Examining the Association Between Implementation and Outcomes

State-wide Scale-up of School-wide Positive Behavior Intervention and Supports

Elise T. Pas, PhD
Catherine P. Bradshaw, PhD

Abstract

Although there is an established literature supporting the efficacy of a variety of prevention programs, there has been less empirical work on the translation of such research to everyday practice or when scaled-up state-wide. There is a considerable need for more research on factors that enhance implementation of programs and optimize outcomes, particularly in school settings. The current paper examines how the implementation fidelity of an increasingly popular and widely disseminated prevention model called, School-wide Positive Behavioral Interventions and Supports (SW-PBIS), relates to student outcomes within the context of a state-wide scale-up effort. Data come from a scale-up effort of SW-PBIS in Maryland; the sample included 421 elementary and middle schools trained in SW-PBIS. SW-PBIS fidelity, as measured by one of three fidelity measures, was found to be associated with higher math achievement, higher reading achievement, and lower truancy. School contextual factors were related to implementation levels and outcomes. Implications for scale-up efforts of behavioral and mental health interventions and measurement considerations are discussed.

Introduction

The common approach to developing educational and prevention programs has been to create a program, test it through a randomized trial, and then offer it to community institutions.[1] This approach has led to the implicit expectation that districts or schools can and will adopt and implement evidence-based programs with a high degree of fidelity; however, implementation is typically poorer in real-world settings than in efficacy trials.[2,3] As a result, there is increasing

Address correspondence to Elise T. Pas, PhD, Johns Hopkins Bloomberg School of Public Health, Johns Hopkins Center for the Prevention of Youth Violence, 624 N. Broadway, Baltimore, MD 21205, USA. Phone: +1-410-5028506; Fax: +1-410-9559088; Email: epas@jhsph.edu.

Catherine P. Bradshaw, PhD, Johns Hopkins Bloomberg School of Public Health, Johns Hopkins Center for the Prevention of Youth Violence, Baltimore, MD, USA. Email: cbradsha@jhsph.edu

interest among federal agencies, researchers, and policy makers in the process by which prevention programs are moved into real-world settings, often referred to as "translational" research.[4] This includes the process by which efficacious practices, interventions, or treatments become implemented effectively in real-world settings.[5] Yet, there has been limited empirical work specifically on the process of translating efficacious practices into various contexts or on how to support implementation and the scaling-up processes.[6]

There is a considerable need for more research on factors that enhance or relate to the adoption and adequate implementation of programs and lead to effective practice and outcomes,[4–7] particularly in school settings where there is a growing emphasis on the implementation of "evidence-based" prevention programs.[8,9] Although this is a positive trend, previous research suggests that, on average, schools are implementing a dozen or more different prevention programs,[10] leaving concerns about the implementation fidelity of these programs.[10,11] In addition to this real-world concern, few rigorous studies of the effectiveness of prevention programs measure or report data on the level of implementation,[12,13] and, therefore, even less is known about implementation of school-based programs when taken to scale.

The current paper applies a Type II translational research approach to examining how the implementation fidelity of an increasingly popular and widely disseminated school-based prevention model called School-Wide Positive Behavioral Interventions and Supports (SW-PBIS)[14] relates to student outcomes. School-level factors which previous research suggests are potentially related to both implementation and the targeted outcomes are considered.[15] A unique feature of this study is the use of data from a state-wide scale-up effort of SW-PBIS, which includes over 870 Maryland public schools. Data for this study are from schools across the state. Maryland is not alone in its efforts to scale up SW-PBIS, as at least 44 states across the United States have developed a state- or district-level infrastructure to support its implementation. The large-scale implementation of SW-PBIS has important implications for behavioral and academic outcomes for students. Issues related to the type of implementation data collected in relation to the extent to which they predict student outcomes are also considered; this issue is of particular importance in scale-up efforts where the resources to collect fidelity data are often limited. First, we provide a brief review of translational research, followed by an overview of SW-PBIS and the infrastructure developed to scale up the model in Maryland.

Translational research

There has been a recent effort to differentiate between two types of translational research: Type I translational research focuses on discovery through clinical trials, whereas Type II examines the process by which efficacious practices, interventions, or treatments become implemented effectively in real-world settings.[5] The current paper focuses on Type II translational research, which "is aimed at enhancing the adoption, implementation, and sustainability of evidence-based or scientifically validated interventions"[4(p 2)] and focuses on achieving broad, population-level effects. While there is an established literature supporting the efficacy (Type I) and effectiveness (one element of Type II research) of a variety of intervention approaches or programs, there has been less empirical work specifically on the process of translating efficacious practices into real-world settings or on how to support implementation and the scaling-up processes.[6] When it comes to scale-up efforts in schools, there is limited empirical research on the extent to which prevention programs are adequately implemented and the association between implementation quality and student outcomes. This illustrates a clear need for additional research on the process of dissemination or planned diffusion[16] of evidence-based programs and whether the effects seen in randomized trials are replicated when brought to scale.[17]

School-wide Positive Behavioral Interventions and Supports

The current paper focuses on the scale-up of SW-PBIS,[14] which is a non-curricular, school-based prevention approach which aims to promote changes in staff behavior in order to positively impact student outcomes such as student discipline, behavior, and academic outcomes. SW-PBIS [14] is based on behavioral, social learning, and organizational behavioral principles. The model is implemented in all school contexts (classroom and non-classroom) with the aim of improving a school's systems and procedures to prevent disruptive behavior and enhance the school's organizational climate. The model follows the three-tiered prevention framework, where a universal system of support is integrated with selective and indicated preventive interventions for students displaying a higher level of need.[18] Two recent randomized controlled effectiveness trials provide evidence of positive outcomes of the universal elements of the SW-PBIS model. Specifically, SW-PBIS has been shown to be effective at reducing student office discipline referrals and suspensions, and improving school climate.[19–22] Teachers in SW-PBIS schools also rate their students as needing fewer specialized support services, and as having fewer behavioral problems (e.g., aggressive behavior, concentration problems, bullying, rejection).[23,24] In addition, there are some favorable results from state-wide evaluations of SW-PBIS.[25,26] Taken together, these studies provide evidence that states can implement SW-PBIS on a large scale and that schools adopting SW-PBIS experience positive effects. Given the wide dissemination of SW-PBIS and previous research documenting its effectiveness, it is a particularly good candidate for Type II translational research focused on implementation quality in scale-up efforts.

