (estimate = 0.09, $SE = 0.04$, $p = 0.03$). Schools with smaller enrollment, less student mobility, and more White students had a higher intercept. Further, schools with greater student mobility also demonstrated greater growth in their fidelity scores (estimate = 0.14, $SE = 0.07$, $p = 0.03$). With regard to class membership, mandated schools were approximately 90% less likely to be in the high-stable fidelity score classes than they were to be in the improving class (estimate = − 2.23, $SE = 0.93$, $p = 0.02$, OR = 0.11), with these covariates accounted for.

Discussion

As greater attention is focused on enhancing educational standards, school safety, and school climate, more states may be inclined to use policy as a lever to improve these outcomes. The paucity of research examining the impact of educational policies mandating the implementation of social and behavioral programs suggests a clear need for more rigorous research to determine whether these approaches are in fact successful at achieving their intended student outcome goals (Sheras & Bradshaw, 2016). Maryland's truancy policy, implemented within the context of a multi-agency state partnership tracking the SW-PBIS scaling-up and implementation (Bradshaw et al., 2012a, b), provided a unique opportunity to evaluate the success of such a mandate

Table 4 Fit indices for growth mixture models of implementation fidelity

Classes	Log likelihood	AIC	BIC	ABIC	VLMR LRT	Entropy	Class sizes
1	− 4890.25	9802.50	9841.71	9806.83	–	–	–
2	− 4844.85	9717.70	9767.60	9723.21	0.03	0.88	13.0%, 87.0%
3	− 4807.34	9648.68	9709.27	9655.38	0.00	0.84	19.2%, 11.5%, 69.3%
4	− 4791.849	9623.70	9694.99	9631.58	0.38	0.86	2.7%, 20.3%, 12.6%, 64.4%

AIC Akaike Information Criterion, *BIC* Bayesian Information Criterion, *ABIC* sample-size adjusted BIC, *VLMR LRT* Vuong-Lo-Mendell-Rubin likelihood ratio test on improving student behavioral outcomes and increasing SW-PBIS adoption and implementation fidelity. Capitalizing on this rare opportunity, we combined archival data from the state and the PBIS Maryland Partnership and conducted RD analyses to estimate the impact of the mandate on both student behavioral and academic outcomes; these outcomes were of primary interest to the state, and thus were the primary focus of the current study. Although there was a significant association between truancy levels at the time of the mandate and in subsequent academic achievement, suspensions, and habitual truancy across multiple years, this association was consistent for mandated and non-mandated schools (i.e., schools below and at or above the 8% cutpoint) for all outcomes except for suspensions. Specifically, the a_2 estimates and p values indicate that the truancy cutpoint of 8% was significantly associated with the suspension rate in every year, such that mandated schools continued to have significantly higher suspensions than those not mandated. As noted in the results, these differences appear to be persisting differences that occurred prior to the mandate. Although the RD analyses aimed at balancing these baseline differences and all of the required statistical assumptions were met for this model (see Jacob et al., 2012; Ryoo & Pullen, 2017), when examining the unadjusted rates of suspensions in the mandated and non-mandated schools, there was still a notable difference in suspension rates; this difference occurred even in the narrowed sample identified in the optimal bandwidth sensitivity analyses, which was restricted to schools with 4–12% truancy rates at the time of mandate. Specifically, in this restricted sample, the non-mandated schools had suspension rates that were 8–10 percentage points lower than the mandated schools. Thus, we conclude that the mandate did not close the suspension gap, although it may have narrowed the gap. It is clear, however, that the mandate did not cause the suspension gap, as the gap was present at baseline and appears to have persisted.

