

Title: *A Multi-Site Cluster Randomized Trial of CharacterStrong's PurposeFull People Elementary SEL and Character Education Program*

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

Madeline Larson, PhD
Department of Research & Development, CharacterStrong SEL
Program
Email: madeline@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 ML 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 [osf.io/qzwkj/].

Abstract

Social-emotional learning (SEL) and character education during the elementary years is important to prevent social, emotional, and behavioral (SEB) difficulties and promote success-enabling factors. Using a cluster-randomized trial, this study evaluated the effects of *PurposeFull People* (*PfP*) – an SEL and character education program – in the context of schools. We conducted a multi-site cluster-randomized trial to examine the effects of *PurposeFull People* (*PfP*; an SEL and character education program) on four student SEB outcomes (behavior expectations, behavior discipline, character traits, and academic engaged time). This study took place in 10 elementary schools (i.e., sites). At each school, 44 teachers/classrooms were randomly assigned to either treatment (*PfP*; $n_{\text{treatment teacher}}=22$) or control condition (business-as-usual). An average of eight students were randomly recruited from each teacher's classroom ($n_{\text{student total}}=354$; $n_{\text{treatment student}}=177$). The four student outcomes were assessed at baseline and 4-month posttest after completion of the *PfP* program. Three-level general and generalized linear mixed models were used to estimate the effects of *PfP* on student outcomes and to probe treatment effect heterogeneity across student subgroups based on baseline status on each outcome. Compared to business-as-usual, the *PfP* produced significant improvements in all four student outcomes. Results about cross-level interactions indicated that the effects of *PfP* on students' character traits and academic engaged time varied based on their baseline status, where students struggling more at

baseline showed a larger improvement than those functioning adequately at baseline. Limitations and implications for future research and practice are discussed.

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

*A Multi-Site Cluster Randomized Trial of CharacterStrong's PurposeFull
People Elementary SEL and Character Education Program*

Elementary schools represent an important context for universal prevention and promotion programming as it is a setting that provides opportunities to support both students' social, emotional, and behavioral health (SEBH) and academic development (Duong et al., 2021). The focus on elementary students' SEBH is important because these factors have been shown to serve as academic enablers that promote higher levels of engagement and performance (Napolitano et al., 2021). In addition, students' SEBH in elementary school is a predictor of life-course outcomes into adulthood (Mascia et al., 2020). Research consistently shows that students with SEBH difficulties are more likely to struggle to establish healthy relationships with others, experience academic difficulties, and develop mental health problems later on in life (Polanin et al., 2021). Due to all of this, the majority of schools in the US are actively integrating Tier 1 universal programs that target students' SEBH (Berkeley et al., 2020).

Social-emotional learning (SEL) and character education are common approaches to Tier 1 universal programming (Jeynes, 2019; Sprague & Walker, 2021); however, these are often fragmented approaches that are not intentionally integrated together (Osher et al., 2019; Sprague et al., 2016). Integrated prevention offers a promising approach by intentionally combining approaches, like SEL and character education, to provide a more comprehensive approach to Tier 1 universal prevention and promotion. The

purpose of this experimental study was to evaluate the effects of *CharacterStrong's PurposeFull People Program* (hereinafter called *PurposeFull People*), which is an integrated Tier 1 SEL and character education program designed for elementary-age students.

Tier 1 Universal Prevention and Promotion

Tier 1 reflects foundational supports that every student receives on a consistent basis to promote specific outcomes of interest. Tier 1 universal promotion is defined as ways of optimizing and enhancing positive social, emotional, and behavioral functioning by delivering support for each student that strengthens assets, competencies, and relationships that result in enhanced well-being and functioning. Tier 1 universal prevention, on the other hand, is defined as the delivery of support for every student to minimize or prevent the development of SEB difficulties that interfere with aspects of daily functioning in school. Schools are increasingly adopting Tier 1 programs and practices to accomplish the goals of prevention and promotion (Arslan et al., 2021). Two common approaches implemented in schools are SEL and character education. SEL focuses on supporting students to acquire and apply specific social and emotional competencies, such as the ability to identify and regulate emotions, show empathy for others, establish and maintain healthy relationships, and make responsible decisions (CASEL, 2022). Research consistently demonstrates that when schools intentionally implement quality SEL curriculum and instruction that positive effects across a wide range of 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).

Character education is one of the longest-standing approaches to Tier 1 universal programming. Character education focuses on cultivating character strengths or traits that support students to explore who they want to be as ethical and responsible members of their school community as well as performance character linked to improved academic engagement and 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 educational outcomes, including grades, test scores, and self-discipline.

While SEL and character education are common approaches used in schools, they are often implemented separately from one another (Elias, 2014; Cook et al., 2015). Although 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.

CharacterStrong's PurposeFull People

CharacterStrong's *PurposeFull People* represents a unique Tier 1 universal SEB program as it integrates traditional SEL skill-focused instruction with character education. *PurposeFull People* was designed to support students in learning specific 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 used to support students to reflect on and describe who they want to be as a student and how to show up as a positive member of their school. Students also develop plans about how to put their character traits into action on a daily and consistent basis. Educators then support students to apply skills and competencies beyond the lessons through specific generalization practices that involve reminders, recognizing and reinforcing using specific language consistent with the character traits, and reviewing situations where students had opportunities to apply character traits. Each of the units 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.

Gaps in the Literature and Study Aims

Several gaps in the extant literature warrant this study. First, although *PurposeFull People* was grounded in empirical evidence on effective teaching of social-emotional competencies and character traits, there is no empirical evidence to support the effectiveness of *PurposeFull People* when delivered as part of routine practice in elementary schools. Thus, comprehensive program evaluation with proper methods (e.g., multi-site cluster-randomized trial) is vital for researchers and educators to make informed decisions about the program's likely 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. 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 (Spybrook et al., 2020).

