Investigating the Effect of School-Wide Positive Behavioral Interventions and Supports on Student Learning and Behavioral Problems in Elementary and Middle Schools

Ji Hoon Ryoo*

University of Virginia

Saahoon Hong, William M. Bart, and Jaehyun Shin

University of Minnesota

Catherine P. Bradshaw

University of Virginia

*Corresponding author: Ji Hoon Ryoo, Ph.D.
Address: 417 Emmet Street South,
215 Bavaro Hall,
Charlottesville, VA 22904-4265
Email: jr3gv@virginia.edu

Ryoo, J. H., Hong, S., Bart, W. M., Shin, J., & Bradshaw, C. P. (2018). Investigating the effect of school-wide positive behavioral interventions and supports on student learning and behavioral problems in elementary and middle schools. *Psychology in the Schools, 55*(6), 629–643. https://doi.org/10.1002/pits.22134

Published in *Psychology in the Schools*

Funding: The research reported here was supported by the Institute of Education Science, U.S. Department of Education, through Grant R305H150027 to Catherine Bradshaw. The opinions expressed are those of the authors and do not represent views of the funders.

Abstract

Many states have scaled up School-Wide Positive Behavioral Interventions and Supports (SW-PBIS) with the goal of improving student behavior and academic outcomes. Although the effects of SW-PBIS on behavioral and discipline outcomes have been promising, the findings for academic achievement have been inconclusive and are often limited to cross-sectional data. This paper examined the longitudinal effect of SW-PBIS on student behavioral problems and academic achievement growth in elementary and middle schools using propensity score matching. We found no statistically significant longitudinal effects of SW-PBIS on either student behavior problems or academic achievement among elementary and middle schools in this quasi-experimental design, even though the positive direction of longitudinal changes would identified. The result raised questions about the efficacy of SW-PBIS utilizing a statewide achievement test when scaled-up to a state-level. Additional research is needed to better understand the impact of SW-PBIS using rigorous scale-up designs.

Keywords: Propensity Score Matching, School-Wide Positive Behavioral Interventions and Supports, Academic Performance

Investigating the Effect of School-Wide Positive Behavioral Interventions and Supports on Student Learning and Behavioral Problems in Elementary and Middle Schools

In the era of accountability, evidence-based and data-driven practices are prominent in school improvement efforts. As school populations become increasingly diverse, social and academic competencies become important concerns in education. To help students to achieve certain levels of social and academic competence, many states have adopted Positive Behavioral Interventions and Supports (PBIS) to improve student outcomes and school climate, and to better serve students who are at risk for academic failure and dropout/expulsion (Norton, 2009). PBIS is defined as "a framework or approach comprised of intervention practices and organizational systems for establishing the social culture, learning and teaching environment, and individual behavior supports needed to achieve academic and social success for all students" (Horner & Sugai, 2010, p. 13). PBIS provides schools with procedures to utilize data and build capacity for improving student behaviors and student learning.

Over 25,000 schools across the U.S. have adopted the universal elements of PBIS, referred to as School-Wide Positive Behavioral Interventions and Supports (SW-PBIS). Specifically, SW-PBIS is a whole-school application of PBIS which aims to enhance student social competency and learning through improved data collection and use systems to improve school functioning and the implementation of evidence-based practices (Sugai, 2007). The current paper explored the behavioral and academic impacts of SW-PBIS using a quasi-experimental design within the state of Minnesota, with the goal of understanding the potential effectiveness of the model when brought to scale state-wide.

Background and Rationale for SW-PBIS

SW-PBIS involves the integration of value-added outcomes, data-based decision making, human ecology and behavioral science, scientific evidence-based practices, and school system changes (Anderson & Kincaid, 2005; Carr, et al., 2002; Carr, et al., 1999; Lafrance, 2010, Lewis & Sugai, 1999; Safran & Oswald, 2003; Sugai, Horner, & Gresham, 2001; Sugai et al., 2000; Utley & Sailor, 2002). Guiding the delivery of SW-PBIS is a model that promotes positive student behavior and social competency (Sugai et al., 2000). In the model, all students are taught school expectations and positive behaviors that are clearly defined. School staff members work together to identify and support students at risk in a preventive continuum of behavioral intervention support.

SW-PBIS provides schools with systemic support for developing positive behaviors and decreasing challenging behaviors that interfere with student learning (Welsh, Parke, Widaman, & O'Neil, 2001). Emerging research has suggested that one potent factor that accounts for decrements in academic test scores is student problem behavior, which results in lost instructional time (Lassen, Steele, & Sailor, 2006; Scott, Nelson & Liaupsin, 2001). It is theorized that implementing SW-PBIS may translate into academic achievements for students by reducing the rates of behavioral problems in the classroom that might lead to decreasing opportunities for learning (Scott & Barrett, 2004). Support for this hypothesis comes from studies showing reductions in the numbers of office discipline referrals (ODRs), improved prosocial behaviors and peer relationships, and increased instructional time, academic achievement, and satisfaction about school safety (Lafrance, 2010). For example, it is estimated that students lose approximately 20 minutes of instructional time per ODR (Horner & Sugai, 2003). In this regard, a school could save approximately 659 instructional hours per year when

SW-PBIS is implemented. As a result of the reduced time spent on discipline issues, school staff members are better able to respond efficiently to student needs.

Prior research has demonstrated equivocal findings in that studies of the effect of SW-PBIS on student achievement and behavior based on statewide-standardized test scores at different grade levels and office discipline referral and out-of-school suspension data have revealed mixed results (e.g., Bradshaw, Mitchell, & Leaf, 2010; Horner et al., 2009). For example, Sailor and colleagues (2006) concluded that three middle schools and one K-3 charter school implementing SW-PBIS in a low-income, multicultural, and urban school district in North Carolina showed improved school-level academic achievement. Lassen and colleagues (2006) examined three-year school discipline data in an inner-city middle school and identified the positive relationships between SW-PBIS with reductions in office discipline referrals and student academic achievement in reading and math.

