Using Complier Average Causal Effect Estimation (CACE) to Determine the Impacts of the Good Behavior Game Preventive Intervention on Teacher Implementers

Juliette K. Berg[1]

Catherine P. Bradshaw[1]

Booil Jo[2]

Nicholas S. Ialongo[3]

[1]University of Virginia, Curry School of Education

[2]Stanford University, Department of Psychiatry and Behavioral Sciences

[3]Johns Hopkins University, Bloomberg School of Public Health

Berg, J. K., Bradshaw, C. P., Jo, B., & Ialongo, N. S. (2017). Using Complier Average Causal Effect Estimation to Determine the Impacts of the Good Behavior Game Preventive Intervention on Teacher Implementers. *Administration and Policy in Mental Health and Mental Health Services Research*, *44*(4), 558–571. https://doi.org/10.1007/s10488-016-0738-1

Published in *Administration and Policy in Mental Health and Mental Health Services Research*

Funding: This research was supported in part by grants from the Institute of Education [R305A080326; R305A130060; R305A130701] and the National Institute of Mental Health [P30 MH08643].

Acknowledgements: The authors would like to thank the administration and the teachers of the Baltimore City School District who are devoted educators of children. We also thank Celene Domitrovich for her contribution to the project.

Author Note: Correspondence regarding this manuscript may be addressed to Juliette Berg, Curry School of Education, University of Virginia, PO Box 400281, Charlottesville, VA 22904. Electronic mail may be sent to berg.juliette@gmail.com.

Abstract

Complier Average Causal Effect (CACE) analysis, a causal inference approach that accounts for levels of teacher implementation compliance, was used to examine one-year impacts of PAX Good Behavior Game (PAX GBG) and Promoting Alternative Thinking Strategies (PATHS) on teacher efficacy and burnout. Teachers in 27 elementary schools were randomized to PAX GBG, an integration of PAX GBG and PATHS, or a control condition. There were positive overall effects on teachers' efficacy beliefs, but high implementing teachers also reported increases in burnout across the school year. This approach may offer new information not captured using the traditional intent-to-treat approach.

Keywords: whole school intervention, implementation, causal inference, elementary school, teacher effects

Using Complier Average Causal Effect Estimation (CACE) to Determine the Impacts of the Good Behavior Game Preventive Intervention on Teacher Implementers

There is increased focus on the use of universal school-based interventions to promote a range of academic and behavioral outcomes for students and possibly staff; however, it is common for there to be variation in the extent to which teachers fully comply with the intended implementation model (Domitrovich et al., 2009). Most intervention studies examining the effects of school-based programs have taken an intent-to-treat (ITT) approach, whereby the researchers estimate the effect of being assigned to the treatment condition (Schochet, Puma, & Deke, 2014). However, in the case of low program implementation, the effect of being assigned to treatment may substantially differ from the effects of the treatment on those who are assigned and participate. More specifically, when implementation is low, there are typically small or null effects on those who do not fully implement, and thus the ITT estimates may understate the effect of intervention (Stuart, Perry, Le, & Ialongo, 2008).

An alternative to traditional ITT analysis is the complier-average causal effect (CACE) analysis approach, which has been successfully used to estimate treatment effects accounting for compliance (see Angrist, Imbens, & Rubin, 1996; Little & Yau, 1998; Jo, 2002; Stuart et al., 2008). Both ITT and CACE approaches are useful for gaining a more complete understanding of intervention effects (Jo, 2002, Stuart et al., 2008). The CACE method has recently been applied to the estimation of treatment effects with noncompliance of several randomized interventions serving families and youth (e.g., Barnard, Frangakis, Hill, & Rubin; 2003; Connell, Dishion, Yasui, & Kavanagh, 2007; Stanger, Ryan, Fu, & Budney, 2011); however, there has been less focus on classroom-based preventive interventions implemented by teachers. In the current study, we used the CACE method to estimate the impacts of a commonly used classroom-based preventive intervention called the Good Behavior Game (see Bradshaw, Zmuda, Kellam & Ialongo, 2009; Ialongo et al., 1999, 2001; Kellam et al., 2008) on teachers' self-efficacy and burnout over the course of a school year. This universal, behavioral management model was combined with a social-emotional learning intervention and implemented in elementary schools. The overall goal of the current study was to provide an example of CACE analysis as applied to a classroom-based intervention. This study represents a novel extension and application of this analytic approach to better understand the impact of the intervention, which had varying levels of implementation across teachers.

Teachers' Compliance With School-based Program Implementation

Teachers are often the primary implementers of classroom-based preventive interventions, yet the degree to which they opt to implement the various components of the intervention often varies (Domitrovich et al., 2009). In fact, many of the efficacy and effectiveness studies of school-based prevention programs have noted considerable variation in implementation quality, which in turn attenuates program impacts on student and staff outcomes (Durlak & Dupree, 2008; Durlak, Weissberg, Dymnicki, Taylor, & Schellinger, 2011; Ringeisen et al., 2003). There is emerging evidence that certain characteristics of implementers, including teacher characteristics of attitudes and beliefs about themselves or the school environment, are associated with variation in implementation (Domitrovich, Bradshaw et al., in press; Payne & Eckert, 2010). Much of the exploration into the effects associated with variation in implementation has been descriptive and post hoc, with limited use of causal inference approaches and a lack of clarity about the direction of effects. Nevertheless, this line of research suggests that poor implementation and characteristics systematically associated with variation in implementation are typically unmeasured or not accounted for in traditional randomized trials employing an ITT approach. This source of bias may in turn result in an under-estimation of intervention effects. Additional research is needed to demonstrate a causal impact of programs, while taking into consideration program implementation. In fact, the assumption has been that increased compliance with intervention implementation translates into better outcomes (Botvin, Baker, Dusenbury, Botvin, & Diaz, 1995; Derzon et al., 2005; Durlak & Dupree, 2008; Rohrbach et al., 1993). The current proposal focused on implementation or compliance, which is defined as “the discrepancy between what is planned and what is actually delivered when an intervention is conducted” (Domitrovich, Bradshaw et al., 2008, p. 7, also see Chen, 1998; Hulleman & Cordray, 2009; O’Donnell, 2008). A common indicator of compliance of school-based interventions is program dosage, which includes the frequency with which the program is implemented, with the expectation that a higher dosage is associated with better outcomes.

Background on the Interventions

The current study tested two evidence-based elementary school prevention programs: the PAX version of the Good Behavior Game (PAX GBG; Embry et al., 2003) and Promoting Alternative Thinking Strategies (PATHS; Greenberg, Kusché, & CPRG, 2011; Kusché, Greenberg, & CPPRG, 2011). Specifically, a randomized controlled trial (RCT) design was used to compare these two intervention models against a control group. The first intervention model was the PAX GBG alone. The second model was the combination of the PAX GBG and PATHS (Domitrovich, Bradshaw, Greenberg, Embry, Poduska, & Ialongo, 2010). PAX GBG focuses on providing teachers with an efficient way of reinforcing the inhibition of aggressive/disruptive and off-task behavior in a “game” like context (Embry et al., 2003). The PATHS curriculum trains teachers to provide explicit instruction to students to promote the development of emotional awareness and communication, self-regulation, social problem solving, and relationship management skills (e.g., interpersonal skills, conflict management) through didactic lessons that take place weekly across the school year (Greenberg & Kusche, 2006). Several large RCTs of GBG have demonstrated positive effects on student peer relations, aggressive/off-task behavior, substance use, and academic outcomes (e.g., Bradshaw et al., 2009; Ialongo et al., 1999, 2001; Kellam et al., 2008). Similarly, prior RCTs of PATHS have yielded positive effects on student social-emotional skills, peer relations, prosocial cognitive functioning, socially-competent behaviors, and behavioral adjustment (e.g., CPPRG, 1999; Greenberg & Kusche, 2006; Greenberg et al., 1995).

