International Journal of Education and Social Science; Vol. 8 No. 5; October 2021
ISSN 2410-5171 (Online), 2415-1246 (Print)
Published by Research Institute for Progression of Knowledge

Evaluating Population-Level Implementations of Evidence-Based Programming: PAX Good Behavior Game and Youth Crime

Gulcan Cil
Oregon Research Institute
Eugene, OR 97403, USA

Jason Fruth
Ohio Research Solutions
Dayton, OH 45409, USA

Anthony Biglan
Oregon Research Institute
Eugene, OR 97403, USA

Abstract

This study analyzes whether the implementation of the PAX Good Behavior Game in elementary schools was associated with lower levels of school-based law enforcement referrals and arrests. It utilizes publicly available administrative data in a quasi-experimental design. School-based law enforcement referrals and arrests in zip codes in which the PAX Good Behavior Game was implemented in schools as part of a statewide population-level public health initiative were compared to those of otherwise similar comparison schools in neighboring zip codes in which the intervention was not yet implemented. For the two geographic regions examined in this study, one region recorded a significant decrease in law enforcement referrals in intervention zip codes compared to comparison zip codes. Neither region recorded significant changes in school arrests among zip codes. The modeling used to detect these changes has meaningful implications for evaluating population-level implementation.

Keywords: Prevention, Evidence-based Program, Evaluation, PAX Good Behavior Game

Subject: Prevention Science

Introduction

This study explores the possibility of using administrative data to evaluate the impact of the PAX Good Behavior Game (PAX GBG) at the population level. PAX GBG and the original intervention it derives from have proven effects in multiple randomized control studies. While randomized control experiments are the gold standard to demonstrate program efficacy, they are often too costly for assessing the impact of a program at a population level. These designs allow for extensive programmatic data collection and analysis under small-scale, experimental conditions, but population-level implementations necessarily lack the resources to carry out the same level of evaluation. Population-level implementations will also only occur with interventions that have already shown proven effects according to programmatic data.

For a classroom-based universal preventive intervention like PAX GBG, local and regional data will be available due to mandatory reporting and public record laws. In the absence of sufficient large-scale programmatic data, these other data collected for administrative purposes can be of significant value. These data used in robust longitudinal quasi-experimental designs with a careful consideration to potential confounds could provide an opportunity for population-level program evaluation.

These distal data points could also represent contextually compelling or stakeholder-specific data more appropriate for gauging the impact of a population-level implementation of interventions – especially those impacting an array of outcomes like PAX GBG.

Randomized-control Trials Involving the Good Behavior Game

PAX GBG is a universal classroom-based preventive intervention applied by teachers as part of the daily management of their classroom (Embry, 2002). This intervention is not a student behavioral curriculum but a set of research-based strategies that teachers implement while carrying out daily tasks associated with managing the classroom including: getting students' attention, selecting students for tasks, transitioning from one task to the next, working as part of a team, limiting problematic behavior, and reinforcing pro-social behavior. PAX GBG is the commercially-available derivative of the good behavior game and script written by Dr. Jaylan Turkkan and used in the first of ongoing comparative effectiveness trials at Johns Hopkins University (Kellam, et al., 1994). The intervention has shown an array of positive longitudinal effects on academics, behavior, mental health, and other lifetime outcomes (Wilson, Hayes, Biglan, & Embry 2014). These outcomes have been tracked as part of three separate cohorts of participants in ongoing studies at Johns Hopkins University as well as a number of other randomized control studies throughout North America and Europe (Kellam, et al., 1994, Ialongo, et al., 1999, & Domitrovich, et al., 2016).

When teachers implemented the Good Behavior Game in the classroom with their students, the research-based strategies designed to improve attention, behavior, and self-regulation decreased problematic behaviors by 75% through direct observation (Embry, 2002, Wilson, Hayes, Biglan, & Embry, 2014). These problematic behaviors included failing to follow directions, talking out of turn, or any off-task, contextually-inappropriate behavior. This improved student behavior provides an opportunity for increased teacher reinforcement and feedback that supports and improves teacher student relationships in a transactional model (Huber, Fruth, Avila-John, & Ramirez, 2016). The improved student behavior also enhanced relationships and peer support among students (Ialongo, et al., 1999). Better student and teacher relationships improved individual classroom climate independent of the entire school's climate (Domitrovich, et al., 2015, Smith, et al., 2013).