Because the mandate specifically indicated that schools needed to be trained in (if not already) and implement SW-PBIS, we also examined the training rates between mandated and non-mandated schools, as well as implementation fidelity growth trajectories over time as secondary and tertiary research questions. From a policy perspective, these two latter research questions are perhaps equally important and interesting, as they enabled us to explore whether such a mandate would actually increase SW-PBIS training and fidelity. The RD analyses suggested that SW-PBIS training rates among mandated and non-mandated schools were not impacted by the mandate. When examining the data more descriptively, the results suggest that the training rates did not differ statistically leading up to and across the 2 years following the mandate. However, mandated schools began to demonstrate trends of higher SW-PBIS training rates by 2012, and in the 2 years afterward. Finally, GGM indicated that three growth trajectories in SW-PBIS fidelity emerged over time. The majority of schools had achieved fidelity with nearly 89% of components implemented and were high-stable in their scores. There were two nearly equal–sized classes that showed improving and declining fidelity patterns. Schools demonstrating the improving growth trajectory began with low fidelity (i.e., IPI score of about 56), but had clinically significant growth, with scores reaching beyond the commonly used 80% threshold for fidelity (see Bradshaw et al., 2009a, b; Pas & Bradshaw, 2012). Schools demonstrating the declining growth trajectory began above that 80% threshold and fell nearly to the other class's starting point. The mandated schools were more likely to be in the improving class than they were to be in the high-stable class; they were no more likely to be in the declining class. Only in the unadjusted model did mandated schools demonstrate a statistically significantly lower intercept. This fidelity finding is important given that schools already trained in SW-PBIS were mandated to improve fidelity; these results suggest that the majority of all secondary schools either already had reached high fidelity or had improved their fidelity, with mandated schools being more likely to improve than decline over time.

Limitations and Future Directions

Given that Maryland was one of the first states to adopt SW-PBIS state-wide (see Barrett et al., 2008; Bradshaw & Pas, 2011; Bradshaw et al., 2012a, b; Pas & Bradshaw, 2012), we are unable to determine whether these findings are generalizable to other states. We also focused specifically on middle and high schools, as they were most directly impacted by the mandate, given their elevated levels of truancy compared to elementary schools. In fact, the state first mandated implementation of SW-PBIS in elementary schools based on suspension data, followed by truancy data in secondary schools (Bradshaw et al., 2012a, b). It is possible that the state mandate was explicitly intended to motivate adoption among secondary schools by state legislators, as there had been some promising evidence of the impact of SW-PBIS prior to its adoption (Barrett et al., 2008); this is why we focused on training and implementation fidelity in our second and third research questions. Nevertheless, there has been less research on the impact of SW-PBIS in secondary schools, as the majority of the efficacy work to date in Maryland and other states has been conducted in elementary schools (for notable examples examining secondary schools, see Bradshaw et al., 2014; Pas et al., 2019; for a review also see Gage et al., 2018).

We were also reliant on implementation fidelity at Tier 1 from the IPI, which was completed by school coaches. Given the IPI is a fidelity measure, it was only available for the SW-PBIS trained schools, and thus, it is unknown whether there were elements of SW-PBIS present in the non-trained schools (see Bradshaw et al., 2008a, b). Other fidelity measures, such as the School-wide Evaluation Tool, have been developed and used by Maryland and other states; but, again, these measures were only administered in the trained schools, and even fewer of these schools had School-wide Evaluation Tool data for the observation period examined in this study. Importantly, prior research on the IPI has actually shown it to be a better predictor of student outcomes, like suspensions, than other widely used fidelity measures (e.g., School-wide Evaluation Tool, Benchmarks of Quality; Pas & Bradshaw, 2012). We only have internal consistency data on the IPI from a prior study by Bradshaw et al. (2009a, b), as the IPI data analyzed in the current paper are archival and based on the summary/subscale scores uploaded by the participating schools into the state's data system. As such, item-level data are not available, and thus, alphas cannot be calculated based on the current sample of schools' subscale scores.