To address extant gaps in the literature, we conducted a multi-site cluster randomized trial (MS-CRT) with 10 elementary schools to evaluate the efficacy of *PurposeFull People* SEL and character education program on student SEB and academic-related outcomes. In addition, this study explored "for whom and under what conditions" the *PurposeFull People* was effective. Three sequential research questions (RQs) guided this study: (1) To what extent do student SEB outcomes at posttest [i.e., character building, academic engaged time (AET), behavior expectations and discipline] correlate with *PurposeFull People*, demographics at all three levels (i.e., students, teachers, schools), and baseline outcomes? (2) Compared to the control condition, did *PurposeFull People* lead to improvement in student SEB outcomes while adjusting for demographics at all three levels and students' baseline status on SEB outcomes? (3) To what extent do students' baseline scores on SEB outcomes moderate the effect of *PurposeFull People* on their posttest scores while adjusting for demographics at all three levels?

Method

Setting and Participants

The setting for this study was rural and urban public elementary schools in the Midwest Region of the United States. These schools were

recruited as part of existing research-practice partnerships. A total of 10 schools were recruited, and school-wide demographics of these schools were diverse in terms of race and socioeconomic status ($M_{Non-White} = 47.60\%$; $M_{FRPL} = 61.20\%$, $M_{SpEd} = 15.20\%$; $M_{Reading\ proficiency} = 56\%$; $M_{drop-out} = 3.40\%$). Forty-four 2nd through 5th Grade teachers were recruited from these schools to participate in the study. Most participating teachers ($n_{teacher} = 44$; grade level: 2 through 5) self-identified as female (86.36%), White (72.73%), and with more than 13 years of teaching experience (38.64%; Table 1). A random sample of 8 students were recruited to participate from each of the 44 classrooms for a total of 354. Most participating students were male (50.85%), White (59.60%), and not receiving Free/Reduced-Priced Lunch (FRPL; 51.56%). The school and participant recruitment process led to a 3-level nested data structure - 354 students were nested within the classes of 44 teachers from 10 schools (Treweek et al., 2018). The demographics of teacher and student samples were consistent with rural and urban demographics of the Midwest of the US (Tipton & Miller, 2022).

Procedures

This study was approved by the participating district's board and program review committee. The authors collaborated with district leaders to identify elementary schools (i.e., sites) to conduct a multi-site cluster randomized trial (MS-CRT). The authors then consulted with stakeholders in the participating schools to identify (a) classes from the 2nd through 5th grades as the target group and (b) four student SEB outcomes of interest to

evaluate the effectiveness of *PfP*. From each school, the authors identified four or five teachers, who were debriefed about the study and then consented to participate ($n_{\text{total teacher}}=44$). Parental and student assent were not obtained given the project was within the purview of general education. Within each school, the classes of participating teachers were paired according to their class-wide student demographics (e.g., race, gender, FRPL) with the nearest-neighbor matching method (Dang et al., 2021). Each teacher (and their class) in a pair was then randomly allocated to either the treatment ($n_{\text{treatment teacher}}=22$, *PurposeFull People*) or control group ("business-as-usual", BAU). Around eight students in each participating teacher's class (range: 7-10 per class) were randomly recruited to obtain their parental consent to participate in this study ($n_{\text{total student}} = 354$, $n_{\text{treatment student}} = 177$; Figure 1). This multistage sampling process was to enhance the cost-effectiveness of the MS-CRT while retaining sample representativeness. This study involved multiple steps across five months as depicted in the CONSORT diagram (Figure 1; Moher et al., 2012). The study procedures followed the CONSORT report guideline (extension for CRTs; Supplementary Material 1). To facilitate causal inference, the multi-step study design entailed a sequential organization of key study components, including pretests of four student SEB outcomes (i.e., character building, AET, behavior expectations, discipline), the training and delivery of *PfP*, and the posttests of student outcomes four months after delivery of *PfP* (Figure 1).

Study Conditions

Before active implementation of *PurposeFull People*, teachers in the treatment group received a 3-hour training conducted by CharacterStrong trainers in a "tell-show-do" format that focused on "why, what, and how" to implement *PurposeFull People* with high fidelity. First, teachers learned about the structure and content of *PurposeFull People*, which emphasize the strategic implementation of core components. Teachers were given the opportunity to (a) observe various aspects of *PurposeFull People*, (b) experience the use of the digital platform accompanying *PurposeFull People*, and (c) engage in small group discussion and role-play to practice implementing core components of the program. Prior to active implementation, the authors assisted teachers in developing an actionable implementation plan. The plan outlined when and how they would deliver *PurposeFull People* and strategies for overcoming pre-identified obstacles that might arise during the active implementation stage (Sanetti et al., 2018). In the 4-month active implementation stage, the authors led structured weekly professional learning community (PLC) meetings with teachers. The meetings focused on reflection, collaboration, and planning with colleagues regarding the implementation of *PurposeFull People*. The PLC meetings utilized a planning template to help teachers identify strategies to enhance their implementation of *PurposeFull People*. Additionally, the authors provided teachers with weekly prompts and reminders to support their effective delivery of *PurposeFull People*. For

comparison, teachers in the control group met with their administrators for a duration equivalent to the training of the treatment group, ensuring that participants' time and attention were balanced between the two groups.

Measures

Character Building

Character reflects both students' understanding of character strengths (e.g., respect, responsibility) and motivation and intention to put their character into action through the choices they make and behaviors they engage in. To evaluate the influence of the character education element in *PurposeFull People*, a 3-item self-report scale was used, taking into consideration the readability of the items for our target age group (i.e., elementary students). This scale is composed of three items, which were rated by the participating students on a 4-point Likert scale ranging from 0 ("not at all true") to 3 ("very true for me"). The average score of the three items was used as the outcome variable for students' character building, which was a measure of intrinsic motivation and behavioral intention to put character and SEL skills into action. Intrinsic motivation is a reliable, valid, and significant construct predictive of behavioral enactment (Eccles & Wigfield, 1998, 2002; Gottfried, 1990). In this study, the student-reported character measure consisted of items from a valid and reliable measure of internal motivation to engage in specific behaviors (Gottfried, 1990). Specifically, students rated their understanding of the importance and motivation to put character strengths into action. In the context of this

study, the scale showed good reliability in the current sample (McDonald's $\omega = .82$; Cronbach's $\alpha = .71$; Dunn et al, 2016).

