International Journal of Psychology

International Journal of Psychology, 2020
Vol. 55, No. S1, 4–15, DOI: 10.1002/ijop.12575

Impacts of school-wide positive behaviour support: Results from National Longitudinal Register Data

Nicolai T. Borgen[1], Lars J. Kirkebøen[2], Terje Ogden[3], Oddbjørn Raaum[4], and Mari-Anne Sørlie[3]

[1]Department of Sociology and Human Geography, University of Oslo, Oslo, Norway
[2]Statistics Norway, Oslo, Norway
[3]Norwegian Center for Child Behavioral Development, Oslo, Norway
[4]Ragnar Frisch Centre for Economic Research, Oslo, Norway

Problem behaviour in schools may have detrimental effects both on students' well-being and academic achievement. A large literature has consistently found that school-wide positive behaviour support (SWPBS) successfully addresses social and behavioural problems. In this paper, we used population-wide longitudinal register data for all Norwegian primary schools and a difference-in-difference (DiD) design to evaluate effects of SWPBS on a number of primary and secondary outcomes, including indicators of externalising behaviour, school well-being, pull-out instruction, and academic achievement. Indications of reduced classroom noise were found. No other effects were detected. Analyses revealed important differences in outcomes between the intervention and control schools, independent of the implementation of SWPBS, and that a credible design like DiD is essential to handle such school differences.

Keywords: SWPBS; Difference-in-difference design; Longitudinal; Intervention; Register data.

Problem behaviour in schools may have detrimental effects on both students' well-being and academic achievement. Research shows that students who display disruptive and aggressive behaviours are at risk of academic problems and marginalisation (Bradshaw, Waasdorp, & Leaf, 2015), and these students can have negative spillover effects for the rest of the classroom (Carrell & Hoekstra, 2010). To meet such challenges, schools have increasingly turned to program interventions (Benner, Nelson, Sanders, & Ralston, 2012), such as school-wide positive behaviour supports (acronymized as SWPBS or SWPBIS).

SWPBS is an evidence-based, data-driven, systematic school-wide program implemented by more than 23,000 schools in the United States (U.S.) and internationally (Gage, Whitford, & Katsiyannis, 2018). The primary aim of SWPBS is to address social and behavioural concerns in schools by reducing disruption and more severe problem behaviour, such as violence and bullying. By improving the learning environment, SWPBS is also assumed to impact academic achievements, although improving academic outcomes has not been a main focus in the intervention (Gage, Sugai, Lewis, & Brzozowy, 2015).

A large literature mainly from the U.S. has consistently found that SWPBS successfully addresses social and behavioural problems, with the most convincing evidence coming from randomised-controlled trials (RCTs) (e.g., Bradshaw, Waasdorp, & Leaf, 2012). However, while RCTs provide credible effect estimates for the participants included in the trial, the results are not always generalizable to the target population or related target populations. Participants who volunteer to participate in RCTs may differ from the target population in important aspects (Stuart, Bradshaw, & Leaf, 2015), such as being more likely to implement interventions in accordance with the model (Pas & Bradshaw, 2012).

Unlike in randomised trials or pilot studies, implementation of evidence-based interventions in regular practice (i.e., scale-up) is often poor (Pas & Bradshaw, 2012). This

Correspondence should be addressed to Nicolai T. Borgen, Department of Sociology and Human Geography, University of Oslo, P. O. Box 1096, Blindern, 0317 Oslo, Norway. (E-mail: n.t.borgen@sosgeo.uio.no).

The standards of the Norwegian Social Science Data Services were followed throughout the conduct of the study. The study was financed by a grant from the Research Council of Norway (grant #238050). The Ragnar Frisch Centre for Economic Research was project manager. All authors have contributed to writing the article. Oddbjørn Raaum, Lars J. Kirkebøen, and Nicolai T. Borgen have analysed and interpreted the data. Terje Ogden and Mari-Anne Sørlie have provided school-level data on the SWPBS intervention. The register data were made available by Statistics Norway.

is problematic because intervention effects are in general best when interventions are implemented in accordance with the original model tested in research, without violations of goals, underlying theory, and guidelines (i.e., fidelity) (Domitrovich et al., 2008). Results in RCTs are accordingly likely to be stronger than in real-world conditions (Bradshaw et al., 2015; Flay et al., 2005). Recently, there have been efforts to examine whether program participants in RCTs differ from the wider population (Stuart & Rhodes, 2017). However, so far no study has used population-wide register data to evaluate SWPBS.

In this paper, population-wide longitudinal Norwegian register data were used to evaluate the effects of SWPBS (called N-PALS in Norway) on a number of primary and secondary outcome variables, including indicators of externalising behaviour, school well-being, pull-out instruction, and academic achievement. An advantage of register data is that they include information on all schools that have ever implemented SWPBS in Norway, not only those who volunteer in research, and also all other schools. Additionally, the register data allow for comparison of SWPBS schools and control schools 3 years prior to and 5 years after the initiation of the intervention, which is a considerably longer evaluation period than in most studies (Madigan, Cross, Smolkowski, & Strycker, 2016).

Longitudinal data on SWPBS schools and control schools allow for use of a difference-in-difference (DiD) design, which have not been previously used to evaluate SWPBS (Mitchell, Hatton, & Lewis, 2018). DiD has become a key method in the evaluation literature, as it enables handling of both unobserved (stable) heterogeneity between schools and (general) changes over time unrelated to the intervention (Smith & Todd, 2005). In the current study, impacts of SWPBS were investigated by comparing different cohorts of students within the same schools, before and after the implementation of SWPBS, and after accounting for changes in student composition and cohort effects.

PREVIOUS RESEARCH ON SWPBS

Studies from the U.S. have shown that SWPBS reduces problem behaviour, office discipline referrals, suspensions, and non-attendance (for a review, see Horner, Sugai, & Anderson, 2010). However, less than a third of the studies are grounded in rigorous designs, with most studies being case studies and cross-sectional studies (Chitiyo, May, & Chitiyo, 2012).