The improved proximal behavioral outcomes from implementing PAX GBG have also been shown to have positive academic effects in real-world quasi-experimental trials. Students who receive PAX GBG in the classroom scored higher on standardized math and reading assessments regardless of the curricula used by the teacher (Weis, Osborne, & Dean, 2015). In the original randomized trails, implementing the Good Behavior Game in the classroom also reduced identification rates for special education (Bradshaw, Zmuda, Kellam, & Ialongo, 2009). Ultimately, children exposed to the Good Behavior Game for at least one year experienced improved behavioral and academic success throughout their schooling resulting in improved high school graduation and college entrance rates (Bradshaw, et al., 2009).

Children exposed to the Good Behavior Game in their classroom for at least one year in the original trials also experienced significant differences with more distal outcomes. Children who had the Good Behavior Game demonstrated less lifetime use of alcohol, tobacco, and illicit drugs including opioids (Kellam, et al., 2014, & Furr-Holden, et al., 2004). The Good Behavior Game also reduced both early and high-risk sexual activity (Kellam, et al., 2014). These data indicate the malleability and strengthening of self-regulation after exposure to the intervention.

Children exposed to the Good Behavior Game during elementary school in the original study also showed reduced lifetime conduct problems and psychiatric disorders (Kellam, et al., 2008, & Wang, et al., 2009, Jiang, et al., 2018, & Petras, Masyn, & Ialongo, 2011). This includes a reduction in lifetime suicide ideation and attempts (Wilcox, et al., 2008, & Katz, et al., 2013). These changes in psychological and physiological make-up of children are indicated by the positive expression of brain-derived neurotrophic factor after at least one year of exposure to GBG (Musci, et al., 2014).

For this particular study, the researchers sought publicly available administrative data to serve as proxies for many of these data points that would be of particular interest to stakeholders but otherwise only available through thorough and costly research of programmatic or evaluation data such as the original and ongoing implementation and tracking at Johns Hopkins University.

Anti-social, aggressive, and violent behavior among school-age children is of particular interest because of its costly and detrimental impact on the classroom. In the original trials, the Good Behavior Game decreased aggressive and externalizing behaviors, especially in school-age boys (Kellam, et. al., 1994, Kellam, et al., 1998). Schools implementing the Good Behavior Game and other strategies now associated with PAX GBG also reported less bullying, fewer violent school injuries, fewer associated nurses' office visits, and were less likely to be charged with crimes as minors (Iaolongo, et al., 1999, Embry, et al., 1996, Brenner, Krug, Dahlberg, & Powell, 1997, Kellam, Reid, & Balster, 2008, & Kellam et al., 2008).

Utilizing Administrative Data and Varied Evaluation Designs

Initial research utilizes numerous supports and associated research efforts to detect efficacy and effectiveness during the development of interventions. This includes strict adherence to research standards to determine real effects as well as accounting for error and variance under the careful watch of numerous researchers. These results are used to enhance interventions in their preparation for generalized and even scaled use. However, the substantial research mechanisms and even programmatic data collection used to detect significant changes due to interventions are not feasible in larger, scaled efforts such as population-level implementations and public health initiatives. Publicly available databases along with essential human resources for data collection and analysis may serve as valuable resources in determining the effects and impact of a population-level implementation of an intervention.

Quasi-experimental designs and the use of publicly available administrative data and subsequent modeling can assist in the evaluation of large-scale initiatives by providing valuable feedback to stakeholders. Though random assignment is not feasible or even appropriate in a population-level implementation, quasi-experimental designs are suitable for detecting trends similar to those found through randomized designs. These databases and research designs provide a mechanism for public officials to track scaled, population-level effects of their implementations to determine if the intervention is having similar effects at scale that it did in efficacy and effectiveness trials.

