School Mental Health

PurposeFull People SEL and Character Education Program: A Cluster Randomized Trial in Schools Implementing Tier 1 PBIS with Fidelity

Manuscript Number:	
Full Title:	PurposeFull People SEL and Character Education Program: A Cluster Randomized Trial in Schools Implementing Tier 1 PBIS with Fidelity
Article Type:	Original Research
Keywords:	Tier 1 PBIS; SEL; character building; Integrated universal prevention; cross-level interaction
Corresponding Author:	Yanchen Zhang, Ph.D. The University of Iowa Iowa City, Iowa UNITED STATES
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	The University of Iowa
Corresponding Author's Secondary Institution:	
First Author:	Yanchen Zhang, Ph.D.
First Author Secondary Information:	
Order of Authors:	Yanchen Zhang, Ph.D.
	Clayton R Cook, PhD
	Brian Smith, PhD
Order of Authors Secondary Information:	
Funding Information:	Not applicable

Title: PurposeFull People SEL and Character Education Program: A Cluster Randomized Trial in Schools Implementing Tier 1 PBIS with Fidelity

Corresponding Author:

Yanchen Zhang, PhD,
Department of Psychological & Quantitative Foundations,
University of Iowa Email: yanchen-zhang@uiowa.edu
Mail: 361 Lindquist Center, Iowa
City, IA 52242 ORCID:
0000-0003-1756-1077

Co-Authors:

Clayton R Cook, PhD
Department of Research & Development, CharacterStrong
SEL Program Email: crcook@umn.edu
Mail: 4227 S. Meridian St., C #320, Puyallup, WA 98373

Brian Smith, PhD
Department of Research & Development, CharacterStrong
SEL Program Email: brian@characterstrong.com
Mail: 4227 S. Meridian St., C #320, Puyallup, WA 98373

Declarations

Funding: Not applicable

Conflicts of interest/Competing interests: YZ is an independent researcher responsible for program evaluation and declares that he has no competing or financial interests. CRC and BS are the developers of the PurposeFull People SEL and Character Education Program, but they did not participate or influence the data collection, analysis, or result reporting.

Availability of data and material: The de-identified dataset is available via the Open Science Framework [DOI 10.17605/OSF.IO/WPK4Q].

Abstract

Social-emotional learning (SEL), character education, and positive behavior intervention and supports (PBIS) are common approaches to Tier 1 universal social, emotional, and behavioral (SEB) supports in schools. There has been emerging evidence supporting the superior effect of integrated prevention that combines approaches to Tier 1 universal SEB programming compared to the traditional fragmented piecemeal approach. We conducted a cluster-randomized trial to examine the additive effect of PurposeFull People (PfP) – an SEL and character education program – in the context of schools implementing Tier 1 PBIS with fidelity. This study took place in five elementary schools implementing Tier 1 PBIS with fidelity. Twenty teachers/classrooms were randomly assigned to treatment (PfP combined with Tier 1 PBIS; ntreatment=10) or control condition (Tier 1 PBIS only). An average of eight students were randomly recruited from each teachers' classrooms (nstudent=161; ntreatment student = 81). Four student SEB outcomes (behavior expectations, behavior discipline, character, and academic engaged time) were assessed at baseline and 4-month posttest. Multilevel ANCOVAs and generalized linear mixed models were used to assess the effectiveness of PfP and the effect variation across subgroups (i.e., student demographics and baseline status). Compared to PBIS alone, PfP combined with PBIS led to significantly larger improvements in students' behavior expectations, behavior discipline, character, and academic engaged time. Cross-level interaction results indicated that the effectiveness of PurposeFull People on students' SEB outcomes varied based on their baseline status, where students struggling most at baseline demonstrated the largest improvement. Limitations and implications for future research and practice are discussed.

Keywords: Tier 1 PBIS, SEL, character building, Integrated universal prevention, cross-level interaction.

PurposeFull People SEL and Character Education Program: A Cluster Randomized Trial in Schools Implementing Tier 1 PBIS with Fidelity

Introduction

Social, emotional, and behavioral (SEB) well-being and performance are critical enablers of academic success (Napolitano et al., 2021). In contrast, students with SEB difficulties during the elementary years are likely to experience challenges establishing and maintaining relationships and performing well in school (Sutherland et al., 2019). To promote SEB well-being and prevent difficulties, there is a need for Tier 1 universal programming in schools. The successful delivery of Tier 1 universal programs in schools is a fundamental component of our nation's strategy to prevent SEB problems and ensure children access supports that enable academic and life success (Nisar et al., 2022). Social-emotional learning (SEL), character education, and positive behavior intervention supports (PBIS) are common approaches to Tier 1 universal programming (Jeynes, 2019; Sprague & Walker, 2021). However, these are often competing and fragmented approaches that are not intentionally integrated (Osher et al., 2019; Sprague et al., 2016). Integrated prevention offers a solution to siloed, fragmented approaches by intentionally combining approaches to increase the yield of Tier 1 universal programming on student outcomes. There is emerging evidence demonstrating the promise of integrated prevention in school mental health (e.g., Cook et al., 2015). This cluster randomized trial aimed to examine the additive effects of PurposeFull People -- an SEL and character education program when integrated in schools already implementing Tier 1 PBIS with adequate fidelity.

Needs for Tier 1 Universal Social, Emotional, and Behavioral Supports

There are several approaches to Tier 1 universal SEB programming in schools. Three of the most common approaches are PBIS, SEL, and character education. Although PBIS represents a comprehensive 3-tiered framework designed to improve SEB and educational outcomes for students in schools (Horner et al., 2015), the foundational Tier 1 universal PBIS practices include (1) defining and teaching a set of behavior expectations, (2) establishing a reinforcement system for engaging in behavior expectations, (3) implementing a system to respond to behaviors that are inconsistent with expectations, and (4) developing a data collection system for future decision making (Horner et al., 2015). Latest estimates indicate that PBIS is being implemented in over 25,000 schools nationwide (PBIS, 2022). Mounting evidence has supported the effectiveness of PBIS in improving student behaviors, mental health, and academic performance across diverse school settings and populations, including schools serving marginalized populations (e.g., low SES, racial/ethnic minorities, juvenile justice, Kim et al., 2018;

Goodman-Scott et al., 2021; Jolivette et al., 2020; Gage et al., 2017).

As a Tier 1 teaching and learning approach, SEL is commonly defined as the process where students are supported to acquire and apply specific competencies (e.g., the ability to identify and regulate emotions), show empathy for others, establish and maintain healthy relationships, and make responsible decisions (CASEL, 2022). Research consistently shows that when schools intentionally implement quality SEL curriculum and instruction, positive effects on a wide range of student outcomes are likely to happen, including academic engagement and performance, positive relationships, well-being, and longer-term follow-up effects (Corcoran et al., 2018; Durlak et al., 2011). These positive effects of SEL were also found in schools serving high-needs or marginalized student groups (Garner et al., 2014; CASEL, 2013; Greenberg, 2010; O'Neill, 2011). Like PBIS, numerous schools nationwide are implementing SEL programs, and state educational agencies have developed policies calling for schools to adopt and implement SEL curricula (Eklund et al., 2018).

Character education is one of the long-standing approaches to Tier 1 universal programming that dates to Smith's seminal work in 2013. Character education focuses on cultivating moral character strengths or traits that support students to become good, ethical people as well as performance character strengths linked to improved academic performance and later work performance, such as perseverance, responsibility, and engagement. While findings for character education have been mixed, a meta-analysis of 52 studies conducted by Jeynes (2019) found that character education was associated with higher levels of behavioral (self-discipline) and educational outcomes (e.g., grades, test scores). Across the diverse school populations sampled in the 52 studies, the positive effects of character education held for students of racial minority and of low SES.

While Tier 1 PBIS, SEL, and character education are common approaches used in schools, they are often implemented in a fragmented and siloed manner (Elias, 2014; Cook et al., 2015). There have been calls to intentionally integrate these approaches given that they come from different yet complementary theoretical underpinnings. Researchers have argued for combining PBIS and SEL into a more comprehensive Tier 1 approach to SEB programming (Bradshaw et al., 2013). Moreover, while SEL and character education have been developed as distinct approaches (Elias et al., 2008), researchers have identified the synergy between the two approaches and argued for the two to be integrated into a unified approach (Elias, 2014). The calls for combining Tier 1 universal SEB support into a more comprehensive approach are consistent with the concept of integrated prevention.

Integrated Prevention

Existing Tier 1 universal SEB programs often differ regarding the underlying theory of change and practical strategies used to drive meaningful changes in student outcomes. These theoretical differences manifest into different practices that have the potential to be advantageous from an integrated prevention standpoint because together the practices help create a more comprehensive approach to Tier 1 universal SEB programming. Leading researchers (e.g., Domitrovich et al., 2010; Osher et al., 2016) have called for the development and evaluation of integrated models of prevention that increase the effect sizes of prevention efforts. Integrated prevention consists of merging different independent universal programs or practices into a complementary approach that enables students to access more comprehensive experiences and support that has the potential to produce better outcomes than can be achieved by either one alone. An integrated approach provides greater theoretical breadth, access to greater exposure regarding time and practices, as well as draws from the strengths of each program to promote synergistic effects. Integrated approaches may also serve to address some of the resource limitations associated when implementing more intensive targeted or intensive interventions by preventing students who require more costly and time-intensive support (Domitrovich et al., 2010). When integrating universal supports it is recommended that they are complementary, rather than redundant, to ensure that program practices do not introduce an unnecessary burden on educators to implement (Fixsen et al., 2009).

There is mounting evidence supporting integrated prevention for school mental health. For example, Cook and colleagues (2015) conducted a study to examine the combined and isolated effects of a social-emotional learning curriculum and Tier 1 positive behavior intervention and supports (PBIS). Findings from this study revealed effects supportive of the concept of integrated prevention, with teachers implementing a combined SEL plus PBIS approach being associated with greater reductions in both externalizing and internalizing problems among students relative to those in SEL only, PBIS only, and business-as-usual conditions. Other studies have demonstrated evidence supporting integrated prevention by combining academic with behavioral supports (Gettinger et al., 2021) or SEL programs with other classroom-based practices (Reinke et al., 2012). Although there is a growing body of evidence supporting integrated universal prevention, more research is needed to extend relevant theoretical and practical knowledge, such as the specific and dynamic mechanisms of change for integrated prevention, the synergistic or potentially agonistic effects of integrating universal programs of different nature, and treatment effect heterogeneity of integrated prevention among diverse service settings and populations (Ialongo et al., 2019; Domitrovich et al., 2010).