Of particular interest are three RCTs, all of which consistently indicate that SWPBS improves behaviour. Benner et al. (2012) found that SWPBS reduced externalising problem behaviours (e.g., aggression, noise), and Horner et al. (2009) found that SWPBS resulted in an improvement in students' perceived school safety. Bradshaw et al. (2012) found that SWPBS had a positive effect on children's behaviour problems, concentration problems, social–emotional functioning, and prosocial behaviour. Using the same data as Bradshaw et al. (2012), Bradshaw, Mitchell, and Leaf (2010) found significant reduction in student suspensions and office discipline referrals and Waasdorp, Bradshaw, and Leaf (2012) found that SWPBS lowered the rates of teacher-reported bullying and peer rejection.

The SWPBS model mainly addresses social and behavioural problems in school (Horner et al., 2010). However, when problem behaviour interferes with teaching, reduced disruption may increase students exposure to classroom instruction and consequently to increased academic engagement and learning (Gage et al., 2015). More than 20 studies have examined the effects of SWPBS on academic achievements. The majority of studies report that SWPBS seems to improve academic outcomes in the U.S., but these studies have used a descriptive design or a case study design which does not account for selection effects (for a review, see Gage et al., 2015).

Experimental studies have not found any effect of SWPBS on academic achievement (Benner et al., 2012; Bradshaw et al., 2010; Horner et al., 2009), while quasi-experiments have reported mixed effects (Caldarella, Shatzer, Gray, Young, & Young, 2011; Freeman et al., 2016; Gage et al., 2015; Gage, Leite, Childs, & Kincaid, 2017; Lane, Wehby, Robertson, & Rogers, 2007; Madigan et al., 2016). However, small sample sizes and a short research period (Madigan et al., 2016) limit the validity of the outcomes. Accordingly, there is still an open question whether the model impacts academic achievement.

Although most of the evidence on SWPBS comes from U.S. studies, two sampling studies with a quasi-experimental design have evaluated SWPBS in Norway (Sørlie & Ogden, 2007, 2015). Behavioural problems in schools are major challenges in Norway, with some surveys indicating more problem behaviour, such as classroom disruption, in Norway than in other western countries, highlighting the need for school interventions (Sørlie & Ogden, 2015). Results after 3 years of implementation in the first 28 SWPBS schools in Norway, indicated positive main effects in the small to large range on both severe and less severe problem behaviours, pull-out instruction, and classroom climate (Sørlie & Ogden, 2015), and on teachers' practice and perceived efficacy (Sørlie, Ogden, & Olseth, 2016).

THE PRESENT STUDY

The present study contributes to the literature by using register data and a DiD design to evaluate the effects of SWPBS on a number of primary and secondary outcome indicators. It complements prior Norwegian studies by evaluating effects for all schools that have

implemented SWPBS. Moreover, no studies have explored potential effects of SWPBS on academic achievement and bullying in Norway. Although these are secondary outcomes, it is of interest to search for possible ripple effects on bullying and achievement. There has been growing concerns regarding bullying, and positive school-wide prevention efforts such as SWPBS has been suggested as an effective intervention to reduce bullying (Waasdorp et al., 2012). Academic outcomes are, together with behaviour and attendance, important indicators of school effectiveness (Freeman et al., 2016).

Based on SWPBSs intentional purpose and prior research, the following research question were formulated for this study:

- Does SWPBS have long-term effects on the prevalence of classroom noise and bullying?
- Does SWPBS have long-term effects on students' academic achievements and well-being in school?
- Does SWPBS result in reduced use of pull-out instruction for special need students?

METHODS

Participants

Compulsory education in Norway starts at the age of six and lasts for 10 years, with primary education in grades 1 to 7 and lower secondary education in grades 8 to 10. Schools are publicly funded and very few are private. Compared to other European countries, Norway has an inclusive school setting and few special schools (Sørlie & Ogden, 2015). Students attend school in their local catchment area, and there is no formal tracking by student ability. All Norwegian primary schools (grades 1–7) were included in the study ($N = 2365$). Each school and student has a unique identifier that allows matching of SWPBS schools to population-wide register data. The study combined school-level data on program implementation (fidelity), obtained from The Norwegian Center for Child Behavioural Development (NCCBD), with school-level and student-level register data.

The *unit of analyses* was "school grade cohort," which closely corresponds to birth cohorts, as there is no grade retention in Norway. Outcomes were observed for each school grade cohort over a 8-year period (3 years before to 5 years after SWPBS was initiated) and linked to when the intervention model was implemented. The student-level variables included students born or immigrated to Norway before the age of six, and who completed lower secondary school the calendar year they turned 15, 16 or 17 years. Approximately, 95% finish compulsory school by the age of 16. The matching of students to schools is based on residential address. As explained in detail in Appendix S1 (appendices available online in the Supporting Information section), uncertainty in this matching may result in a slight bias in the effects on standardised national tests, while effects on classroom noise, bullying, well-being, and special needs education are unaffected. Sensitivity analysis in Appendix S1 suggests that effect size is similar using predicted and actual school (Figure A1.1).

The intervention

SWPBS is a structured yet flexible whole-school approach with the main goal to prevent and reduce school problem behaviour, and to promote an inclusive learning environment that can facilitate safety and the psycho-social functioning and learning of all students (Sørlie & Ogden, 2015). The focus is on positive, systematic, data-driven, educative, and reinforcement-based practices conducted within a framework of research-based, collective (school-wide), proactive and predictable approaches. The core features of SWPBS build on decades of research in education, mental health, and behaviour analysis, which in the SWPBS model is organised as a school-wide approach with multiple tiers of support and interventions, and systems to improve fidelity and sustainability (Horner et al., 2014). The prevention model involves all staff and students, and takes approximately 3 to 5 years to fully implement.

The SWPBS is organised according to the principle of matching interventions to students' risk level (Sørlie & Ogden, 2015). More specifically, the intervention model relies on a three-tiered system of evidence-based preventions and supports. Tier I interventions (universal, primary prevention) apply to everyone and all settings in the school with the goal to "prevent problems by defining and teaching consistent behavioural expectations across the school setting and recognizing students for expected and appropriate behaviors" (Lohrmann, Forman, Martin, & Palmieri, 2008, p. 256). Tier II interventions (selected, secondary prevention) are designed for students at moderate risk for severe behaviour problems and who might not respond sufficiently to the universal interventions. The interventions are standardised and mostly delivered in short-term organised small-groups. Tier III (indicated, tertiary prevention) targets the few students with or at high risk of conduct disorder. The interventions at this level are intensive, highly individualised, and multi-modal.