In contrast, New Hampshire's two-year study reported a positive relationship between SW-PBIS and student achievement in mathematics but not in reading/language (Muscott, Mann, & LeBrun, 2008). A Maryland-based randomized controlled effectiveness trial reported no significant impact of SW-PBIS on academic achievement when examining the effect of SW-PBIS on academic performance among 37 elementary schools (21 SW-PBIS schools and 16 non SW-PBIS schools); however, they did report significant effects on suspensions and a significant reduction in office discipline referrals (Bradshaw, Mitchell, & Leaf, 2010). It is possible that simply focusing on social behavior may not translate into improved academic outcomes unless coupled with an enhanced academic curriculum or program (Bradshaw et al., 2010; Horner et al., 2009).

Taken together, the effectiveness of SW-PBIS on student academic achievement and behavior problems has been inconclusive and often limited to cross-sectional data; thus, additional research is needed to examine the association between SW-PBIS and both academic and behavioral outcomes when scaled-up state-wide. The current study aimed to address these gaps by investigating the longitudinal outcomes associated with SW-PBIS using a quasi-experimental design. Specifically, we examined standardized test scores for mathematics and reading, as well as out-of-school suspensions when SW-PBIS was scaled-up across elementary and middle schools in a single state.

Method

Data from this study comes from the state of Minnesota, which launched a large-scale SW-PBIS effort in 2005. In Minnesota, SW-PBIS schools receive training over a two-year period. The Minnesota SW-PBIS state leadership team provided all of the training and technical assistance with the goal of supporting high fidelity of implementation of SW-PBIS across the state. This study examined the longitudinal effects of Minnesota SW-PBIS on school-level student achievement as measured by Minnesota comprehensive assessment–series II (MCA-II) test scores from the 2007-08 school year to the 2009-10 school year, with the MCA-II being replaced by the MCA-III in 2011. A total of 585 schools in 183 districts had been trained and will be in the model by 2018 (MN PBIS, 2016).

For the current study, we examined data across two cohorts over three years: cohort one included elementary schools starting with the 3rd grade in the Fall of 2007. Cohort two included middle schools starting with the 5th grade in the Fall of 2007. At the baseline year (2007-08), we used school-level data including that of 2006-07 MCA-II achievement scores at all grades as well as that of 2007-08 school demographic and school characteristics to reduce the selection bias of schools. We collected MCA-II achievement scores for the cohort groups from the 2007-08 to 2009-10 school years. We also collected school-level in-school and out-of-school suspension rate data for the 2009-10 and 2011-12 school years. Different from the achievement data, suspension rate data are school-level data including the cohort groups and thus, we compared the suspension rate data beyond the study period from 2007-08 to 2009-10.

Implementation Fidelity Measures

The *School-Wide Evaluation Tool* (SET; Sugai, Lewis-Palmer, Todd, & Horner, 2001) is a research tool designed to assess and evaluate the critical features of SW-PBIS. The SET consists of 29 items organized into the following seven subscales that represent the seven key features of SW-PBIS: (1) Expectations Defined, (2) Behavioral Expectations Taught, (3) System for Rewarding Behavioral Expectations, (4) System for Responding to Behavioral Violations, (5) Monitoring and Evaluation, (6) Management, and (7) District-Level Support. The total scale and the subscales are reported with a percentage ranging from 0% to 100% with the higher scores indicating greater fidelity of SW-PBIS. Based on the prior research conducted by Horner et al. (2004), the SET has strong psychometric properties with respect to internal consistency (0.96), inter-observer reliability varying 98.4% to 100%, test-retest reliability varying 93% to 100%, and construct validity (the SET correlated with the Effective Behavior Support: Self-Assessment Survey, $r = 0.75$, $p \leq 0.01$).

In MN SW-PBIS, an independent SET evaluator completed the SET during a two and one half hour site visit to a SW-PBIS school; the assessor observed classrooms and common areas and conducted informal interviews with staff and students. The Minnesota Department of Education recommends that all SW-PBIS schools complete the SET in the Fall (September or October) during the first year of training, and annually in the Spring (April or May) thereafter, until the school achieves full implementation. Full implementation is defined as having a score of 80 or better overall and 80 or better in the domain of Behavioral Expectation Taught, or “80/80”. In this quasi-experimental design, it was not possible to track retrospectively the SET scores of two cohort groups, but the MN PBIS team reported that the average SET score for SW-PBIS schools starting implementation between 2005 and 2009 was 84 (MN PBIS, 2012).

The *Benchmarks of Quality* (BoQ; Kincaid, Childs, & George, 2005) is a self-assessment tool that schools use to assess implementation fidelity for tier 1 of SW-PBIS. The BoQ consists of 53 Benchmarks of Quality that addresses 10 critical elements: (1) PBIS team, (2) Faculty Commitment, (3) Effective Procedures for Dealing with Discipline, (4) Data Entry and Analysis Established, (5) Expectations and Rules Developed, (6) Reward/Recognition Program Established, (7) Lesson Plans for Teaching Expectations/Rules, (8) Implementation Plan, (9) Crisis Plan, and (10) Evaluation. Cohen, Kincaid, and Childs (2007) reported the following psychometric results for the BoQ: internal consistency (0.96), test-retest reliability (0.94), inter-rater reliability (0.87), and concurrent validity with the SET ($r = 0.51$, $p < 0.05$).