SW-PBIS scale-up in Maryland

Maryland has developed a coordinated system for implementation of SW-PBIS. Over the past 12 years, a collaboration between the Maryland State Department of Education, Sheppard Pratt Health System, and Johns Hopkins University[25–27] has trained a total of 877 schools (e.g., elementary, middle, high, alternative, special) in SW-PBIS, of which 740 (84 % of trained schools) are actively implementing and participating in the state initiative. This is made possible through the state-wide infrastructure, which includes a variety of core elements for dissemination,[7,28–30] including a consortium of stakeholders (e.g., educators, researchers, policymakers) who jointly coordinate, train, and support schools in the implementation of SW-PBIS. There are multiple levels of coordination (for details, see Barrett et al.[25]) to promote high quality implementation. Similar systems of support have been utilized in other translational efforts to disseminate programs and achieve high fidelity (for examples, see Bloomquist et al.,[31] Fixsen et al.,[32] Spoth and Greenberg[33]). The Maryland Initiative also maximizes the dissemination of SW-PBIS through the promotion of exchange between school practitioners, who may be more effective in shaping their colleagues' opinions about SW-PBIS than the consortium,[34] and by utilizing coaches and district leaders.[16,34] Finally, there is ongoing data collection, evaluation, and technical assistance provided by the partners regarding implementation and outcomes.[29,30–33] The data from the current study come from the state's evaluation efforts.

Linking implementation with outcomes in scale-up efforts

While several efficacy studies of behavioral or mental health prevention programs have documented an association between implementation and outcomes,[35–39] there have been relatively few studies which have examined the link between implementation and outcomes within the context of state-wide scale-up efforts. For example, research on the Triple P-Positive Parenting Program, which targets changes in child behavior through training parents to alter the home environment,[40] has found that the intensity of the program as well as the format (e.g., self-directed

versus group format) was significantly associated with parent and child outcomes.[41] Similarly, an evaluation of the Promoting Alternative Thinking Strategies (PATHS) social–emotional curriculum reported a significant interaction between implementation and contextual factors, like administrator support, on student behavioral and emotional outcomes.[42] Taken together, the available research suggests a need for more empirical research on the association between implementation quality and outcomes when interventions are brought to scale.[43]

Role of contextual factors

When examining the association between implementation and outcomes, it is important to adjust for contextual factors, which may influence implementation, as well as the outcomes. For example, there is literature suggesting that a high rate of disorder or disorganization can impede successful implementation of programs and can negatively impact program outcomes, [15,20,44] whereas a climate which encourages the adherence to implementation (or fidelity) may improve implementation.[16,45] A previous study of SW-PBIS in elementary schools found that implementation fidelity was associated with school-level factors, such as the percent of certified teachers in the school;[27] however, the available district-level predictors were not associated with implementation. Also relevant was the number of years since training, such that schools that implemented the model longer achieved higher levels of fidelity[27] (see Rohrbach et al.[30] and Rogers[34]). Together, these findings suggest that it is important to account for school-level contextual factors when examining the association between implementation and outcomes in scale-up efforts.

Overview of the current study

The current paper examined how the level of implementation of SW-PBIS related to student outcomes, while adjusting for school-level contextual factors, which are associated with both implementation quality and student outcomes. The data come from the state-wide evaluation of SW-PBIS, which is led by the PBIS Maryland Consortium. A variety of data elements are collected by the PBIS Maryland Consortium, including the implementation quality of SW-PBIS. The data reported in this paper focus on program implementation in the spring of 2009 (i.e., from the 2008 to 2009 year) and student outcomes in spring 2010 (i.e., from the 2009 to 2010 year), while controlling for predictor variables, which preceded each individual school's year of training. Data from elementary and middle schools were examined, including traditional K-5 or K-6 elementary schools, K-8 schools, and middle schools with grades 5 or 6 to 8. High schools implementing SW-PBIS were excluded because the assessment of student outcomes varied substantially for this school level (i.e., different standardized testing approach).

The outcomes of interest were student achievement on the Maryland School Assessment (MSA) for math and reading, truancy rates (i.e., percent of students absent greater than 20 days in the school year), and suspensions (i.e., the total number of suspension events divided by the total number of students times 100). Baseline data for each outcome (i.e., achievement, truancy, and suspensions in the year prior to the school's training in SW-PBIS) were controlled for. The level of implementation of SW-PBIS was assessed by three measures: the Implementation Phases Inventory (IPI),[46] the School-wide Evaluation Tool (SET),[47] and the Benchmarks of Quality (BoQ),[48,49]

It was hypothesized that higher levels of implementation would be associated with higher levels of achievement and lower rates of truancy and suspensions. Based on the literature identifying potentially important contextual factors on implementation quality and student outcomes,[15] a set of school-level variables (i.e., student enrollment; students per teacher; rates of mobility and teacher certification; and years since training) was controlled for. It was hypothesized that large school size and student to teacher ratio, lower rates of teacher certification, and high student mobility would be associated with poorer SW-PBIS implementation and outcomes, based on the use of these variables as proxies for disorder.[15]

On the other hand, we hypothesized that the longer the schools implemented SW-PBIS, the higher their implementation quality would be.[50]

Method

Participants

Eligibility

Within the state of Maryland, there are 24 districts, all of which participate in the SW-PBIS Initiative. The focus is on traditional (i.e., non-special education, non-alternative) elementary and middle school, since the initiative had a strong support system for these schools relative to high schools or non-traditional schools. There were 474 schools (i.e., traditional elementary or Kindergarten to grade 5 [K-5] or K-6 schools, traditional middle or grades 5 or 6–8 schools, and K-8 schools) across the 24 districts which were trained in SW-PBIS in 2008 or earlier. Of these schools, 421 (or 88.1 %) submitted data regarding implementation on at least one measure and therefore were eligible for inclusion in the analyses. The sample included 269 elementary schools, 140 middle schools, and 12 K-8 schools. School-level demographics for the sample are reported in Table 1.