Since 2002, SWPBS has been implemented in 244 primary schools in Norway (9%). The core model components and implementation structure of the Norwegian SWPBS model (called N-PALS) are equal to the U.S. version. The core components are: (a) SWPBS strategies including teaching of school rules, positive

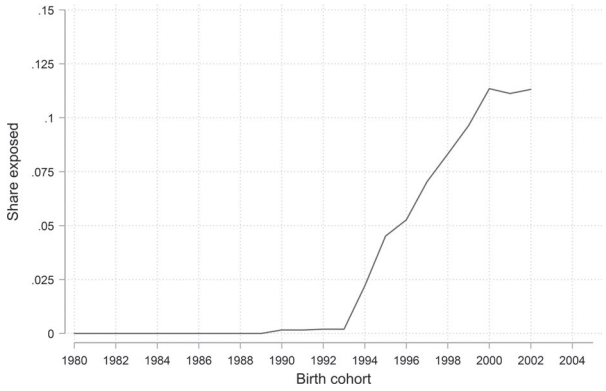

Figure 1. Percentage of students exposed to SWPBS during grade 5 to 7 across birth cohorts.

expectations, systematic encouragement of positive behaviour, (b) monitoring of student behaviour using the school-wide-information system, (c) school-wide corrections with mild and immediate consequences, (d) time-limited small group instruction for students at risk, (e) individual interventions and support plans for high-risk students, (f) classroom management skills for teachers, and (g) parent information and collaboration strategies. Except for minor adaptations of the training materials (e.g., pictures, videos, response cards, concepts), no changes were made when SWPBS was transferred to Norway. N-PALS does not include any evidence-based interventions to promote academic performance, similar to the original model.

The school's readiness for implementation was initially assessed, and approval from at least 80% of the staff was required. Each school appointed a representative team (five persons) who were trained on a monthly basis to plan, inform, carry out, monitor, and report on the intervention outcomes at their school. The teams were locally trained and supervised by a coach for 2 years (10 sessions/2 hours per year). The coaches were trained (1 year) and certified by the national implementation team at NCCBD. All training was nationally standardised and free of charge (except travel costs). The school team trained the rest of the staff in key features and intervention components and attended four half-day regional booster session per year. The schools used various web-based feedback systems based on nationally standardised assessment tools to secure data-based decisions and fidelity.

In this study, students are considered exposed to the intervention if they attend grades 4 to 7 in a school that implements SWPBS. The share of exposed students has increased from close to zero for the 1994 cohort to about 11% of those born in 2000 and later (Figure 1). Across the sample of nine cohorts, about 8% were ever exposed to the program.

Measures

Primary outcomes

Prior studies have used reliable multi-item assessment scales to capture changes in more and less severe externalising behaviours within and outside the classroom context (Sørlie & Ogden, 2015). Because such measures are not available in registers, single items from annual nation-wide surveys among all seventh graders in Norway (>90% response rate) were selected as proxy variables. While Sørlie and Ogden (2015) used teacher assessments, this study used student-reported frequency of *classroom noise* and *bullying* in the second semester of seventh grade (age 13), obtained from the Pupil Survey administrated by the Norwegian Directorate for Education and Training. The item classroom noise was measured on a five-point scale (*1 = fully agree with classroom order, 5 = fully disagree with classroom order*), and was for the analyses standardised to $M = 0$, $SD = 1$. Bullying was measured by asking students how often they had been bullied by peers at school during the last months. Bullied was defined as the share of students being bullied at least 2–3 times a month.

Secondary outcomes

Student-reported well-being in school from the Pupil Survey was measured in the seventh grade with one item (*5 = enjoy school very much, 1 = does not enjoy school at all*) and standardised to $M = 0$, $SD = 1$. Academic performance refers to the schools' average score on standardised national tests in literacy, English (foreign language), and numeracy. Standardised national testes in literacy, English, and numeracy were standardised for each student ($M = 0$, SD 1), averaged for each student, and then school averages were computed. All tests were performed early in the eighth grade (age 13), the first semester after leaving a SWPBS school. Data on special needs education were obtained from the nation-wide compulsory education information system Grunnskolens Informasjonssystem (GSI), administrated by The Norwegian Directorate for Education and Training. These data were reported by school staff and are measured in this study at the overall school level. The share of special education students and the share receiving most of their instruction outside ordinary class (i.e., pull-out instruction) were included as secondary outcome indicators.

Treatment (intervention)

The treatment indicator tracks the position of each school grade cohort relative to the year of program implementation. Students finishing primary school (grade 7) just before SWPBS was implemented were labelled −1. The next cohort, exposed to SWPBS for 1 year (grade

TABLE 1
Descriptive statistics. Outcomes and student composition

	N	M	SD_{Total}	$SD_{Between}$	SD_{Within}
Panel A: Student outcomes					
Classroom noise	11,591	2.61	0.477	0.269	0.404
Bullied (=1)	17,374	0.074	0.068	0.033	0.061
Academic performance	19,245	−0.002	0.305	0.227	0.205
Well-being	17,375	4.26	0.276	0.156	0.234
Pull-out instruction (=1)	21,080	0.053	0.040	0.028	0.029
Special education (=1)	21,079	0.070	0.038	0.031	0.023
Panel B: Student composition					
Girls (=1)	11,591	0.489	0.107	0.054	0.096
Fathers' education	11,590	4.46	0.623	0.545	0.303
Mothers' education	11,591	4.66	0.593	0.500	0.327
Fathers' earnings	11,590	441	100	89	44.6
Mothers' earnings	11,591	228	60.3	50.1	33.6
Native Norwegians (=1)	11,591	0.919	0.129	0.119	0.043
First-generation immigrants (=1)	11,591	0.024	0.038	0.023	0.030
Second-generation immigrants (=1)	11,591	0.056	0.112	0.106	0.031

Note: All results with student weights. The *SD*s in the $SD_{Between}$ and SD_{Within} columns show the degree to which the outcome variables vary between and within schools, respectively. Dummy variables are indicated with (=1), and the mean of the dummy variables equal the proportion of cases with a value of 1.