In Minnesota, schools that have reached a score of 80 or above on the SET and have completed the 2-year training sequence are eligible to complete the BoQ every year for two years and only complete a SET every third year. Therefore, even being eligible to complete the BoQ is an indication of implementation fidelity. Schools receiving an overall score of 70% or above met the first criterion in the application process to be recognized as a sustaining exemplar Minnesota PBIS school. Approximately 92% of SW-PBIS schools starting implementation between 2005 and 2009 that qualified to complete the BoQ scored 70 or above; whereas, the remaining 8% scored below 70% (MN PBIS, 2012).

Research Design

We employed a quasi-experimental design by selecting the control group for each SW-PBIS cohort group using propensity score matching method (PSMM; Rosenbaum & Rubin, 1983; Rubin, 2007). Before selecting the control groups in PSMM, multiple imputation (MI) was used to avoid possible threats to valid statistical inference due to missing values on the covariates. Using SPSS 24 (IBM Corp.) with fully conditional specification (MCMC) of 1000 maximum iterations, five imputed datasets were generated. The outcomes of interest for this study were MCA-II test data and suspension rate data for SW-PBIS trained elementary and middle schools and the propensity score matched non-SW-PBIS schools.

Propensity Score Matching Method (PSMM)

The annual application process by the Minnesota Department of Education (MDE) permitted the selection of SW-PBIS schools and all schools participated in MCA-II testing. The numbers of schools implementing SW-PBIS were relatively small as of the Fall of 2007, i.e., 16 out of 950 elementary schools and 17 out of 896 middle schools, which shows that the data are unbalanced. Though such a selection process may cause a selection bias, it is a common procedure in educational programs. However, such selection bias may result in invalid statistical inference and prevent us from drawing any causal inference from the controlled study. To reduce the selection bias and achieve a quasi-experimental design from this observational and unbalanced data, the study employed the propensity score matching method (Guo & Fraser, 2015).

Specifically, the propensity scores were computed such that if two schools had the same propensity scores, their covariates come from same distribution. Thus, this procedure enabled us to reduce potential bias that may result from imbalanced covariate distributions across the two groups, which tends to produce unbiased estimates of the treatment effects (Mitra & Reiter, 2012). Based on the recommendations by Rosenbaum and Rubin (1983), we applied optimal matching (Gu & Rosenbaum, 1993; Rosenbaum, 1989) using the MatchIt package in R (Ho, Imai, King, & Stuart, 2011); this approach is similar to nearest neighbor matching using a greedy algorithm involving propensity scores obtained by logistic regression. Another benefit of optimal matching is that it results in matched samples that minimize a global distance measure based on the propensity score. Gu and Rosenbaum (1993) have also provided evidence of the efficiency and validity of this approach for creating matched samples.

We selected a matched sample using variables representing the following seven school characteristics: (1) the numbers of students taking the MCA-II, (2) ethnicity, (3) gender, (4) students eligible for special education service, (5) limited English proficiency service, (6) free/reduced priced meals, and (7) migrant services. The collection of MCA-II test data began in the year, 2007-08, including 2006-07 MCA-II scores for the 3rd through 5th grades in the elementary school cohort and 2006-07 MCA-II scores for the 6th through 8th grades in the middle school cohort. Use of the PSMM resulted in the formation of two cohort samples: one consisting of 32 elementary schools (3.37% of the total of 950 elementary schools; 16 SW-PBIS schools and 16 non SW-PBIS schools) and the other consisting of 34 middle schools (3.79% of the total of 896 middle schools; 17 SW-PBIS schools and 17 non SW-PBIS schools).

There was an improvement of the variable *Distance*, a global distance from propensity scores, from 68.84% to 77.47% for the five imputed datasets in the elementary cohort and from 54.42% to 64.15% for the five imputed datasets in the middle cohort. Out of the 28 covariates used in PSMM, the number of improved covariates ranged from 21 to 25 for the elementary cohort and from 18 to 26 for the middle cohort.

Figures 1 and 2 provide comparisons of the distributions of propensity scores before and after matching on the first imputed dataset for each cohort group. Points under *Matched Treatment Units* and *Matched Control Units* are data selected by PSMM. Regardless of the distributions of the raw data, each figure shows that the distribution of the matched sample is fairly similar to that of the treatment sample.

Insert Figures 1 and 2 here.

The improvements of covariates of the first imputed dataset for each elementary and middle school cohorts were summarized in the supplementary materials (Tables 1 and 2, respectively). In elementary schools, 25 of the 28 covariates (i.e., all except the numbers of students in two ethnicity covariates, Asia/Pacific Islander and Hispanic, and Limited English Proficiency) were associated with improvements. In middle schools, 26 of the 28 covariates (i.e., all except the numbers of students in Limited English Proficiency and Migrant service) were associated with improvements. The improvement on the selection bias of each covariate was summarized in the supplementary Table 1 for elementary schools and the supplementary Table 2 for middle schools.

Selected Samples and Variables

Using the propensity score matching method, we analyzed five imputed datasets for 32 elementary schools from 950 elementary schools and five imputed datasets for 34 middle schools from 896 middle schools. School achievement scores for three years, including two training years were analyzed. The outcomes of interest included ***mathematics and reading school average scores*** for the MCA-II collected in 2008, 2009, and 2010. MDE annually reports school-level aggregated scores as long as the minimum of test-takers is greater than nine. This study used scale scores ranging from 0 to 99. These scale scores are statistical conversions of raw scores to maintain a consistent metric across test forms and to permit comparison of scores across all test administrations within a particular grade and subject.

Schools can also use scale scores to compare the knowledge and skills of groups of students within a grade and subject across years. These comparisons permit the assessment of the impact of changes or differences in instruction or curriculum, which is consistent with the focus of this study. We also obtained ***in-school suspension and out-of-school suspension*** data from the U.S. Department of Education's Office of Civil Rights data collection (http://ocrdata.ed.gov/).