Although most of the focus has been on impacts of PAX GBG and PATHS, as well as other whole-school social-emotional programs, on students, teachers implementing these programs may experience benefits such as increased efficacy in managing their classrooms and reduced emotional exhaustion and other forms of burnout (Bradshaw, Koth, Bevans, Ialongo, & Leaf, 2008; Bradshaw, Koth, Thornton, & Leaf, 2009; Han & Weiss, 2005). On the other hand, the additional burden placed on teachers to implement the program may unintentionally cause some teachers to experience increased burnout and stress. Impacts on teachers, whether positive or negative, may be secondary effects of the program's impacts on students, may stem from teachers' involvement in the training component of the intervention, or may be a function of the supports accompanying the intervention (see Domitrovich et al., 2015 for a more extensive discussion). Further, impacts on students may be a function of how engaged teachers are in implementing the components of the intervention. Greater involvement could result in positive effects as teachers learn to better manage their classrooms, or in negative effects as teachers' burden increases. The effects on teachers of implementing classroom-based interventions, positive or negative, likely have important implications for how effective they are at producing positive effects on students. Yet few studies have specifically tested the impacts of student-focused classroom-based interventions on teacher outcomes (Domitrovich et al., 2015).

Study Design

Data for this study came from 27 elementary schools in a large urban, east coast public school district. Schools were recruited and principals agreed to participate in a randomized controlled trial of two intervention models and to potentially receive one year of training and coaching. Schools were then randomized (i.e., cluster randomized trial) to one of three conditions: the PAX GBG only (9 schools), the integration of PAX GBG and PATHS (referred to as PATHS to PAX) (9 schools), and a control condition (9 schools) where teachers conducted their usual practice. The study took place over the course of one school year. A novel aspect of the design of the current study was the plan to contrast the PAX GBG classroom management model when implemented alone with an integrated training, combining PAX GBG with the PATHS social-emotional learning program. In the current study, we were particularly interested in impacts on teacher outcomes, rather than the traditional impacts solely on students. In fact, as noted above, both PAX GBG and PATHS have the potential to positively impact teacher outcomes of burnout and efficacy, as a function of their positive impact on classroom management and student behavior; however, no studies have taken into consideration compliance when examining these effects. Specifically, our prior analysis of data from this trial using an ITT approach suggested that teachers in the integrated condition reported feeling more efficacious and feeling more personal accomplishment relative to control teachers; however, they did not report reduced levels of emotional exhaustion or depersonalization (Domitrovich et al., 2015).

In the current study, we operationalized implementation compliance as the teachers' use of the PAX GBG "games" in the classroom using records of how many games they played throughout the school year and for how long they played each game. More specifically, compliance was defined as being above a cut point on both the number of games played and the total number of minutes of games played. Based on our prior ITT findings (Domitrovich et al., 2015), we expected to find stronger effects on teacher efficacy and personal accomplishment among intervention teachers who sufficiently complied with the program components because these were the teachers who stood to gain the most from the intervention. Furthermore, we anticipated that these effects would be most pronounced in the integrated condition. We also expected to find intervention effects on emotional exhaustion and depersonalization, although the direction of effects was less clear to us. On the one hand, teachers who were provided the tools to handle behavior management challenges and to improve children's social skills and who felt more efficacious in doing so could in turn experience reductions in emotional exhaustion and depersonalization. However, the potential burden of implementing a new program could put additional strains on teachers and increase burnout (Domitrovich et al., 2010; Han & Weiss, 2005), particularly among teachers who spent the most time integrating program components into their daily practice. Thus it is especially important to examine the program impacts on teachers using a CACE analysis, in light of the potential added burden of implementing a multicomponent program (Domitrovich et al., 2010).

Overview of the CACE Approach

The overall goal of the current study was to estimate the effects of the interventions on teachers while accounting for compliance with assigned intervention. In order to do this, we needed to compare outcomes for teachers in the treatment group who complied with implementing the intervention to the outcomes for teachers in the control group who would potentially do the same if assigned to the intervention group. The obvious challenge with this approach is that potential participation under the treatment condition is observed in the treatment group but not in the control group. Angrist et al. (1996) provided a framework for this approach, which outlined a process for a two-arm trial with binary compliance in the potential outcomes framework (also see Frangakis & Rubin, 2002; Holland, 1986). They defined four compliance types on the basis of individuals' treatment assignment status (1=treatment, 0=control) and potential treatment receipt status (1=received, 0=not received). Compliers are those who receive or participate in the treatment when assigned to the treatment group and do not receive or participate when assigned to the control group. In an ITT analysis, the effect of treatment assignment is the same as the effect of full participation for the compliers. Always-takers are those who will always receive the treatment, no matter what group they are assigned to. Never-takers are those who will never receive, regardless of the treatment assignment. Defiers are those who will not receive if assigned to the treatment group and will receive if assigned to the control group. Since the compliance type defined this way is independent of treatment assignment status, the difference between the treatment and control condition within each compliance type can be interpreted as a causal effect (Frangakis & Rubin, 2002). Similar to Angrist et al. (1996), our primary interest is to estimate the causal treatment effect for compliers (i.e., CACE). A nuance of the application of CACE to the current study is that the participants here are the teachers who are in effect *delivering* the intervention, rather than *receiving* it from another source; this focus on implementation, coupled with the use of a continuous compliance indicator for which we set a threshold of high and low compliance (as compared to a traditional categorical approach to compliance) make this application of CACE particularly unique.

Assumptions to identify CACE. Given that treatment receipt behavior of each individual can be observed only under the condition he or she is assigned to, CACE cannot be calculated directly comparing his or her outcome under the treatment and under the control conditions. Holland (1986) called this the fundamental problem of causal inference. However, under certain conditions, we are able to identify the causal effect at the average level. In particular, the set of conditions (assumptions) used in Angrist et al. (1996) have been widely used to identify CACE. One core assumption is *ignorable treatment assignment*, which provides the basis for causal inference as it guarantees the comparability between treatment arms. In our case, this assumption is automatically satisfied as schools are randomized to intervention and control conditions. A second assumption is the stable unit treatment value (SUTVA), which means that the potential outcome of each individual is not affected by the treatment assignment status of other individuals. This is a questionable assumption in school settings because teachers in the same school are highly likely to interact with one another. To minimize this interaction across different treatment arms, we employed cluster randomization, where the unit of randomization is school. Previous studies suggested that by employing cluster randomized trials (CRT), interaction or contamination among individuals becomes a more manageable problem (Jo, Asparouhov, Muthén, Ialongo, & Brown, 2008; Sobel, 2006). That is, now we only need to worry about interactions among teachers within schools, which can be handled using statistical techniques such as multilevel analysis or generalized estimating equations. We cannot prevent interactions among teachers across different intervention arms, although the likelihood will remain about the same as that observed when no nesting exists. A third assumption is monotonicity, which assumes that there are no defiers. This is a reasonable assumption in our case because teachers in the control group did not have access to the intervention. It is also assumed that there are at least some compliers, meaning that the offer of the intervention induces at least some teachers to participate. This is a reasonable assumption in our study.