Moreover, we were limited in the extent to which we could explore variation in outcomes associated with differential levels of fidelity. As noted above, prior to this mandate, schools in the state had volunteered for SW-PBIS training and had to meet a series of readiness and buy-in criteria (e.g., forming a PBIS team, provide a 3-year commitment, solicit buy-in from 80% of staff). These are unmeasured variables that could have impacted training, fidelity, and outcomes associated with the mandate. While many of these constructs are beyond the scope of the current study, they are potential confounding variables that were not available for inclusion in the current study. Future studies should explore factors such as fidelity (e.g., Mercer et al., 2017), social validity (e.g., Lane et al., 2009), and time (Gage et al., 2018) in greater detail with regard to mandated implementation. Additionally, we focused exclusively on Tier 1 implementation because that was the emphasis of the mandate. There may have been other more intensive Tier 2 or Tier 3 interventions in place within these schools; however, we lack data on training on the more advanced tiers, as well as implementation fidelity.

We did not examine nesting at the district level, as the schools were nested within just 24 districts, the number of schools within districts varied considerably, and the current models were already relatively complex statistically. However, prior exploration of district-level factors and their association with schools seeking training in or adopting SW-PBIS yielded relatively few significant findings (i.e., percent of schools trained in PBIS in the district and district size), and no such associations were found with fidelity scores (Bradshaw & Pas, 2011). Although examining school-level moderators of effects on outcomes is beyond the scope of the current study, the field may benefit from future exploration of these factors. For example, such future directions might explore the extent to which the effect of the mandate differed for middle schools and high schools, schools in rural and urban settings, or schools with and without Title 1 status. This is a potential area to explore further in future research.

Though the use of the RD design and analyses was a strength, biases can remain in the resulting estimates when it is difficult to formulate a balanced design. Sensitivity analyses with optimal bandwidths suggest that the student outcome findings were robust. And when taken together with the main analyses, they help to optimize this study's validity (i.e., statistical conclusion, internal, and external). We also considered other non-experimental analytic approaches, such as propensity score matching; however, there are similar challenges in ensuring balance (King & Nielsen, 2016), particularly given the relatively high proportion of schools mandated as compared to those that were not mandated. An additional complexity associated with this study was that some schools in both the mandated and non-mandated condition had received training in SW-PBIS prior to the mandate's implementation. Although future analyses may explore the extent to which differential truancy thresholds may have been associated with differential effects, or the optimal threshold for achieving outcomes, our interest in this study was on the effect of the mandate, which was set at 8%, rather than identifying an optimal threshold for the policy.

Admittedly, the SW-PBIS logic model suggests that both training in SW-PBIS and fidelity of SW-PBIS implementation would *precede* improvement in student outcomes. However, it is important to highlight that this study focused specifically on whether student outcomes, training, and fidelity differed for the schools mandated by the state as compared to those that were not mandated. Importantly, this study was *not* designed or intended to estimate the effects of SW-PBIS as a preventive intervention. As such, we were unable to formulate specific conclusions regarding the effectiveness of SW-PBIS based on these findings, particularly in light of the fact that only a proportion of the mandated schools actually implemented SW-PBIS and some implemented prior to the mandate. Rather, the focus of this study was on the effect of a *mandate*, in terms of the student outcomes achieved, as well as the resulting training and fidelity of SW-PBIS. It is possible that some schools may have improved outcomes or implementation even without being mandated; as such, it may be possible that just having a law mandating implementation of SW-PBIS could have generated improved outcomes for some schools, even if they were not mandated. Moreover, one might question if the mandated schools were motivated to improve outcomes (perhaps truancy) through other means than SW-PBIS implementation, to essentially "get off the mandated list"; this would suggest that the mandate, in and of itself, was associated with improvements, but not necessarily through implementation of SW-PBIS. However, this did not appear to be the case, given the RD analysis indicated that the mandated schools did not improve student behavioral or academic outcomes relative to the non-mandated schools.