The key predictor variable is the status of implementation of SW-PBIS. All other school variables served as covariates and included the numbers of students taking MCA-II tests, ethnicity, gender, special education, limited English proficiency, free/reduced priced meals, and mobility. Tables 1 and 2 provide the means and standard deviations of matched samples conditional on SW-PBIS and overall means of matched samples. The mean growth curve patterns in the matched samples are plotted in Figures 3 - 6.

Insert Tables 1 and 2 & Figures 3 to 6 here

Statistical Analyses

Five imputed datasets of each cohort, obtained by PSMM, consist of multiple responses, including mathematics and reading test scores over 3 years and suspension rate data at 2009-10 and 2011-12. The track of the two cohorts was only possible for three years due to the transition from elementary to middle or from middle to high schools. As a result, it was not possible to track academic achievement for more than three years.

The analysis of each subject proceeded separately in each cohort group using the linear mixed model (LMM; Laird & Ware, 1982). The selection of the best-fitted models resulted from an examination of the polynomial functions using a step-up approach modeling from the simplest functional form to the complex functional form, using the likelihood ratio test and fit indices (Ryoo, 2011). The model selection of cohort achievement indicated at most linear growth with the use of polynomial functions. We examined the best-fitted models up to a linear growth model with a linear random effect allowing individual school's influence on both initial achievement and growth pattern.

All of the best-fitted models over subjects, mathematics and reading, and cohort groups, elementary and middle, took the form of the following linear model with linear random effects:

$$y_{ij} = (\beta_0 + b_{0i}) + \beta_1 \cdot swpbis_i + (\beta_2 + b_{1i}) \cdot grade_{ij} + \beta_3 \cdot grade_{ij} \cdot swpbis_i + \varepsilon_{ij} \qquad (1)$$

where $i = 1,2,\cdots,N$ represents subjects and $j = 1,2,\cdots,n_i$ represents time points. β_k, $k = 0,1,2,3$ indicate the regression coefficients of group mean changes explaining the within-subject variance while b_{0i} and b_{1i} indicate the coefficients for random effects explaining the between-subject variance. All analyses made use of the `lme4` package in the R program (Bates, Mächler, Bolker, & Walker, 2015).

For the effect of SW-PBIS on the in- and out-of-school suspension data, we ran a series of Mann-Whitney tests (Mann & Whitney, 1947) as a nonparametric alternative to a two-sample *t*-test with five imputed datasets for each subject (math and reading) and grade level (elementary and middle). It should be noted that the suspension rate data serve as a proxy of student behavioral data for these cohorts, because the data are school-level data. Therefore, we examined the long-term school level effect of SW-PBIS on student behavior by considering the in- and out-of-school suspension rate data of 2011-12 in addition to that of 2009-10. Mann-Whitney test is a best fitting test because the suspension rate data is ordinal or continuous as a percentage, independently observed, and not normally distributed although its population distributions of SW-PBIS and non SW-PBIS can be assumed to be same. All Mann-Whitney tests were conducted using SPSS 24 (IBM Corp.).

Results

Model Evaluation

By considering the R^2 statistic, indicating overall model fit, for the LMM (Vonesh and Chinchilli, 1997) defined by $R^2 = \left[cor\left(y_{ij}, \hat{y}_{ij}\right)\right]^2$, we examined how reliable the result of $\bar{\beta}_3$ is in terms of model selection. The R^2 in the LMM can be interpreted as the variance explained by the fitted model (Long, 2012) similar as in regression analysis. In elementary schools, the averages of five R^2s were 0.91 for both math and reading, which means that 91% of the response variance was accounted by the model. In middle schools, the averages of five R^2s were 0.96 for both math and reading, which means that 96% of the response variance is accounted by the model. Overall, the $R^2 \geq 0.50$ was large, which indicates that the overall model fit was good.

Effect of SW-PBIS on Student Academic Outcomes

Since we used multiple imputation in the propensity score matching method, the results came from data analyses performed five times that is the number of multiple imputation datasets

(Rubin, 1987). β_3 s in Equation (1) are the parameters of interest that indicate the effects of SW-PBIS with their standard errors $(s.e.)$ being estimated. The overall estimate ($\overline{\beta}_3$) is obtained from the average of the $\hat{\beta}_3$ s, the overall standard error ($\overline{s.e.}$) is obtained from the average of the $(s.e.)$s, and the between-imputation variance ($\overline{B}$) is obtained from $\frac{1}{5-1}\sum_{j=1}^{5}\left(\hat{\beta}_{j3}-\overline{\beta}_3\right)^2$.

The significance test of the null hypothesis $\overline{\beta}_3 = 0$ involves a comparison of the t-test statistics whose distribution is Student's t-distribution with the total variance being

$$T = \overline{s.e.} + \left(1 + \frac{1}{5}\right)\overline{B}$$

and degrees of freedom being

$$df = (5-1) \times \left(1 + \frac{5 \cdot \overline{s.e.}}{(5+1)\overline{B}}\right)^2$$

Based on the t-test derived from the above, we found that the elementary school cohort equation has a negative regression coefficient, $\overline{\beta}_3 = -0.83$, for the subject of Mathematics but a positive regression coefficient, $\overline{\beta}_3 = 0.91$, for the subject of reading. On the other hand, the middle school cohort equation had a positive regression coefficient, $\overline{\beta}_3 = 0.24$ for Mathematics but a negative regression coefficient, $\overline{\beta}_3 = -0.17$, for Reading. However, none of those coefficients were statistically significant. The results are summarized in Table 3.

Insert Table 3 here

Effect of SW-PBIS on Student Behavior

Mann-Whitney tests indicated that the changes of suspension rates of SW-PBIS cohort group were not significantly different from those of the comparison group. As seen in Table 4, five imputed datasets for each subject (math and reading) over two grade levels (elementary and middle) were analyzed and Mann-Whitney test results; these analyses did not show any significant positive effects on the school level suspension rate. The analyses indicated that the suspension rates among SW-PBIS schools were higher at 2009-10, and although the gap of suspension rates between SW-PBIS and non SW-PBIS narrowed, the effect was not statistically significant. That is, the initially higher suspension rates of SW-PBIS schools were similar to those of non SW-PBIS schools by 2011-12.