Measures

Implementation of SW-PBIS using the Implementation Phases Inventory (IPI)

The IPI [46] assesses the presence of 44 key elements of SW-PBIS following a "stages of change" theoretical model, whereby schools move through a series of four stages: *preparation* (Cronbach's alpha [α]=.65;* e.g., "PBIS team has been established," "School has a coach"), *initiation* (α=.80; e.g., "A strategy for collecting discipline data has been developed," "New personnel have been oriented to PBIS"), *implementation* (α=.90; e.g., "Discipline data are summarized and reported to staff," "PBIS team uses data to make suggestions regarding PBIS implementation"), and *maintenance* (α=.91; e.g., "A set of materials has been developed to sustain PBIS," "Parents are involved in PBIS related activities"). The schools' PBIS intervention support coach reviewed each of the 44 items on the scale and indicated the extent to which each core feature was in place at the school on a 3-point scale from 0 (*not in place*) to 2 (*fully in place*). Schools received a percentage of implemented elements for each stage, such that a higher score indicated greater implementation. The IPI was developed in conjunction with the PBIS Maryland State Leadership Team to track different phases of implementation; it reflects the core elements of universal SW-PBIS (in the preparation, initiation, and implementation stages), as well as some more advanced features, such as preparing for parental involvement and implementation of selected and indicated preventive interventions (in the maintenance stage). A previous study of the psychometric properties of the IPI found it to have adequate internal consistency (α=.94) and reliability (test–retest correlation of .80).[46]

Implementation of SW-PBIS using the School-wide Evaluation Tool (SET)

The SET [47] is conducted by an external evaluator and consists of seven subscales that assess the degree to which schools implement the key features of SW-PBIS.[51] The scales assessed include: (a)

---

*This scale measures pre-implementation readiness and therefore the variability and the internal consistency on this scale are low.

Expectations Defined; (b) Behavioral Expectations Taught; (c) System for Rewarding Behavioral Expectations; (d) System for Responding to Behavioral Violations; (e) Monitoring and Evaluation; (f) Management; and (g) District-Level Support. Each item of the SET is scored on a 3-point scale from 0 (*not implemented*) to 2 (*fully implemented*). A scale score reflecting the percentage of earned points is calculated, such that higher scores reflect greater implementation fidelity. The SET was created by the developers of SW-PBIS; it is the most commonly used measure of the core features of the universal SW-PBIS model. Previous studies have documented the reliability and validity of the SET.[52,53]

Table 1
School and district demographic characteristics

	All Schools (*N*=421)
Implementation	
IPI Scale Scores (0–100 %)	
Preparation	95.69 (7.42)
Initiation	91.80 (11.13)
Implementation	88.18 (16.38)
Maintenance	76.85 (22.33)
SET Scale Scores (0–100 %)	
Expectations defined	94.66 (12.35)
Expectations taught	96.38 (8.23)
Rewarding system	97.01 (8.68)
Violations system	92.28 (12.29)
Decision making	97.94 (6.94)
Leadership	95.09 (7.49)
District support	96.97 (11.74)
BoQ Score (0–100 %)	88.58 (11.59)
School demographic characteristics	
Years since initial SW-PBIS Training	3.84 (2.17)
Enrollment	579.24 (233.94)
Student/teacher ratio	19.62 (3.72)
% of certified teachers	38.73 (14.62)
Student mobility (%)	24.48 (16.28)
Baseline measures	
Reading performance (% passing MSA)	68.74 (18.87)
Math performance (% passing MSA)	63.46 (21.69)
Truancy rate (%)	9.00 (7.22)
Suspensions (%)	10.10 (15.71)
2009–2010 outcome measures	
Reading performance (% passing MSA)	83.85 (9.36)
Math performance (% passing MSA)	80.31 (12.96)
Truancy rate (%)	7.66 (5.62)
Suspensions (%)	9.50 (12.09)

Means are reported, with standard deviations in parentheses
MSA Maryland School Assessment, *SET* School-wide Evaluation Tool, *IPI* Implementation Phases Inventory, *BoQ* Benchmarks of Quality

Implementation of SW-PBIS using the Benchmarks of Quality

The BoQ [48,49] is completed by multiple PBIS team members and the coach and consists of 53 individual benchmarks assessing 10 areas of implementation (i.e., PBIS team, faculty commitment, effective disciplinary procedures, data entry and analysis plan, expectations and rules, the recognition system, lesson plans for teaching expectations, implementation plan for PBIS, classroom systems, and evaluation). Team members and the PBIS coach each independently complete a rating of each item on a 3-point scale (0=not in place, 1=needs improvement, and 2=in place) and their responses are combined, such that the most frequently endorsed rating for each item is the final score. An overall percentage of implementation was calculated by adding all earned points and dividing by the total possible points. In completing the BoQ, multiple team members and the coach provide ratings, which are then averaged into a single score for the school. Only the overall BoQ score is provided to the state, and thus only this score is available for analysis in the current study. This is the only implementation measure in this study which incorporates scores from multiple raters. The BoQ has documented adequate internal consistency, test–retest reliability, inter-rater reliability, and concurrent validity with the SET.[49]

School-level demographic characteristics

Data on the year in which the schools were trained were provided by the PBIS Maryland Consortium. These data were used to calculate the years since training (i.e., number of years implementing SW-PBIS) as well as to determine which year's data should be used for the school-level covariates. This variable ranges from 1 to 10 years, reflecting training in the summers of 1999 through 2008, respectively. The demographic information regarding the schools was provided by the Maryland State Department of Education. Data regarding school size (e.g., student enrollment, student/teacher ratio [i.e., number of students per teacher]), percent of certified teachers (i.e., those certified to teach in the state of Maryland by completing the required coursework such as a Bachelor's degree from a pre-approved teacher preparation program and have passed a basic skills and content area test), and student mobility (i.e., the percent of students who entered the school, plus the percentage who withdrew from the school, divided by total student enrollment) were obtained to serve as predictors, as were outcome data (i.e., MSA math and reading, truancy rates, and suspensions). The school covariates reflect data from the year preceding a school's training in SW-PBIS (e.g., if a school was trained in summer 2007, then the data from the 2006 to 2007 school year were used; if the school was trained in summer 2005, data from 2004 to 2005 were used, and so on). This same procedure was used for the baseline data of each outcome (i.e., achievement, truancy, suspensions). The outcome variables were from the 2009 to 2010 school year in all cases (see Table 1 for a full listing of demographic and SW-PBIS information for this sample of schools). The inter-correlations among these variables are reported in Table 2.