Student Behavior Expectations

PurposeFull People includes content and supports to enable students to reflect on their character and then their 'character into action' by engaging in prosocial behaviors that help create a positive culture and climate. Given the focus on character and behavior, a 3-item teacher-reported scale was created to assess the extent to which a student exhibits behaviors consistent with behavioral expectations. The first two items assessed the degree to which a student's demonstrated behavioral expectations inside and outside of class in school. The 3rd item assessed how often a student exhibited behaviors that went above and beyond behavioral expectations. Teachers rated the three items for each participating student in their classrooms using a 4-point Likert scale ranging from 0 "never" to 3 "almost all the time". The average score of the three items formed the outcome variable for students' behavior expectation. Teacher-reported measures of student behavioral expectations using ratings and points have been shown to be a valid measure of student response to intervention (e.g., Campbell & Anderson, 2011; Park & Blair, 2020). Moreover, operational definitions of behavioral expectations serve as the foundation of behavior report cards which have been shown to correlate with observational methods, demonstrating evidence of convergent validity (Vannest et al., 2010). In the context of this study, the scale had good reliability in the

current sample (McDonald's $\omega = .82$; Cronbach's $\alpha = .82$).

Student Behavior Discipline

Student behavior discipline was gathered from 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).

Academic Engaged Time (AET)

Academic engaged time was the primary academic-related outcome assessed in this study. The operational definition of AET was: Paying attention to and participating in instruction and/or learning activities during class time (e.g., ask and answer questions, work with others in small group activities, stay on-task for independent work). The Direct Behavior Rating: Single-Item Scale (DBR-SIS; Chafouleas et al., 2011) was used by the participating teachers to assess the AET of the participating students in their classrooms. Specifically, a teacher rated the AET of each student based on a scale of "0%" through "100% of the observed time". The outcome variable of a student's AET was the average score of three DBR-SIS from three different core instruction times pre-determined by the teachers.

Existing literature has established the good reliability and validity of DBR-SIS as a pragmatic tool to assess AET (Smith et al., 2018).

Class-Wide Fidelity of PurposeFull People

The intervention fidelity of *PurposeFull People* at classroom level was assessed with a 5-item scale that was completed via interviews with teachers implementing *PfP*. The items were rated on a 3-point Likert scale ranging from 0 "did not implement consistently" to 2 "fully implemented". The items were designed to assess the adherence and dosage of the implementation of the core components of *PfP* that are hypothesized to associate with increased likelihood of promoting student character building and behaviors. 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, role model, 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. The fidelity data (the sum score of 5 items) was collected from teachers in the *PurposeFull People* condition only, which indicated that teachers adequately adhered to the core components of *PurposeFull People* with acceptable dosage ($M = 9.1$, $SD = 1.10$; range: 7-10).

Demographics

In all models, we controlled for some theoretically-relevant demographics at all three levels (i.e., students, teachers, and schools; Table 1) as confounders to enhance the power and precision of coefficient estimates. The participating schools extracted and provided the authors the information from their school administrative, including demographics of students (gender, race, FRPL), teachers (grade level, gender, and teaching experience), and school enrollments (proportions of students who met reading proficiency benchmark, did not self-report as being White or non-Hispanic, received special education services or FRPL, dropped out in the academic year).

Data Analytic Plan

First, Chi-square tests and independent-sample *t*-tests were performed to assess the baseline equivalences in participants' demographics and the pretest scores of four SEB outcomes between study conditions (Table 1). After the assurance of baseline equivalences, we fitted a series of 3-level ML-ANCOVAs to estimate the teacher-level effect of study condition (*PurposeFull People* vs. business-as-usual) on each of the three student-level continuous outcome variables (character building, behavior expectation, or AET; Wan, 2021). For behavioral discipline (a binary student-level outcome variable), we fitted a series of 3-level generalized linear mixed models (GLMM; Li & Redden, 2015) to estimate the cluster-specific effect of study condition. The multilevel models (MLMs, i.e., ML-ANCOVA and GLMM) accommodated (a) the 3-level nested data structure

(student nested in teachers and then in schools), (b) class-level random allocation and treatment delivery, (c) cross-level interactions between class-level treatment and student-level baseline status on outcomes. To obtain conservative estimates, we used listwise deletion to handle missingness (< 0.05% and only in demographic variables) in all MLMs (Baraldi & Enders, 2010).

Before model building, we ran descriptive statistics on all variables planned to enter the MLMs, which generally supported the sample adequacy for MLMs [e.g., normal distribution for continuous outcomes, Bernoulli distribution for binary outcome, and significant intra-class correlations of outcomes at both class and school levels (ICCs); Tables 2, 3, 4, 5]. For RQ1, we calculated the correlation matrices of variables at each of the three levels (student, teacher, and school) to explore the directions and strengths of associations among key variables (Tables 2, 3, 4). For RQs 2 and 3, we followed a stepwise modeling approach to configure the ML-ANCOVAs and GLMMs in the same way (Heck & Thomas, 2020; Supplementary Material 2 detailed MLM equations). For RQ2, we built a main effect model for each of the four student SEB outcomes (Table 6). In the level-1 (student-level) equation, the posttest of an outcome served as the level-1 dependent variable controlling for students' baseline status on the same outcome and student demographics. In the level-2 (teacher-level) equation, the level-1 intercept was set as random to vary across teachers/classes, while being predicted by study condition controlling for

teacher demographics. In the level-3 (school-level) equation, the level-2 intercept was set as random to vary across schools, while being predicted by school demographics. For RQ2, the level-1 slope of student baseline status on an outcome was set to be predicted by the level-2 study condition to probe their cross-level interaction (Table 7; Gardner, 2023).