Conclusions and Implications

These findings suggest that the state's efforts to promote broad dissemination of SW-PBIS through the use of a state policy focused on truancy did not improve student outcomes but may have begun to shift or "nudge" the mandated schools to access training and reach fidelity of SW-PBIS. Although the RD analysis does not reflect significant differences in training, the descriptive analyses suggested that by 2014, the majority of SW-PBIS trained secondary schools had been mandated. Despite this, not all of the mandated schools had received training in SW-PBIS by 2014 (i.e., 4 years following the initial rollout of the mandate). Moreover, the fidelity analyses suggested that mandated schools were more likely to demonstrate the improving fidelity growth trajectory, with poor initial fidelity, but improving after the mandate to surpass the benchmark for adequate fidelity.

Taken together, the results suggest that the mandate may have had some modest effect on increasing training and fidelity of SW-PBIS. It is possible that the window of time examined in this study was not sufficient for full implementation of the policy or full implementation of SW-PBIS, much less student outcomes to be achieved. Nevertheless, the results suggest that even when implementing SW-PBIS within the context of a mandate, schools can improve fidelity and achieve adequate to high fidelity over time. It is quite possible that there may have been limited implementation of the policy with regard to accountability for follow through in training of the mandated schools. As such, researchers should be thoughtful when partnering with policymakers who are interested in mandating the use of different programs or practices, as such mandates may not prove to be effective, particularly if the policy and/or related programming do not result in the intended uptake of the evidence-based program. Additional procedures may be needed in future mandated implementation of SW-PBIS or other prevention models to further ensure that there are sufficient supports and accountability measures to optimize training, uptake, and implementation of the mandated program. Future studies may also explore whether the program itself was implemented as intended, which in turn translated into improved student outcomes, or if the threat of the mandate sufficiently motivated the schools to improve outcomes through other means.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11121-021-01257-0.

Acknowledgements The authors would like to acknowledge the Maryland PBIS Management Team, which includes the Maryland State Department of Education, Sheppard Pratt Health System, and the 24 local school districts. We give special thanks to Philip Leaf, Katrina Debnam, Elizabeth Stuart, Susan Barrett, and Jerry Bloom for their support of this project.

Funding The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305H150027 (PI: C. Bradshaw) to the University of Virginia.

Compliance with Ethical Standards

Ethics Approval All procedures performed in studies involving human participants were in accordance with ethical standards of the institutional and national research committee and the 1964 Helsinki declaration and its later amendments or comparable ethical standards. The study was approved under exempt status by the universities' IRB.

Informed Consent Not applicable, as it is analysis of archival data.

Conflict of Interest The authors declare that they have no conflicts of interest. The lead author on this paper is also the editor of the journal *Prevention Science*; however, the peer review of this manuscript was managed by another associate editor.

Disclaimer The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education or the Maryland State Department of Education.

References

Barrett, S. B., Bradshaw, C. P., & Lewis-Palmer, T. (2008). Maryland statewide PBIS initiative: Systems, evaluation, and next steps. *Journal of Positive Behavior Interventions, 10,* 105–114.

Bradshaw, C. P., Debnam, K. J., Koth, C. W., & Leaf, P. J. (2009). Preliminary validation of the Implementation Phases Inventory for assessing fidelity of schoolwide positive behavior supports. *Journal of Positive Behavior Interventions, 11,* 145–160. https://doi.org/10.1177/1098300708319126

Bradshaw, C. P., Debnam, K. J., Lindstrom Johnson, S., Pas, E. T., Hershfeldt, P., Alexander, A., et al. (2014). Maryland's evolving system of social, emotional, and behavioral interventions in the public schools: The Maryland Safe and Supportive Schools Project. *Adolescent Psychiatry, 4,* 194–206. https://doi.org/10.2174/221067660403140912163120

Bradshaw, C. P., Koth, C. W., Bevans, K. B., Ialongo, N. S., & Leaf, P. J. (2008a). The impact of school-wide Positive Behavioral Interventions and Supports (PBIS) on the organizational health of elementary schools. *School Psychology Quarterly, 23*, 462–473.