Procedure

As a requirement of the PBIS Maryland Initiative, the IPI is completed bi-annually (fall and spring) by a district-appointed technical assistance provider (i.e., a SW-PBIS Coach, which in Maryland is often a school psychologist or counselor) and submitted electronically to the PBIS Maryland Consortium through the www.PBISMaryland.org web site. As noted above, the SET is completed by an external district assessor, and the BoQ is completed by the school's SW-PBIS team; these data elements are completed annually in the spring and are also submitted

Table 2
Correlations among school-level baseline variables (N=421 Schools)

School-level Baseline measures	1	2	3	4	5	6	7	8	9	10	11
1. Years since training	1										
2. Student enrollment	.142**	1									
3. Student/teacher ratio	.169**	.205**	1								
4. % of certified teachers	.091	-.063	.154**	1							
5. Student mobility (%)	.055	-.234**	-.121**	-.168**	1						
6. Reading achievement	-.616**	-.046	.011	.217**	-.439**	1					
7. Math achievement	-.556**	-.193**	.036	.260**	-.419**	.897**	1				
8. Truancy	-.067	.145**	-.034	-.301**	.354**	-.402**	-.501**	1			
9. Suspensions	-.224**	.219**	-.025	-.181**	.155**	-.069	-.201**	.555**	1		
10. SET	.097	-.056	.078	.185**	.002	.022	.033	-.075	-.023	1	
11. IPI	.266**	-.011	.111*	.258**	-.082	.001	.022	-.177**	-.128**	.393**	1
12. BoQ	.289**	-.029	.066	.222**	-.151**	.031	.066	-.170**	-.072	.266**	.586**

The school demographic data come from the year prior to the first training in PBIS (e.g., these data are from the 2006–2007 school year for schools trained in the summer of 2007)

SET School-wide Evaluation Tool, *IPI* Implementation Phases Inventory, *BoQ* Benchmarks of Quality

*p<.05, **p<.01

electronically through the Consortium's website. The non-identifiable school-level data have been approved for analysis in this study by the Johns Hopkins Bloomberg School of Public Health Institutional Review Board.

Analyses

The M*plus* 6.1 statistical software [54] was used to fit a structural equation model (SEM) [55] in order to test the hypothesized associations between fidelity and student outcomes, while adjusting for covariates. Specifically, an SEM using maximum likelihood robust (MLR) estimation was fit. A confirmatory factor analysis (CFA) was conducted on the measurement model of implementation (i.e., the four IPI scales and the seven SET scales). A latent variable SEM approach was taken to allow for a reduction in the fidelity data dimensionality; to allow for a parsimonious, more interpretable model; and to eliminate concerns of multicollinearity due to the high correlations among the subscales within each measure (i.e., the IPI and SET).[56] The BoQ was modeled as a third manifest (i.e., observed) indicator of implementation fidelity, as only the overall score was available to the researchers. The two implementation factors and the observed BoQ scores were then used to predict student outcomes (i.e., math and reading academic performance, truancy, and suspensions), while adjusting for the school-level covariates (i.e., years since training, school enrollment, the student/teacher ratio, the percent of certified teachers, and student mobility) and the baseline outcome measures.

As schools were nested within districts, the clustering of schools within districts was accounted for using the Huber–White corrections to adjust the standard errors;[54] however, district-level covariates were not modeled due to the relatively small number of districts (i.e., 24), and because prior research using this dataset suggested that district covariates generally were not significantly associated with implementation.[27] Given that all of the schools in the study were from a single state, no state variables could be modeled. Model fit was determined through inspection of the Root Mean Square Error of Approximation (RMSEA), Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), and the Standardized Root Mean Residual (SRMR).[57] A value between .05 and .08 on the RMSEA is considered acceptable fit; a CFI and TLI of greater than .90 is considered acceptable; and the SRMR should be less than or equal to .08.[58]

Results

SEM results

As described above, an SEM was fit to test the primary hypotheses regarding the association between the different sources of implementation fidelity data and student outcomes, while adjusting for school-level demographic factors and the baseline student outcomes. First, the measurement model was fit to verify the factor structure of the two latent implementation variables (i.e., the IPI and SET). The CFA indicated that these latent variables had adequate fit with an RMSEA=.045, CFI=.962, TLI=.951, and SRMR=.116 (see the Measurement Model section in Table 3 for factor loadings).

Model fit

The measurement model was incorporated into the hypothesized structural model, which included the BoQ as a third manifest implementation fidelity indicator. Student outcomes were regressed on the implementation variables, adjusting for the five school-level contextual factors and the baseline student outcome variables (see Fig. 1). This model had adequate fit with an RMSEA=.070, CFI=.897, TLI=.859, and SRMR=.186. The modification indices were examined for potential aspects of the model that could be improved, but none were substantively relevant. Therefore, the model reported in Table 3

and Figure 1 was selected as the final model. The substantive findings from that model are reported below.

Relationship between school-level contextual factors and implementation

The years since training (Standardized Coefficient [Std. Coeff.]=.289, $p=.002$) and percent of certified teachers (Std. Coeff. =.187, $p=.002$) were both positively related to the IPI factor. Schools with a greater number of years since their training in SW-PBIS and a higher percent of standard certified teachers had better implementation. The student enrollment (i.e., number of students in the school), student to teacher ratio, and mobility were not significantly related the IPI factor. None of the modeled covariates were related to the SET factor. Similar to the IPI factor, the years since training (Std. Coeff.=.131, $p=.015$) and percent of certified teachers (Std. Coeff.=.151, $p=.011$) were also positively related to the observed score on the BoQ. In addition, student mobility (Std. Coeff.=−.245, $p=.007$) was negatively related to the observed score on the BoQ, indicating that lower rates of mobility were associated with higher implementation scores based on the BoQ. Indicators of school size (i.e., student enrollment and student to teacher ratio) were not related to the implementation levels assessed by the BoQ. This model accounted for a significant proportion of variance in the IPI factor (i.e., $R^2=.156$) and the observed BoQ score (i.e., $R^2=.116$), but not the SET factor.

Relationship between school-level contextual factors and outcomes

As expected, the baseline measure of the math and reading achievement and truancy outcomes were positively related to their respective outcomes, such that earlier high achievement on math (Std. Coeff.=.720, $p<.001$) and reading (Std. Coeff.=.634, $p<.001$), and higher rates of truancy (Std. Coeff.=.671, $p<.001$) were associated with higher levels of these outcomes in 2010. Surprisingly, the relationship between baseline suspensions and the suspension outcome only approached significance at the $p=.10$ level (Std. Coeff.=.400). The number of years since training (Std. Coeff.=.248, $p<.001$), student enrollment (Std. Coeff.=−.085, $p=.001$), and mobility (Std. Coeff.=−.193, $p=.001$) were significantly related to math achievement such that a greater number of years since training in SW-PBIS, smaller school size, and lower mobility all related to higher math achievement in 2010. Only years since training (Std. Coeff.=.169, $p<.001$) and mobility (Std. Coeff.=−.268, $p<.001$) were significantly related to reading achievement in 2010. Higher levels of mobility (Std. Coeff.=.157, $p=.002$) were related to higher levels of truancy in 2010. Finally, student enrollment (Std. Coeff.=.197, $p=.004$) was related to suspension rates in 2010. The relationships between student enrollment and math achievement and suspensions implicitly demonstrates that middle schools generally had lower rates of achievement and higher suspensions than elementary schools, as elementary schools on average are smaller.