CharacterStrong's PurposeFull People

CharacterStrong's PurposeFull People was a Tier 1 universal prevention program that was designed to integrate traditional SEL skills-focused instruction with character education for elementary students (kindergarten through 5th grades) with vertically aligned content. The PurposeFull People was designed to support students to learn 10-character traits (courage, respect, perseverance, gratitude, honesty, kindness, empathy, responsibility, cooperation, and creativity) that also serve as organizing units for specific lessons. These character traits are defined in collaboration with school personnel and students through a community agreements approach (Berkowitz, 2002; Pattaro, 2016). Specifically, a comprehensive list of character traits was used to spur conversations among school personnel students regarding what they look like, sound like, and feel like to put them into action through choices and behavior (Brown, 2005). This way, school personnel's and students' perspectives and voices are used to select and define the character traits that match their languages, values, beliefs, and cultural backgrounds. The final list of character traits is used to support students to reflect on and describe who they want to be as a student. Students also develop plans about how to put their character traits into action on a daily and consistent basis. Educators then support students in following through specific generalization practices that involve reminders, recognizing and acknowledging using language consistent with the character traits, and reviewing opportunities where students can apply the character traits in action.

Each of the lessons is organized by a 5-part instructional sequence: (a) start intentionally, (b) engage relationally, (c) respond with empathy, (d) value practiced consistently, and (e) exit intentionally. Within this lesson structure, students are taught specific social-emotional skills including social skills (active listening, conflict resolution) and emotion regulation skills (name-it-to-tame it, taking a deep breath). There are also supplemental lessons designed to crosswalk the character traits and social-emotional skills with school-wide and classroom behavioral norms or expectations, which make them amenable for integrated prevention with Tier 1 PBIS. Specifically, social-emotional learning and character education provided students with the internal skills, mindsets, and characters that support positive behavior. Conversely, Tier 1 PBIS provided explicit behavioral expectations that enabled students to use their characters and social-emotional skills to regulate themselves, which in turn led to consistently exhibited behavioral expectations and an inclusive and supportive environment for all. Every week, on-site coaches supported the educators to teach and practice these integrations with their students so they understood that (a) social-emotional skills help people regulate self in response to situations to exhibit expected behaviors, and (b) character education helps people think about who they are and how they want to show up as an important member of the school community

by consistently exhibiting expected behaviors.

Gaps in the Literature and Study Aims

Several gaps in the extant literature warrant this study. First, although PurposeFull People was grounded in SEL research and learning theories, no empirical evidence exists to support the effectiveness of PurposeFull People when delivered as part of routine practice in elementary school settings. Thus, comprehensive program evaluation with proper methods (e.g., a cluster-randomized trial) is vital to aid researchers, practitioners, and school leaders to make data-based decisions about the expected effectiveness (Parker et al., 2021). Second, there is a need to continue to add evidence to the literature on integrated prevention. This study provides a unique approach to integrated prevention as PurposeFull People was designed to integrate SEL and character education. Also, by evaluating the effects of PurposeFull People in the context of schools already implementing Tier 1 PBIS with high fidelity, there is an opportunity to examine the additive effects of integrating SEL and character education with Tier 1 PBIS. Third, prevention researchers suggest that the effectiveness of a universal program likely varies according to children's baseline status (Greenberg & Abenavoli, 2017). Children with higher needs at baseline are likely to be more responsive to universal programs than peers who are higher in strengths and low on difficulties at baseline (i.e., demonstrating a larger or steeper change over time; Calhoun et al., 2020). Thus, it is important to extend main effect analyses by exploring if the treatment effect varies across subgroups based on baseline status on outcomes as well as child- and teacher-level moderators (e.g., demographics, fidelity; Spybrook et al., 2020).

To address extant gaps in the literature, this study conducted a cluster-randomized trial (CRT) in elementary schools to evaluate the efficacy of PurposeFull People SEL and character education program on student SEB outcomes when delivered in schools implementing Tier 1 PBIS with fidelity. In addition, this study explored "for whom and under what conditions" the PurposeFull People was effective. Three research questions guided the study design, data collection, and analyses:

(1) Compared to the Tier 1 PBIS only condition, did the PurposeFull People plus Tier 1 PBIS condition lead to a greater improvement in student SEB outcomes (expected behaviors, character, academic engaged time, and discipline referral) while controlling for demographics, baseline status on outcomes, and PBIS fidelity?

(2) To what extent do student and/or teacher demographics moderate the effect of the PurposeFull People plus Tier 1 PBIS on student SEB outcomes?

(3) To what extent do students' baseline status on outcomes moderate the effect of the PurposeFull People plus Tier 1 PBIS on their outcomes at posttest?

Method

Setting and Participants

This study was part of the consultation project between the authors and a Midwest school district already implementing PBIS about extending their universal SEL programming. Participating teachers and students were randomly recruited from five elementary schools in that district. The schools were diverse regarding race and socioeconomic backgrounds of students (MNon-White = 67.2%; MFRPL =79%, MSpEd=17.6%). At the time of this study, the schools were implementing Tier 1 PBIS with high fidelity (M = 85, SD = 4.67) as evidenced by the Tier 1 section of the PBIS Tiered Fidelity Inventory (TFI, a gold standard fidelity measure for PBIS; McIntosh et al., 2017). Most participating teachers (nteacher= 20; Grades 4 and 5) self-identified as female (85%), White (70%), and with less than 12 years of work experience (Table 1). Most participating students (nstudent = 161) were male (51.55%), White (44.72%), and received free/reduced-priced lunch (68.75%). The 2-level random recruitment process generated a nested data structure where 161 students were nested in 20 teachers' classrooms (Treweek et al., 2018). The sample demographics were generally consistent with the school population of the Midwest (Tipton & Miller, 2022).

Procedures

This study was first approved by the school district's research review committee and school board. The authors collaborated with district leaders to randomly select five elementary schools for a cluster-randomized trial. Given the available resource and existing programs, the leaders from the participating schools identified upper elementary grades (4th and 5th grades) that they wanted to focus on for the study. Four teachers were randomly selected from each participating school. After consenting, the 20 teachers (and their corresponding classrooms) were paired within schools based on class-wide student demographics (i.e., ethnicity, free/reduced-priced lunch, gender) using the nearest-neighbor matching approach (Dang et al., 2021). Each teacher within a pair was then randomly assigned to either the active control (ncontrol=20, Tier 1 PBIS) or treatment group (ntreatment=20, PurposeFull People combined with Tier 1 PBIS). To enhance trial cost-efficiency and retain sample representativeness, 7-10 students from each teacher's classroom were randomly selected for obtaining parental consent to complete the pre- and posttests (ncontrol student = 81, ntreatment student = 80; Consented rate = 90.45%; see Figure 1).

This study involved multiple steps across five months as depicted in the CONSORT diagram

(Figure 1; Moher et al., 2012). We followed the CONSORT report guideline for cluster trials (Supplementary Material 1). To enable proper causal inference, the multi-step design separated and sequenced the study components of the baseline pretests of four outcomes (Behavior Expectations, Behavior Discipline, Character, and Academic Engaged Time), PurposeFull People training and implementation, as well as the 4-month posttests of four outcomes (Figure 1). In this study, the school- and classroom-level fidelity of PBIS were background variables and not a part of the study design or analytic models. Thus, the PBIS fidelity data were collected only at baseline before implementing PurposeFull People. To assess the effects of the PurposeFull People on outcomes of interest to local educators, the four outcomes were determined in collaboration with the participating schools and then were assessed via a multi-method multi-informant approach.

Study Conditions

Before active implementation, teachers in the treatment group (PurposeFull People combined with PBIS) received 3-hour training from the authors in a tell-show-do approach about "why, what, and how" to implement PurposeFull People, including how it complements and integrates with Tier 1 PBIS. Specifically, teachers received instruction in PurposeFull People structure and content, with an emphasis on the core components to deliver it with fidelity. Teachers had an opportunity to see aspects of the program delivered as well as gain firsthand experience using the digital platform and role-play implementing aspects of the curriculum. After training and before active implementation, the authors supported teachers to create an implementation plan that spelled out when and how they intended to implement PurposeFull People as well as identifying barriers to implementation and overcoming those barriers to stay the course with implementation (Sanetti et al., 2018). During the 4-month active implementation, teachers participated in structured weekly professional learning community (PLC) meetings led by the authors that involved reflection, collaboration, and planning with others around PurposeFull People implementation. The PLC protocol included a planning template to support teachers to identify ways in which they would improve the delivery of PurposeFull People with fidelity and integrate it intentionally with PBIS. Teachers also received weekly prompts and reminders from the authors to facilitate their delivery of PurposeFull People and Tier 1 PBIS.

Teachers in the Tier 1 PBIS only control group met with their school administrators for the same duration as the PurposeFull People training session to control for time and attention. Moreover, these teachers received the same amount of follow-up support in the form of weekly PLCs dedicated to reflecting on, collaborating, and planning improvements in Tier 1 PBIS implementation and weekly prompts and reminders about the implementation of Tier 1 PBIS practices (e.g., teaching, model, cue, recognize/acknowledge, and respond to behavior).

Measures

Student Behavior Expectations

A 3-item scale was designed to assess the degree to which students exhibited behaviors consistent with school-wide behavioral expectations established in the participating school district's existing PBIS program (consistent across all schools). The first two items assessed the degree to which students exhibited expected behaviors in classroom or non-classroom settings, while the last item assessed the degree to which students exhibited behaviors above and beyond the behavioral expectations (e.g., be respectful, safe, and responsible). Teachers rated all items were rated for the participating students in their classrooms on a 4-point Likert scale (0 = “never” to 3 = “almost all the time”). The mean of the three items was used as the outcome variable for behavioral expectation in data analysis. This scale showed acceptable internal consistency (McDonald’s $\omega = .65$; Cronbach's α is not used due to its deficiency in small scales and unrealistic assumption; Dunn et al, 2016) and test-retest reliability (Pearson’s $r = .67$, $p < .01$).