Power and sensitivity analyses were performed with the *Optimal Design* version 3.01. To obtain the most conservative estimates, we ran the power analysis for the multilevel interaction models (RQ3) that have the most complex configuration. Given our sample sizes at three levels ($n_{\text{school}} = 10$, $n_{\text{teacher}} = 44$, $n_{\text{student}} = 354$), an Alpha level of .05, an ICC of .10 or .30, 70% and 50% of variance explained by level-2 and level-3 covariates, respectively, we would have sufficient power ($\geq 80\%$) to find a minimum detectable effect size (Delta) as small as 0.32 or 0.34, respectively (Figure 2). We reported two types of effect sizes for adequate internal and external validities of our findings (Tymms, 2004). To compare across variables and models within this study, we estimated standardized fixed effect sizes (partial Cohen's d ; Brysbaert et al., 2018). To compare with other studies, we estimated the standardized mean differences effect size (SMDES; Morris, 2008), which is interpreted as the units of standard deviations of the pre-posttest mean difference between the treatment and the control conditions. The number of hypothesis tests performed in this study may inflate the Type I error rate of p -values (i.e., elevated false discovery rate [FDR]). Hence, we estimated the FDR-corrected p -values (i.e., q -values with

a preset level of significance of $q = .05$) for all coefficients (Wason & Robertson, 2021). Analyses were performed with HLM version 6.08 and SPSS version 26.

Results

RQ 1: Sample Adequacy and Correlation Among Key Variables

The Chi-square tests and independent-sample t-tests revealed that the class-level random assignment led to probabilistically equivalent groups in terms of baseline outcomes, teacher and student demographics (Table 1). Of note, no baseline equivalence test was needed for school-level variables because randomization happened within schools for MS-CRT. The ICCs at school and class levels were significant for all four student SEB outcomes, which indicated that students from the same class/school shared significant similarity in their SEB outcomes at posttest. These results supported the need to use MLMs to analyze the data in this study. Moreover, the correlation matrix at student level showed a strong association among study conditions and the posttest scores of four student outcomes, while the matrices at class and school levels showed moderate associations among demographics and class/school-level aggregates of student outcomes (Tables 2, 3, 4).

RQ 2: Main Effect Models

The results indicated consistent and positive effects of *PurposeFull People* on all four student SEB outcomes, after adjusting for student baseline status on the same outcome and demographics at all three levels

(Table 6). Specifically, posttest outcomes of students in the treatment condition were significantly larger than those of students in the control group regarding character building ($b = 0.5$, partial Cohen's $d = 3.59$, $q = .01$; SMDES = 0.93), AET ($b = 5.90$, partial Cohen's $d = 3.42$, $q = .01$; SMDES = 0.46), behavior expectation ($b = 0.25$, partial Cohen's $d = 1.93$, $q = .01$; SMDES = 0.41) and discipline (odds ratio = 0.14, $q = .01$). The significant coefficients of students' baseline status on all outcomes suggested that a student with higher scores at baseline tended to end up with higher scores at posttest.

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

Multilevel moderation analyses revealed significant cross-level interactions between class-level treatment and student-level baseline status of character building and AET (Table 7; Figures 3 and 4, respectively), but not for behavior expectation or discipline. To facilitate interpretation of the treatment by baseline status moderation effect, we plotted the class-specific regression lines between students' baseline and posttest scores for character building and AET, which were grouped by two smoothed regression lines comparing the treatment versus control groups (Loader, 2012). Visual inspection of the plots indicated that the class-specific associations (i.e., dotted/dashed lines) between students' baseline and posttest scores for character building and AET considerably separated and clustered based on study conditions. For students showing higher levels of

baseline status in character building and AET, the posttest scores were similar between students in the treatment (dashed lines) and control groups (dotted lines). This is evidenced by the converging trend of all lines in the high baseline region (upper right corner of the plots). Conversely, for students showing low levels of baseline status in character building and AET, the posttest scores of students from the treatment group were significantly higher than students from control who had the same baseline scores. This is evidenced in the widening gap between lines in the low baseline region (bottom left corner of the plots). These interaction effects suggest that delivery of *PurposeFull People* led to a significant buffering effect on character building and AET for students with low baseline status. The buffering effect of *PurposeFull People* may remedy the common phenomenon of "the poor get poorer, rich get richer" in the education sector (Gardner, 2023).

Discussion

This experimental study evaluated the effects of *CharacterStrong's PurposeFull People Elementary SEL and character education* program on student outcomes. This study was conducted in schools in different geographic areas serving students from diverse backgrounds. Findings from this study revealed positive effects of *PurposeFull People* on student outcomes measured across different methods and informants. Specifically, positive effects were observed for teacher reports of student behaviors and academic engaged time, behavior discipline according to school

administrative data, and student self-reported understanding and motivation related to character strengths. This study's findings raise important implications for both science and practice, and come with specific limitations that are important to acknowledge.

Main Effects of *PurposeFull People*

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 who they want to be in the context of character strengths as well as receive support to put specific character strengths into action through their choices and behavior (Smith, 2013). The SEL part supports students to acquire and use certain social and emotional skills that enable students to better manage thoughts, feelings, and behaviors during social and academic situations. These social and emotional skills enable students to better regulate themselves in certain situations where it may be more challenging to put their character into action (e.g., acting kind towards others). Students' application of character strengths 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 the character education or SEL aspects of *PurposeFull People* was emphasized. 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 approaches are integrated together.

The findings observed in this study are likely due in part to the implementation support provided. The implementation support provided by *CharacterStrong* were intentionally designed to operate on mechanisms of behavior change (e.g., motivation, social norms, self-efficacy) These implementations 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 use of *PurposeFull People* across all teachers i the treatment condition, which is consistent with research on implementation strategies that drive successful implementation (Cook et al., 2019).

There are numerous SEL and character education programs designed for the elementary level. *PurposeFull People* is unique in that it integrates SEL and character education as well as resources on specific generalization practices that support students to apply what they learn beyond the lessons themselves. More research is needed to replicate and extend the findings from this study about *PurposeFull People* and whether it offers relative

advantage over other programs. Generally, there is a need in the prevention science 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). Such research could also attend to cost-benefit and effectiveness analyses to determine the return on investment.