Relationship between implementation and student outcomes

Controlling for the direct effects of baseline measures and the school-level covariates, the IPI factor was significantly related to the math, reading, and truancy outcomes. Specifically, higher implementation, as indicated on the IPI, was associated with higher math achievement (Std. Coeff. =.146, $p=.042$), higher reading achievement (Std. Coeff.=.171, $p=.006$), and lower truancy rates (Std. Coeff.=−.088, $p=.056$). Neither the IPI factor, SET factor, nor the observed BoQ scores were related to suspensions. The SET factor and BoQ also were not related to any of the other outcomes. This model accounted for a significant proportion of variance (i.e., R^2) for three of the outcomes. Specifically, the R^2 values are as follows: math achievement=.750, reading achievement=.707, truancy=.651. The R^2 value for suspension ($R^2=.296$) approached significance at the $p=.10$ level.

Table 3

Standardized path coefficients for structural equation model of implementation and outcomes

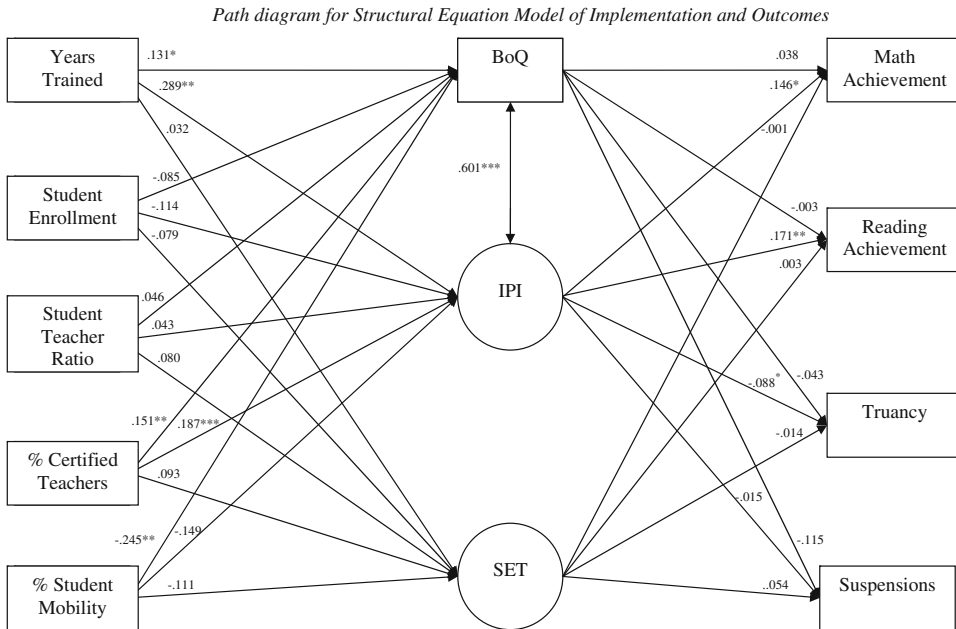| Independent variable	Math	Reading	Truancy	Suspensions	IPI	SET	BoQ
Dependent variables							
Measurement model							
Preparation	–	–	–	–	.652***	–	–
Initiation	–	–	–	–	.869***	–	–
Implementation	–	–	–	–	.873***	–	–
Maintenance	–	–	–	–	.825***	–	–
Expectations defined	–	–	–	–	–	.382***	–
Expectations taught	–	–	–	–	–	.853***	–
Rewarding expectations	–	–	–	–	–	.641***	–
Responding to Beh violation	–	–	–	–	–	.378***	–
Monitoring and evaluation	–	–	–	–	–	.450***	–
Management	–	–	–	–	–	.664***	–
District support	–	–	–	–	–	.120	–
Structural model							
IPI factor	.146*	.171**	−.088*	−.015	–	–	–
SET factor	−.001	.003	−.014	.054	–	–	–
BoQ (manifest)	.038	− .003	−.043	−.115	–	–	–
Years since trained	.248***	.169***	.036	.162	.289**	.032	.131*
Student enrollment	−.085***	.022	−.025	.197**	−.114	−.079	−.085
Student/teacher ratio	.039	.020	−.007	−.035	.043	.080	.046
% certified teachers	.056	.013	−.046	.004	.187***	.093	.151**
% Mobility	−.193**	−.268***	.157**	.002	−.149	−.111	−.245**
Baseline math	.720***	–	–	–	–	–	–
Baseline reading	–	.634***	–	–	–	–	–
Baseline truancy	–	–	.671***	–	–	–	–
Baseline suspensions	–	–	–	.400	–	–	–

Each outcome was regressed on its baseline corresponding measure only (N=421 Schools)

–, not applicable

*p<.05, **p<.01, ***p<.001

Path diagram for Structural Equation Model of Implementation and Outcomes

Discussion

This paper examined the relationship between implementation, as measured by three different instruments, on student outcomes, using data from a state-wide dissemination of a widely used, school-based prevention program. The availability of three indicators of implementation quality and multiple student outcome data provided a unique opportunity to explore these associations. These findings also shed light on how the choice of an implementation measure can influence the pattern of findings. Below, the specific substantive findings based on the SEM results are considered, as are some implications of these findings for future studies of the association between implementation fidelity and outcomes in scale-up efforts.