Student Behavior Discipline

Student behavior discipline was gathered using school administrative data. The data available were a binary outcome variable (yes vs. no) to indicate whether a student had at least one office discipline referral (ODR) in the past four months. The 4-month window ensured that ODR that happened at baseline (Fall to Winter before PurposeFull People) would not be counted toward the posttest (Winter to Spring after PurposeFull People). In education literature, ODR (in a count or binary format; Rocque, 2010) is a common and reliable indicator of disciplinary consequences of student problem behaviors at the class or school levels (Irvin et al., 2004).

Character Building

To assess the impact of the character education component of PurposeFull People, a 3-item self-report scale was developed with attention to the items' readability for the intended age range (i.e., elementary students). The scale consisted of three items “I understand why character is important,” “I understand how my school’s behavior expectations are connected to my character as a student,” and “I am motivated to show positive character traits at school.” Each item was rated on a 4-point Likert scale (0 = “not at all true for me” to 3 = “very true for me”). The mean of the three items was used as the outcome variable for student character building in data analysis. This scale demonstrated acceptable internal consistency (McDonald’s $\omega = .85$) and test-retest reliability (Pearson’s $r = .71$, $p < .01$).

Academic Engaged Time (AET)

The Direct Behavior Rating: Single-Item Scale (DBR-SIS; Chafouleas et al., 2011) was completed by participating teachers to assess AET for each of the randomly selected students in their classrooms. AET was defined as a student being on-task by paying attention to and participating in the academic instruction or task at hand, including asking and answering questions, participating with others in small groups, and staying on-task during independent work. Teachers assessed the percentage of AET for a student on a single scale ranging from 0% to 100% of the time. For data analysis, the outcome variable of AET was the mean of three DBRs completed during pre-determined core instruction times. Prior research has demonstrated the reliability and validity of DBR-SIS for AET (Smith et al., 2018).

School-Level Intervention Fidelity of Tier 1 PBIS

The Tiered Fidelity Inventory (TFI) is a comprehensive and effective tool to determine the extent to which a school is implementing PBIS (Algozzine et al., 2014). Given the scope of this study, only the Tier 1 TFI data were gathered by trained raters for this study. The Tier 1 ratings occur across 15 different dimensions, capturing implementation components such as teaching behavioral expectations, responding to well-defined problem behaviors, teaming, and data collection. Each dimension is rated on a 3-point Likert scale (0 = “not implemented” to 2= “fully implemented”, which results in a total possible score of 30. The TFI demonstrated acceptable or great psychometrics (test-retest reliability = .99, internal consistency = .96, and significant correlation with other measures of intervention fidelity; McIntosh et al., 2017)

Teacher-Level Intervention Fidelity of Tier 1 PBIS

The TFI does not capture teacher/classroom-level intervention fidelity of Tier 1 PBIS. Hence, we created a 10-item self-report measure of individual teachers’ Tier 1 PBIS fidelity in their classrooms based on prior research to distill Tier 1 PBIS into common practice elements (Filter et al., 2022). Example practice elements included (a) teach and model 3-5 schoolwide behavior expectations based upon the behavior expectation matrix, (b) anticipate and deliver pre-statements to encourage behavioral expectations, and (c) refer to visuals (e.g., posters that serve as visual cues) to review and encourage schoolwide behavior expectations. Each item represents a practice element and was rated by teachers on a 3-point Likert scale (0 = not implementing to 2 = fully implementing). The scale score is the sum of all items (Max = 20). In this sample, teachers implemented Tier 1 PBIS with relatively high fidelity (M = 17.4, SD= 1.62).

Teacher *PurposeFull People* Fidelity

PurposeFull People fidelity was measured using a 5-item measure that was completed via interviews with teachers. The scale was used to capture the delivery of the core components of *PurposeFull People* that are hypothesized to be linked to an increased probability of promoting positive changes in student outcomes. These included (a) delivered the character trait lessons as planned and at least two times per week, (b) delivered the SEL skill lessons to teach the social skills and emotion regulation skills, (c) implemented generalization practices consistently (remind, recognize, and review) to support students to apply character traits and SEL skills beyond the lessons, (d) delivered the integration lessons linking character traits to PBIS expectations, and (e) attended weekly PLCs to drive continuous improvement in implementation. These items were rated on a 3-point scale ranging from 0 = did not implement consistently to 2 = fully implemented. These data were gathered for teachers in the *PurposeFull People* condition only, and results indicated that teachers adequately adhered to the core components of *PurposeFull People*, with a mean of 9.3 out of 10 (range 8-10).

Data Analysis

Before analysis, Chi-square tests and independent-sample t-tests were performed to confirm the baseline equivalences between study conditions regarding participant demographics, PBIS fidelity, and the pretest values of four outcomes (Table 1). A series of 2-level ML-ANCOVAs (Wan, 2021) was built to compare the treatment effect of PurposeFull People combined with PBIS against that of PBIS alone on each of the continuous outcomes (behavioral expectation, character, or AET). Two-level generalized linear mixed models (GLMM; Li & Redden, 2015) were built to estimate the cluster-specific results for the binary outcome of behavioral discipline. The multilevel models (MLMs, i.e., ML-ANCOVA and GLMM) accommodated the nested data structure (student nested in teachers/classes), class-level treatment, and cross-level interactions between treatment and key covariates while accounting for and explaining the between-class variations of the outcomes (e.g., teachers may make ODRs differently).

With level-specific centering of covariates (i.e., group-mean center at student level and grand-mean center at classroom level; Hamaker & Muthen, 2020), the descriptive statistics [e.g., no missing values, distribution, and intra-class correlations (ICCs); Table 1] supported the sample adequacy for MLMs. The ML-ANCOVAs and GLMMs were configured the same way following a stepwise modeling approach (Heck & Thomas, 2020; equations in Supplementary Material 2). For RQ1, the main effect model was built for each of the four outcomes (Table 2). In the individual-level (level-1) equations, the posttest of an outcome served as the level-1 dependent variable controlling for individual baseline outcomes and student demographics. The level-1

intercept was set as random to vary across classes, which was predicted by treatment (PurposeFull People and PBIS vs. PBIS alone) controlling for teacher demographics, class-level baseline outcome, school- and classroom-wide PBIS fidelity at the level-2 equation. For RQ2, we added treatment into all level-2 equations of the fixed effects of student demographics to probe their cross-level interactions (Table 3). At the level-2 equation of the random intercept, we created interaction terms by multiplying treatment with teacher demographics and class-wide PBIS fidelity to probe their interactions (Table 4). For RQ3, the level-1 slope of baseline outcome was predicted by level-2 treatment to probe their cross-level interaction (Table 5).

Power analysis was performed with the Optimal Design 3.01. Given our sample sizes, nine covariates, Alpha level of .05, an ICC of .01, and 70% variance explained by level-2 covariates, we will have sufficient power (>= 80%) to detect a minimum detectable effect size as small as 0.23. We interpreted results based on statistical significance and standardized mean differences (SMDES; Morris, 2008). The SMDES is interpreted as the units of standard deviations of the pre-posttest mean difference between the treatment and the control conditions. Analyses were performed with HLM ver. 6.08 and SPSS ver. 26.

Results

RQs 1 and 2: Main Effect and Subgroup Interaction Models

Chi-square and independent-sample t-tests revealed that random assignment created probabilistically equivalent groups at both student and teacher levels regarding baseline outcomes, school-/class-wide PBIS fidelity, teacher and student demographics (Table 1). For RQ 1, results revealed consistent positive main effects of PurposeFull People combined with PBIS on all four outcomes, while controlling for baseline, PBIS fidelity, and demographics (Table 2). Specifically, post-test scores in the treatment group were significantly improved over those in the control group regarding behavioral expectation ($b = 0.08$, $p < .01$; SMDES = 0.35), character ($b = 0.22$, $p < .05$; SMDES = 0.67), AET ($b = 2.84$, $p < .05$; SMDES =0.36), and behavior discipline (odds ratio= 0.12, $p < .05$). The significant coefficients of baseline across all outcomes suggested that a higher baseline was associated with a better posttest score. Unlike the main effect models, the nonsignificant results of demographic-by-treatment-interaction models (RQs 2; Tables 3 and 4) indicated that none of the class-wide PBIS fidelity, teacher demographics, or student demographics significantly moderated the treatment effect. The nonsignificant interactions implied that PurposeFull People combined with PBIS functioned similarly across student and teacher demographic subgroups regardless of Tier 1 PBIS fidelity.

RQ 3: Cross-level Interaction Between Treatment and Baseline Outcome

Moderation analyses revealed significant cross-level interactions between class-level treatment and student-level baseline values of AET, character, and behavioral expectation (Table 5; Figures 2, 3, and 4, respectively). To facilitate interpretation, plots of the class-specific slopes of baseline for each of the three outcomes and two smoothed regression lines for the treatment or control groups were created (Loader, 2012). The class-specific associations between baseline and posttest of outcomes (i.e., separate lines) were significantly clustered based on study conditions. Among students starting with a high level of baseline outcome, the posttest scores between treatment and control groups were similar as evidenced by the converging lines in the high baseline region (upper right corner in figures). But among students starting with a low level of baseline outcome, the posttest scores of students from treatment classes were significantly higher than those of students from control classes as evidenced in the gap between lines in the low baseline region (bottom left corner). Because the control group received PBIS alone, the interaction effects implied that the combination of PurposeFull People and PBIS had a significant "booster" (i.e., additive) effect above and beyond PBIS alone in improving the AET, character, and behaviors of students struggling with low baselines.