Methods

Outcome Data and Measure

This study used school-level data on two measures of antisocial behavior in schools: (a) the rates of school referrals to law enforcement and (b) the rate of school arrests. The data came from the U.S. Department of Education Civil Rights Data Collection, or CRDC (U.S. Department of Education, 2018), biennially from 2011-2012 school year to 2015-2016 school year. Starting in 2011-2012 school year, all public schools in the U.S. were required to report to CDRC. These data are of particular interest to stakeholders and funders responsible for improving the public health on local, regional, and national scales. These data also correspond to commonly tracked programmatic data from implementations of PAX GBG.

Sample Population

We identified 23 total zip codes across two separate sites in Northwest Ohio and Central Ohio that have high rates of PAX GBG implementations based on the relatively high concentration of individuals trained, and assumed that all schools in these zip codes had PAX GBG implementations. There were a total of 49 schools across these two sites. There were 31 schools in 9 Northwest Ohio PAX GBG zip codes and 18 schools in 5 Central Ohio PAX GBG zip codes. Table 1 below shows the list of these zip codes and the number of schools in each zip code.

Table 1
List of zip codes with schools associated with each code.

	Zip codes	Number of schools
Northwest	43402	8
	43443	1
	43447	3
	43450	3
	43522	1
	43551	7
	43619	2
	44817	3
	45872	3
Central	43040	7
	43045	3
	43064	4
	43067	1
	43344	3

We matched schools implementing PAX GBG with comparison schools with no PAX GBG implementation. The candidate schools for the comparison group included all schools in the neighboring areas, namely, Hancock and Seneca counties in the Northwest region, and Marion and Delaware counties in the Central region. The matching was performed using one-to-one propensity score matching with no replacement based on the demographic composition in each school including proportion of enrolled students identifying as White, Black, Hispanic, and other ethnic/racial categories, proportion of students with disabilities, and proportion of students with limited English proficiency. The matching criteria also include the socio-economic characteristics of the school zip code including median household income, educational background of the household, and percentage identified as urban. Zip code level socio-economic information are from U.S. Census Bureau, Missouri Census Data Center (U.S. Census Bureau, 2016) based on 2010 Census and 2008-2012 American Community Survey data.

Analysis

We utilized a generalized difference-in-differences, or a controlled interrupted time series model to estimate the impact of PAX GBG implementation on rates of school referrals to law enforcement and school arrests. We estimated the impact of PAX GBG implementation in each of the two regions using interaction terms and the pooled data as given here:

$$Y_{sy} = \beta_1 GBGNW_s PostNW_y + \beta_2 GBGCentral_s PostCentral_y + X_{sy} \cdot \delta + \alpha_s + \alpha_y + \epsilon_{sy} \quad [1]$$

Here, Y_{sy} is the number of school referrals to law enforcement, or the number of school arrests, per 1,000 students in school s year y. We ran separate regressions for the two outcome variables. The key predictors of the model were the two interaction terms between "PAX GBG communities" in Northwest Ohio and Central Ohio and the corresponding "post-treatment" indicators for the two regions. The rapid adoption of PAX GBG started in 2013 in Northwest Ohio and 2014 in Central Ohio. Therefore, the post-treatment dummy is equal to one for the period after 2013 for Northwest Ohio treatment and comparison schools and equal to one after 2014 for Central Ohio schools.

The model includes year fixed effects (α_y) to flexibly allow for year-to-year changes in the outcome that are common to all schools. A standard difference-in-differences model would include indicators for the treatment group schools to account for the baseline differences in the outcome variables between the treatment and the comparison. In this generalized model, these indicators are subsumed by the school fixed effects (α_s). School fixed effects account for baseline differences in the referrals and arrests more flexibly by allowing them to vary at the school-level rather than imposing homogeneity within treatment or comparison groups.

Finally, the adjusted models include time-varying school level variables (X_{sy}) such as racial composition of enrolled students, proportion of students with disabilities, and proportion of students with limited English proficiency (LEP). Time-invariant school characteristics were accounted for by school fixed effects.

The strength of this analysis design compared to a cross-sectional comparison between schools with and without PAX GBG implementation is that this model takes advantage of pre-treatment data to account for the differences in outcomes between the treatment and comparison schools that existed before the PAX GBG implementation as well as any changes in outcomes that may have occurred in all schools over time – including comparison schools. It also includes a matched comparison group to mitigate the selection bias.