The findings indicate that the IPI factor was significantly related to reading and math achievement and truancy such that higher implementation was associated with subsequent higher achievement and lower rates of truancy. Interestingly, suspensions were not related to either of the implementation factors or the observed scores on the BoQ. In addition, the relationship between the suspension outcome and baseline suspension was not significant and the proportion of variance explained in the suspension outcome was lower than for the other outcomes and only approached significance. The findings for suspension were surprising given that suspensions are considered to

be a proximal outcome of SW-PBIS.[21] It is important to note, however, that suspensions are the one outcome in which subjectivity plays a role, as adult behavior affects the rate of suspensions in a school (e.g., the supervision of adults to realize a negative behavior has occurred; the choice of a teacher to refer a student to the office; and the choice of a principal to suspend). In addition, there have been efforts to explicitly decrease suspension rates in the state of Maryland; in comparing the baseline and 2010 rates, one sees that there is a drop in average suspension rates (i.e., from 11 % to 9 %) and in the variability (i.e., standard deviation of 17 to 11). Therefore, there may be overall shifts in suspensions that are associated with accountability efforts. Finally, there is some evidence from effectiveness studies that the ability to detect effects of an intervention varies by the measurement approach. For example, it is common in school-based studies for some measures (e.g., teacher-reported measures) to generate larger effect sizes.[59] A meta-analysis of the Triple P program also found that the type of measurement was associated with the detection of effects.[41,60] The suspension measure may be less sensitive to implementation effects, within the context of a state-wide scale-up effort.

The SET factor and observed BoQ scores were not significantly related to any of the outcomes. Of concern in the case of the SET was that there was a potential ceiling effect of scores, such that the average score was about 95 % and there was little variability in these scores (see Table 1). The SET's restriction of range likely led to its inability to discriminate between schools' outcomes.

The scores on the BoQ were lower on average and showed greater variability than the SET, and were similar to IPI average scores. The SET also correlated with the IPI and BoQ at a lower magnitude than the IPI and BoQ correlate with one another (see Table 2). In addition, the proportion of variance explained in the SET factor was not significant and included a non-significant factor loading for one of the scale scores. The non-significant scale, *District-Level Support,* is a two-item scale asking whether the district provides coaching support and funding for SW-PBIS.

Perhaps the differences detected in the predictive validity are the result of these three measures assessing slightly different aspects of SW-PBIS implementation. For example, the IPI is inherently different than the other two measures, as it takes a "stages of change" approach, ranging from start-up activities to more advanced implementation and sustainability. This may make it more appropriate for assessing fidelity over multiple years of implementation. In contrast, the SET focuses primarily on the start-up activities and initial phases of implementation and is the only measure completed by an outsider to the school. In fact, recent research on the SET suggests that this measure is most reliable at the elementary (versus middle and high) school level and may be most appropriate for administration in schools which have just begun implementing SW-PBIS. Nevertheless, the SET is still the most widely used measure. The IPI and BoQ were more recently developed, and thus would benefit from further research on their psychometric properties and predictive validity. In particular, replications of this study should be conducted, particularly in other states where there may be different levels of infrastructure to support SW-PBIS implementation. These areas have implications for the data collection practices in the state, as currently all three measures are collected by the state on SW-PBIS schools.

Limitations and future directions

There are several limitations to consider when interpreting these findings. Type II translational research is often characterized as "messy," [5] as it is difficult to implement carefully controlled designs when examining the real-world process of program implementation. In addition, the measures of contextual factors used were school-level proxies for disorder,[61] rather than specific survey measures of students or staff members. On the other hand, multiple ratings of the implementation quality of SW-PBIS were used, which is a unique and strong aspect of this study.

The outcome and implementation data come from one time point (i.e., the Spring of 2010 and the 2008 to 2009 school year, respectively) which is 11 years after the state initiative began. This highlights two other common obstacles in translational research, including the extended amount of time it may take to disseminate an intervention or approach [30,34] and the difficulty of assessment of a "moving target".[5(p 212)] By focusing on two school years (baseline and outcome), the data analysis is simplified and made more interpretable; however, future analyses should take into consideration patterns of implementation over time, beginning with the first year of implementation. The fact that different numbers of schools joined the initiative at different points across the 12-year effort makes such analyses complicated, as the data would need to be aligned by implementation year, rather than calendar year. Despite this, all school demographics modeled were from the year preceding the training year and the number of years since training was accounted for. It is possible that the associations between implementation fidelity and outcomes would vary at different points in the scaling-up process (e.g., if measured earlier in the statewide scale-up or in the future). This is an area for further research.

Given that this study occurred in one state, it is unknown whether the findings would generalize to other states where there may be different levels of support provided to schools implementing SW-PBIS and the data collected regarding implementation, the school context (e.g., requirements for teacher certification, varying school sizes or levels of mobility), and the tests of achievement outcomes would be different. Finally, non-implementing schools were not examined in this study, as implementation data were not collected from these schools. Therefore, it is unknown whether schools not trained in SW-PBIS implement similar strategies to trained schools or whether the use of SW-PBIS is superior to non-use within the context of a state-wide scale-up (i.e., randomized controlled trials have established its effectiveness on a smaller scale). Similarly, we were not able to track the implementation of other programs in combination with SW-PBIS (e.g., bullying prevention); this is, however, an important area for further research, as previous studies suggest that schools are likely implementing multiple prevention programs simultaneously.[10]

As noted above, the findings for the SET were less informative than the other two measures of implementation. This was a bit surprising given that the SET is the most established and most widely used measure of SW-PBIS fidelity, whereas the IPI and BoQ are newer measures. On the other hand, these two latter measures were developed in part to address some of the concerns regarding the SET related to a potential ceiling effects [53] and the burden of administration by an outside assessor. The current findings suggest that the IPI has the best predictive validity of the three measures examined. This finding also highlights a practical barrier in conducting scale-up efforts and evaluating their effectiveness; what is practical (e.g., only collecting data using one implementation measure) and desired may not result in a comprehensive understanding of the outcomes. On the other hand, collecting multiple measures of implementation from different sources is often seen as burdensome and redundant by schools but may be important.

Implications for Behavioral Health

Numerous authors have concluded that findings from efficacy and effectiveness trials rarely directly translate when broadly disseminated.[62,63] Instead, programs need to be evaluated under real-world conditions, be practically important, and have adequate supports in place (e.g., manuals, technical assistance) to ensure implementation and then also must be evaluated in a scale-up effort [1]. The current study is one attempt to fill this research gap, as it relates to a school-based prevention framework targeting positive behavioral supports and improving school climate and orderliness. Currently, there are two randomized trials documenting positive effects of SW-PBIS on student office discipline referrals, student discipline problems, and school climate.[20–24] Research is also under way to determine the extent to which the trial findings generalize to the broader set of schools within the state.[17] The current study represents an important next step in the research on

state-wide dissemination of school-based prevention programs and highlights the importance of developing an infrastructure to collect data on implementation quality and program outcomes when prevention efforts are brought to scale.