Moderation Effects Based on Baseline Status

The moderation effects between treatment group and students' baseline status is one of the most noteworthy findings to discuss. Findings indicated that *PurposeFull People* demonstrated the strongest effects for students who had the greatest need and room for improvement at baseline. This is particularly important when considering the characteristics of students with the greatest needs at baseline. The characteristics of students (e.g., students living in poverty) who had the greatest need and room for improvement at baseline were those that have been associated with increased exposure to adverse experiences, such as community-based adverse childhood experiences (e.g., crime) and discrimination and racism (Baugh et al., 2019; Berger, 2019). Our findings are consistent with findings from other studies revealing similar interaction effects between treatment condition and baseline status (Low et al., 2015).

Moreover, the significant moderation effect is consistent with recommendations by Greenberg and Abenavoli's (2017) to consider students' baseline status when evaluating the effects of universal prevention

programs. The findings help develop a more precise understanding regarding *with whom* universal prevention programs work the best (Thayer et al., 2019). Findings that highlight with whom programs like *PurposeFull People* work best is important when considering school-based implementation efforts focused on closing opportunity and outcome gaps for students. It is equally important to understand *under what conditions/contexts* and *how* specific programs like *PurposeFull People* work. While descriptive statistics suggest that *PurposeFull People* worked similarly in different geographic contexts, it is an important direction for future research on *PurposeFull People* to better explore under what conditions/contexts and how it works to effect change on student outcomes.

Implications for Research and Practice

This study has several 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. Hence, implementation supports

were provided to guide teachers to make proper distinctions and connections between the components of the two different approaches.

Study Limitations

Several limitations in this study are important to acknowledge when interpreting findings and pinpoint directions for future research. First, this study was conducted with 10 urban and rural public elementary schools in the Midwest US, which may impact the generalizability of the study findings to other areas and regions in the US (e.g., private school in the East Region of the US). Future research should replicate and extend our study with more schools and grade-levels from different geographic regions while recruiting students from diverse backgrounds. Relatedly, although student or teacher race was not significant predictors of student outcomes in this study, future research with a larger and more racially-ethnically diverse school sample may yield different results (e.g., significant interactions between student race and treatment). It can inform future follow-up studies about equity-focused designs and culturally responsive refinement of the content and process of *PurposeFull People* (Marhayani, 2016). For instance, an enhanced "character trait" selection approach grounded in values generated from vastly different cultural groups in the local schools may turn out as an effective and culturally responsive approach.

Second, this study used a multi-method, multi-informant approach to outcome evaluation, but the primary outcomes (i.e., student behavior expectation and characters) were assessed via "ad hoc" scales. These scales

were designed to match and assess the core components and targets of change in *PurposeFull People*, which was appropriate for the preliminary nature of this study and RQs (Lee et al., 2020). But this approach may impact the interpretation and generalizability of our findings. For instance, the ad hoc measures did not undergo comprehensive measurement validation (although we evaluated and found their reliability to be adequate in the current sample). Also, these brief scales did not cover students' use of other social-emotional skills or indicators of wellbeing (e.g., educational purpose, sense of belonging; Nisar et al., 2022), which make suitable targets to explore in follow-up studies on the fine-grained effects of *PurposeFull People* with established comprehensive measures of student SEB outcomes.

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).

Conclusion

This study provides evidentiary support for the positive effects of *PurposeFull People* on elementary students' character strengths, AET, behavior expectations and discipline. These findings inform future research to examine the effects of integrated approaches like *PurposeFull People* that combine SEL and character education versus standalone approaches. We hope the findings spur other researchers to replicate the findings from this study, including the use of specific implementation supports that aim to drive high-fidelity implementation. We encourage researchers to continue to conduct more nuanced research to identify with whom and under what conditions universal prevention programs work. We know there is a need for researchers to conduct comparative studies that examine the relative effects of different SEL and character education programs. Last, we encourage policymakers to create and support policies that call for the integration of SEL and character education programs with academics to ensure students access the experience and support they need to be successful inside and outside of school.

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>
- Agboola, A., & Tsai, K. C. (2012). Bring character education into classroom. *European journal of educational research*, 1(2), 163-170.
- 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.
- Arslan, G., Allen, K. A., & Tanhan, A. (2021). School bullying, mental health, and wellbeing in adolescents: Mediating impact of positive psychological orientations. *Child Indicators Research*, 14, 1007-1026.
- 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.
- Berkowitz, M. W., Bier, M. C., & McCauley, B. (2017). Toward a science of character education. *Journal of Character Education*, 13(1), 33-51.
- Berkeley, S., Scanlon, D., Bailey, T. R., Sutton, J. C., & Sacco, D. M. (2020). A snapshot of RTI implementation a decade later: New picture, same

- story. *Journal of Learning Disabilities*, 53(5), 332-342.
- 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.
- Campbell, A., & Anderson, C. M. (2011). Check-in/check-out: A systematic evaluation and component analysis. *Journal of Applied Behavior Analysis*, 44(2), 315-326.
- 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.
- 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,*
- Duong, M. T., Bruns, E. J., Lee, K., Cox, S., Coifman, J., Mayworm, A., & Lyon, A. R. (2021). Rates of mental health service utilization by children and adolescents in schools and other common service settings: A systematic review and meta-analysis. *Administration and Policy in Mental Health and Mental Health Services Research, 48,* 420-439.
- 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.
- 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.
- 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
- 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*,
- 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.
- Hayashi, A., Liew, J., Aguilar, S. D., Nyanamba, J. M., & Zhao, Y. (2022). Embodied and social-emotional learning (SEL) in early childhood: Situating culturally relevant SEL in Asian, African, and North American contexts. *Early Education and Development, 33*(5),
- 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.
- 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-
- Jolivet, 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.
- Lee, M. K., Basford, J. R., Heinemann, A. W., & Cheville, A. (2020). Assessing whether ad hoc clinician-generated patient questionnaires provide psychometrically valid information. *Health and Quality of Life Outcomes*, 18, 1-17.
- 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.
- Lickona, T., Schaps, E., & Lewis, C. (2007). CEP's Eleven Principles of