Results

The descriptive characteristics of treatment and comparison schools prior to PAX GBG implementation are provided in Table 2 below. There is no statistically significant difference between the treatment and the comparison schools in terms of the composition of students by race/ethnicity, disability status, and LEP status. The average number of referrals to law enforcement per 1,000 students appears to be higher, and the number of school arrests per 1,000 students are lower in PAX GBG schools relative to the comparison schools but these differences are not statistically significant.

Table 2
Demographic composition of PAX GBG and comparison schools, prior to PAX GBG implementation.

	All schools	**PAX GBG schools**	**Comparison schools**	**Difference**	***p***
% White	88.24 (7.49)	88.34 (6.79)	88.15 (8.17)	0.19	0.856
% Black	1.68 (1.88)	1.66 (1.83)	1.71 (1.93)	-0.05	0.855
% Hispanic	4.99 (4.45)	5.25 (4.41)	4.74 (4.50)	0.51	0.427
% other non-White	5.08 (3.64)	4.75 (2.96)	5.41 (4.20)	-0.66	0.211
% with disability	13.66 (6.90)	13.98 (8.35)	13.35 (5.10)	0.63	0.531
% with limited English proficiency	0.68 (1.00)	0.72 (1.05)	0.65 (0.96)	0.07	0.642
Referrals to law enforcement (per 1,000)	0.61 (3.22)	0.90 (4.41)	0.32 (1.14)	0.58	0.212
School arrests (per 1,000)	0.01 (0.20)	0.00 (0.00)	0.03 (0.28)	-0.03	0.321
# of schools	98	49	49		

The point estimates for β_1 and β_2 in equation 1 which represent the changes in rates of school referrals to law enforcement and the changes in rates of school arrests associated with PAX GBG implementation are presented in panels A and B of Table 3 respectively. The results in Panel A of Table 3 imply that PAX GBG implementation in Northwest Ohio was associated with a decline in school referrals to law enforcement. There were 1.7 fewer referrals per 1000 students in the Northwest Ohio PAX GBG schools in the post-PAX GBG period relative to non-PAX GBG schools (p-value=0.018). This change is after accounting for the school characteristics and the number of referrals that would be considered "normal" for each school (school fixed effects) and is above and beyond any change that had been experienced by all schools that year (year fixed effects). No statistically significant changes in referrals to law enforcement were detected in Central Ohio PAX GBG schools.

Possible explanations for this result include a comparatively small number of communities in the Central Ohio region implementing PAX GBG and included in this study, having only one post-treatment data point for these schools, and having a relatively small number of referrals in these schools. Results in Panel B of Table 3 indicate that there were no changes in rates of school arrests associated with PAX GBG.

Table 3

Estimated impact of PAX GBG implementation on law enforcement referrals and school arrests.

	Panel A				**Panel B**			
	Referrals to law enforcement				**School arrests**			
	Unadjusted		**Adjusted**		**Unadjusted**		**Adjusted**	
	Change in referrals (β)	p	Change in referrals (β)	p	Change in arrests (β)	p	Change in arrests (β)	p
PAX GBG Northwest	-1.69	0.016	-1.71	0.018	-0.02	0.713	-0.02	0.706
PAX GBG Central	0.46	0.579	0.42	0.616	0.09	0.121	0.09	0.123
N (school*year)	287				287			
# of schools	98				98			

Note: Only the point estimates for β_1 and β_2 in equation 1 are presented. Both unadjusted and adjusted models include year fixed effects and school fixed effects. Adjusted model includes school-level variables described above.

Discussion

The results of this quasi-experimental comparison provide limited support for the benefit of the PAX Good Behavior Game in reducing antisocial behavior given the data utilized. Schools in Northwest Ohio that implemented PAX GBG had significantly fewer referrals to law enforcement than matched schools that did not implement PAX GBG. However, there was not a significant difference between these two sets of schools in the number of arrests. In the smaller number of schools in central Ohio, there were no differences between PAX GBG and non-PAX GBG schools on referrals or arrests. This may be because there were fewer schools and thus less statistical power.