7), was labelled 1. Cohort 4 was the first cohort exposed through grades 4 to 7. For each school grade cohort outcomes in grade 7 or 8 were analysed.

Control variables

The following variables from register data were included as covariates; student composition within schools using gender, mother's and father's level of education (nine dummies for each parent, from no education to PhD), mother's and father's earnings and earnings squared (in 1000 Norwegian Krone), immigrant background (six dummies), and interactions between school county (dummies) and school cohort (dummies). Descriptive statistics are presented in panel B of Table 1.

Fidelity

The effective behaviour support self-assessment survey (EBS, 46 items) was routinely completed each year (from 2008 onwards) by teachers and staff in all intervention schools and used as measure of perceived program fidelity, and explained in more detail in Appendix S4.

Analytic approach

Schools implementing interventions like SWPBS may differ from other schools (e.g., higher levels of problem behaviour, more proactive school management). To account for selection of schools into treatment and time effects common to all schools, a DiD design was preferred. This design compares changes in outcomes within schools following implementation of SWPBS with corresponding changes in other schools. A linear regression model controlling for unobserved persistent differences between schools, general time trends, and time-varying differences in student composition was estimated (see Appendix S2 for details). An advantage of DiD is that it accounts for all time-invariant differences between schools, such as stable school traits, teacher characteristics, and student characteristics, irrespective of proxies for these differences.

The key identifying assumption is that the outcomes would have evolved similarly over time in both intervention and control schools absent of implementing SWPBS (net of time-varying covariates). This "parallel trends"-assumption is untestable, but its credibility can be tested indirectly by comparing trends for SWPBS and non-SWPBS schools prior to implementation. Before implementation, there may be between-school differences, but these should be stable.

The DiD model was estimated with the year prior to implementation as reference category. After estimating the regression model, a linear combination of coefficients was used to rescale the coefficients to the difference from an average of 1, 2, and 3 years before implementation. A summary measure of the effects of SWPBS is provided by comparing the post period (2–5 years after program initiation, reflecting the minimum time considered necessary to fully implement SWPBS) with the difference from the pre-period (1–3 years before implementation). Averaging effects over several years increases the statistical power when studying persistent effects.

RESULTS

Data description

Table 1 shows descriptive statistics for outcomes (panel A) and control variables (panel B). School-by-grade cells

TABLE 2
Average fixed effects difference between SWPBS schools and control schools during pre-intervention years

	(1)	*(2)*
Classroom noise	−0.048***	−0.023
Bullied	0.0094***	0.0027**
Academic performance	0.016*	0.0074
Well-being	−0.004	0.002
Pull-out instruction	0.0050***	0.0014
Special education	−0.0006	−0.0012
Student composition controls	No	Yes

Note: Results in column (1) are without any control variables while the results in column (2) includes the following student composition variables: Share female, fathers' and mothers' education and earnings, immigrant background, birth cohort, school county X birth cohort.
*$p < .10$, **$p < .05$, ***$p < .01$

are weighted with number of students. The outcomes (except test scores) are shown in original units without standardisation. Substantially, fewer number of observations for classroom noise reflect that registration of this question started later than bullying and well-being, meaning that the number of available cohorts are fewer. On average, 7.4% of the students were bullied, 7.0% received special education, and 5.3% received most instruction outside regular class (i.e., the mean of the dummy variables). While student composition, test scores, and share of special education students mostly varied between schools, the remaining outcomes varied as much or more over time within schools.

Compared to other Norwegian schools, the intervention schools tended to have less classroom noise, more bullying, higher test scores, and more pull-out instruction *prior* to SWPBS (Table 2). Significant difference in bullying remained after controlling for student composition, highlighting the need for a research design able to correct for stable differences in outcomes not related to observable proxies.

Validity of research design

To check the validity of the identification strategy, we analysed whether there was evidence of differential changes in SWPBS and control schools before initiation of SWPBS. If such "placebo effects" were significant, this would indicate confounding variation and the research design could not be trusted. Figure 2 presents the estimated "placebo effects" (pre-implementation cohorts) and estimated program effects. The placebo effect estimates were small and mostly insignificant, indicating little evidence of differential changes and systematic trends in the intervention schools (relative to control schools) before implementation. We concluded that the research design seemed valid, and that the effect estimates are informative.

Main program effects

The presumably valid estimates of program effects are presented in Figure 2 by number of years after initiation of SWPBS, which correspond to years exposed (coefficients in Table A3.1, Appendix S3). For the primary outcomes, there were indications of reduced classroom noise. The estimate after 2 years was significant at the 5% level, but the estimates for subsequent cohorts, while systematically negative, were not significant. No intervention effect was observed on bullying or on any of the secondary outcome indicators.

In order to increase precision and power, we merged outcome periods and estimated average effects for years 1 to 5 and 2 to 5 after initiation of SWPBS. Results are presented in Table 3. Given that SWPBS is expected to take several years to implement, the estimates for years 2 to 5 are the most relevant. For classroom noise, we found an average reduction of 5.7% of an *SD* for years 1 to 5 and of 7.5% for years 2 to 5. The former is not significant, while the latter is significant at the 10% level with the 95% confidence interval (CI) ranging from −16.3 to +1.4. For bullying, we found a non-significant increase of, respectively, 0.6 and 0.7% points (CI for years 2–5 ranged from −0.2 to +1.5). Estimated average effects for test scores, well-being, special education, and pull-out instruction were all insignificant.

Effects in schools with high fidelity

From prior research, it was expected that high fidelity was required to produce an effect of SWPBS (e.g., Bradshaw et al., 2010; Sørlie & Ogden, 2015). Results from the EBS indicated that only 30% of the Norwegian schools had implemented SWPBS with sufficient fidelity (80%) within 3 years after initiation of SWPBS. However, Figure 3 shows that there was no evidence of differential effects when effect estimates in schools with high- and low-fidelity scores were compared (for detailed results on fidelity and details on Figure 3, see Appendix S4).