- Effective Character Education. Character Education Partnership.
- 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
- Marhayani, D. (2016). Development of character education based on local wisdom in indogenous people Tengahan Sedangagung. *Journal of Education, Teaching and Learning, 1*(2), 66.
- Mascia, M. L., Agus, M., & Penna, M. P. (2020). Emotional intelligence, self-regulation, smartphone addiction: which relationship with student well-being and quality of life?. *Frontiers in Psychology, 11*, 375.
- 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.
- Predy, L., McIntosh, K., & Frank, J. L. (2014). Utility of number and type of

- office discipline referrals in predicting chronic problem behavior in middle schools. *School Psychology Review*, 43(4), 472-489.
- 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.
- Polanin, J. R., Espelage, D. L., Grotzinger, J. K., Spinney, E., Ingram, K. M., Valido, A., ... & Robinson, L. (2021). A meta-analysis of longitudinal partial correlations between school violence and mental health, school performance, and criminal or delinquent acts. *Psychological Bulletin*, 147(2), 115.
- Reinke, W. M., Herman, K. C., Darney, D., Pitchford, J., Becker, K., Domitrovich, C., & Jalongo, 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. *The 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.
- Wigelsworth, M., Verity, L., Mason, C., Qualter, P., & Humphrey, N. (2022). Social and emotional learning in primary schools: A review of the current state of evidence. *British Journal of Educational Psychology*, 92(3), 898-924.
- 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.
- Zhang, Y., Cook, C. R., & Smith, B. (2023). PurposeFull People SEL and Character Education Program: A Cluster Randomized Trial in Schools

Implementing Tier 1 PBIS with Fidelity. *School Mental Health*, 1-18.

Figure 1

The CONSORT multi-site cluster trial extension diagram delineates the design, timeline of enrollment, randomization, pretest, treatment implementation, and posttest of four outcomes.