These results imply further replication to determine whether PAX GBG reliably reduces antisocial behavior that otherwise requires significant educational intervention. Here we clarify the limitations of this study with an eye toward how the empirical analysis of the impact of PAX GBG could be strengthened. This also provides opportunities to assess the appropriateness and validity of the publicly available data sets as proxies for the original efficacy and effectiveness outcomes.

Although the study used a matched sample of PAX GBG schools and comparison schools, there may still be a selection bias associated with a potential non-random adoption of PAX GBG. This results from the real-world selection, funding, and application of PAX GBG as a commercially-available public health intervention that would naturally not follow the strict randomized assignment of an efficacy trial. There may be other systematic differences between the schools that adopted PAX GBG and those that did not. If those differences, predispositions, or administrative tendencies are correlated with the outcome variables, the results would be biased and potentially provide iatrogenic or even synergistic effects.

The timing of the adoption may also not necessarily be random. The analysis accounts for the baseline differences in outcome variables between the treatment and comparison schools and the common changes in outcomes over time. However, there is only one pre-treatment and two post-treatment data points for the schools in Northwest Ohio and two pre-treatment and one post-treatment data points for the schools in Central Ohio. These data sources did not provide enough pre-treatment time series data to test the validity of the implicit assumption of similar pre-treatment trends in the PAX GBG and comparison schools. This study could also be strengthened with ongoing follow-up and additional data points to analyze longer term impact.

There are a relatively small number of schools that have any referrals to law enforcement and even a smaller number of schools with any school arrests. Accordingly, we rely on only a small sample of schools for the source of variation in the dependent variables. Although all schools are required to report data to CRDC, there is no mechanism to verify the validity or uniformity in reporting. This may lead to possible measurement errors in the outcome data, which in turn would result in inflated standard errors. More importantly, if such potential inaccuracies in reporting are higher or lower in one group of schools, PAX GBG or comparison, relative to the other group, it may lead to bias in point estimates.

Research on this and other versions of the Good Behavior Game have demonstrated reduced disruptive behavior in classrooms and reduced youth arrests in original studies (Kellam et al., 1998, Wilson et al., 2014, & Kellam et al., 2008). These are consistent in showing the benefit of the Good Behavior Game for reducing antisocial behavior. Further assessment of the impact of PAX GBG in a real-world or scaled implementation would be a step in the right direction for public health with the ultimate goal to produce population-level reductions in antisocial behavior and other outcomes in future studies with access to other publicly available data sets.

References

Bradshaw, C. P., Zmuda, J. H., Kellam, S. G., & Ialongo, N. S. (2009). Longitudinal impact of two universal preventive interventions in first grade on educational outcomes in high school. *Journal of Educational Psychology. 101(4).* 926-937.

Brener, N. D., Krug, E. G., Dahlberg, L. L., & Powell, K. E. (1997). Nurses' logs as an evaluation tool for school-based violence prevention programs. *Journal of School Health. 67(5).* 171-4.

Domitrovich, C. E., Pas, E. T., Bradshaw, C. P., Becker, K. D., Keperling, J. P., Embry, D. D., & Ialongo, N. (2015). Individual and school organizational factors that influence implementation of the PAX Good Behavior Game intervention. *Prevention Science.*

Domitrovich, C.E., et al., *How Do School-Based Prevention Programs Impact Teachers? Findings from a Randomized Trial of an Integrated Classroom Management and Social-Emotional Program.* Prev Sci, 2016. *17*(3): p. 325-37.

Embry, D. D. (2002). The Good Behavior Game: A best practice candidate as a universal behavioral vaccine. *Clinical Child & Family Psychology Review. 5*(4). 273-297.

Embry, D. D., Flannery, D. J., Vazsonyi, A. T., Powell, K. E., & Atha, H. (1996). Peacebuilders: a theoretically driven, school-based model for early violence prevention. *American Journal of Preventative Medicine. 12(5).* 91-100.

Huber, M., Fruth, J.D., Avila-John, A. & Ramirez, E. (2016). Teacher self-efficacy and student outcomes: A transactional approach to prevention. *Journal of Education and Human Development*. 5(1).

Ialongo, N., et al., (1999). *Proximal impact of two first-grade preventive interventions on the early risk behaviors for later substance abuse, depression, and antisocial behavior.* American Journal of Community Psychology, *27*(5): p. 599-641.