DISCUSSION

In this article, a credible non-experimental research design and population-level longitudinal registry data were used to study school-level effects of SWPBS in Norwegian primary schools. Indications of reduced classroom noise (primary outcome) were found. This is in line with several previous studies, including RCTs and credible non-experimental designs, where reduced student problem behaviour follows from implementation of the SWPBS model (e.g., Benner et al., 2012; Bradshaw et al., 2012). While several previous studies are based on teacher-assessed outcomes, the present study is notable in

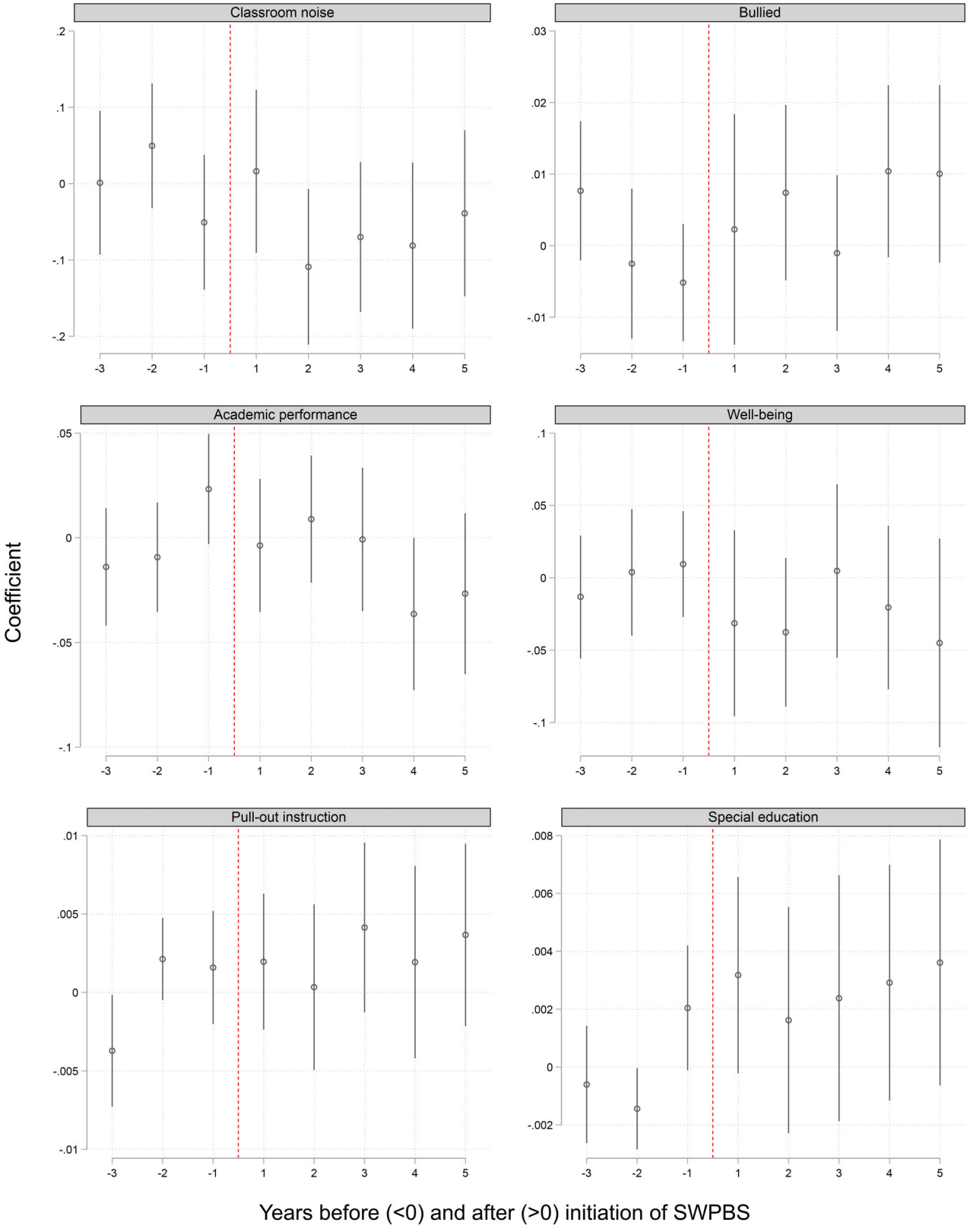

Figure 2. DiD effect estimates (after) and pre-program heterogeneity with 95% CI. Note: Outcome metrics: Standardised for classroom noise, academic performance, and well-being. Observed share for bullied, pull-out instruction, and special education. The dotted line separates coefficients before (−3 to −1) and after (1 to 5) initiation of SWPBS. [Colour figure can be viewed at wileyonlinelibrary.com].

TABLE 3
Effect estimates

	(1) Classroom noise	*(2) Bullied*	*(3) Academic performance*	*(4) Well-being*	*(5) Pull-out instruction*	*(6) Special education*
1–5 years	−0.057 (0.042)	0.0058 (0.0042)	−0.012 (0.012)	−0.026 (0.023)	0.0024 (0.0023)	0.0027 (0.0018)
2–5 years	−0.075* (0.044)	0.0067 (0.0044)	−0.014 (0.013)	−0.025 (0.023)	0.0025 (0.0024)	0.0026 (0.0019)
N	11,464	17,242	19,079	17,242	20,882	20,881

Note: Cluster robust standard errors in parentheses. Outcome metrics: Standardised for classroom noise, academic performance, and well-being. Actual share for bullied, pull-out instruction, and special education. Student composition controls: Share female, fathers' and mothers' education and earnings, immigrant background, birth cohort, school county X birth cohort
$^{*}p < .10$, $^{**}p < .05$, $^{***}p < .01$

finding indications of effects on classroom noise reported by students.

The present study is unable to detect effects on a range of other outcomes. Contrary to results from a previous study (Waasdorp et al., 2012), no effects were found on bullying (primary outcome). Likewise, the effects on students' well-being and academic performance as well as pull-out instruction (secondary outcomes) were all close to zero and insignificant. The lack of effect on test scores in the current study is in line with results from previous high-quality studies (e.g., Bradshaw et al., 2010; Horner et al., 2009), and no academic support is included in SWPBS in Norway.