Finally, the exclusion restriction assumes that always-takers and never-takers in the control group will not benefit from the program and therefore the distribution of outcomes is the same in the treatment and control groups for these two types. In our context, this means that there is no effect of assignment for never-takers. As teachers in the control group did not have access to the intervention, the stratum of always-takers does not apply to our study. Given our simplified setting with only compliers and never-takers, we will use non-compliers to refer to never-takers. The assumption of exclusion restriction may need to be relaxed in school-based interventions where it is quite possible that teachers are affected by the intervention assignment even if they do not participate. For example, teachers may be affected by the training at the beginning of the year even if they do not end up implementing the program in their classrooms according to our definition of implementation. Since compliance in our case is not a dichotomous variable (i.e., teachers can vary in their frequency and quality of program implementation) for whether a teacher participates or not, the cut-off for determining participation will affect the likelihood that the exclusion restriction is met. Given the possible deviation from the exclusion restriction, we additionally conducted CACE estimation assuming an alternative assumption that the intervention effects are additive (Jo et al., 2008), meaning that the intervention effects do not change depending on the values of covariates. In addition, we conducted CACE estimation using two different cut points of our original continuous compliance indicator of program dosage. Comparing the CACE estimates assuming the exclusion restriction and the additivity and with two different cut points served as sensitivity analyses because we cannot that we meet the exclusion restriction.

Method

Sample

The current study sample included 350 K-5 teachers across 27 schools. Schools, and therefore teachers, were enrolled in three cohorts (i.e., for one year each, in three consecutive years) and provided consent for their voluntary participation. The sample was generally evenly split across the three cohorts (31% cohort 1, 34% cohort 2, and 35% cohort 3) and across the three conditions (25% PAX GBG, 29% PATHS to PAX, 37% control). The majority of students in the schools was African American (88% on average) and received free and reduced meals (i.e., FARMs; 85%). The vast majority of the teacher sample was female (i.e., 88%). Less than half was 30 or younger (41.4%), and taught students in grades 3 through 5 (44.1%). Just over half of the teachers had a graduate degree (56.4%). See Table 1 for additional details on the sample as well as average scores on the key measures administered in this study.

Measures

All outcome measures were assessed using a teacher self-report measure administered four times (i.e., fall baseline and three follow-ups) over the course of the school year.

Teacher Burnout. Teachers were asked to report on their level of *emotional exhaustion* (9 items, e.g., I feel used up at the end of the workday, $\alpha = .92$), *personal accomplishment* (8 items, e.g., I deal very effectively with the problems of my students, $\alpha = .85$), and *depersonalization* (3 items, e.g., I've become more callous towards people since I took this job, $\alpha = .64$) from the *Maslach Burnout Inventory* (MBI; Maslach et al., 1997). Responses were rated on a 7-point Likert scale from *never* to *every day,* with higher scores indicating greater burnout.

Teacher Efficacy. Teachers reported on a 5-point scale their self-efficacy in two domains. The *Behavior Management Self-Efficacy Scale* (Main & Hammond, 2008) assessed teachers' self-efficacy in promoting classroom behavior management (14 items; e.g., I am able to use a variety of behavior management techniques; α = .94). The *Social-Emotional Learning Efficacy Scale* (Domitrovich & Poduska, 2008) assessed teachers' self-efficacy in promoting social-emotional skills in students (8 items; e.g., I am able to teach children to show empathy and compassion for each other; α = .93).

Compliance. Both teacher completed weekly logs (in which they recorded the number of games they played) and the number of minutes they spent on each game were indicators of compliance. The number of games and the number of minutes played were each summed, for a total score for each measure across the school year. The compliance cut point will affect the exclusion restriction (Stuart et al., 2008); therefore there is a trade-off when deciding where to set the cut point. For example, if the cut point is 5 games, the assumption is that teachers who led less than 5 games would not be affected by the intervention. Setting the cut point too low may lead to great variation in the degree to which compliers implemented the program. However, with a higher cut point the sample size among compliers becomes small and implies a larger estimated CACE, in turn reducing the quality of CACE estimates. Therefore, compliance was defined in two ways: a *medium compliance* cut point for teachers who fell above the 50th percentile on both the number of games played and the minutes played (n= 81 total treatment teachers), and a *high compliance* cut point was defined as teachers who fell above the 75th percentile on both number of games and minutes played (n=29 total treatment teachers).

Covariates. Several baseline variables that were correlated with whether or not treatment teachers were classified as compliers or not were included as covariates in all models (Domitrovich, Bradshaw et al., 2015). A teacher information form was completed at baseline to collect information on teacher demographics (e.g., gender, age, education, degree attained), professional development experiences, and information regarding other social-emotional and classroom management interventions being used by the teacher. Teacher gender, age, graduate degree attainment, the grade level taught, cohort, and school mobility were included in the current study. In addition, several other baseline scales were included as covariates. A total mindfulness score was computed as the mean of 20 items (e.g., When I am in the classroom I have difficulty staying focused on what is happening in the present; $\alpha = .84$) from the *Mindfulness in Teaching Scale* (Frank, Jennings, & Greenberg, 2014). The *Openness to Innovation* subscale from the *Trust in Schools* measure (Bryk & Schnieder, 2002) was computed as the mean score of three items (e.g., Take responsibility for improving the school; $\alpha = .84$). Baseline depersonalization and emotional exhaustion were also included as covariates in all models where they were not the outcome.

Estimation of CACE

CACE models were estimated separately for each of the treatment conditions relative to control (i.e., integrated v. control and PAX GBG vs. control). Linear growth curve models with intercept and slope parameters were used to estimate the initial level and change of each outcome over the school year. In this longitudinal framework, CACE was defined as the effect of intervention assignment for compliers on the change (slope) in each outcome (e.g., Jo & Muthén, 2003; Jo, Wang, & Ialongo, 2009). As described above, compliance was defined in the current study using two different cut points (50^{th} and 75^{th} percentile), as a sensitivity analysis for both the cut point and the potential deviation from the exclusion restriction. Specifically, the CACE models were identified in two different ways. First, we assumed the exclusion restriction. In this model, the slope was regressed on treatment assignment in the complier class but not in the non-complier class. In this case, we are assuming that non-compliers are not affected by treatment assignment. However, as discussed earlier, this assumption might have been violated in our trial. Given these possibilities of deviation from the exclusion restriction, we additionally conducted CACE estimation assuming that the intervention effects are additive (Jo et al., 2008). That is, we assumed that the intervention effects do not change depending on the covariates. In the model with the additive treatment effect assumption (instead of the exclusion restriction), the slope was regressed on treatment in both the complier and non-complier classes. In both models, the intercept and slope were regressed on the pre-treatment covariates.

Missing data on the compliance measure caused 32 cases (19 in the integrated condition and 13 cases in the PAX GBG condition) to drop out (thought listwise deletion), resulting in a total sample size of 318 teachers. An additional set of teachers was excluded in the analysis phase due to missing data on the covariates. Specifically, in the models comparing the integrated condition to the control condition, 25 teachers were dropped, resulting in a sample size of 185 teachers. In the models comparing the PAX GBG condition to the control condition, 34 teachers were dropped, resulting in a total sample size of 202. In principle, we could incorporate all cases including the ones with incomplete information. However, we employed listwise deletion, given that there is little research which provides guidelines for handling simultaneous complications of noncompliance, clustering, and missing data. In this study, we focused on handling of noncompliance and clustering, and ignored biases introduced by dropping teachers with missing data, which is a limitation of the study. We used maximum likelihood (ML) estimation with the expectation maximization (EM) algorithm (Little & Rubin, 2002) for CACE estimation, which can be conveniently implemented using the mixture modeling feature in Mplus (Muthén & Muthén, 1998-2015). In this framework, compliance status was defined by a categorical latent variable, with one class referring to the compliers and the other class referring to the non-compliers. Given our simplified setting where there are only compliers and never-takers, the compliance class membership was completely observed in the treatment group whereas completely unobserved in the control group. The unknown compliance type of individuals in the control condition was handled as missing data via the EM algorithm. Characteristics of teachers were used as predictors of the latent complier class membership (Domitrovich et al., 2015).