Bradshaw, C. P., Koth, C. W., Thornton, L. A., & Leaf, P. J. (2009b). Altering school climate through school-wide Positive Behavioral Interventions and Supports: Findings from a group-randomized effectiveness trial. *Prevention Science, 10*, 100–115. https://doi.org/10.1007/s11121-008-0114-9

Bradshaw, C. P., Mitchell, M. M., & Leaf, P. J. (2010). Examining the effects of schoolwide Positive Behavioral Interventions and Supports on student outcomes: Results from a randomized controlled effectiveness trial in elementary schools. *Journal of Positive Behavior Interventions, 12*, 133–148.

Bradshaw, C. P., & Pas, E. T. (2011). A state-wide scale-up of Positive Behavioral Interventions and Supports (PBIS): A description of the development of systems of support and analysis of adoption and implementation. *School Psychology Review, 40*, 530–548.

Bradshaw, C. P., Pas, E., Bloom, J., Barrett, S., Hershfeldt, P., Alexander, A. C., & A. E., & Leaf, P. . (2012a). A state-wide collaboration to promote safe and supportive schools: The PBIS Maryland Initiative. *Administration and Policy in Mental Health and Mental Health Services Research, 39*(4), 225–237. https://doi.org/10.1007/s10488-011-0384-6

Bradshaw, C. P., Reinke, W. M., Brown, L. D., Bevans, K. B., & Leaf, P. J. (2008b). Implementation of school-wide Positive Behavioral Interventions and Supports (PBIS) in elementary schools: Observations from a randomized trial. *Education and Treatment of Children, 31*, 1–26. https://doi.org/10.1353/etc.0.0025

Bradshaw, C. P., Waasdorp, T. E., & Leaf, P. J. (2012b). Effects of school-wide Positive Behavioral Interventions and Supports on child behavior problems. *Pediatrics, 130*, 1136–1145. https://doi.org/10.1542/peds.2012-0243

Calonico, S., Cattaneo, M. D., Farrell, M. H., & Titiunik, R. (2016). rdrobust: Robust data- driven statistical inference in regression-discontinuity designs [Computer software manual]. Retrieved from https://CRAN.R-project.org/package1⁄4rdrobust (R package version 0.93).

Dimmery, D. (2016). rdd: Regression discontinuity estimation [Computer software manual]. Retrieved from https://CRAN.R-project.org/package1⁄4rdd (R package version 0.57).

Fagan, A., Bumbarger, B., Barth, R., Bradshaw, C. P., Rhoades Cooper, B., & Supplee, & Walker, D. . (2019). Scaling up evidence-based interventions in US public systems to prevent behavioral health problems: Challenges and opportunities. *Prevention Science, 20*, 1147–1168. https://doi.org/10.1007/s11121-019-01048-8

Gage, N. A., Grasley-Boy, N., George, H. P., Childs, K., & Kincaid, D. (2019). A quasi-experimental design analysis of the effects of school-wide Positive Behavior Interventions and Supports on discipline in Florida. *Journal of Positive Behavior Interventions, 21*, 50–61. https://doi.org/10.1177/1098300718768208

Gage, N. A., Lee, A., Grasley-Boy, N., & Peshak George, H. (2018). The impact of school-wide positive behavior interventions and supports on school suspensions: A statewide quasi-experimental analysis. *Journal of Positive Behavior Interventions, 20,* 217–226.

Glasgow, R. E., Vinson, C., Chambers, D., Khoury, M. J., Kaplan, R. M., & Hunter, C. (2012). National Institutes of Health approaches to dissemination and implementation science: Current and future directions. *American Journal of Public Health, 102,* 1274–1281.

Hatzenbuehler, M. L., Schwab-Reese, L., Ranapurwala, S. I., Hertz, M. F., & Ramirez, M. R. (2015). Associations between antibullying policies and bullying in 25 states. *JAMA Pediatrics, 169*(10), e152411. https://doi.org/10.1001/jamapediatrics.2015.2411

Horner, R. H., Sugai, G., & Anderson, C. M. (2010). Examining the evidence base for schoolwide positive behavior support. *Focus on Exceptional Children, 42,* 1–14.