Discussion

This study preliminarily evaluated the effectiveness of the PurposeFull People SEL and character education program in the context of schools implementing Tier 1 PBIS with fidelity. This was the first cluster randomized trial evaluating the effects of PurposeFull People, and it provided an opportunity to extend the growing body of research on integrated school-based prevention at the Tier 1 universal level. Our findings revealed the positive effects of PurposeFull People on selected student outcomes measured across different methods and informants. Specifically, positive effects were observed for teacher reports of student behaviors consistent with established school-wide behavioral expectations and norms, reductions in behavior discipline according to school administrative data, improvements in student self-reported understanding and motivation related to character traits, and increases in teachers' direct behavior ratings of academic engaged time. These findings support not only the efficacy of PurposeFull People but provide evidentiary support for the concept of integrated prevention.

The Positive Effect of PurposeFull People on Student SEB Outcomes

There is a need for innovative approaches that integrate SEL and character education (Elias, 2014). PurposeFull People was intentionally designed to bring these two traditions together into a unified approach rather than conceptualize and implement them as separate and potentially

competing approaches (Elias, 2014). The character part of PurposeFull People supports students to explore whom they want to be in the context of character traits as well as receive support to put into action specific character traits through their choices and behavior (Smith, 2013). The SEL part supports students to acquire and use certain social and emotional skills that enable them to better manage thoughts, feelings, and behaviors during social and academic situations to consistently display character traits and stay on track toward achieving important short-term and long-term goals for themselves. Students' application of character traits and social-emotional skills are supported through educators' use of generalization practices (pre-corrective reminders, reinforcement, and feedback) that help students apply what they learn beyond the lessons themselves, which is critical to interventions that focus on acquisition and application of knowledge and skills (McIntosh & MacKay, 2008). What is unclear from this study design is whether similar effects would have been observed if only character traits or SEL skills were taught in the context of Tier 1 PBIS. Future studies should explore SEL and character education in isolation and together to identify their unique contributions as well as the additive effects when the two approaches are integrated.

The findings observed in this study are likely due in part to the implementation support and the inclusion of specific lessons to cross-walk the character traits and SEL skills taught in PurposeFull People with school-wide behavioral expectations. The implementation supports included training, post-training implementation planning, weekly semi-structured PLCs to drive continuous improvement, and weekly email prompts and reminders. The combination of these implementation supports appeared to result in high-fidelity implementation across all the teachers, which is consistent with research on implementation strategies that drive successful implementation (Cook et al., 2019). The specific lessons dedicated to integrating PurposeFull People with Tier 1 PBIS are also unique as most Tier 1 programs do not include specific lessons and implementation supports aimed to intentionally support integration with other Tier 1 universal SEB supports. SEL and character education is a crowded space with numerous programs being marketed to schools. There is the potential that PurposeFull People has an advantage over other SEL or character education programs by integrating the two paradigms into one and including resources to integrate with other Tier 1 approaches such as PBIS. However, more research is needed to replicate and extend the findings from this study about PurposeFull People and whether it has a competitive advantage over other programs. Generally, there is a need in the prevention science and school mental health literature for comparative experimental research that evaluates the relative effects and advantages on implementation and student outcomes by comparing different SEL programs against one another (Duncan et al., 2017).

Integration of SEL, Character Education, and PBIS

It is plausible to expect that the integration of multiple Tier 1 universal SEB programs would not

produce any additive benefits and may exceed the capacity of educators to implement the programs in tandem with fidelity (Garwood, 2022). Research similar to this study is needed to help guide educational decision-makers in knowing whether the integration of Tier 1 universal supports can be reasonably implemented with fidelity and leads to enhanced student outcomes. Findings suggested that teachers were able to maintain fidelity with Tier 1 PBIS practice elements while also reaching adequate fidelity with the delivery of the core components of PurposeFull People. What remains unclear is whether the sequencing of integrated approaches matters. For example, would similar effects be observed if Tier 1 PBIS was implemented after PurposeFull People was adopted and implemented? Or would educators be able to reach fidelity in both if Tier 1 PBIS and PurposeFull People were adopted and implemented at the same time? Answering these types of research questions would help guide decision-making around the integration of Tier 1 universal supports and how best to sequence implementation efforts in a way to build a more comprehensive approach to Tier 1 universal programming and achieve successful implementation over time.

Baseline Status Moderating the Effect of Tier 1 Universal Programs

It is noteworthy to discuss the interaction effects revealed in this study indicating that baseline status on outcome measures served as a moderator of the effect size of PurposeFull People on student outcomes. Findings suggested that PurposeFull People demonstrated the strongest effects for students who had the greatest need and room for improvement at baseline. It is important to consider the characteristics associated with students struggling with the greatest needs at baseline for universal prevention programming. In school-based literature, the structured and systemic oppression in the educational and societal systems where students learn and develop imposed greater levels of risk for students from certain demographic backgrounds, such as living in poverty or chaotic households, exposure to trauma to adverse childhood experiences, and coming from racial and linguistic minoritized groups (Baugh et al., 2019; Berger, 2019). Our findings are consistent with the literature on SEL programming that has found similar interaction effects between treatment and student subgroups (Low et al., 2015).

Moreover, the significant moderation effect of student baseline status in our study echoed Greenberg and Abenavoli's (2017) recommendation that the evaluation of universal prevention programs should dive into subgroup analysis such as comparing the differential effects of a program across subgroups based on baseline status or demographics. While this study was not specifically designed and powered for subgroup analyses, our findings did support such an approach to develop a more precise understanding regarding for whom certain Tier 1 universal SEB programs work (Thayer et al., 2019). This finding also sheds light on the extended utility of Tier 1 universal prevention programs (e.g., PurposeFull People) to be strategically used as Tier 2 small-group interventions for students with greater needs (e.g., low baseline in the target

outcomes) or low-resource schools. For instance, school mental health professionals can consult with teachers to select relevant components or practices from the full manual of PurposeFull People to deliver to a subgroup of students in their classroom identified as high-need/risk via universal screening. Given the wide availability of evidence-based Tier 2 SEB interventions, future research can explore the additive effect of integrating existing Tier 2 SEB interventions and selective character education components from PurposeFull People delivered as Tier 2 small group interventions.

Implications and Limitations for Practice and Research

The findings from this study and similar ones on integrated universal prevention highlighted the importance of avoiding siloed or fragmented approaches. Together, these findings can inform mental health practitioners, school leaders, and policymakers to invest and try out integrated approaches to Tier 1 universal prevention programming for diverse school settings and student populations. Furthermore, practitioners and leaders pay attention to implementation facilitators and barriers to the successful integration of universal programs of different yet complementary theoretical underpinnings, such as program differentiation and connection (Carroll et al., 2007). For instance, PurposeFull People are designed to be crosswalked with PBIS. So implementation supports were provided to guide teachers to make proper distinctions and connections between the components of the two different approaches.

Several limitations in this study warrant caution in result interpretation and call for future research. First, this study was conducted with upper elementary students (4 to 5th graders) from a limited number of elementary schools in the Midwest US, which may impact the generalizability of the findings to other grades in elementary schools in other US regions or foreign countries. Future research should replicate and extend our study with more elementary schools from different geographic locations serving diverse demographics of students. Second, while this study used a multi-method, multi-informant approach to outcome evaluation, the focal outcomes of student behavior and characters were assessed via ad hoc measures for this study. Plus, these ad hoc measures did not undergo comprehensive measurement validation, although we evaluated preliminary psychometrics in the current sample and found acceptable reliabilities. Given the scope of the research questions, this study did not include specific measures of student use of social-emotional skills or indicators of social-emotional well-being (e.g., emotional well-being, perceptions of belonging; Nisar et al., 2022). Also, the teachers' self-report of fidelity may yield unreliable results compared to structured observations. Future studies need to include a more comprehensive set of validated measures to assess effects across a wider range of student outcomes and teacher implementation behaviors (i.e., fidelity).

Third, potential contamination may exist between study conditions because teachers from different conditions but the same school may communicate. However, this would unlikely

jeopardize the internal validity of the findings (e.g., underestimate of effect sizes due to treatment diffusion) because (a) teachers were organized into PLCs according to their conditions, and (b) teachers had to have full access to the PurposeFull People curriculum to implement it to the extent that student outcomes improve (Teerenstra et al., 2006). Moreover, contamination always attenuates treatment effect estimates, but we still identified significant positive effects despite the potential attenuation (Rhoads, 2011). Last, it is important to recognize that the study design and analysis were both at the class level. It may differ from the school-level implementation of integrated prevention where common language and practices are adopted across all educators in the building. Also, most SEL curricula, like PurposeFull People, are supposed to be implemented for all grade levels. Thus, school-level implementation of PurposeFull People may yield different results. Future research should use schools as the unit of treatment and analysis to test school-wide integration of PurposeFull People and PBIS.

Conclusion

This study provided preliminary evidence supporting the efficacy of PurposeFull People in elementary schools. It added to the emerging evidence supportive of integrated prevention, which encourages researchers and practitioners to consider how to effectively integrate different Tier 1 universal approaches into a coordinated and comprehensive whole. The hope is that the findings from this study can inform future research to examine the effects of PurposeFull People in different schools as well as the potential competitive advantage of combining SEL and character education versus standalone approaches. Also, the findings may spur more nuanced research about how best to sequence and scale up integrated universal prevention programs to achieve better and more equitable outcomes for all students (Fagan et al., 2019). Ultimately, our findings corroborated with previous studies (e.g., Cook et al., 2015) to support that SEL and Tier 1 PBIS work synergistically to promote better outcomes for students than either one alone. Rather than treating existing Tier 1 programs as separate and implementing them in a fragmented manner, emerging literature has warranted the need to integrate them to enhance the overall effectiveness to promote the complete mental health of the increasingly diverse school population (e.g., Domitrovich et al., 2010; Elias et al., 2015; Ialongo et al., 2019).

Reference

Algozzine, R. F., Barrett, S., Eber, L., George, H., Horner, R. H., Lewis, T. J., . . .Sugai, G. (2014). SWPBIS Tiered Fidelity Inventory. Eugene, OR: OSEP Technical Assistance Center on Positive Behavioral Interventions and Supports. http://www.pbis.org

Anthony, C. J., Ogg, J., & Jenkins, L. N. (2021). Academic enablers as dynamic moderators: Exploring academic enablers and achievement across elementary school. *Journal of School Psychology*, *86*, 15-31.