Jiang, D., et al., (2018). *A Comparison of Variable- and Person-Oriented Approaches in Evaluating a Universal Preventive Intervention.* Prev Sci.

Kellam, S. G., Hendricks Brown, C., Poduska, J., Ialongo, N., Wang, W., Toyinbo, P., Petras, H., Ford, C., Windham, A., & Wilcox, H. (2008). Effects of a universal classroom behavior management program in first and second grades on young adult behavioral, psychiatric, and social outcomes. *Drug and Alcohol Dependence. 95(1)* S5-S28.

Kellam, S.G., Ling, X., Merisca, R., Brown, C.H., & Ialongo, N. (1998). The effect of the level of aggression in the first grade classroom on the course and malleability of aggressive behavior into middle school. *Development and Psychopathology*, *10*(02), 165-185.

Kellam, S. G., Rebok, G. W., Ialongo, N., & Mayer, L. S. (1994). The course and malleability of aggressive behavior from early first grade into middle school: Results of a developmental epidemiologically-based preventive trial. *Journal of Child Psychology and Psychiatry, 35*(2), 259-281.

Kellam, S. G., Reid, J., & Balster, R. L. (2008). Effects of a universal classroom behavior program in first and second grades on young adult outcomes. *Drug and Alcohol Dependence. 95(0 1).* S1-S4.

Kellam, S. G., Wang, W., Mackenzie, A. C., Brown, C. H., Ompad, D. C., Or, F., Ialongo, N. S.,

Poduska, J. M., & Windham, A. (2014). The impact of the Good Behavior Game, a universal classroom-based preventive intervention in first and second grades, on high-risk sexual behaviors and drug abuse and dependence disorders into young adulthood. *Preventive Science. 15(1).* S6-S18.

Musci, R. J., Bradshaw, C. P., Maher, B., Uhl, G. R., Kellam, S. G., & Ialongo, N. S. (2014).

Reducing aggression and impulsivity through school-based prevention programs: a gene by intervention interaction. *Prevention Science. 15(6).* 831-840.

Petras, H., Masyn, K., & Ialongo, N. (2011). The developmental impact of two first grade preventive interventions on aggressive/disruptive behavior in childhood and adolescence: an application of latent transition growth mixture modeling. *Preventive Science. 12(3).* 300-313.

Poduska, J. M., Kellam, S. G., Wang, W., Brown, C. H., Ialongo, N. S., & Toyinbo, P. (2008). Impact of the Good Behavior Game, a universal classroom-based behavior intervention, on young adult service use for problems with emotions, behavior, or drugs or alcohol. *Drug and Alcohol Dependence*, *95*, S29-S44.

Smith, E. P., Osgood, D. W., Caldwell, L., Hynes, K., & Perkins, D. F. (2013). Measuring collective efficacy among children in community-based afterschool programs: exploring pathways toward prevention and positive youth development. *American Journal of Community Psychology*. *52(1-2).* 7-40.

U.S. Department of Education, Office of Civil Rights, Civil Rights Data Collection (2018). Retrieved from: https://ocrdata.ed.gov.

U.S. Census Bureau, Missouri Census Data Center (2016). ZCTA Master Data Set. (2018).Retrieved from: https://mcdc.missouri.edu/geography/ZCTAs-2010.html.

Wang, Y., Browne, D. C., Petras, H., Stuart, E. A., Wagner, F. A., Lambert, S. F., Kellam, S. G., & Ialongo, N. S. (2009). Depressed mood and the effect of two universal first grade preventive interventions on survival to the first tobacco cigarette smoked among urban youth. *Drug and Alcohol Dependence. 100(3).* 194-203.

Weis, R., Osborne, K., & Dean, E. (2015) Effectiveness of a Universal, Interdependent Group Contingency Program on Children's Academic Achievement: A Countywide Evaluation, *Journal of Applied School Psychology*. 31(3). 199-218.

Wilson, D., Hayes, S., Biglan, A., & Embry, D. (2014). Evolving the future: toward a science of intentional change. *Brain and Behavioral Sciences*. *37(4).* 395-416.