Insert Table 4 here

Discussion

The main purpose of this study was to use a quasi-experimental design to estimate the impacts of SW-PBIS on statewide standardized test scores and suspension rates in a scale-up effort within the state of Minnesota. After creating well-matched samples using propensity scores, we found no significant effect of SW-PBIS on academic achievement over multiple years. Although elementary SW-PBIS schools were more likely to experience decreasing achievement than middle schools, those results were not statistically significant. Taken together, these findings suggested that there were no significant effects of SW-PBIS on academic performance or suspension rates in Minnesota following the scale-up effort. These findings add to the body of mixed findings regarding the effects of SW-PBIS on student outcomes (e.g., Bradshaw et al., 2010; Horner et al., 2009; Welsh et al., 2001).

There are several limitations to this study that affect the generalizability of its findings. First, the sample data were from one cohort group starting in the Fall of 2007 that was drawn from five available cohort groups starting with either the 3rd or the 5th grade of the Fall of the years, 2005, 2006, 2007, 2008, and 2009. This sampling method resulted in different school characteristics across years and with not all SW-PBIS schools being selected for the cohort group. This may prevent us from generalizing the findings to all SW-PBIS schools.

Schools in comprehensive school reform efforts demonstrated strong effects across different poverty levels, when implemented for five years or more (Borman, Hewes, Overman, & Brown, 2003). A data collection period of longer duration may be necessary to detect longitudinal changes on school improvement. On the other hand, our findings are based on longitudinal data analysis over three years, to avoid discontinuity of cohort groups due to transition from elementary to middle or from middle to high schools as well as to avoid discontinuity of MCA II after 2010. Prior research on SW-PBIS has rarely examined longitudinal effects on both academic and behavioral outcomes (cf. Bradshaw et al., 2010).

Second, this study used a statewide standardized test that was aligned to Minnesota academic standards. As the number of schools facing repeated failure in the state-based accountability system increased, these schools require more comprehensive school reform efforts, not limited to SW-PBIS. Efforts to implement SW-PBIS in schools could benefit from SW-PBIS aligned with other school reform initiatives. Such initiatives include Curriculum-Based Measurement for Progress Monitoring, Response to Intervention, teacher mentoring program, Reading First, the First Move scholastic chess program, and other programs for enhancing academic performance.

Third, our selection of optimal matching may not provide the best-selected sample from the PSMM, since there was no one dominating option in the PSMM. Instead of relying on the selection of the matching methods, we applied multiple imputation to obtain more reliable results. But the imputation process may not be the best way of reducing the selection bias of imbalanced data (Mitra & Reiter, 2012). However, it is not yet known if other methods would provide a better result so that we selected the most common and stable way of reporting the result in the study (Rubin, 1987).

Fourth, in the case of multiple dependent variables, a frequent suggestion is to consider the joint models for longitudinal data (Fitzmaurice, Davidian, Verbeke, & Molenburghs, 2008). One such model is the multivariate linear mixed model, but analysis of the test data with the program SabreR in R indicated that the multivariate linear mixed model did not fit the data due to the high correlation between mathematics and reading outcomes (Berridge & Crouchley, 2011; Long, 2012). Thus, we fit the LMM for two separate datasets for mathematics and reading.

The unit of analysis for this study was school-level achievement, which might provide a narrow view of the effects of SW-PBIS on student learning. In general, student achievement levels and their areas of social competence vary. Students could benefit differently from a SW-PBIS school.

An alternative statistical approach could include a multilevel perspective with students nested within schools. The multilevel perspective may help us to examine mediator-moderator components as well. These components are critical in identifying the strength and direction of the linkages between student characteristics at the individual level and school climates at the school level (Somech, 2010). However, that was not feasible in this study due to small sample size. Finally, SW-PBIS may not have been implemented as it was intended. Thus, inquiry is needed to identify and address the barriers and challenges that schools face that can affect the implementation of SW-PBIS with fidelity.

Despite these limitations, our study suggests that MN SW-PBIS did not result in significant impacts on academic and behavior outcomes over time. The characteristics of SW-PBIS schools and their resources should be revisited to determine the extent to which SW-PBIS is aligned with school efforts to improve student behaviors among students in needs. In addition, the results lead us to reflect on the main characteristics of SW-PBIS and their goals. SW-PBIS was initially intended to improve behavioral outcomes, with the goal of increasing opportunities for learning. As a result, without significant impacts on behavioral outcomes, SW-PBIS had no effect on improvement on academic performance.

Conclusions and Implications

Over the last decades, schools have become increasingly diverse in socioeconomic status, ethnicity, religion, culture, and language. In a diverse society, SW-PBIS was a systematic approach to establish a school-wide culture for the social competence. This study sheds light on the linkage between student behaviors and academic achievement growth in SW-PBIS.

SW-PBIS itself may not be enough to demonstrate student success in academic without positive changes in behavioral competence. There is a growing emphasis on implementing school-wide systematic approaches to improve student academic performance. In this manner, the results raised questions about the efficacy of SW-PBIS when scaled-up, as it may be more advantageous to layer on more intensive interventions within the school-wide framework.

Public schools, educators, and policy makers should support the incorporation of data-driven practices into SW-PBIS. Research on the academic effects of such integrated models would help us to better understand the interaction between academic and behavioral excellence in student learning. SW-PBIS may engender greater educational benefits in schools if it is integrated with other approaches intended to promote student learning.