In principle, between school and within school level parameters can be formally modeled taking into account compliance in the context of cluster randomized trials (i.e., multilevel modeling). However, in practice, the number of clusters is often small (9 schools per condition in our study). Fairly large numbers of clusters (preferably 50 or more) are necessary to yield accurate CACE estimates when taking a formal multilevel approach (Jo, Asparouhov, Muthén, Ialongo, & Brown, 2008). Instead, we used the sandwich estimator in conjunction with the ML-EM mixture (TYPE=MIXTURE COMPLEX) to adjust the standard errors for the clustering of teachers within schools.

Results

Descriptive statistics on the study variables for the two sample conditions are reported in Table 1, whereas Table 2 indicated baseline differences on the covariates between compliers and non-compliers. Specifically, complier teachers in the integrated condition were less burnt out at baseline compared to non-compliers, using either the medium or high compliance cut point. Complier teachers in the integrated condition also had higher mindfulness scores at baseline (high compliance cut point only). Complier teachers in the PAX GBG condition were less likely to have a graduate degree compared to non-compliers. Complier teachers in both conditions were in schools with less mobility regardless of the compliance cut point. In order to interpret the magnitude of effects across the different models, effect sizes (ES) were calculated by dividing the outcome difference across the two conditions by the square root of the total variance obtained from a fully unconditional model.

ITT Estimates

Teachers in the integrated condition reported increases in SEL efficacy (ES = .15) and BM efficacy (ES = .11) relative to teachers in the control condition. In addition, teachers in the integrated condition reported increases in personal accomplishment, one dimension of burnout, relative to teachers in the control condition (ES = .09). There were no significant impacts on change in depersonalization (ES = .01) or emotional exhaustion (ES = .02). The PAX GBG condition did not impact teachers' SEL efficacy (ES = .04), BM efficacy (ES = .02), personal accomplishment (ES = .03), depersonalization (ES = .01), or emotional exhaustion (ES = .01). Results are presented in the right-hand column of Table 3.

CACE Estimates

Effects of each intervention condition relative to the control condition with compliance are reported in the left-hand columns of Table 3, with the left-most columns reporting the medium compliance cut point with and without the exclusion restriction. Using the medium compliance cut point, complier teachers in the integrated condition showed statistically significant increases in SEL efficacy (ES = 0.13) and depersonalization (ES = 0.11 to 0.13) across the school year (with and without the exclusion restriction), and those in the PAX GBG condition also showed increases in depersonalization without the exclusion restriction only (ES = 0.13). Complier teachers showed increases in emotional exhaustion in both conditions without the exclusion restriction only (PAX GBG condition ES = 0.10; integrated condition ES = 0.13). Without the exclusion restriction, the effects of treatment assignment on the slopes of the outcomes were also estimated for the non-compliers. Non-complier teachers in both conditions reported decreases in emotional exhaustion (PAX GBG ES = 0.22; integrated ES = 0.32). Non-compliers in the integrated condition reported increases in personal accomplishment (ES = 0.25). The effect of being assigned to treatment among non-compliers on these outcomes was stronger than the effect of treatment among compliers. In addition, non-compliers in the integrated condition reported decreases in depersonalization (ES = 0.17), whereas the effect among compliers was positive.

The right-hand columns show results from the models using the high compliance cut point with and without the exclusion restriction. In most instances the results were similar, with the exception of personal accomplishment when comparing PAX GBG to the control condition. Using the high compliance cut point, complier teachers in both conditions showed increases in BM efficacy (PAX GBG ES = 0.32; integrated ES = 0.39) across the school year with and without the exclusion restriction. Those in the integrated condition showed increases in emotional exhaustion with and without the exclusion restriction (ES = 0.24 to 0.26). Personal accomplishment increased among high complier teachers in the PAX GBG condition with and without the exclusion restriction (ES = 0.19 to 0.69), but decreased among those in the integrated condition without the exclusion restriction only (ES = 0.20). In addition, in contrast to compliers, non-compliers in the integrated condition reported increases in SEL efficacy (ES = 0.10) and personal accomplishment (ES = 0.14). Finally, results were similar when models were estimated using a sandwich estimator, and were therefore not sensitive to adjusting the standard errors to account for the clustering of teachers in schools.

Covariate Associations with Compliance

Table 4 shows results from the logistic regression predicting high compliance from the models without the exclusion restriction. When comparing the integrated condition to the control, gender (odds ratio [OR] = 5.38), cohort (OR = 3.94), mobility (OR = 1.15), and mindfulness (OR = 0.03) significantly predicted compliance when personal accomplishment was the outcome. Gender (OR = 21.19), cohort (OR = 13.10), mobility (OR = 1.30), and depersonalization (OR = 2.59) predicted compliance when emotional exhaustion was the outcome. Mindfulness predicted compliance with regard to BM efficacy (OR = 0.24), such that compliers were more likely to be higher on mindfulness at baseline. The covariates did not significantly predict SEL efficacy or depersonalization. When comparing PAX GBG to control, mobility predicted compliance for the SEL efficacy outcome (OR = 1.14). Cohort predicted compliance for the BM efficacy (OR = 0.55) and depersonalization (OR = 0.38) outcomes.

Discussion

Many district, school, structural, training, and teacher factors can facilitate or impede the implementation of school-based prevention programs, particularly those that are dependent on teachers' use in the classroom (Domitrovich et al., 2009; Han & Weiss, 2005). Comparing average outcomes of schools randomized to a program group and a control group should produce unbiased estimates of a program's impacts, assuming that randomization was successful in creating equivalent groups, but the contrasts between the conditions are diminished as a result of variation in treatment received (Weiss, Bloom, & Brock, 2013). In the current study, average outcomes across treatment conditions should have produced unbiased estimates of the impacts on teacher efficacy and burnout of two evidence-based prevention programs -- PAX GBG and PATHS -- intended to build students' social-emotional skills, reduce aggressive behaviors, and help teachers manage their classrooms. But teachers varied in the degree to which they implemented the programs, which is a common occurrence in school-based programming (Domitrovich et al., 2008; Fixsen, Naoom, Blasé, Friedman, & Wallace, 2005). Ignoring this noncompliance may result in decreased power to detect average effects (Jo et al., 2008). This source of variation likely diminishes the contrasts between treatment conditions and, in turn, attenuates program impacts. CACE estimation is one approach to accounting for this source of variation. In this study, we applied the CACE framework to account for teacher compliance in implementing a major component of the PAX GBG program. The interventions tested within this study are in fact similar in many ways to other classroom-based prevention programs that largely rely on teachers for implementation (e.g., 4R's, Second Step).

Overall, the CACE estimation approach was helpful in understanding treatment-control differences when accounting for variation in treatment conditions due to teacher compliance. This approach revealed impacts on teachers that were distinctly different than those produced using an ITT approach. First, some intervention effects on teacher efficacy and burnout were stronger among teachers who complied compared to teachers overall. Specifically, we found positive effects on social-emotional and behavioral management efficacy among complier teachers in both of the intervention conditions. Teachers on average had greater increases in efficacy in the integrated condition than in the control condition. In the case of behavior management efficacy, these effects seemed to be concentrated among those most likely to comply with the program model. In the case of social-emotional efficacy, the effects among compliers and among the teacher sample overall were similar.