Horner, R. H., Sugai, G., Smolkowski, K., Eber, L., Nakasato, J., Todd, A. W., & Esperanza, J. (2009). A randomized, wait-list controlled effectiveness trial assessing school-wide Positive Behavior Support in elementary schools. *Journal of Positive Behavior Interventions, 11*, 133–144.

Horner, R. H., Todd, A. W., Lewis-Palmer, T., Irvin, L. K., Sugai, G., & Boland, J. B. (2004). The school-wide evaluation tool (SET): A research instrument for assessing school-wide positive behavior behavior support. *Journal of Positive Behavior Interventions, 6,* 3–12. https://doi.org/10.1177/10983007040060010201

Imbens, G. W., & Kalyanaraman, K. (2009). *Optimal Bandwidth Choice for the Regression Discontinuity Estimator* (NBER Working Paper 14726). Cambridge, MA: National Bureau of Economic Research.

Jacob, R., Zhu, P., Sommers, M., & Bloom, H. (2012). *A practical guide to regression discontinuity*. New York, NY: MDRC. Available at http://dev.mdrc.org/sites/default/files/RDD%20Guide_Full%20rev%202016_0.pdf

King, G., & Nielsen, R. (2016). *Why propensity scores should not be used for matching*. Retrieved from https://gking.harvard.edu/publications/why-propensity-scores-should-not-be-used-formatching

Kittelman, A., McIntosh, K., & Hoselton, R. (2019). Adoption of PBIS within school districts. *Journal of School Psychology, 76*, 159–167.

Lane, K. L., Kalberg, J. R., Bruhn, A. L., Driscoll, S. A., Wehby, J. H., & Elliott, S. (2009). Assessing social validity of school-wide positive behavior support plans: Evidence for the reliability and structure of the Primary Intervention Rating Scale. *School Psychology Review, 38*, 135–144.

Lee, D. S. (2008). Randomized experiments from non-random selection in U.S. House elections. *Journal of Econometrics, 142*, 675–697. https://doi.org/10.1016/j.jeconom.2007.05.004

Lee, A., & Gage, N. A. (2020). Updating and expanding systematic reviews and meta-analyses on the effects of school-wide positive behavior interventions and supports. *Psychology in the Schools, 57,* 783–804. https://doi.org/10.1002/pits.22336

Lee, D. S., & Lemieux, T. (2010). Regression discontinuity designs in economics. *Journal of Economic Literature, 48*(2), 281–355. https://doi.org/10.1257/jel.48.2.281

Lindstrom Johnson, S., Alfonso, Y. N., Pas, E. T., Debnam, K. J., & Bradshaw, C. P. (in press). Scaling-up PBIS: The distribution of costs across educational stakeholders. *School Psychology Review.*

Madigan, K., Cross, R. W., Smolkowski, K., & Strycker, L. A. (2016). Association between schoolwide positive behavioural interventions and supports and academic achievement: A 9-year evaluation. *Educational Research and Evaluation, 22*(7–8), 402–421. https://doi.org/10.1080/13803611.2016.1256783

McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics, 142*(2), 698–714. https://doi.org/10.1016/j.jeconom.2007.05.005

Mercer, S. H., McIntosh, K., & Hoselton, R. (2017). Comparability of fidelity measures for assessing Tier 1 school-wide Positive Behavioral Interventions and Supports. *Journal of Positive Behavior Interventions, 19*(4), 195–204.

Muthén, L. K., & Muthén, B. O. (2002–2018). *Mplus user's guide*. Retrieved from http://www.statmodel.com/ugexcerpts.shtml

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of latent classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling: A Multidisciplinary Journal, 14*(4), 535–569.

OSEP Technical Assistance Center on Positive Behavioral Interventions and Supports (2019). Positive Behavioral Interventions & Supports [Website]. Retrieved from www.pbis.org.