References

Anderson, C.M., & Kincaid, D. (2005). Applying behavior analysis to school violence and discipline problems: Schoolwide positive behavior support. *Behavior Analyst, 28*, 49-63.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using {lme4}. *Journal of Statistical Software, 67*, 1-48. doi:10.18637/jss.v067.i01

Berridge, D. M., & Crouchley, R. (2011). *Multivariate generalized linear mixed models using R*. Boca Raton, FL: Chapman & Hall/CRC.

Borman, G., Hewes G., Overman, L., & Brown, S. (2003). Comprehensive school reform and achievement: A meta-analysis. *Review of Educational Research, 73*(2), 125-230.

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*(3), 133-148.

Carr, E. G., Dunlap, G., Horner, R. H., Koegel, R. L., Turnbull, A. P., & Sailor, W. (2002). Positive behavior support: Evolution of an applied science. *Journal of Positive Behavior Interventions, 4*, 4-16.

Carr, E. G., Horner, R. H., Turnbull, A. P., Marquis, J. G., Magito-McLaughlin, D., McAtee, M. L., et al. (1999). *Positive behavior support as an approach for dealing with problem behavior in people with developmental disabilities: A research synthesis* (American Association on Mental Retardation Monograph Series. No. 3). Washington, DC: American Association on Mental Retardation.

Cohen, R., Kincaid, D., & Childs, K. E. (2007). Measuring school-wide positive behavior support implementation: Development and validation of the Benchmarks of Quality.

Journal of Positive Behavior Interventions, 9(4), 203-213.

Fitzmaurice, G., Davidian, M., Verbeke, G., & Molenberghs, G. (2008), *Longitudinal data analysis*. Boca Raton, FL: Chapman & Hall/CRC.

Gu, X., & Rosenbaum, P. R. (1993). Comparison of multivariate matching methods: structures, distances, and algorithms. *Journal of Computational and Graphical Statistics, 2*, 405-420.

Heckman, J. J., Ichimura, H., & Todd, P. E. (1997). Matching as an economic evaluation estimator: Evidence from evaluating a job training program. *Review of Economic Studies, 64*, 605-654.

Heckman, J. J., Ichimura, H., & Todd, P. E. (1998). Matching as an econometric evaluation estimator. *Review of Economic Studies, 65*, 261-294.

Ho, D. E., Imai, K., King, G., & Stuart, E. A. (Forthcoming), "MatchIt: Nonparametric Preprocessing for Parametric Causal Inference," *Journal of Statistical Software*, http://gking.harvard.edu/matchit.

Horner, R., & Sugai, G. (2003, July). PBS lessons learned: Scaling up and progress indicators. Paper presented at the PBIS Leadership Conference, Naperville, IL.

Horner, R., & Sugai, G. (2010). Implementation Blueprint and Self-Assessment: Positive Behavioral Interventions and Supports. *Center on Positive Behavioral Interventions and Supports, Office of Special Education Programs, US Department of Education*.

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*(3), 133-144.

Horner, R., Todd, A., Lewis-Palmer, T., Irvin, L. K., Sugai, G., & Boland, J. (2004). The

School-wide Evaluation Tool (SET): A research instrument for assessing school-wide positive behavior supports. *Journal of Positive Behavior Interventions, 6*, 3-12.

IBM Corp. Released 2016. IBM SPSS Statistics for Windows, Version 24.0. Armonk, NY: IBM Corp.

Kincaid, D., Childs, K., & George, H. (2005). *School-wide benchmarks of quality*. Unpublished instrument, University of South Florida, Tampa, Florida.

Lafrance, J. A. (2010). *Examination of the fidelity of school-wide positive behavior support implementation and its relationship to academic and behavioral outcomes in Florida.* Retrieved from http://gateway.proquest.com/openurl%3furl_ver=Z39.88-2004%26res_dat=xri:pqdiss%26rft_val_fmt=info:ofi/fmt:kev:mtx:dissertation%26rft_dat=xri:pqdiss:3383666 (Dissertation Abstract: 2010-99090-466)

Laird, N. M., & Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics, 38*(4), 963-974.

Lassen, S. R., Steele, M. M., & Sailor, W. (2006). The relationship of school-wide positive behavior support to academic achievement in an urban middle school. *Psychology in the schools, 43*(6), 701-712.

Lewis, T. J., & Sugai, G. (1999). Effective behavior support: A systems approach to proactive schoolwide management. *Focus on Exceptional Children, 31*(6), 1-24.

Long, J. D. (2011). *Longitudinal Data Analysis for the Behavioral Sciences Using R*. Thousand Oaks, CA: SAGE Publications

Luiselli, J. K., Putnam, R. F., Handler, M. W., & Feinberg, A. (2005). Whole-school positive behavior support: Effects on student discipline problems and academic performance. *Educational Psychology, 25,* 183–-198.

Mann, H. B., & Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *Annals of Mathematical Statistics*, *18*, 50-60.

Mitra, R. & Reiter, J. P. (2012). A comparison of two methods of estimating propensity scores after multiple imputation. *Statistical Methods in Medical Research*. Advance online publication. doi:10.1177/0962280212445945

MN PBIS (2011). *Minnesota Positive Behavioral Interventions and Supports*. Retrieved August 22, 2011, from http://www.pbismn.org/documents/MNSW_PBIS_Schools.pdf

MN PBIS (2016). Minnesota SW-PBIS Schools: Cohorts 1-12. Retrieved February 1, 2017, from http://www.pbismn.org/documents/MNSW_PBIS_Schools_2016_17.pdf

Muscott, H. S., Mann, E. L., &LeBrun, M. R. (2008). Positive behavioral interventions and supports in New Hampshire effects of large-scale implementation of schoolwide positive behavior support on student discipline and academic achievement. *Journal of Positive Behavior Interventions, 10*(3), 190-205.