The present study is an example of a study where changes in at-scale outcomes do not match what one would expect from smaller-scales studies (Eisner & Malti, 2015). In general, it is hard to evaluate what other developments contribute to aggregate changes in outcomes. The DiD design explicitly accounts for sources of change across cohorts shared by program and control schools. Our effect estimates reflect observed outcomes compared to the outcomes one would have expected in the absence of SWPBS, including that the schools may initiate other interventions or act differently in other ways. The lack of substantial effects may partly be due to that many control schools implement other programs (Bradshaw et al., 2010). In the current study, data on other programs were not available. Thus, we do not know whether SWPBS replaces other, equally effective programs.

Another possible explanation of the limited effects in the present study is that few schools implement with fidelity. Low fidelity is a widespread problem in SWPBS schools, with only two out of 10 effectiveness studies reporting high fidelity among schools (Chitiyo et al., 2012). However, in a previous study including 28 of the first SWPBS schools in Norway, 75% implemented with high fidelity after 3 years (Sørlie & Ogden, 2015). High fidelity is also reported in RCTs from the United States (Pas & Bradshaw, 2012). The present study estimates average intervention effects of SWPBS in all Norwegian schools which have implemented SWPBS ($N = 244$). Only 30% of the schools implemented with fidelity after 3 years, suggesting that fidelity is a major challenge in scale-up of SWPBS.

When fidelity is low in many schools, the estimated average effect will reflect this. We examined whether the effects were stronger for high-fidelity schools without finding evidence of differential effects, but the small number of such schools made these estimates less precise. However, reliable assessment of fidelity and investigating differential effects by fidelity is more problematic than acknowledged in the literature. For example, schools with high perceived fidelity or that voluntarily continue to answer surveys may be schools with better outcomes and/or greater initial motivation for change. Restricting analyses based on fidelity scores or survey response may confound fidelity and other characteristics of schools, and thus give biased estimates.

Other possible explanations for the partly conflicting results with previous studies relate to data and methods. While most studies have used teacher-assessed outcome variables, particularly office discipline referrals (Chitiyo et al., 2012), the outcomes in the present study were student-assessed survey data, test scores, and register data. For example, Sørlie and Ogden (2015) found intervention effects based on teacher assessments but not on student-assessed outcomes. On methods, we study long-run effects and take pre-existing hard to observe differences in to account. Pre-existing school differences were found to affect the effect estimates unless they were adequately accounted for, suggesting that different empirical strategies across studies may partly explain the contrast between effects reported by the current and prior studies.

Strengths and limitations

Few studies have evaluated SWPBS using rigorous designs (Chitiyo et al., 2012). The major strength of the present study is national coverage and longitudinal data with very limited attrition. The register data were collected yearly by school authorities in a consistent way for all schools and students in grades 7 and 8, independent of SWPBS. The response rates for the survey-based student-reported outcomes are very high compared to other surveys. The test score data have similar high participation rates. We also study whether there is differential attrition in SWPBS and controls schools, finding