The estimation of program impacts on burnout (i.e., personal accomplishment, depersonalization, and emotional exhaustion) while accounting for compliance revealed a different story than the estimation of average impacts across all teachers. Specifically, program effects on personal accomplishment were stronger among teachers most likely to comply in both intervention conditions. On average, there were no significant differences between treatment conditions and the control condition in growth or change of emotional exhaustion or depersonalization across the school year. However, accounting for compliance seemed to uncover some opposing findings among compliers and non-compliers and some increases in burnout among compliers. Specifically, being in the complier group in the integrated condition led to greater reports of emotional exhaustion, whereas being in the non-complier group within an intervention school was associated with reduced emotional exhaustion. Furthermore, being a complier led to slightly greater reports of depersonalization in both conditions, while being a non-complier in an integrated school was associated with reduced depersonalization. Overall program impacts on emotional exhaustion and depersonalization may have been masked by these opposing findings. In addition, the trends using the medium and high compliance cut points were somewhat similar, but there were several differences. The effects using the high compliance cut point were notably stronger for all the outcomes except depersonalization. In addition, there were a few cases where the direction of the effect was different (e.g., SEL efficacy in the integrated condition and emotional exhaustion in the PAX GBG condition).

Limitations

The initial sample size was small, and the sample of teachers became smaller when split into compliance groups. Estimation of program impacts using the high compliance cut point yielded an especially small sample size and likely rendered the estimates unstable. In addition, despite the fact that schools and not teachers were randomized, the small sample of schools in each condition prevented us from employing a multilevel modeling approach. Multilevel mixture modeling using the EM estimation approach is computationally demanding and treatment effects accounting for compliance are poorly estimated when the number of clusters are small (Jo et al., 2008). Therefore, we were not able to accommodate both the clustering of teachers in schools and compliance.

The interpretation of our findings is limited by our measure of compliance. In most prior applications of the CACE approach, full compliance was known, either based on a theoretical idea of what level of compliance is needed to benefit from the program (e.g., Stuart et al., 2008) or because compliance was defined as electing to receive the intervention or not (e.g., Connell et al., 2007; Cowen, 2008). In our case, we did not have a target level of dosage to measure perfect compliance and compliance was on a continuum from low to high. As evidenced by the findings, the cut point we used for compliance made a difference. Thus, when using the CACE approach to account for service delivery by teachers rather than program uptake of participants, as it has most often been used, it would be useful to have a more precise definition of full compliance. In addition, CACE estimation relies on a set of assumptions, some of which are difficult to meet when applied to school-randomized trials. As discussed earlier, a violation of the exclusion restriction is imaginable in the current study because teachers are likely affected by the intervention even if they did not participate and because our compliance measure is a continuous variable from which we established artificial cut points. The models assuming additivity that relax the exclusion restriction are more likely to suffer from a violation of normality, however. Given these trade-offs, we conducted the CACE estimation with two different cut points and with and without the assumption of the exclusion restriction. We gained more confidence in our findings because our results generally held through sensitivity testing in which we relaxed the exclusion restriction (Jo, 2002). However, the results were somewhat sensitive to the cut point used for compliance. This is not surprising given that compliance was higher and more concentrated using the high compliance cut point rather than the medium compliance cut point.

On the other hand, the sample size was significantly reduced using the high compliance cut point. Further tests of violations of the identifying assumptions are not possible. It is important to keep in mind that bias from violation of these assumptions will be problematic in any application of CACE estimation (Jo, 2002).

Conclusion and Implications

Most universal school-based interventions are tested using an ITT approach. Average treatment effects are useful for understanding whether school-based interventions can work under real world conditions. Estimating variation in implementation and impacts can help unpack under what conditions and for whom programs are effective. This can be helpful in targeting interventions and informing design and implementation of evidence-based interventions (Schochet, Puma, & Deke, 2014). CACE estimation is one approach for taking into account implementation or compliance when estimating causally estimated treatment impacts. Applied to a case study of two evidence-based interventions implemented in the classroom, we believe that this approach was helpful in uncovering effects that differed from the traditional ITT approach. The findings suggest the possibility that the highest implementing teachers benefited from the interventions in that they felt more efficacious in their instruction across the school year than teachers in the control group. On the other hand, the results raise the possibility that the increased demand put on teachers in the intervention schools may have increased burnout for some teachers over the year. The current findings suggest the possibility that the implementation of an intervention can increase stress and burnout for certain teachers, even as the design intends for the intervention to be integrated into the regular curriculum and seeks to minimize the amount of additional burden placed on the teacher. Another possibility is that certain teachers who are engaging most in the intervention are becoming more aware and learning to recognize their own emotional responses. As a result of this increased emotional awareness, they may be perceiving and reporting greater feelings of burnout. However, we do not know the extent to which the level of burnout reported may translate into significant or clinical impairment. Regardless, the extent to which these emotional symptoms may be affecting students still needs to be addressed. Given previous research on the negative associations between teacher burnout and student outcomes (Maslach et al., 1997; Pas & Bradshaw, 2014), it is possible that any increased burden placed on teachers could attenuate the effects of the interventions on children (for further discussion of the effects of teacher burnout on students see Abenavoli et al., 2013). The question also becomes how we can provide the necessary supports for teachers to implement classroom-based interventions without the generation or perception of increased stress. Teachers' implementation of social-emotional learning programs involves the development of a similar set of skills among adults. One possibility is that prevention programs that foster social-emotional skill-building in children could also provide the supports and skill-building for teachers to manage and develop their own emotional responses; this is particularly important as greater demands are placed on teachers to incorporate these lessons into their everyday classroom routines.

References

Abenavoli, R. M., Jennings, P. A., Greenberg, M. T., Harris, A. R., & Katz, D. A. (2013). The protective effects of mindfulness against burnout among educators. *The Psychology of Education Review, 37,* 57-69.

Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association, 91,* 444–455.

Barnard, J., Frangakis, C. E., Hill, J. L., & Rubin, D. (2003). Principal stratification approach to broken randomized experiments. *Journal of American Statistical Association, 98*, 299-323. DOI: 10.1198/016214503000071

Botvin, G. J., Baker, E., Dusenbury, L., Botvin, E. M., & Diaz, T. (1995). Long-term follow-up results of a randomized drug abuse prevention trial in a White middle-class population. *JAMA*, *273*, 1106-1112.

Bradshaw**,** C.P., Koth, C.W., Bevans, K.B., Ialongo, N., & Leaf, P.J. (2008). The impact of school-wide Positive Behavioral Interventions and Supports (PBIS) on the organizational health of elementary schools. *School Psychology Quarterly, 23*, 462-473. doi:10.1037/a0012883

Bradshaw, C.P., Koth, C.W., Thornton, L.A., & Leaf, P.J. (2009). Altering school climate through school-wide Positive Behavioral Interventions and Supports: Findings from a group-randomized effectiveness trial. *Prevention Science, 10*(2), 100-115. doi: 10.1007/s11121-008-0114-9

Bryk, A.S. & Schnieder, B. (2002). *Trust in schools: A core resource for improvement*. New York: Russell Sage.

Chen, H.T. (1998). Theory-driven evaluations. *Advances in Educational Productivity*, *7*, 15-34.

Conduct Problems Prevention Research Group (1999). Initial impact of the Fast Track prevention trial for conduct problems: II. Classroom effects. *Journal of Consulting and Clinical Psychology, 67*, 648-657.