In addition, the importance of how implementation is measured is highlighted. This includes consideration of how measures used in randomized controlled trials may translate into the real-world setting in terms of their reliability and validity, and how the utility of the measures may change over time. The purpose for which a measure is used and whether it continues to be an effective measure over time are also important factors to consider. Although this study revealed significant associations between one measure of implementation and student outcomes, it also demonstrated non-significant associations between outcomes and two other measures. This highlights the importance of developing implementation measures which are studied across time and in large-scale initiatives, and are shown to be reliable and have predictive validity. These findings have broader implications for behavioral science, as they suggest a need for better implementation measures that are sensitive to both to the foundational pieces needed when first beginning the implementation of a new program, as well as the evolving efforts over time which may be harder to detect.

References

1. Flay BR, Biglan A, Boruch RF, et al. Standards of evidence: Criteria for efficacy, effectiveness and dissemination. *Prevention Science.* 2005;6(3):151–175.

2. Dusenbury L, Brannigan R, Falco M, et al. A review of research on fidelity of implementation: Implications for drug abuse prevention in school settings. *Health Education Research.* 2003;18(2):237–256.

3. Ringwalt CL, Ennett S, Johnson R, et al. Factors associated with fidelity to substance use prevention curriculum guides in the nation's middle schools. *Health Education & Behavior.* 2003;30(3):375–391.

4. SPR MAPS II Task Force. Type 2 translational research: Overview and definitions. 2008; http://preventionscience.org/SPR_Type%202%20Translation%20Research_Overview%20and%20Definition.pdf.

5. Woolf SH. The meaning of translational research and why it matters. *Journal of the American Medical Association.* 2008;299:211–213.

6. Woolf SH. Potential health and economic consequences of misplaced priorities. *Journal of the American Medical Association.* 2007;297:523–526.

7. Spoth R. Translating family-focused prevention science into effective practice: Toward a translational impact paradigm. *Current Directions in Psychological Science.* 2008;17(6):415–421.

8. Ringwalt C, Vincus AA, Hanley S, et al. The prevalence of evidence-based drug use prevention curricula in U.S. middle schools in 2008. *Prevention Science.* 2011;12(1):63–69.

9. Sloboda Z, Pyakuryal A, Stephens PC, et al. Reports of substance abuse prevention programming available in schools. *Prevention Science.* 2008;9(4):276–287.

10. Gottfredson GD, Gottfredson DC, Czeh ER, et al. *National study of delinquency prevention in schools.* Ellicott City, MD: Gottfredson Associates; 2000.

11. Gottfredson GD, Jones EM, Gore TW. Implementation and evaluation of a cognitive–behavioral intervention to prevent problem behavior in a disorganized school. *Prevention Science.* 2002;3(1):43–56.

12. Domitrovich CE, Greenberg MT. The study of implementation: Current findings from effective programs that prevent mental disorders in school-aged children. *Journal of Educational & Psychological Consultation.* 2000;11(2):193–221.

13. Durlak JA. *Successful prevention programs for children and adolescents*. New York: Plenum; 1997.

14. Sugai G, Horner RH, Gresham FM. Behaviorally effective school environments. In: Shinn MR, Walker HM, Stoner G, eds. *Interventions for academic and behavior problems: II. Preventive and remedial approaches.* Bethesda, MD US: National Association of School Psychologists; 2002:315–350.

15. Domitrovich CE, Bradshaw CP, Poduska JM, et al. Maximizing the implementation quality of evidence-based preventive interventions in schools: A conceptual framework. *Advances in School Mental Health Promotion.* 2008;1(3):6–28.

16. Schoenwald SK, Hoagwood K. Effectiveness, transportability, and dissemination of interventions: What matters when? *Psychiatric Services.* 2001;52(9):1190–1197.

17. Stuart E, Cole S, Bradshaw CP, et al. The use of propensity scores to assess the generalizability of results from randomized trials. *The Journal of the Royal Statistical Society, Series A.* 2011; 174(2):369–386.

18. O'Connell ME, Boat T, Warner KE. *Preventing mental, emotional, and behavioral disorders among young people: Progress and possibilities.* Washington, DC: The National Academies Press. Committee on the Prevention of Mental Disorders and Substance Abuse Among Children, Youth, and Young Adults: Research Advances and Promising Interventions; 2009.

19. Bradshaw CP, Koth CW, Bevans KB, et al. The impact of school-wide Positive Behavioral Interventions and Supports (PBIS) on the organizational health of elementary schools. *School Psychology Quarterly.* 2008;23(4):462–473.

20. Bradshaw CP, Koth CW, Thornton LA, et al. Altering school climate through school-wide Positive Behavioral Interventions and Supports: Findings from a group-randomized effectiveness trial. *Prevention Science.* 2009;10(2):100–115.

21. Bradshaw CP, Mitchell MM, Leaf PJ. 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.* 2010;12(3):133–148.

22. Horner RH, Sugai G, Smolkowski K, et al. A randomized, wait-list controlled effectiveness trial assessing School-Wide Positive Behavior Support in elementary schools. *Journal of Positive Behavior Interventions.* 2009;11(3):133–144.

23. Bradshaw CP, Waasdorp TE, Leaf PJ. Effects of School-Wide Positive Behavioral Interventions and Supports on child behavior problems and adjustment. Manuscript submitted for publication. 2012.

24. Waasdorp TE, Bradshaw CP, Leaf PJ. 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.* 2012;116(2):149–156.

25. Barrett SB, Bradshaw CP, Lewis-Palmer T. Maryland statewide PBIS initiative: Systems, evaluation, and next steps. *Journal of Positive Behavior Interventions.* 2008;10(2):105–114.

26. Muscott HS, Mann EL, LeBrun MR. 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.* 2008;10(3):190–205.

27. Bradshaw CP, Pas ET. 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.* 2011;40(4):530–548.

28. Adelman HS, Taylor L. Toward a scale-up model for replicating new approaches to schooling. *Journal of Educational & Psychological Consultation.* 1997;8(2):197.

29. Kreuter MW, Bernhardt JM. Reframing the dissemination challenge: a marketing and distribution perspective. *American Journal of Public Health.* 2009;99(12):2123–2127.

30. Rohrbach LA, Grana R, Sussman S, et al. Type II Translation: Transporting prevention interventions from research to real-world settings. *Evaluation & the Health Professions.* 2006;29(3):302–333.