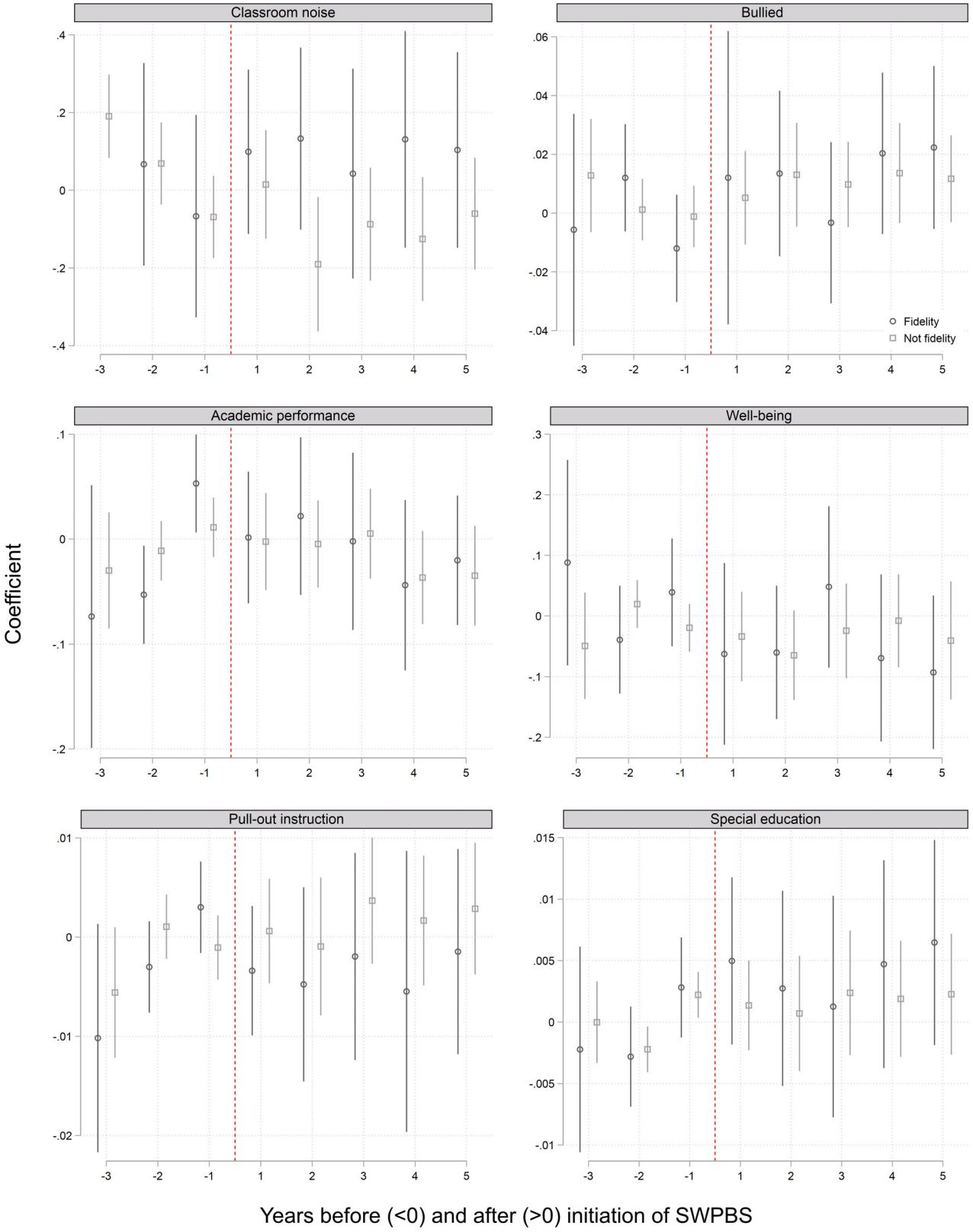

Figure 3. Comparing schools implementing with high fidelity (≥80%) and those that do not (<80%). Note: Outcome metrics: Standardised for classroom noise, academic performance, and well-being. Observed share for bullied, pull-out instruction, and special education. The dotted line separates coefficients before (−3 to −1) and after (1 to 5) initiation of SWPBS. Results shown with 95% CI. [Colour figure can be viewed at wileyonlinelibrary.com].

no evidence of this, and no indication of biased effect estimates due to differential attrition (see Appendix S6).

The data include school outcomes (as reported by students) several years after the initiation of SWPBS, and allow for investigation of more long-lasting effects of SWPBS than in previous studies (Madigan et al., 2016). The key benefit of the DiD design used in this paper is the avoidance of bias from factors changing over time unrelated to SWPBS as well as unobserved persistent differences between schools. A less comprehensive design would have given misleading estimates (see Appendix S5). For example, after implementing SWPBS, the intervention schools had more classroom noise and higher levels of bullying than other schools. However, this was the case also before implementing SWPBS, and thus not an effect of the program.

Although we conclude that the data are informative, and the design is valid, there are several important limitations. The present study focuses on average (school-level) effects and the design possesses sufficient power to rule out relatively small average effects. For example, for average effects on bullying over years 2 to 5, significant effects were detectable if they exceeded 0.2% of a *SD*. However, there may be larger effects for subgroups of students (e.g., students at elevated risk). Even if the student response rates are high, at-risk students may be overrepresented among those not answering. Similarly, at-risk students may be overrepresented among the relatively few students that do not take the standardised tests. Thus, while the estimates are informative about average effects, they do not include evidence on effects for particularly targeted students.

In terms of reliability and validity, there are both benefits and limitations from studying variables collected for other purposes than program evaluation. Moreover, there are pro and cons of student-reported outcomes compared to teacher-reported data. To ensure comparability between survey years, outcomes are based on single items and this may reduce reliability. Regarding validity, the variables studied may not be sensitive to potential changes caused by SWPBS. The same variables are used by Norwegian educational authorities to monitor bullying, well-being, and academic performance. The bullying variable is based on the Olweus Bullying Questionnaire and has been shown to have high reliability and validity (Olweus, 2013). Research also shows that test scores predicts students' later outcomes. However, analysis of the validity of classroom noise and well-being are scarce. Additionally, while effects on variables collected for other purposes would suggest generalizable effects beyond the SWPBS constructs, there remains a question to what extent the outcome variables match the program objectives.

It may also be hard to get consistent measures of effects based on subjectively assessed outcomes, using reports from either students or teachers. In general, analyses of the Norwegian student and teacher surveys mostly show high internal reliability across items within survey and topic, and moderate positive correlations between student and teacher responses on similar items across surveys. There are further complications studying effects of SWPBS related to expectations and perceptions. For example, even if disruptive behaviour is objectively reduced, students and teachers may adjust their expectations and the effect estimates will be bias towards zero. Additionally, both teachers' and students' perception of disruptive behaviour may change as an effect of the program, which can give a positive or negative bias (Gage et al., 2018).

CONCLUSION

In this paper, population-wide register data and a DiD design were used to evaluate school-level effects in a scale-up of SWPBS in Norway on classroom noise, bullying, well-being, academic achievement, and special needs education. While some evidence of reduced classroom noise was found, there were no significant effects on other outcomes. No effect on academic achievement is in line with previous high-quality studies, while for other outcomes less favourable effects were found than in previous studies, including RCTs from the US. Intervention effects are likely stronger in effectiveness trials where the program is evaluated under more optimal conditions of delivery (i.e., higher fidelity), which is one likely explanation for the less favourable outcomes in the present study. Another explanation is that outcome variables selected from register data are less suited for measuring primary outcome effects of SWPBS. The DiD design is generally considered to be a credible identification strategy when randomization is not feasible. The register data allow for studying all schools, irrespective of implementation quality or motivation for answering program-provided surveys, providing precise and arguably unbiased estimates. Schools were followed for several years before and after initiation of SWPBS. This enabled both analyses of longer-term effects than previous studies, investigation of differences between schools prior to implementation, and evaluation of the credibility of the DiD design.

We found that DiD was valid in this particular case, and that less comprehensive designs would provide misleading results. Changes in outcomes across student cohorts, unrelated to SWPBS, imply that a before-after comparison within SWPBS would give biased estimates of program effects. Likewise, comparing intervention schools with other (control) schools disregarding pre-existing differences would also have provided biased effect estimates. The current study exemplifies the relevance of the DiD design as well as a the usefulness and limitations of using register data in future evaluations of school interventions.

Manuscript received May 2018
Revised manuscript accepted January 2019
First published online April 2019

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Appendix S1. School and student identification.

Appendix S2. The DiD model.

Appendix S3. Coefficients from Figure 2.

Appendix S4. Fidelity.

Appendix S5. School heterogeneity illustrated by fixed effects distributions.

Appendix S6. Response rate.

Appendix S7. Raw differences between SWPBS schools and control schools.

REFERENCES

Benner, G. J., Nelson, J. R., Sanders, E. A., & Ralston, N. C. (2012). Behavior intervention for students with externalizing behavior problems: Primary-level standard protocol. *Exceptional Children*, *78*(2), 181–198.

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.