Connell, A. M., Dishion, M. Y, Kavanagh, K. (2007). An adaptive approach to family intervention: Linking engagement in family-centered intervention to reductions in adolescent behavior. *Journal of Consulting and Clinical Psychology, 75*, 568-579. DOI: 10.1037/0022-006X.75.4.568

Cowen, J. M. (2008). School choice as a latent variable: Estimating the "complier average causal effect" of vouchers in Charlotte. *The Policy Studies Journal, 36*, 301-315.

Derzon, J. H., Sale, E., Springer, J. F., & Brounstein, P. (2005). Estimating intervention effectiveness: Synthetic projection of field evaluation results. *Journal of Primary Prevention, 26*, 321-343.

Domitrovitch, C. E., Bradshaw, C. P., Berg, J., Pas, E. T., Becker, K., Musci, R…Ialongo, N. (2015). How do school-based prevention programs impact teachers: Findings from a randomized trial of combined classroom management and social-emotional programs. *Journal of Prevention Science*, under review.

Domitrovich, C. E., Gest, S. D., Gill, S., Jones, D. J., & DeRouise, R. S. (2009). Teacher factors related to the professional development process of the Head Start REDI intervention. *Early Education and Development*, *20*, 402-430.

Domitrovich, C. E., Bradshaw, C. P., Poduska, J. M., Hoagwood, K. E., 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*, 6-28.

Domitrovich, C., Bradshaw, C., Greenberg, M., Embry, D., Poduska, J., & Ialongo, N. (2010). Integrated models of school-based prevention: Logic and theory. *Psychology in the Schools, 47,* 71-88.

Domitrovich, C. & Poduska, J. (2008). The Social-Emotional Learning Efficacy Scale. Unpublished technical report

Durlak, J. A., & DuPre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology, 41*, 327-350.

Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. B. (2011). The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions. *Child Development, 82*, 405-432.

Embry, D., Staatemeier, G., Richardson, C., Lauger, K., & Mitich, J. (2003). The PAX Good Behavior Game (1st ed.). Center City, MN: Hazelden.

Fixsen, D., Naoom, S., Blasé, K., Friedman, R., & Wallace, F. (2005). 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 (FMHI Publication #231).

Frangakis, C. E., Rubin, D. B. (2002). Principal stratification in causal inference. *Biometrics, 58,* 21-29.

Frank, J. L., Jennings, P.A., & Greenberg, M.T. (2014). *Validation of the Mindfulness in Teaching Scale.* Manuscript submitted for publication.

Greenberg, M., & Kusche, C. (2006). Building social and emotional competence: The PATHS Curriculum. In S.R. Jimerson and M.J. Furlong (Eds.), *Handbook of school violence and* *school safety: From research to practice* (pp. 395-412). Mahwah, NJ: Erlbaum.

Greenberg, M., Kusche, C., Cook, E., & Quamma, J. (1995). Promoting emotional competence in school-aged children: The effects of the PATHS curriculum. *Development and Psychopathology, 7*, 117-136.

Greenberg, M.T., Kusché, C.A., & Conduct Problems Prevention Research Group (2011). Grade level PATHS (Grades 3-5). South Deerfield, MA: Channing-Bete Co.

Han, S. & Weiss, B. (2005). Sustainability of teacher implementation of school-based mental health programs. *Journal of Abnormal Child Psychology*, *33,* 665-679.

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association, 81*, 945–970.

Hulleman, C. S., & Cordray, D. S. (2009). Moving from the lab to the field: The role of fidelity and achieved relative intervention strength. *Journal of Research on Educational Effectiveness, 2*, 88-110.

Ialongo, N., Wethamer, L., Kellam, S., Brown, C., Wang, S., & Lin, Y. (1999). Proximal impact of two first-grade preventive interventions on the early risk behaviors for later substance abuse, depression, and anti-social behavior. *American Journal of Community Psychology, 27*, 599–641.

Ialongo, N., Poduska, J., Werthamer, L., & Kellam, S. (2001). The distal impact of two first-grade preventive interventions on conduct problems and disorder in early adolescence. *Journal of Emotional and Behavioral Disorders*, *9,* 146-160.

Jo, B. (2002). Estimation of intervention effects with noncompliance: Alternative model specifications. *Journal of Educational and Behavioral Statistics, 27*, 385-409.

Jo, B., Asparouhov, T. Muthén, B. O. (2008). Intention-to-treat analysis in cluster randomized trials with noncompliance. *Statistics in Medicine, 27*, 5565-5577. DOI: 10.1002/sim.3370

Jo, B., Asparouhov, T., Muthén, B. O., Ialongo, N. S., & Brown, C. H. (2008). Cluster randomized trials with treatment noncompliance. *Psychological Methods, 13*, 1-18.

Jo, B., & Muthén, B. (2003). Longitudinal studies with intervention and noncompliance: Estimation of causal effects in growth mixture modeling. In N. Duan & S. Reise (Eds.), *Multilevel modeling: Methodological advances, issues, and applications* (pp. 112-139). Mahwah, NJ: Lawrence Erlbaum Associates.

Jo, B., Wang, C. P., & Ialongo, N. S. (2009). Using latent outcome trajectory classes in causal inference. *Statistics Interface, 2*, 403-412.

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

Kusché, C.A., Greenberg, M.T., & Conduct Problems Prevention Research Group (2011). Grade level PATHS (Grades 1-2). South Deerfield, MA: Channing-Bete Co.

Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data*. Hoboken, NJ: John Wiley & Sons, Inc.

Little, R. J. A., & Yau, L. H. Y. (1998). Statistical techniques for analyzing data from prevention trials: Treatment of no-shows using Rubin's causal model. *Psychological Methods, 3,* 147–159.

Main, S., & Hammond, L. (2008). Best practice or most practiced? Pre-service teachers' beliefs about effective behavior management strategies and reported self-efficacy. *Australian Journal of Teacher Education, 33,* 28-39.

Maslach, C., Jackson, S.E., & Leiter, M.P. (1997). Maslach burnout inventory: Third edition. In: Wood RJ, editor. *Evaluating stress: A book of resources*. Lanham, MD US: Scarecrow Education, p. 191-218.

O'Donnell, (2008).C. L. (2008). Defining, conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in K-12 curriculum intervention research. *Review of Educational Research, 78*, 33–84.

Pas, E., & Bradshaw, C. P., (2014). What affects teacher ratings of student behaviors? The potential influence of teachers' perceptions of the school environment and experiences. *Prevention Science,* 15, 940-950. Doi: 10.1007/s11121-013-0432-4

Payne, A. A., & Eckert, R. (2010). The relative importance of provider, program, school, and community predictors of the implementation quality of school-based prevention programs. *Prevention Science, 11*(2), 126-141.

Rohrbach, L. A., Graham, J. W., & Hansen, W. B. (1993). Diffusion of a school-based substance abuse prevention program: Predictors of program implementation. *Preventive Medicine, 22*, 237-260.

Schochet, P. Z., Puma, M., & Deke, J. (2014). *Understanding variation in treatment effects in education impact evaluations: An overview of quantitative methods* (NCEE 2014-4017). Washington, DC: U.S. Department of Education, Institute of Education Sciences, national Center for Education Evaluation and Regional Assistance, Analytic Technical Assistance and Development. Retrieved from http://ies.ed.gov/ncee/edlabs.

Sobel, M. E. (2006). What do randomized studies of housing mobility demonstrate: causal inference in the face of interference. *Journal of the American Statistical Association*, *101*, 1398–1407.