Norton, L. C. (2009). The impact of positive behavior interventions & supports (PBIS) on student behavior and academic achievement (Order No. 3421365). Available from ProQuest Dissertations & Theses A&I. (751613282). Retrieved from http://login.ezproxy.lib.umn.edu/login?url=http://search.proquest.com.ezp1.lib.umn.edu/docview/751613282?accountid=14586

Reynolds, C. R., Skiba, R. J., Graham, S., Sheras, P., Conoley, J. C., & Garcia-Vazquez, E. (2008). Are zero tolerance policies effective in the schools? An evidentiary review and recommendations. *The American Psychologist, 63*(9), 852-862.

Rosenbaum, P. R. (1989). Optimal matching for observational studies. *Journal of the American Statistical Association, 84*(408), 1024-1032.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects, *Biometrika, 70*, 41-55.

Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. NY: John Wiley & Sons

Rubin, D. B. (2007). The design versus the analysis of observational studies for causal effects: Parallels with the design of randomized trial. *Statistics in Medicine, 26*, 20-36.

Ryoo, J. (2011). Model selection with the linear mixed model for longitudinal data. *Multivariate Behavioral Research, 46*(4), 598-624.

Safran, S. P., & Oswald, K. (2003). Positive behavior supports: Can schools reshape disciplinary practices? *Exceptional Children, 69*, 361-373.

Sailor, W., Zuna, N., Chol, J. H., Thomas, J., McCart, A., & Roger, B. (2006). Anchoring schoolwide positive behavior support in structural school reform. *Research and Practice for Persons with Severe Disabilities, 31*(1), 18-30.

Scott, T. M., & Barrett, S. B. (2004). Using staff and student time engaged in disciplinary procedures to evaluate the impact of school-wide PBS. *Journal of Positive Behavior Interventions, 6*(1), 21-27.

Scott, T. M., Nelson, C. M., & Liaupsin, C. J. (2001). Effective instruction: The forgotten component in preventing school violence. *Education and Treatment of Children, 24*(3), 309-322.

Simonsen, B., Eber, L., Black, A. C., Sugai, G., Lwandowski, H., Sims, B., & Myers, D. (2012). Illinois Statewide Positive Behavioral Interventions and Supports: Evolution and impact on student outcomes across years. *Journal of Positive Behavioral Interventions, 14*(1), 5-6.

Somech, A. (2010). Participative decision making in schools: A mediating-moderating analytical framework for understanding school and teacher outcomes. *Educational Administration, 46*(2), 174-209.

Sugai, G. (2007). Promoting behavioral success in schools: Commentary on exemplary practices. In McDougal, J., & Miller, D. *Special Practitioner's Edition: Promoting behavioral success*. *Psychology in the Schools, 44, 113-118.*

Sugai, G., Horner, R. H., Dunlap, G., Hieneman, M., Nelson, C. M., Scott, T., Liaupsin, C., Sailor, W., Turnbull, A. P., Turnbull, H. R., Wickham, D., Wilcox, B., & Ruef, M. (2000). Applying positive behavior support and functional behavioral assessment in schools. *Journal of Positive Behavior Interventions, 2*(3), 131-143.

Sugai, G., Horner, R. H., & Gresham, F. (2001). Behaviorally effective school environments. (pp. 315-350). In M. R. Shinn, G. Stoner, & H. M. Walker (Eds), *Interventions for academic and behavior problems: Preventive and remedial approaches.* National Association of School Psychologists. Silver Spring, MD.

Sugai, G., Lewis-Palmer, T., Todd, A., & Horner, R. H. (2001). School-wide evaluation tool. Eugene: University of Oregon.

Utley, C. A., & Sailor, W. (2002). Positive behavior support and urban school improvement: A Special Section of the *Journal of Positive Behavior Interventions* [Guest editorial]. *Journal of Positive Behavior Interventions, 4,* 195.

Vonesh, E. G., & Chinchilli, V. M. (1997). *Linear and nonlinear models for the analysis of repeated measurements*. New York: Marcel Dekker.

Welsh, M., Parke, R. D., Widaman, K., & O'Neil, R. (2001). Linkages between children's social and academic competence: A longitudinal analysis. *Journal of School Psychology, 39*(6), 463-482.

Table 1

Means and standard deviations of the elementary school test scores conditional on SW-PBIS

		Mathematics			Reading		
		2008	2009	2010	2008	2009	2010
Imputed Data - 1	SW-PBIS	52.52 (7.91)	48.78 (8.28)	45.41 (10.84)	52.36 (11.83)	48.82 (8.81)	50.62 (7.87)
	No SW-PBIS	52.67 (5.85)	49.42 (5.25)	47.95 (7.82)	56.01 (10.08)	50.02 (6.47)	51.41 (6.53)
	Total	52.59 (6.89)	49.07 (6.97)	46.60 (9.48)	54.13 (11.01)	49.36 (7.74)	50.99 (7.17)
Imputed Data – 2	SW-PBIS	52.52 (7.91)	48.78 (8.28)	45.41 (10.84)	52.36 (11.83)	48.82 (8.81)	50.62 (7.87)
	No SW-PBIS	52.67 (5.83)	49.66 (5.30)	46.31 (7.53)	55.11 (8.74)	49.94 (6.75)	51.93 (6.81)
	Total	52.60 (6.85)	49.21 (6.90)	45.83 (9.30)	53.73 (10.35)	49.38 (7.75)	51.23 (7.30)
Imputed Data - 3	SW-PBIS	52.52 (7.91)	48.78 (8.28)	45.41 (10.84)	52.36 (11.83)	48.82 (8.81)	50.62 (7.87)
	No SW-PBIS	50.48 (6.60)	48.21 (7.13)	44.92 (7.44)	52.04 (10.56)	48.34 (8.02)	49.71 (6.04)
	Total	51.53 (7.27)	48.49 (7.62)	45.18 (9.26)	52.20 (11.07)	48.58 (8.30)	50.19 (6.97)
Imputed Data – 4	SW-PBIS	52.52 (7.91)	48.78 (8.28)	45.41 (10.84)	52.36 (11.83)	48.82 (8.81)	50.62 (7.87)
	No SW-PBIS	53.97 (5.70)	51.45 (6.33)	49.85 (7.49)	56.04 (9.96)	51.46 (7.91)	54.07 (6.91)
	Total	53.25 (6.84)	50.11 (7.38)	47.50 (9.54)	54.20 (10.94)	50.14 (8.35)	52.24 (7.52)
Imputed Data - 5	SW-PBIS	52.52 (7.91)	48.78 (8.28)	45.41 (10.84)	52.36 (11.83)	48.82 (8.81)	50.62 (7.87)
	No SW-PBIS	52.48 (6.13)	48.50 (6.28)	47.23 (8.26)	53.45 (10.90)	48.64 (8.71)	50.96 (7.17)
	Total	52.50 (7.03)	48.65 (7.33)	46.23 (9.65)	52.87 (11.24)	48.74 (8.62)	50.77 (7.44)