Baugh, A. D., Vanderbilt, A. A., & Baugh, R. F. (2019). The dynamics of poverty, educational attainment, and the children of the disadvantaged entering medical school. *Advances in Medical Education and Practice*, *10*, 667.

Berger, E. (2019). Multi-tiered approaches to trauma-informed care in schools: A systematic review. *School Mental Health*, *11*(4), 650-664.

Bradshaw, C. P., Bottiani, J., Osher, D., & Sugai, G. (2013). Integrating positive behavioral interventions and supports (PBIS) and social emotional learning. In M. D. Weist, N. A. Lever, C. P. Bradshaw, & J. Owens (Eds.), Handbook of school mental health: Advancing practice and research (2nd ed., pp. 101–118). New York, NY: Springer.

Brown, J. (2005). Character education: A resurgence to safety, a perspective in curricula trend. *Academic Leadership: The Online Journal, 3(2), 4.*

Berkowitz, M. W. (2002). The science of character education. *Bringing in a new era in character education, 508, 43-63.*

Calhoun, B., Williams, J., Greenberg, M., Domitrovich, C., Russell, M. A., & Fishbein, D. H. (2020). Social emotional learning program boosts early social and behavioral skills in low-income urban children. *Frontiers in Psychology, 11, 561196.*

Chafouleas, S. M., Kilgus, S. P., Riley-Tillman, T. C., Jaffery, R., & Harrison, S. (2011). Direct Behavior Rating Single Item Scales (DBR-SIS): An evaluation of the impact of various components to rater training on accuracy of rating academic engagement and disruption. *Manuscript under review.*

Chicago, IL: CASEL. CASEL (2022). CASEL guide: Effective social and emotional learning programs: Middle and high school edition. Chicago, IL: CASEL.

Carroll, C., Patterson, M., Wood, S., Booth, A., Rick, J., & Balain, S. (2007). A conceptual framework for implementation fidelity. *Implementation science, 2(1), 1-9.*

Cook CR, Lyon AR, Locke J, Waltz T, & Powell BJ (2019). Adapting a compilation of implementation strategies to advance school-based implementation research and practice. *Prevention Science*, 20(6), 914–935.

Cook, C. R., Frye, M., Slemrod, T., Lyon, A. R., Renshaw, T. L., & Zhang, Y. (2015). An integrated approach to universal prevention: Independent and combined effects of PBIS and SEL on youths' mental health. *School Psychology Quarterly*, *30*(2), 166.

Corcoran, R.P., Cheung, A.C.K., Kim, E., & Xie, C. (2018). Effective universal school-based social and emotional programs for improving academic achievement: A systematic review and meta-analysis of 50 years of research Educational Research Review, 25

Dang, Z., Deng, C., Yang, X., Wei, K., & Huang, H. (2021). Nearest neighbor matching for deep clustering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 13693-13702).*

Domitrovich, C. E., Bradshaw, C. P., Greenberg, M. T., Embry, D., Poduska, J. M., & Ialongo, N. S. (2010). Integrated models of school-based prevention: Logic and theory. *Psychology in the Schools*, *47*(1), 71-88.

Duncan, R., Washburn, I. J., Lewis, K. M., Bavarian, N., DuBois, D. L., Acock, A. C., ... & Flay, B. R. (2017). Can universal SEL programs benefit universally? Effects of the positive action program on multiple trajectories of social-emotional and misconduct behaviors. *Prevention science*, *18*(2), 214-224.

Dunn, T. J., Baguley, T., & Brunsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British journal of psychology, 105(3), 399-412.*

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.

Eklund, K., Kilpatrick, K. D., Kilgus, S. P., & Haider, A. (2018). A systematic review of state-level social-emotional learning standards: Implications for practice and research. School Psychology Review, 47(3), 316–326.

Elias, M. J. (2014). The future of character education and social-emotional learning: The need for whole school and community-linked approaches. *Journal of Character Education*,

Elias, M. J., Leverett, L., Duffell, J. C., Humphrey, N., Stepney, C., & Ferrito, J. (2015). Integrating SEL with related prevention and youth development approaches.

Elias, M.J., Parker, S.J., Kash, V.M., Weissberg, R.P., & O'Brien, M.U. (2008). Social and emotional learning, moral 176 Narvaez 176 education, and character education: A comparative analysis and a view toward convergence. In L.P. Nucci & D. Narvaez (Eds.), Handbook of moral and character education (pp. 248–266). New York: Routledge.

Fagan, A. A., Bumbarger, B. K., Barth, R. P., Bradshaw, C. P., Cooper, B. R., Supplee, L. H., & Walker, D. K. (2019). Scaling up evidence-based interventions in US public systems to prevent behavioral health problems: Challenges and opportunities. *Prevention Science*, *20*(8), 1147-1168.

Filter, K. J., Johnson, L. D., Ford, A. L., Sowle, C. A., Bullard, S. J., Cook, C. R., ... & Dupuis, D. (2022). An Expert Consensus Process to Distill Tier 1 PBIS into Core Practice Elements Essential to Frontline Implementation. *Education and Treatment of Children,*

Fixsen, D. L., Blase, K. A., Naoom, S. F., & Wallace, F. (2009). Core implementation components. *Research on social work practice*, *19*(5), 531-540.

CASEL—Collaborative for Academic, Social, and Emotional Learning (2013). The 2013 CASEL Guide: effective social and emotional learning programs—Preschool and Elementary School Edition. Chicago, IL: Author.

Garner, P. W., Mahatmya, D., Brown, E. L., & Vesely, C. K. (2014). Promoting desirable outcomes among culturally and ethnically diverse children in social emotional learning programs: A multilevel heuristic model. Educational Psychology Review, 26(1), 165

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, 158–167.

Greenberg, M. T. (2010). School-based prevention: current status and future challenges. Effective Education, 2, 27–52.

Goodman-Scott, E., McMahon, G., Kalkbrenner, M. T., Smith-Durkin, S., Patel, S., Czack, A., & Weeks, N. (2021). An Ex Post Facto Study Examining Implementation of Positive Behavioral Interventions and Supports Across School and Community Variables From an Inclusive Innovation Perspective. *Journal of Positive Behavior Interventions, 10983007211013784*.

Garwood, J. D. (2022). Special Educator Burnout and Fidelity in Implementing Behavior Support Plans: A Call to Action. *Journal of Emotional and Behavioral Disorders*,

Gettinger, M., Kratochwill, T. R., Eubanks, A., Foy, A., & Levin, J. R. (2021). Academic and Behavior Combined Support: Evaluation of an integrated supplemental intervention for early elementary students. *Journal of School Psychology*, *89*, 1-19.

Greenberg, M. T., & Abenavoli, R. (2017). Universal interventions: Fully exploring their impacts and potential to produce population-level impacts. *Journal of Research on Educational Effectiveness*, *10*(1), 40-67.

Hagermoser Sanetti, L. M., Williamson, K. M., Long, A. C., & Kratochwill, T. R. (2018). Increasing in-service teacher implementation of classroom management practices through consultation, implementation planning, and participant modeling. *Journal of Positive Behavior Interventions*, *20*(1), 43-59.

Heck, R. H., & Thomas, S. L. (2020). An introduction to multilevel modeling techniques: MLM and SEM approaches. Routledge.

Horner, R. H., & Sugai, G. (2015). School-wide PBIS: An example of applied behavior analysis implemented at a scale of social importance. *Behavior analysis in practice*, *8*(1), 80-85.

Ialongo, N. S., Domitrovich, C., Embry, D., Greenberg, M., Lawson, A., Becker, K. D., & Bradshaw, C. (2019). A randomized controlled trial of the combination of two school-based universal preventive interventions. *Developmental psychology*, *55*(6), 1313.

Ialongo, N. S., Domitrovich, C., Embry, D., Greenberg, M., Lawson, A., Becker, K. D., & Bradshaw, C. (2019). A randomized controlled trial of the combination of two school-based universal preventive interventions. *Developmental psychology, 55(6), 1313.*

Irvin, L. K., Tobin, T. J., Sprague, J. R., Sugai, G., & Vincent, C. G. (2004). Validity of office discipline referral measures as indices of school-wide behavioral status and effects of school-wide behavioral interventions. *Journal of Positive Behavior Interventions*, *6*(3),

Jeynes, W. H. (2019). A meta-analysis on the relationship between character education and student achievement and behavioral outcomes. Education and Urban Society, 51 (1), 33–

Jolivette, K., & Swoszowski, N. C. (2020). Improving youth and staff supports: FW-PBIS framework implementation in juvenile correctional facilities. *Remedial and Special Education, 41(2), 67-69.*

Kim, J., McIntosh, K., Mercer, S. H., & Nese, R. N. (2018). Longitudinal associations between SWPBIS fidelity of implementation and behavior and academic outcomes. *Behavioral Disorders, 43(3), 357-369.*

Li, P., & Redden, D. T. (2015). Comparing denominator degrees of freedom approximations for the generalized linear mixed model in analyzing binary outcome in small sample cluster-randomized trials. *BMC Medical Research Methodology, 15(1), 1-12.*

Loader, C. (2012). Smoothing: local regression techniques. In Handbook of computational statistics (pp. 571-596). Springer, Berlin, Heidelberg.

Low, S., Cook, C. R., Smolkowski, K., Buntain-Ricklefs, J. (2015). Promoting social–emotional competence: An evaluation of the elementary version of Second Step. *Journal of School Psychology*, 53(3), 463–477. doi:10.1016/j.jsp.2015.09.002

McIntosh, K., & MacKay, L. D. (2008). Enhancing generalization of social skills: Making social skills curricula effective after the lesson. *Beyond Behavior*, 18, 18–25.

McIntosh, K., Massar, M. M., Algozzine, R. F., George, H. P., Horner, R. H., Lewis, T. J., & Swain-Bradway, J. (2017). Technical adequacy of the SWPBIS tiered fidelity inventory. *Journal of Positive Behavior Interventions*, *19*(1), 3-13.