Stanger, C., Ryan, S. R., Fu, H., & Budney, A. J. (2011). Parent training plus contingency management for substance abuse families: A Complier Average Causal Effects (CACE) analysis. *Drug and Alcohol Dependence*, *118*, 119-126. doi:10.1016/j.drugalcdep.2011.03.007

Stuart, E. A., Perry, D. F., Le, H., & Ialongo, N. S. (2008). Estimating intervention effects of prevention programs: Accounting for noncompliance. *Prevention Science*, *9*, 288-298. DOI 10.1007/s11121-008-0104-y

Weiss, M. J., Bloom, H. S., & Brock, T. (2013). *A conceptual framework for studying the sources of variation in program effects*. MDRC. New York: MDRC.

Table 1. *Descriptive Information on Teacher Participants and Schools*

Teacher Characteristics (%)	**Total**		**PATHS to PAX**		**PAX GBG**		**Control**	
Female	88.7		90.2		88.0		88.3	
Taught grades 3-5	41.5		45.1		38.0		42.2	
Age ≤30 Years	38.7		39.0		45.4		32.8	
Has graduate degree	52.8		52.4		56.5		50.0	
Cohort								
Cohort 1	32.1		32.9		37.0		27.3	
Cohort 2	34.3		41.5		31.5		32.0	
Cohort 3	33.6		25.6		31.5		40.6	
Teacher Self-Report				**Mean (*SD*)**				
Openness to Innovation Time 1	3.75	(.86)	3.79	(.86)	3.79	(.84)	3.70	(.89)
Mindfulness Time 1	3.98	(.43)	3.97	(.37)	3.99	(.46)	3.99	(.44)
Behavioral Management Efficacy								
Time 1	3.84	(.63)	3.78	(.58)	3.93	(.61)	3.82	(.68)
Time 4	4.03	(.60)	4.16	(.55)	4.00	(.59)	3.96	(.63)
Social-Emotional Learning Efficacy								
Time 1	3.60	(.67)	3.51	(.60)	3.72	(.64)	3.56	(.73)
Time 4	3.76	(.68)	3.92	(.65)	3.78	(.63)	3.63	(.73)
Emotional exhaustion								
Time 1	3.39	(1.39)	3.38	(1.34)	3.39	(1.44)	3.40	(1.40)
Time 4	3.17	(1.48)	3.07	(1.45)	3.14	(1.45)	3.27	(1.54)
Personal Accomplishment								
Time 1	5.90	(0.87)	5.77	(.91)	5.97	(.82)	5.92	(.88)
Time 4	5.96	(0.83)	6.09	(.76)	5.91	(.79)	5.91	(.91)
Depersonalization								
Time 1	2.21	(1.32)	2.08	(1.13)	2.35	(1.35)	2.20	(1.42)
Time 4	2.39	(1.38)	2.36	(1.30)	2.39	(1.32)	2.42	(1.49)
School-level Variables								
Mobility rate	35.38	(8.20)	37.70	(9.30)	34.04	(7.53)	34.38	(8.14)

Table 2. *Baseline Differences Between Compliers and Noncompliers Randomized to the Treatment Condition*

	With medium compliance cut-off			With high compliance cut-off		
Covariate	Non-complier	Complier	Omnibus test (χ^2 or t-test)	Non-complier	Complier	Omnibus test (χ^2 or t-test)
PATHS to PAX v. Control						
Female	91.7%	88.2%	0.27	92.8%	76.9%	3.11
Late Elementary School	48.9%	46.9%	0.03	46.9%	53.8%	0.21
New teacher	44.4%	37.5%	0.37	42.2%	38.5%	0.06
Graduate degree	52.3%	62.5%	0.79	58.7%	46.2%	0.69
Cohort			6.17*			0.70
Cohort 1	43.8%	17.6%		34.8%	23.1%	
Cohort2	35.4%	50.0%		40.6%	46.2%	
Cohort 3	20.8%	32.4%		24.6%	30.8%	
Mobility (school)	39.1%	34.8%	2.51*	37.90%	34.50%	1.46
Mindfulness	3.91 (.40)	4.05 (.32)	-1.69	3.92 (.37)	4.21 (.28)	-2.63*
Depersonalization	2.34 (1.16)	1.71 (.99)	2.51*	2.20 (1.18)	1.46 (.52)	3.59**
Openness to innovation	3.71 (.85)	3.90 (.88)	-0.92	3.74 (.85)	4.05 (.88)	-1.21
Emotional exhaustion	3.69 (1.33)	2.94 (1.25)	2.46*	3.56 (1.36)	2.52 (.86)	2.64*
Pax v. Control						
Female	86.9%	89.4%	0.15	88.0%	87.5%	0.00
Late E.S.	42.4%	34.0%	0.77	37.8%	43.8%	0.20
New teacher	39.7%	56.5%	2.93	45.5%	56.3%	0.63
Graduate degree	69.0%	45.7%	5.75*	59.1%	56.3%	0.05
Cohort			5.26			7.75*
Cohort 1	45.9%	25.5%		42.4%	6.3%	
Cohort 2	29.5%	34.0%		28.3%	50.0%	
Cohort 3	24.6%	40.4%		29.3%	43.8%	
Mobility (school)	36.2%	32.3%	3.09**	35.20%	30.60%	2.57*
Mindfulness	3.98 (.45)	4.00 (.447)	-0.18	3.98 (.46)	4.03 (.44)	-0.37
Depersonalization	2.43 (1.35)	2.25 (1.37)	0.65	2.38 (1.44)	2.19 (.78)	0.76
Openness to innovation	3.68 (.80)	3.91 (.88)	-1.37	3.79 (.80)	3.81 (1.04)	-0.12
Emotional exhaustion	3.51 (1.54)	3.26 (1.31)	0.83	3.36 (1.49)	3.55 (1.16)	-0.47

Note. * $p<.05$; ** $p<.01$

Table 3. *CACE Effects with Covariates*

	CACE											
	With medium compliance cut point											
	With ER				Without ER							
	Compliers				Compliers				Noncompliers			
	Slope	(SE)		ES	Slope	(SE)		ES	Slope	(SE)		ES
SEL efficacy												
P2P v. control	0.09	(0.04)	*	0.13	0.09	(0.04)	*	0.13	0.10	(0.04)	*	0.14
PAX v. control	0.00	(0.04)		0.00	-0.02	(0.06)		0.02	0.07	(0.07)		0.10
BM efficacy												
P2P v. control	0.04	(0.04)		0.06	0.04	(0.04)		0.07	0.04	(0.06)		0.06
PAX v. control	0.05	(0.75)		0.07	0.04	(0.03)		0.06	-0.04	(0.03)		0.06
Depersonalization												
P2P v. control	0.15	(0.07)	*	0.11	0.16	(0.07)	*	0.12	-0.23	(0.12)	*	0.17
PAX v. control	0.16	(0.10)		0.11	0.19	(0.08)	*	0.13	-0.18	(0.12)		0.13
Personal accomplishment												
P2P v. control	-0.06	(0.08)		0.06	-0.07	(0.06)		0.07	0.23	(0.06)	***	0.25
PAX v. control	-0.07	(0.04)		0.07	-0.07	(0.04)		0.07	0.00	(0.06)		0.00
Emotional exhaustion												
P2P v. control	0.08	(0.10)		0.06	0.18	(0.08)	*	0.13	-0.44	(0.12)	***	0.32
PAX v. control	0.11	(0.07)		0.07	0.15	(0.07)	*	0.10	-0.32	(0.14)	*	0.22

Note. *p < .05. **p < .01. ***p < .001.
ES = Effect Size;
P2P = PATHS to PAX;
ER=Exclusion Restriction

Table 3 (cont'd)