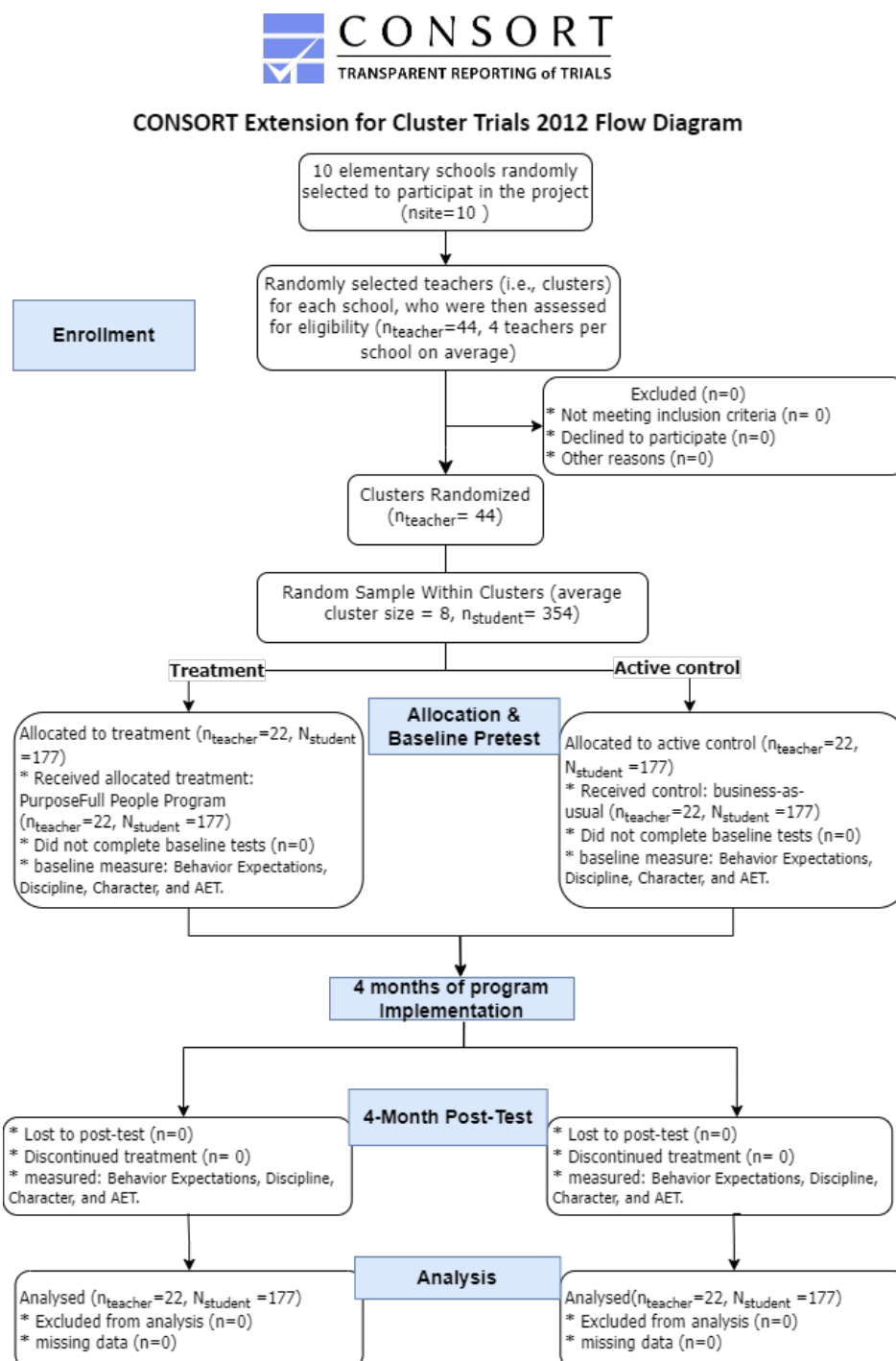


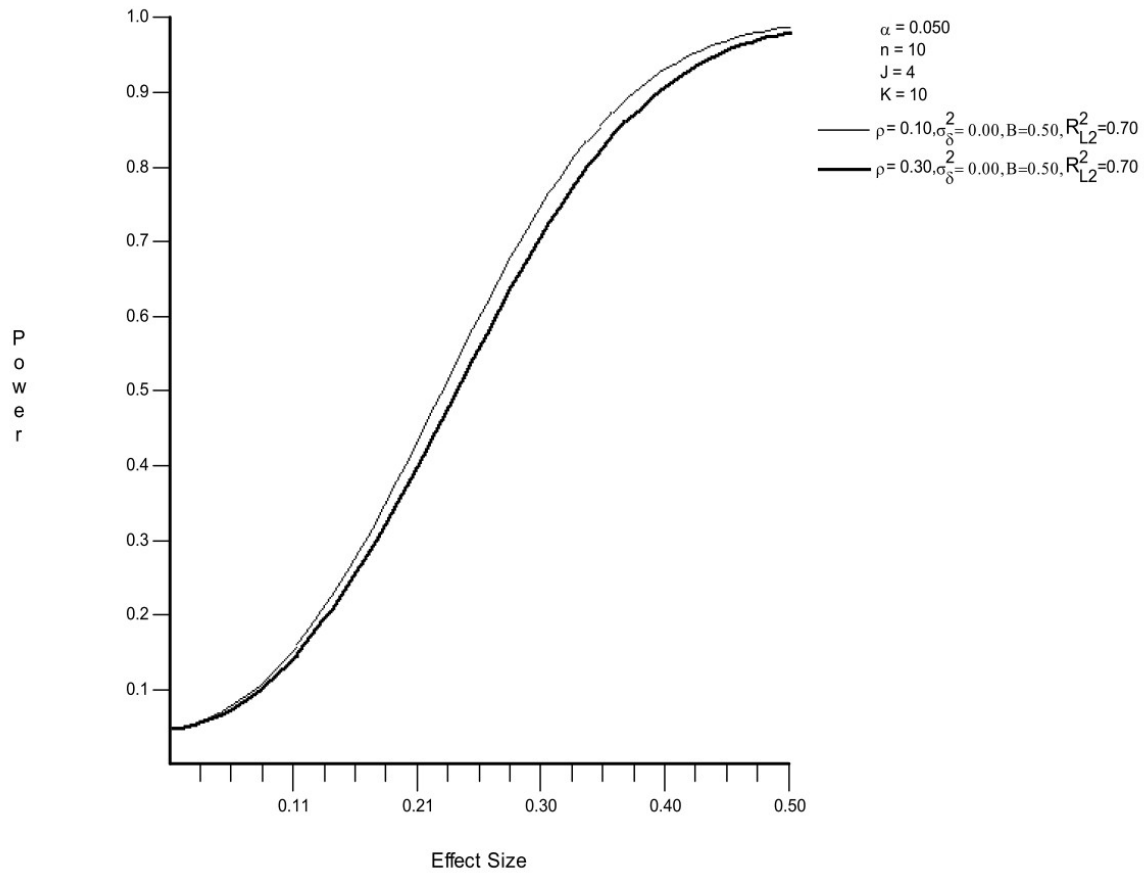
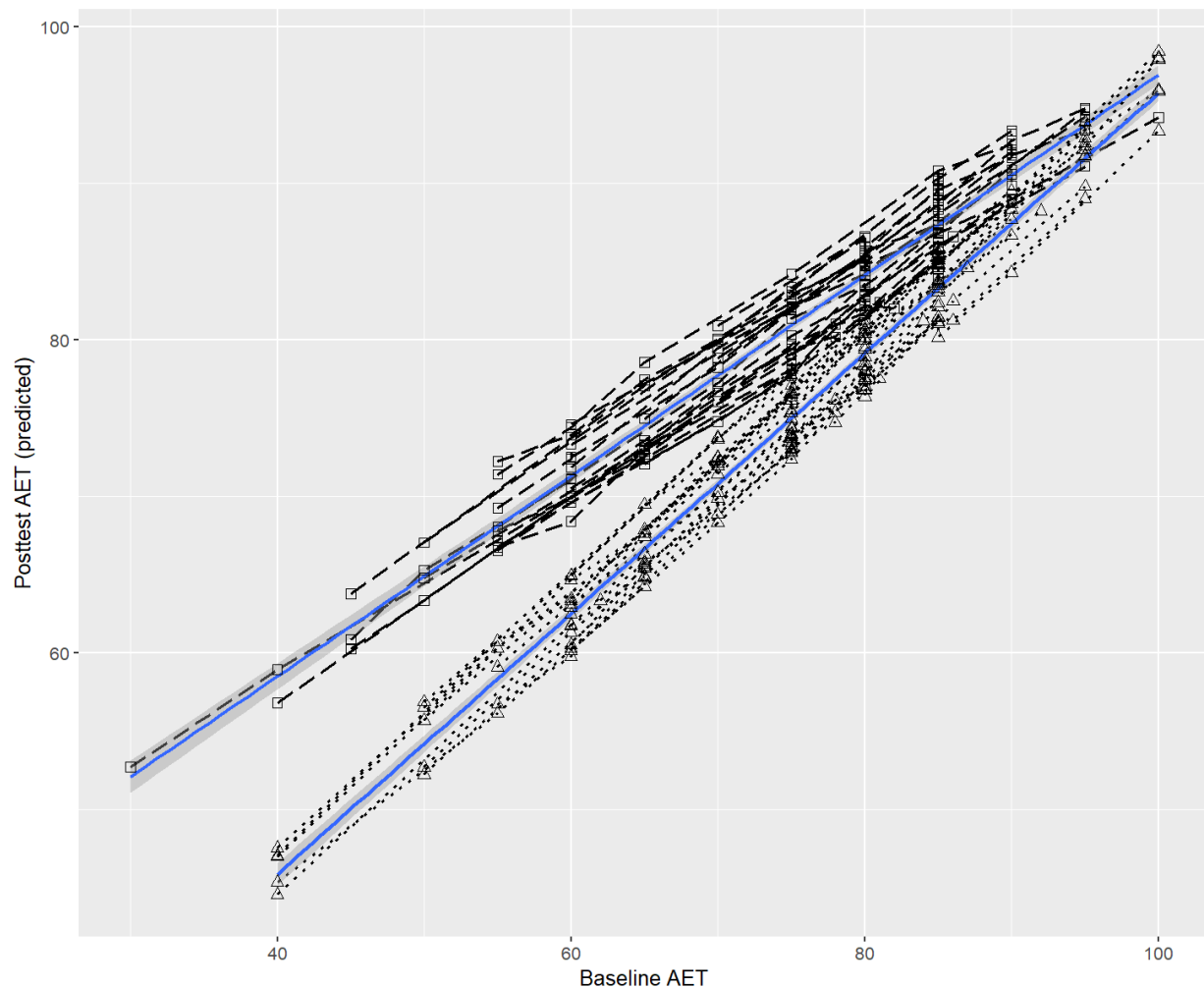
Figure 2*Power Analysis*