Morris, S. B. (2008). Estimating effect sizes from pretest-posttest-control group designs. *Organizational research methods*, *11*(2), 364-386.

Napolitano, C. M., Sewell, M. N., Yoon, H. J., Soto, C. J., & Roberts, B. W. (2021, June). Social, emotional, and behavioral skills: An integrative model of the skills associated with success during adolescence and across the life span. In *Frontiers in Education* (Vol. 6, p. 679561). Frontiers Media SA.

Nisar, H., Elgin, D., Bradshaw, C., Dolan., V., Frey, A., Horner, R., Owens, J., Perales, K., & Sutherland, K. (2022). Promoting Social and Behavioral Success for Learning in Elementary Schools: Practice Recommendations for Elementary School Educators, School and District Administrators, and Parents. 2M Research Services. Contract No. 92990019F0319. Arlington, TX.

O'Neill, J. M., Clark, J. K., & Jones, J. A. (2011). Promoting mental health and preventing substance abuse and violence in elementary students: a randomized control study of the Michigan Model for Health. *Journal of School Health, 81, 320–330.*

Osher, D., Kidron, Y., Brackett, M., Dymnicki, A., Jones, S., & Weissberg, R. P. (2016). Advancing the science and practice of social and emotional learning: Looking back and moving forward. Review of Research in Education, 40(1), 644–681.

Osher, D., Mayer, M. J., Jagers, R. J., Kendziora, K., & Wood, L. (2019). Keeping students safe and helping them thrive: A collaborative handbook on school safety, mental health, and wellness (Vols. 2). Praeger/ABC-CLIO.

Parker, K., Nunns, M., Xiao, Z., Ford, T., & Ukoumunne, O. C. (2021). Characteristics and practices of school-based cluster randomised controlled trials for improving health outcomes in pupils in the United Kingdom: a methodological systematic review. *BMC Medical Research Methodology*, *21*(1), 1-17.

Reinke, W. M., Herman, K. C., Darney, D., Pitchford, J., Becker, K., Domitrovich, C., & Ialongo, N. (2012). Using the classroom check-up model to support implementation of PATHS to PAX. *Advances in School Mental Health Promotion*, *5*(3), 220-232.

Rhoads, C. H. (2011). The implications of "contamination" for experimental design in education. *Journal of Educational and Behavioral Statistics*, *36*(1), 76-104.

Rocque, M. (2010). Office discipline and student behavior: Does race matter?. *American Journal of Education, 116(4), 557-581.*

Smith, B. H. (2013). School-based character education in the United States. *Childhood Education*, *89*(6), 350-355.

Smith, R. L., Eklund, K., & Kilgus, S. P. (2018). Concurrent validity and sensitivity to change of Direct Behavior Rating Single–Item Scales (DBR-SIS) within an elementary sample. School Psychology Quarterly, 33, 83–93.

Sprague, J. R., & Walker, H. M. (2021). *Safe and healthy schools: Practical prevention strategies*. Guilford Publications.

Sprague, J. R., Whitcomb, S., & Bear, G. (2019). Mechanisms for promoting and integrating school-wide discipline approaches. In M. Mayer & S. R. Jimerson (Eds.), School safety and violence prevention: Science, practice, policy (pp. 95–120). APA

Spybrook, J., Zhang, Q., Kelcey, B., & Dong, N. (2020). Learning from cluster randomized trials in education: An assessment of the capacity of studies to determine what works, for whom, and under what conditions. *Educational Evaluation and Policy Analysis, 42(3),*

Sutherland, K. S., Conroy, M. A., McLeod, B. D., Kunemund, R., & McKnight, K. (2019). Common practice elements for improving social, emotional, and behavioral outcomes of young elementary school students. *Journal of Emotional and Behavioral Disorders*, *27*(2), 76-85.

Teerenstra, S., Melis, R. J. F., Peer, P. G. M., & Borm, G. F. (2006). Pseudo cluster randomization dealt with selection bias and contamination in clinical trials. *Journal of clinical epidemiology*, *59*(4), 381-386.

Thayer, A. J., Campa, D. M., Weeks, M. R., Buntain-Ricklefs, J., Low, S., Larson, M., & Cook, C. R. (2019). Examining the differential effects of a universal SEL program on student functioning through the dual continua model of mental health. T*he Journal of Primary Prevention,* 40, 405–427. https://doi.org/10.1007/s10935-019-00557-0

Pattaro, C. (2016). Character education: Themes and researches. An academic literature review. *Italian Journal of Sociology of Education, 8(1).*

Tipton, E. & Miller, K. (2022) The Generalizer. Webtool hosted at https://thegeneralizer.org.

Treweek, S., Pitkethly, M., Cook, J., Fraser, C., Mitchell, E., Sullivan, F., ... & Gardner, H. (2018). Strategies to improve recruitment to randomised trials. *Cochrane Database of Systematic Reviews*, (2).

U.S. Department of Education, Office of Special Education and Rehabilitative Services. (2021). Supporting Child and Student Social, Emotional, Behavioral, and Mental Health Needs. Washington, DC.

Wan, F. (2021). Statistical analysis of two arm randomized pre-post designs with one post-treatment measurement. *BMC medical research methodology*, *21*(1), 1-16.

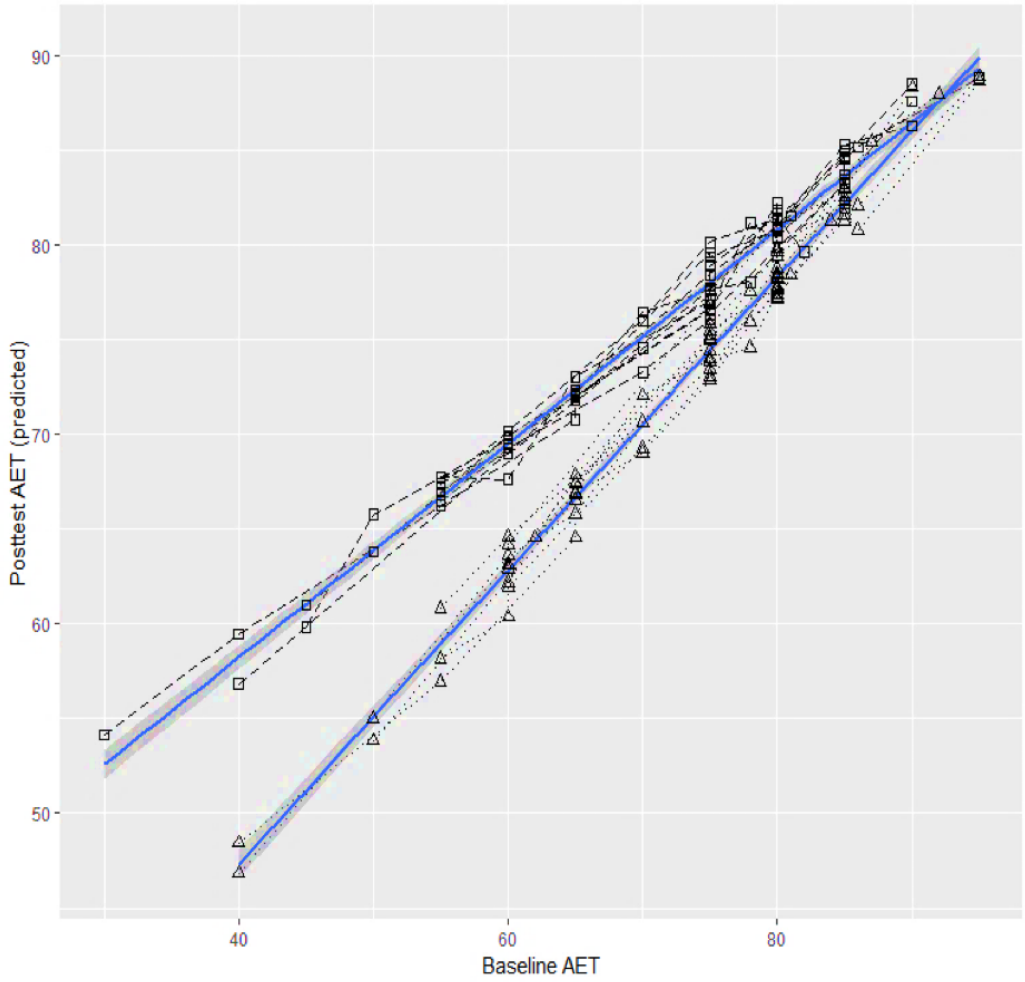

Figure 2. The interaction effect between study conditions and baseline AET on posttest AET. Black lines represent class-specific associations between baseline and posttest AET, long-dash lines with square dots = treatment classes, dotted lines with triangle dots = control classes. Blue solid lines = the smoothed linear regression lines fitted separately for each study condition (upper line = treatment groups, lower line = control group), grey shades surrounding the solid lines = 95% confidence region