	CACE													ITT			
	With high compliance cut point																
	With ER				Without ER												
	Compliers				Compliers				Noncompliers								
	Slope	(SE)		ES	Slope	(SE)		ES	Slope	(SE)		ES	Slope	(SE)		ES
SEL efficacy																
P2P v. control	-0.32	(0.26)		0.48	-0.36	(0.28)		0.54	0.07	(0.02)	**	0.10	0.10	(0.02)	***	0.15
PAX v. control	0.12	(0.08)		0.11	0.02	(0.04)		0.02	0.01	(0.03)		0.02	0.03	(0.02)		0.04
BM efficacy																
P2P v. control	0.25	(0.07)	***	0.39	0.24	(0.07)	***	0.39	0.02	(0.02)		0.04	0.07	(0.02)	**	0.11
PAX v. control	0.21	(0.04)	***	0.32	0.21	(0.05)	***	.32	-0.03	(0.02)		0.03	0.01	(0.02)		0.02
Depersonalization																
P2P v. control	0.16	(0.12)		0.12	0.16	(0.13)		0.12	-0.13	(0.10)		0.09	0.02	(0.05)		0.01
PAX v. control	0.08	(0.17)		0.06	0.12	(0.08)		0.08	-0.11	(0.08)		0.07	-0.01	(0.05)		0.01
Personal accomplishment																
P2P v. control	-0.11	(0.07)		0.12	-0.18	(0.04)	***	0.20	0.12	(0.04)	**	0.14	0.08	(0.03)	*	0.09
PAX v. control	0.17	(0.08)	*	0.19	0.61	(0.12)	***	0.69	-0.07	(0.03)	*	0.08	-0.02	(0.03)		0.03
Emotional exhaustion																
P2P v. control	0.36	(0.16)	*	0.26	0.32	(0.15)	*	0.24	-0.25	(0.11)	*	0.18	-0.02	(0.05)		0.02
PAX v. control	-0.42	(0.44)		0.29	-0.42	(0.54)		0.29	0.01	(0.06)		0.01	-0.02	(0.04)		0.01

Note. $^*p < .05$, $^{**}p < .01$, $^{***}p < .001$.
ES = Effect Size;
P2P = PATHS to PAX;
ER=Exclusion Restriction

Table 4. *Logistic Regression of High Compliance on Baseline Covariates (Compliers vs. Never-takers)*

	SEL Efficacy		Behavior Management Efficacy		Depersonalization		Personal Accomplishment		Emotional Exhaustion	
Covariate	Estimate	*SE*	Estimate	*SE*	Estimate	*SE*	Estimate	*SE*	Estimate	*SE*
PATHS to PAX v. control										
Intercept	4.99	4.802	-10.555	4.396	-4.095	4.27	-4.11	3.988	6.892	5.545
Female	-0.91	0.81	-1.294	0.78	-1.734	1.304	**-1.683**	**0.766**	**-3.054**	**1.512**
Late elementary school	-0.26	0.609	0.274	0.58	-0.326	0.563	-0.658	0.662	-0.029	0.683
New teacher	0.23	0.676	-0.045	0.534	0.542	0.918	0.32	0.847	0.197	0.676
Graduate degree	-0.39	0.631	-0.049	0.497	-0.711	0.749	-0.023	0.689	-0.901	0.825
Cohort	0.34	0.519	0.676	0.419	-0.543	0.949	**-1.371**	**0.512**	**-2.569**	**0.874**
Mobility (school)	-0.09	0.052	0.077	0.045	-0.117	0.109	**-0.141**	**0.05**	**-0.264**	**0.074**
Mindfulness	-0.11	1.04	**1.445**	**0.632**	2.561	1.606	**3.391**	**1.004**	2.578	1.768
Depersonalization	-0.33	0.426	0.39	0.304	N/A	N/A	-0.397	0.337	**-0.95**	**0.398**
Openness to innovation	0.31	0.312	0.001	0.323	0.403	0.442	0.041	0.368	0.213	0.428
Emotional exhaustion	-0.45	0.286	-0.329	0.224	-0.449	0.329	-0.481	0.255	N/A	N/A
PAX v. control										
Intercept	3.40	3.736	-6.736	3.643	-6.667	3.731	0.509	4.468	-2.754	3.691
Female	-1.21	0.666	0.505	0.86	0.802	0.751	0.251	0.786	-0.075	0.786
Late elementary school	0.60	0.615	-0.086	0.475	0.097	0.468	-0.341	0.524	-0.329	0.651
New teacher	-0.05	0.572	-0.285	0.558	-0.447	0.682	-0.209	0.679	-0.509	0.716
Graduate degree	-0.73	0.575	-0.645	0.614	-1.323	0.722	-0.655	0.635	-0.758	0.86
Cohort	0.04	0.478	**0.599**	**0.290**	**0.971**	**0.445**	-0.025	0.358	1.009	0.652
Mobility (school)	**-0.13**	**0.044**	0.033	0.033	0.015	0.034	-0.045	0.046	0.012	0.033
Mindfulness	-0.39	0.871	0.624	0.628	0.909	0.65	-0.096	0.734	0.031	1.033
Depersonalization	-0.04	0.249	-0.056	0.209	N/A	N/A	-0.254	0.236	-0.271	0.258
Openness to innovation	0.63	0.511	-0.144	0.311	0.182	0.272	-0.097	0.315	-0.092	0.679
Emotional exhaustion	-0.05	0.235	0.264	0.177	-0.291	0.189	0.261	0.232	N/A	N/A

Note. The logistic regression represents the prediction of the covariates on the compliance class. Bolded estimates were statistically significant at $p < .05$. N/A=covariate was left out in the case where it was the outcome.

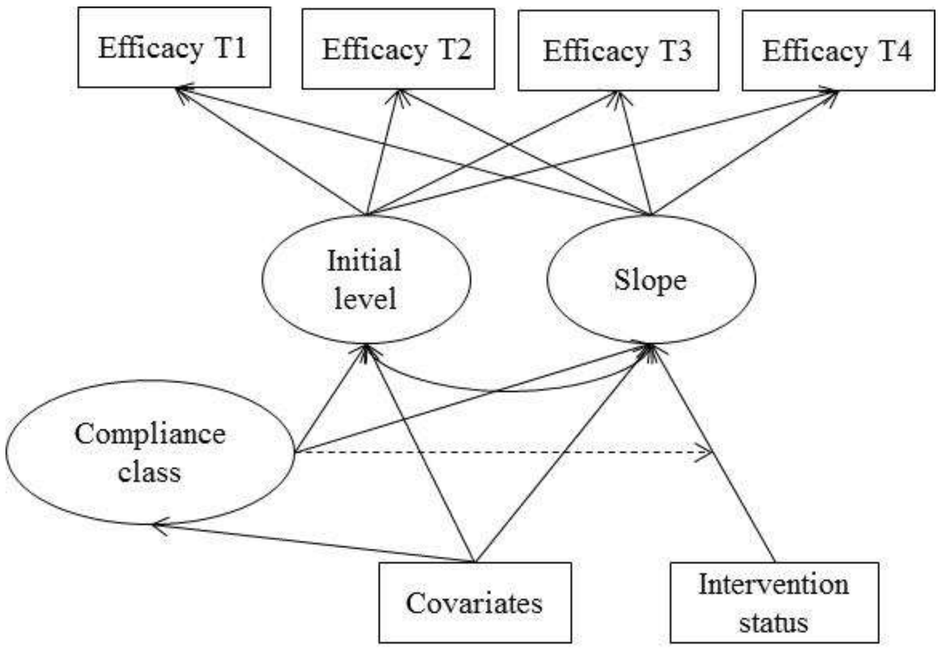

Figure 1. *Compliance Average Causal Effect model with covariate*