31. Bloomquist ML, August GJ, Horowitz JL, et al. Moving from science to service: Transposing and sustaining the early risers prevention program in a community service system. *The Journal of Primary Prevention.* 2008;29(4):307–321.

32. Fixsen DL, Naoom SF, Blase KA, et al. *Implementation research: A synthesis of the literature.* Tampa: University of South Florida, Louis de la Parte Florida Mental Health Institute, The National Implementation Research Network; 2005.

33. Spoth RL, Greenberg, MT. Toward a comprehensive strategy for effective practitioner–scientist partnerships and larger-scale community health and well-being. *American Journal of Community Psychology.* 2005;35:107–126.

34. Rogers EM. Diffusion of preventive innovations. *Addictive Behaviors.* 2002;27(6):989–993.

35. Botvin GJ, Baker E, Dusenbury L, et al. Long-term follow-up results of a randomized drug abuse prevention trial in a White middle-class population. *Journal of the American Medical Association.* 1995;273(14):1106–1112.

36. Derzon JH, Sale E, Springer JF, et al. Estimating intervention effectiveness: Synthetic projection of field evaluation results. *The Journal of Primary Prevention.* 2005;26(4):321–343.

37. Durlak JA, DuPre EP. Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology.* 2008;41(3–4):327–350.

38. Hulleman CS, Cordray, DS. Moving from the lab to the field: The role of fidelity and achieved relative intervention strength. *Journal of Research on Educational Effectiveness.* 2009;2(1):88–110.

39. Ialongo NS, Werthamer L, Kellam SG, et al. 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.* 1999;27(5):599–641.

40. Sanders MR. Triple P-Positive Parenting Program: Towards an empirically validated multilevel parenting and family support strategy for the prevention of behavior and emotional problems in children. *Clinical Child & Family Psychology Review.* 1999;2(2):71–90.

41. Nowak C, Heinrichs N. A comprehensive meta-analysis of Triple P-Positive Parenting Program using hierarchical linear modeling: Effectiveness and moderating variables. *Clinical Child and Family Psychology Review.* 2008;11(3):114–144.

42. Kam C-M, Greenberg MT, Walls CT. Examining the role of implementation quality in school-based prevention using the PATHS curriculum. *Prevention Science.* 2003;4(1):55–63.

43. Jowers KL, Bradshaw CP, Gately S. Taking school-based substance abuse prevention to scale: District-wide implementation of Keep A Clear Mind. *Journal of Alcohol and Drug Education.* 2007;51(3):73–91.

44. Gottfredson GD, Gottfredson DC, Payne AA, et al. School climate predictors of school disorder: Results from a national study of delinquency prevention in schools. *Journal of Research in Crime and Delinquency.* 2005;42(4):412–444.

45. Glisson C, Green P. The effects of organizational culture and climate on the access to mental health care in Child welfare and Juvenile Justice systems. *Administration and Policy in Mental Health and Mental Health Services Research.* 2006;33(4):433–448.

46. Bradshaw CP, Debnam K, Koth CW, et al. Preliminary validation of the Implementation Phases Inventory for assessing fidelity of schoolwide positive behavior supports. *Journal of Positive Behavior Interventions.* 2009;11(3):145–160.

47. Sugai G, Lewis-Palmer T, Todd A, et al. School-wide Evaluation Tool (SET). Eugene, OR: Center for Positive Behavioral Supports, University of Oregon; 2001.

48. Kincaid D, Childs K, George H. School-wide Benchmarks of Quality (Revised). Tampa, Florida: University of South Florida; 2010.

49. Cohen R, Kincaid D, Childs KE. Measuring school-wide positive behavior support implementation: Development and validation of the benchmarks of quality. *Journal of Positive Behavior Interventions.* 2007;9(4):203–213.

50. Bond GR, Drake RE, McHugo GJ, et al. Strategies for improving fidelity in the National Evidence-Based Practices Project. *Research on Social Work Practice.* 2009;19(5):569–581.

51. Horner RH, Todd AW, Lewis-Palmer T, et al. The School-Wide Evaluation Tool (SET): A research instrument for assessing School-Wide Positive Behavior Support. *Journal of Positive Behavior Interventions.* 2004;6(1):3–12.

52. Bradshaw CP, Reinke WM, Brown LD, et al. Implementation of School-Wide Positive Behavioral Interventions and Supports (PBIS) in elementary schools: Observations from a randomized trial. *Education & Treatment of Children (West Virginia University Press).* 2008:1–26.

53. Vincent C, Spaulding S, Tobin TJ. A reexamination of the psychometric properties of the School-Wide Evaluation Tool (SET). *Journal of Positive Behavior Interventions.* 2010;12(3):161–179.

54. Muthén LK, & Muthén BO. *Mplus user's guide.* Los Angeles, CA: Muthén and Muthén; 1997–2009.

55. Bollen KA. *Structural equations with latent variables.* Oxford, England: John Wiley & Sons; 1989.

56. Kline RB. *Principles and practice of structural equation modeling.* 2nd ed. New York, NY: Guilford; 2005.

57. Yu CY, Muthén BO. *Evaluation of model fit indices for latent variable models with categorical and continuous outcomes.* Los Angeles, CA; 2001.

58. McDonald RP, Ho MR. Principles and practice in reporting structural equation analyses. *Psychological Methods.* 2002;7(1):64–82.

59. Wilson SJ, Lipsey MW. School-Based Interventions for Aggressive and Disruptive Behavior: Update of a Meta-Analysis. *American Journal of Preventive Medicine.* 2007;33(2):S130–S143.

60. Thomas R, Zimmer-Gembeck MJ. Behavioral outcomes of Parent–Child Interaction Therapy and Triple P-Positive Parenting Program: A review and meta-analysis. *Journal of Abnormal Child Psychology.* 2007;35(3):475–495.

61. Birnbaum AS, Lytle LA, Hannan PJ, et al. School functioning and violent behavior among young adolescents: A contextual analysis. *Health Education Research.* 2003;18(3):389–403.

62. Hoagwood K, Burns BJ, Kiser L, et al. Evidence-based practice in child and adolescent mental health services. *Psychiatric Services.* 2001;52(9):1179–1189.

63. Sandler I, Ostrom A, Bitner MJ, et al. Developing effective prevention services for the real world: A prevention science development model. *American Journal of Community Psychology.* 2005;35:127–142.