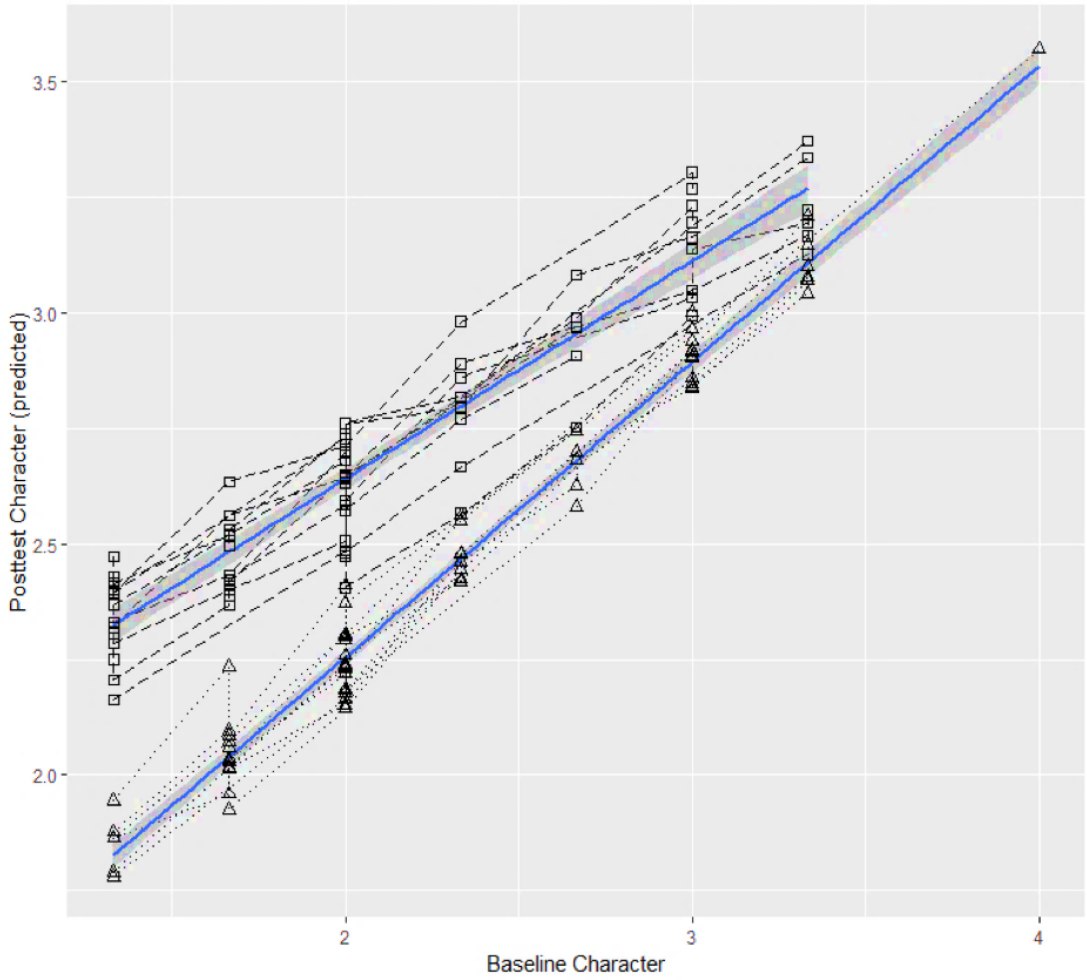

Figure 3. The cross-level interaction effect between study conditions and baseline Characters on the posttest Characters. Black lines represent class-specific associations between baseline and posttest Characters, long-dash lines with square dots = treatment classes, dotted lines with triangle dots = control classes. Blue solid lines = the smoothed linear regression lines fitted separately for each study condition (upper line = treatment groups, lower line = control group), grey shades surrounding the solid lines = 95% confidence region.

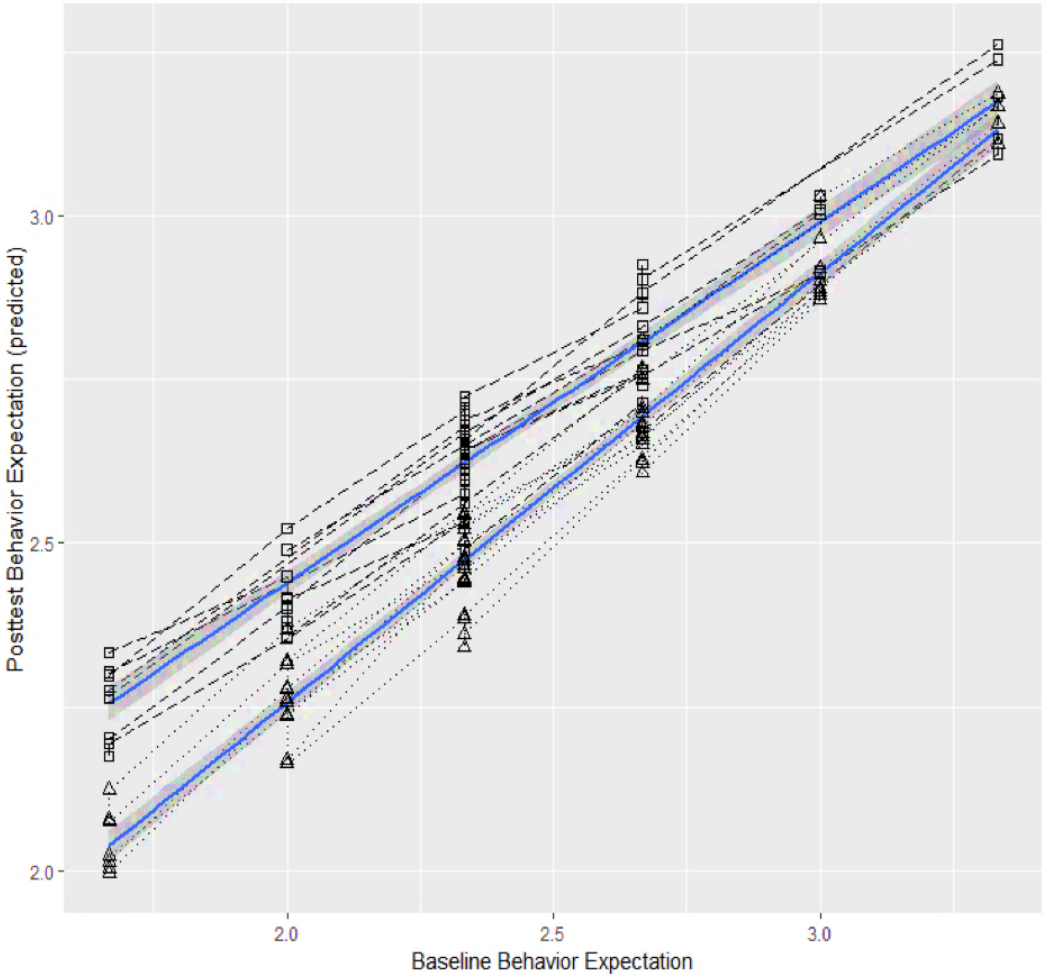

Figure 4. The cross-level interaction effect between study conditions and baseline behavior expectation on the posttest behavior expectation. Black lines represent class-specific associations between baseline and posttest behavior expectation, long-dash lines with square dots = treatment classes, dotted lines with triangle dots = control classes. Blue solid lines = the smoothed linear regression lines fitted separately for each study condition (upper line = treatment groups, lower line = control group), grey shades surrounding the solid lines = 95% confidence region.

Table 1

Descriptive Statistics of Participants and Baseline Equivalences (Nteacher =20, Nstudent=161)

Categorical Variables	Categories	n	%	Baseline Equivalence (χ2)
Teacher at level 2				
Condition	Active control	10	50%	-
	Treatment	10	50%	
Gender	Female	17	85%	$\chi2$ (1, 20) = .39, p = .53
	Male	3	15%	
Race/Ethnicity	Asian	1	5%	$\chi2$ (4, 20) = 0, p = 1
	Black	2	10%	
	Hispanic/Latino	2	10%	
	Mixed	1	5%	
	White	14	70%	
Teaching experience (years)	<= 6	7	35%	$\chi2$ (1, 20) = 1.83, p = .18
	7 - 12	7	35%	
	13+	6	30%	
Grade level	4	9	45%	$\chi2$ (1, 20) = 1.82, p = .18
	5	11	55%	
Student at level 1				
Condition	Active control	81	50.31%	-
	Treatment	80	49.69%	
Gender	Male	83	51.55%	$\chi2$ (1, 161) = .01, p = .94
	Female	78	48.45%	
Race/Ethnicity	White	72	44.72%	$\chi2$ (4, 161) =.5, p = .48
	Hispanic/Latino	25	15.53%	
	Black	19	11.8%	

	Mixed	33	20.5%	
	Other	12	7.45%	
FRPL	Yes	110	68.75%	$\chi 2$ (1, 161) = 0, p = 1
	No	50	31.25%	
Baseline Discipline	Yes	52	32.5%	$\chi 2$ (1, 161) = 0, p = 1
	No	108	67.5%	

Continuous Variables	M (SD) in treatment	M (SD) in control	M Difference	ICC of Posttest	Baseline Equivalence (t-tests)
Behavioral Expectation	2.38(0.43)	2.42(0.44)	-0.04	.065	t (159) = 0.59, p= .55
Character	2.11(0.62)	2.27(0.63)	-0.16	.0001	t (159) = 1.61, p= .11
AET	71.5(13.43)	72.86(11.67)	-1.36	.001	t (159) = 0.69, p= .49
TFI	85(4.67)	85(4.67)	0	-	t (18) = 0, p= 1
Class-wide PBIS Fidelity	17.8(1.62)	18.1(1.52)	-0.30	-	t (18) = 0.43, p=.67

Note. M=mean SD= standard deviation, FRPL= free/reduced price lunch, AET=academic engaged time, ICC = intra-class correlations. "-" = not applicable.

Table 2

Fixed Effect Estimates of ML-ANCOVAs and GLMM Controlling for Baseline Outcome, Fidelity, and Demographics

Teacher-level fixed effects	Behavior Expectation			Character			AET			Discipline (Yes vs. No)		
	b	SE	*p*	*b*	SE	*p*	*b*	SE	*p*	*OR*	95% CI	*p*
Intercept	3.03***	0.14	< .001	2.73***	0.33	< .001	80.05***	5.29	< .001	0	(0, 1)	1
Treatment	0.08**	0.03	.008	0.22*	0.07	.01	2.84*	1.14	.028	0.12*	(0.019, 0.7)	.023
School-lv TFI	0.01*	0.003	.024	0.01	0.01	.459	0.05	0.13	.704	1.07	(0.84, 1.38)	.546
Class-lv PBIS Fidelity	-0.01	0.02	.627	-0.03	0.02	.204	-0.51	0.39	.213	1.33	(0.59, 2.99)	.466
Experience	-0.004	0.003	.130	-0.01	0.01	.331	0.07	0.08	.395	1.1	(0.94, 1.29)	.195
Ethnicity	0.04	0.05	.408	-0.002	0.08	.977	1.25	1. 5	.338	1.84	(0.24, 13.95)	.526
Gender	-0.03	0.04	.396	0.1	0.1	.341	-2.19	1.6	.196	2.12	(0.24, 18.59)	.465
Grade	-0.09*	0.03	.01	-0.04	0.07	.588	-1.86	1.12	.121	0.78	(0.15, 4.1)	.749
Student-level fixed effects												
Ethnicity	-0.04	0.05	.362	-0.04	0.05	.478	-0.18	0.82	.831	0.8	(0.15, 4.21)	.789
Gender	-0.05	0.05	.323	-0.0s9	0.05	.051	1.55*	0.75	.041	1.86	(0.42, 8.23)	.414

FRPL	-0.04	0.05	.388	-0.07	0.05	.191	1.08	0.89	.224	0.63	(0.13, 3.17)	.577
Baseline outcome	0.62***	0.04	< .001	0.58***	0.04	< .001	0.65***	0.03	<.001	3.9 x10^{16}	(0, 1)	1

Note. * < .5, ** < .01, *** < .001. Discipline is binary analyzed by GLMM (1=has been disciplined in past 4 months, 0 = never been disciplined in past 4 months). Treatment = PfP combined with PBIS versus the PBIS only Condition; b = multilevel regression coefficients; SE = standard error, p = the significance of the approximate t-ratio test. OR= odds ratio, 95%CI = 95% confidence interval for OR.