Table 2

Means and standard deviations of the middle school test scores conditional on SW-PBIS

		Mathematics			Reading		
		2008	2009	2010	2008	2009	2010
Imputed Data - 1	SW-PBIS	49.59 (8.72)	49.38 (8.01)	48.63 (7.03)	51.17 (7.82)	51.09 (7.50)	52.40 (6.36)
	No SW-PBIS	50.40 (6.50)	47.67 (8.50)	45.33 (7.33)	52.40 (6.22)	50.72 (8.43)	51.87 (5.21)
	Total	50.00 (7.59)	48.69 (8.14)	47.40 (7.22)	51.80 (7.00)	50.94 (7.78)	52.20 (5.88)
Imputed Data – 2	SW-PBIS	49.59 (8.72)	49.38 (8.01)	48.63 (7.03)	51.17 (7.82)	51.09 (7.50)	52.40 (6.36)
	No SW-PBIS	48.24 (7.53)	46.81 (5.54)	46.73 (6.83)	50.38 (7.41)	50.52 (5.76)	53.86 (5.56)
	Total	48.90 (8.07)	48.29 (7.11)	47.86 (6.92)	50.77 (7.54)	51.27 (6.74)	52.99 (6.01)
Imputed Data - 3	SW-PBIS	49.59 (8.72)	49.38 (8.01)	48.63 (7.03)	51.17 (7.82)	51.09 (7.50)	52.40 (6.36)
	No SW-PBIS	50.62 (5.23)	49.45 (6.51)	48.25 (6.58)	53.09 (5.84)	51.65 (7.06)	53.21 (6.45)
	Total	50.09 (7.16)	49.41 (7.34)	48.48 (6.76)	52.11 (6.91)	51.32 (7.23)	52.72 (6.31)
Imputed Data – 4	SW-PBIS	49.59 (8.72)	49.38 (8.01)	48.63 (7.03)	51.17 (7.82)	51.09 (7.50)	52.40 (6.36)
	No SW-PBIS	51.05 (7.07)	51.84 (6.10)	50.60 (5.89)	51.55 (7.89)	53.62 (5.80)	55.09 (5.66)
	Total	50.34 (7.86)	50.38 (7.31)	49.46 (6.58)	51.37 (7.77)	52.11 (6.89)	53.49 (6.15)
Imputed Data - 5	SW-PBIS	49.59 (8.72)	49.38 (8.01)	48.63 (7.03)	51.17 (7.82)	51.09 (7.50)	52.40 (6.36)
	No SW-PBIS	50.40 (6.41)	50.53 (8.12)	47.50 (7.43)	51.75 (7.63)	51.99 (8.98)	52.23 (6.99)
	Total	49.98 (7.60)	49.84 (7.97)	48.19 (7.10)	51.46 (7.64)	51.45 (8.02)	52.30 (6.51)

Table 3

Results of five linear mixed models corresponding with five multiple imputation datasets

	Elementary cohort		Middle cohort	
	Mathematics	Reading	Mathematics	Reading
	β_3 (s.e.)	β_3 (s.e.)	β_3 (s.e.)	β_3 (s.e.)
Imputed Data - 1	-1.24 (0.89)	1.47 (0.94)	1.08 (0.55)	0.41 (0.54)
Imputed Data - 2	-0.05 (0.88)	1.33 (1.00)	0.16 (0.57)	-0.13 (0.56)
Imputed Data - 3	-0.82 (0.78)	0.53 (0.96)	-0.27 (0.47)	-0.30 (0.49)
Imputed Data - 4	-1.13 (0.79)	0.54 (0.92)	0.32 (0.52)	-0.41 (0.60)
Imputed Data - 5	-0.90 (1.00)	0.69 (0.93)	-0.09 (0.59)	-0.43 (0.59)
Test statistics	-0.83	0.91	0.24	-0.17
Df	74.65	94.61	28.20	94.50
p-value	0.59	0.64	0.19	0.14

Table 4

Mann-Whitney test results corresponding to five multiple imputation datasets

	Elementary School Mann-Whitney test statistics (p-value)		Middle School Mann-Whitney test statistics (p-value)	
	Mathematics	Reading	Mathematics	Reading
Imputed Data - 1	29 (0.558)	38 (0.859)	74 (0.307)	83 (0.357)
Imputed Data - 2	34 (0.594)	29 (0.914)	86 (0.921)	25 (0.090)
Imputed Data - 3	56 (0.792)	34 (0.922)	89 (0.511)	50 (0.213)
Imputed Data - 4	44 (0.650)	37 (0.790)	82 (0.516)	61 (0.183)
Imputed Data - 5	42 (0.806)	45 (0.705)	81 (0.486)	60 (0.292)