Pas, E. T., & Bradshaw, C. P. (2012). Examining the association between implementation and outcomes: State-wide scale-up of school-wide Positive Behavioral Intervention and Supports. *Journal of Behavioral Health Services and Research, 39*(4), 417–433. https://doi.org/10.1007/s11414-012-9290-2

Pas, E. T., Ryoo, J. H., Musci, R., & Bradshaw, C. P. (2019). The effects of a state-wide scale-up of school-wide Positive Behavior Intervention and Supports on behavioral and academic outcomes: A quasi-experimental examination. *Journal of School Psychology, 73*, 41–55. https://doi.org/10.1016/j.jsp.2019.03.001

Positive Behavioral Interventions and Support Program, Title 13A Maryland State Board of Education § 2–205 and 7–304.1 (2008).

Ramaswamy, V., Desarbo, W. S., Reibstein, D. J., & Robinson, W. T. (1993). An empirical pooling approach for estimating marketing mix elasticities with PIMS data. *Marketing Science, 12*, 103–124.

R Core Team. (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Ryoo, J. H., & Pullen, P. C. (2017). Regression discontinuity design for longitudinal data, cross-sectional data, and intervention research. In J. M. Kauffman, D. P. Hallahan, & P. C. Pullen (Eds.), *Handbook of special education.* (2nd ed., pp. 137–150). Routledge.

Ryoo, J. H., Wang, C., Swearer, S., Hull, M., & Shi, D. (2018). Longitudinal model building using latent transition analysis: an example of using school bullying data. *Frontiers in Psychology – Quantitative Psychology and Measurement, 9*, 675. https://doi.org/10.3389/fpsyg.2018.00675

Sheras, P. L., & Bradshaw, C. P. (2016). Fostering policies that enhance positive school environment. *Theory into Practice, 55,* 129–135. https://doi.org/10.1080/00405841.2016.1156990

Spoth, R. L., Rohrbach, L. A., Greenberg, M. T., Leaf, P., Brown, C. H., Fagan, A. A., … Society for Prevention Research Type 2 Translational Task Force. (2013). Addressing core challenges for the next generation of Type 2 translation research and systems: The translation science to population impact (TSci Impact) framework. *Prevention Science, 14*, 319–351.

Stigler, M., & Quast, B. (2015). rddtools: Toolbox for regression discontinuity design (‘rdd’) [Computer software manual]. https://CRAN.R-project.org/package1⁄4rddtools (R package version 0.4.0).

Stuart, E., Bradshaw, C. P., & Leaf, P. (2015). Assessing the generalizability of randomized trial results to target populations. *Prevention Science, 16*, 475–485. https://doi.org/10.1007/s11121-014-0513-z

Sugai, G., & Horner, R. H. (2002). The evolution of discipline practices: School-wide positive behavior supports. *Child & Family Behavior Therapy, 24*(1–2), 23–50. https://doi.org/10.1300/J019v24n01_03

Sugai, G., & Horner, R. H. (2006). A promising approach for expanding and sustaining school-wide Positive Behavior Support. *School Psychology Review, 35*(2), 245–259.

Thistlewaite, D. L., & Campbell, D. T. (1960). Regression-discontinuity analysis: An alternative to the ex-post facto experiment. *Journal of Educational Psychology, 51*, 309–317.

Thoemmes, F., Liao, W., & Jin, Z. (2017). The analysis of the regression-discontinuity design in R. *Journal of Educational and Behavioral Statistics, 42*(3), 341–360. https://doi.org/10.3102/1076998616680587

Trochim, W. (2001). Regression-discontinuity design. *International Encyclopedia of the Social and Behavioral Sciences, 19*, 12940–12945.

United States Department of Education (USDOE) (2019). *Chronic absenteeism in the nation’s schools.* https://www2.ed.gov/datastory/chronicabsenteeism.html#intro

Waasdorp, T. E., Bradshaw, C. P., & Leaf, P. J. (2012). The impact of school-wide Positive Behavioral Interventions and Supports (SWPBIS) on bullying and peer rejection: A randomized controlled effectiveness trial. *Archives of Child and Adolescent Medicine, 116*, 149–156.