Table 3

ML-ANCOVAs and GLMMs for the Cross-Level Interactions Between Student Demographic and Treatment

Teacher-level fixed effects	Behavior Expectation			Character			AET			Discipline (Yes vs. No)		
	b	SE	*p*	*b*	SE	*p*	*b*	SE	*p*	*OR*	95% CI	*p*
Intercept	3***	0.24	< .001	2.8***	0.33	< .001	80.47***	5.35	<.001	0	(0, 1)	1
Treatment	0.12	0.11	.288	0.14	0.12	.260	2.47	1.96	.233	1.5	(0.03, 73.9)	.824
School-lv TFI	0.01	0.01	.148	0.01	0.01	.446	0.05	0.13	.699	1.13	(0.84, 1.53)	.396
Class-lv PBIS Fidelity	0.01	0.01	.795	-0.04	0.02	.171	-0.53	0.39	.198	1.72	(0.58, 5.08)	.296
Experience	-0.01	0.03	.306	-0.01	0.01	.263	0.07	0.08	.425	1.18	(0.96, 1.44)	.103
Ethnicity	0.04	0.06	.489	-0.01	0.08	.922	1.24	1.24	.338	4.92	(0.33, 73.52)	.224
Gender	-0.04	0.07	.618	0.1	0.1	.334	-2.2	1.58	.191	6.21	(0.35, 110.06)	.192

Grade	-0.09	0.05	.096	-0.05	0.07	.528	-1.91	1.11	.110	0.48	(0.06, 3.75)	.451
Student-level fixed effects												
Ethnicity	-0.04	0.07	.548	-0.13	0.07	.075	-0.67	1.16	.563	2.38	(0.28, 20.65)	.429
Ethnicity X lv2 treatment	0	0.1	.999	0.18	0.1	.070	0.97	1.63	.554	0.01	(0, 1.3)	.064
Gender	-0.12	0.07	.071	-0.09	0.06	.172	1.96	1.07	.069	3.3	(0.46, 23.61)	.233
Gender X lv2 treatment	0.14	0.09	.123	0.001	0.09	.989	-0.82	1.51	.585	0.06	(0.001, 5.42)	.221
FRPL	0.03	0.08	.657	-0.06	0.08	.470	0.93	1.26	.464	0.44	(0.05, 3.8)	.451
FRPL X lv2 treatment	-0.15	0.11	.166	-0.04	0.11	.730	0.31	1.79	.863	4.47	(0.06, 365.54)	.503
Baseline outcome	0.6***	0.05	< .001	0.58***	0.04	< .001	0.65***	0.03	< .001	5.96×10^{21}	(0, 1)	1

Note. . * < .5, ** < .01, *** < .001. Discipline is binary by GLMM (1=has been disciplined in past 4 months, 0 = never been disciplined in past 4 months). Treatment = PfP combined with PBIS versus the PBIS only conditions; b = multilevel regression coefficients; SE = standard error, p = significance of the approximate t-ratio test. OR= odds ratio, 95%CI = 95% confidence interval for OR.

Table 4

ML-ANCOVAs and GLMMs for the Teacher-Level Interactions Between Teacher Demographics and Treatment

Teacher-level fixed effects	Behavior Expectation			Character			AET			Discipline (Yes vs. No)		
	b	SE	*p*	*b*	SE	*p*	*b*	SE	*p*	*OR*	95% CI	*p*
Intercept	3.1***	0.48	< .001	1.96*	0.6	.016	64.27***	8.17	< .001	0	(0, 7.56)	.899
Treatment	0.79	1.27	.554	2.25	1.6	.204	35.23	21.76	.149	0	(0, 4.46)	.918
School-Lv TFI	0.01	0.01	.159	-0.001	0.01	.913	0.03	0.16	.835	0.98	(0.84, 1.13)	.835
Class-lv PBIS Fidelity	0.03	0.05	.591	0.01	0.06	.850	0.45	0.84	.609	0.71	(0.38, 1.33)	.681
Experience	0.004	0.01	.680	-0.01	0.01	.643	0.23	0.14	.143	1.05	(0.94, 1.17)	.879
Ethnicity	0.12	0.13	.377	-0.15	0.17	.393	0.46	2.25	.843	0.47	(0.03, 7.18)	.740
Gender	0.01	0.16	.934	0.06	0.2	.756	1.76	2.73	.540	0.53	(0.1, 2.8)	.768
Grade	-0.12	0.1	.289	0.13	0.13	.348	1.36	1.73	.458	1.17	(0.26, 5.41)	.898
Experience X Treatment	-0.01	0.01	.256	0.01	0.01	.531	-0.1	0.18	.576	0.97	(0.84, 1.11)	.987
Ethnicity X Treatment	-0.12	0.17	.489	0.29	0.22	.217	2.73	2.92	.382	2.3	(0.07, 76.26)	.973
Gender X Treatment	-0.04	0.21	.866	0.05	0.26	.847	-6.57	3.52	.103	11.02	(1.43, 84.83)	.908

Grade X Treatment	0.09	0.13	.531	-0.16	0.17	.377	-4.85	2.24	.067	0.33	(0.05, 2.2)	.979
PBIS Fidelity X Treatment	-0.05	0.06	.391	-0.08	0.07	.280	-0.47	0.95	.633	2.85	(1.44, 5.64)	.913
Student-level fixed effects												
Ethnicity	-0.04	0.05	.440	-0.03	0.05	.492	-0.18	0.82	.831	0.89	(0.22, 3.64)	.905
Gender	-0.05	0.05	.336	-0.08	0.05	.073	1.59	0.75	.035	1.41	(0.37, 5.41)	.649
FRPL	-0.04	0.06	.495	-0.05	0.05	.371	1.35	0.89	.133	0.48	(0.12, 1.92)	.393
Baseline outcome	0.61***	0.05	< .001	0.57***	0.04	< .001	0.65***	0.03	< .001	5.27 x 10^6	(0, 1)	.893

Note. * < .5, ** < .01, *** < .001. Discipline is binary analyzed by GLMM (1=has been disciplined in past 4 months, 0 = never been disciplined in past 4 months). Treatment = PfP combined with PBIS versus the PBIS only Condition; b = multilevel regression coefficients; SE = standard error, p = significance of the approximate t-ratio test. OR= odds ratio, 95% CI = 95% confidence interval for OR.

Table 5

ML-ANCOVAs and GLMMs for the Cross-Level Interactions Between Student Baseline Outcome and Treatment

Teacher-level fixed effects	Behavior Expectation			Character			AET			Discipline (Yes vs. No)		
	b	SE	*p*	*b*	SE	*p*	*b*	SE	*p*	*OR*	95% CI	*p*
Intercept	3.04***	0.15	< .001	2.73***	0.17	< .001	80.17***	3.98	< .001	0	(0, 1)	.994
Treatment	0.08**	0.03	.008	0.22**	0.04	< .001	2.83*	0.91	.01	0.86	(0, 1)	1
School-lv TFI	0.01*	0.003	.024	0.01	0.004	.149	0.05	0.07	.449	1.07	(0.83, 1.38)	.548
Class-lv PBIS Fidelity	-0.01	0.02	.627	-0.03	0.02	.08	-0.51	0.27	.081	1.33	(0.59,3)	.468
Experience	-0.004	0.003	.130	-0.01	0.003	.098	0.07	0.05	.166	1.1	(0.94, 1.29)	.196
Ethnicity	0.04	0.05	.419	-0.004	0.05	.943	1.23	0.79	.146	1.83	(0.24, 14.02)	.528
Gender	-0.03	0.04	.405	0.1**	0.03	.005	-2.17	2.06	.314	2.13	(0.24, 18.79)	.465
Grade	-0.09*	0.03	.01	-0.04	0.04	.344	-1.87*	0.81	.038	0.78	(0.15, 4.12)	.749

Student-level fixed effects												
Ethnicity	-0.05	0.05	.309	-0.04	0.06	.461	-0.25	0.78	.751	0.8	(0.15, 4.22)	.791
Gender	-0.04	0.05	.383	-0.08*	0.04	.026	1.57*	0.72	.03	1.85	(0.42, 8.23)	.415
FRPL	-0.04	0.05	.407	-0.07	0.06	.273	1.02	1.06	.337	0.63	(0.13, 3.16)	.575
Baseline outcome	0.68***	0.04	< .001	0.66***	0.02	< .001	0.76***	0.05	< .001	189	(0, 1)	.994
Baseline outcome X lv2 Treatment	-0.13*	0.06	.04	-0.17*	0.07	.016	-0.21*	0.09	.015	0.13	(0, 1)	1

Note. * < .5, ** < .01, *** < .001. Discipline is binary (1=has been disciplined in past 4 months, 0 = never been disciplined in past 4 months) with GLMM. Treatment = PfP combined with PBIS versus the PBIS only Condition; b = multilevel regression coefficients; SE = standard error, p = significance of the approximate t-ratio test. OR= odds ratio, 95%CI = 95% confidence interval for OR.