Bradshaw, C. P., Waasdorp, T. E., & Leaf, P. J. (2012). Effects of school-wide positive behavioral interventions and supports on child behavior problems. *Pediatrics*, *130*(5), e1136–e1145.

Bradshaw, C. P., Waasdorp, T. E., & Leaf, P. J. (2015). Examining variation in the impact of school-wide positive behavioral interventions and supports: Findings from a randomized controlled effectiveness trial. *Journal of Educational Psychology*, *107*(2), 546–557.

Caldarella, P., Shatzer, R. H., Gray, K. M., Young, K. R., & Young, E. L. (2011). The effects of school-wide positive behavior support on middle school climate and student outcomes. *RMLE Online*, *35*(4), 1–14.

Carrell, S. E., & Hoekstra, M. L. (2010). Externalities in the classroom: How children exposed to domestic violence affect everyone's kids. *American Economic Journal: Applied Economics*, *2*(1), 211–228.

Chitiyo, M., May, M. E., & Chitiyo, G. (2012). An assessment of the evidence-base for school-wide positive behavior support. *Education and Treatment of Children*, *35*, 1–24.

Domitrovich, C. E., Bradshaw, C. P., Poduska, J. M., Hoagwood, K., Buckley, J. A., Olin, S., … Ialongo, N. S. (2008). Maximizing the implementation quality of evidence-based preventive interventions in schools: A conceptual framework. *Advances in School Mental Health Promotion*, *1*(3), 6–28.

Eisner, M. P., & Malti, T. (2015). Aggressive and violent behavior. In M. E. Lamb & R. M. Lerner (Eds.), *Handbook of child psychology and developmental science: Socioemotional processes* (pp. 794–841). Hoboken, NJ: John Wiley & Sons Inc.

Flay, B. R., Biglan, A., Boruch, R. F., Castro, F. G., Gottfredson, D., Kellam, S., … Ji, P. (2005). Standards of evidence: Criteria for efficacy, effectiveness and dissemination. *Prevention Science*, *6*(3), 151–175.

Freeman, J., Simonsen, B., McCoach, D. B., Sugai, G., Lombardi, A., & Horner, R. (2016). Relationship between school-wide positive behavior interventions and supports and academic, attendance, and behavior outcomes in high schools. *Journal of Positive Behavior Interventions*, *18*(1), 41–51. https://doi.org/10.1177/1098300715580992

Gage, N. A., Leite, W., Childs, K., & Kincaid, D. (2017). Average treatment effect of school-wide positive behavioral interventions and supports on school-level academic achievement in Florida. *Journal of Positive Behavior Interventions*, *19*(3), 158–167. https://doi.org/10.1177/1098300717693556

Gage, N. A., Sugai, G., Lewis, T. J., & Brzozowy, S. (2015). Academic achievement and school-wide positive behavior supports. *Journal of Disability Policy Studies*, *25*(4), 199–209.

Gage, N. A., Whitford, D. K., & Katsiyannis, A. (2018). A review of schoolwide positive behavior interventions and supports as a framework for reducing disciplinary exclusions. *The Journal of Special Education*, *52*(3), 142–151.

Horner, R. H., Kincaid, D., Sugai, G., Lewis, T., Eber, L., Barrett, S., … Boezio, C. (2014). Scaling up school-wide positive behavioral interventions and supports: Experiences of seven states with documented success. *Journal of Positive Behavior Interventions*, *16*(4), 197–208.

Horner, R. H., Sugai, G., & Anderson, C. M. (2010). Examining the evidence base for school-wide positive behavior support. *Focus on Exceptional Children*, *42*(8), 1.

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.

Lane, K. L., Wehby, J. H., Robertson, E. J., & Rogers, L. A. (2007). How do different types of high school students respond to schoolwide positive behavior support programs? Characteristics and responsiveness of teacher-identified students. *Journal of Emotional and Behavioral Disorders*, *15*(1), 3–20.

Lohrmann, S., Forman, S., Martin, S., & Palmieri, M. (2008). Understanding school personnel's resistance to adopting schoolwide positive behavior support at a universal level of intervention. *Journal of Positive Behavior Interventions*, *10*(4), 256–269.

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

Mitchell, B. S., Hatton, H., & Lewis, T. J. (2018). An examination of the evidence-base of school-wide positive behavior interventions and supports through two quality appraisal processes. *Journal of Positive Behavior Interventions*, *20*(4), 239–250.

Olweus, D. (2013). School bullying: Development and some important challenges. *Annual Review of Clinical Psychology*, *9*, 751–780.

Pas, E. T., & Bradshaw, C. P. (2012). Examining the association between implementation and outcomes. *The Journal of Behavioral Health Services & Research*, *39*(4), 417–433.

Smith, J. A., & Todd, P. E. (2005). Does matching overcome LaLonde's critique of nonexperimental estimators? *Journal of Econometrics*, *125*(1), 305–353.

Sørlie, M.-A., & Ogden, T. (2007). Immediate impacts of PALS: A school-wide multi-level programme targeting behaviour problems in elementary school. *Scandinavian Journal of Educational Research*, *51*(5), 471–492.

Sørlie, M.-A., & Ogden, T. (2015). School-wide positive behavior support–Norway: Impacts on problem behavior and classroom climate. *International Journal of School & Educational Psychology*, *3*(3), 202–217.

Sørlie, M.-A., Ogden, T., & Olseth, A. R. (2016). Examining teacher outcomes of the school-wide positive behavior support model in Norway: Perceived efficacy and behavior management. *SAGE Open*, *6*(2), 1–13.

Stuart, E. A., Bradshaw, C. P., & Leaf, P. J. (2015). Assessing the generalizability of randomized trial results to target populations. *Prevention Science*, *16*(3), 475–485.

Stuart, E. A., & Rhodes, A. (2017). Generalizing treatment effect estimates from sample to population: A case study in the difficulties of finding sufficient data. *Evaluation Review*, *41*(4), 357–388.

Waasdorp, T. E., Bradshaw, C. P., & Leaf, P. J. (2012). The impact of schoolwide positive behavioral interventions and supports on bullying and peer rejection: A randomized controlled effectiveness trial. *Archives of Pediatrics & Adolescent Medicine*, *166*(2), 149–156.