Figure 3

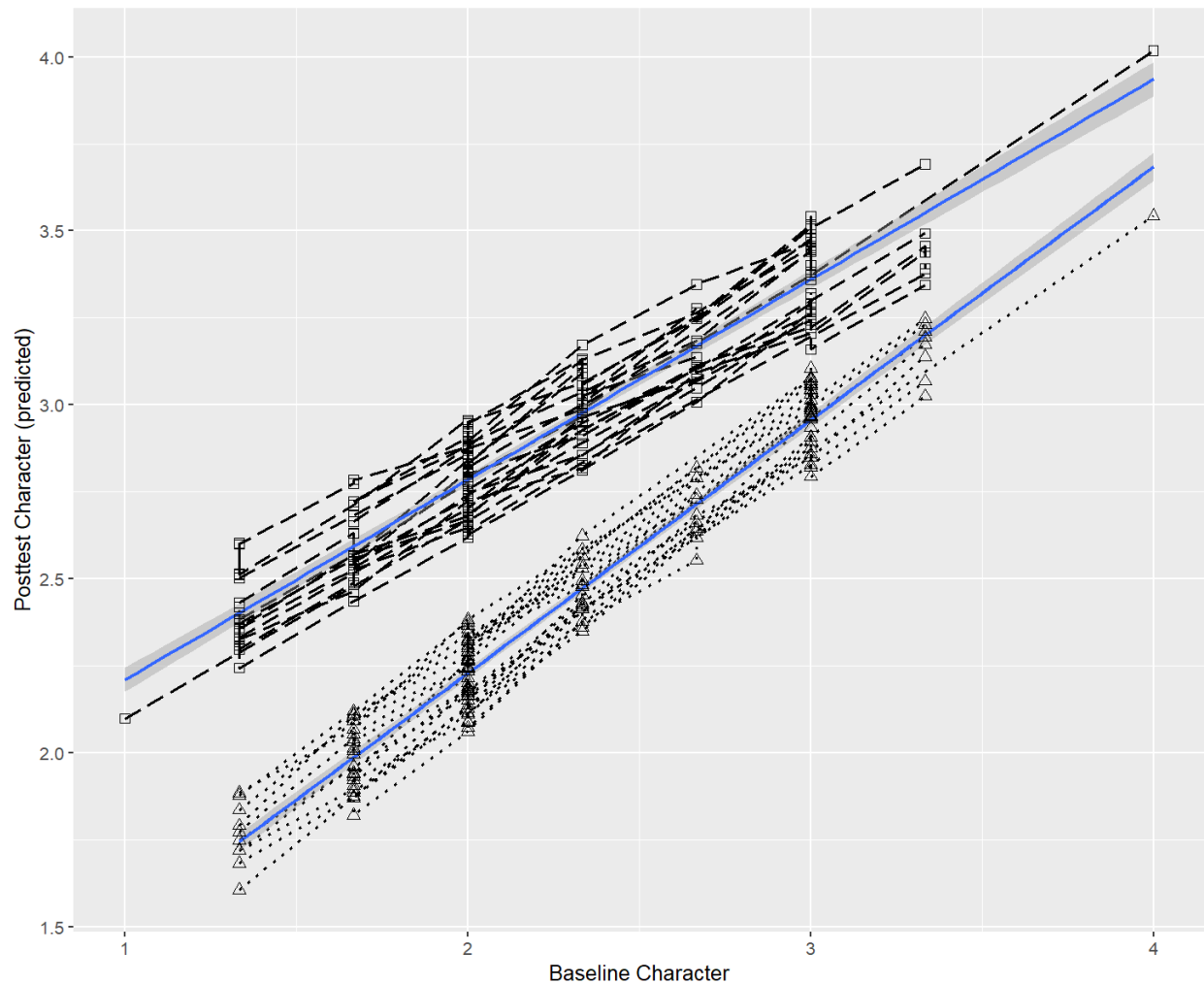
Significant interaction effect between study conditions and baseline AET on posttest AET.



Note. 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 4

The significant cross-level interaction effect between study conditions and baseline Character Building on the posttest Character Building.



Note. Black lines represent class-specific associations between baseline and posttest Character Building, 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 and Baseline Equivalences (Nschool=10, Nteacher=44, Nstudent=354)

Level 3: School				
Continuous Variables	<i>M</i>	<i>SD</i>	Minimum	Maximum
% Reading Proficiency	56.00%	13.23%	29.00%	73.00%
%Non-White	47.60%	24.36%	10.00%	75.00%
%FRPL	61.20%	26.67%	18.00%	91.00%
%Special Education	15.20%	3.97%	10.00%	24.00%
%Drop-out	3.40%	2.17%	0.00%	7.00%
Level 2: Teacher/Classroom				
Categorical Variables	Categori es	n	%	Baseline Equivalence (χ^2)
Study condition	Control	22	50.00%	-
	Treatme nt	22	50.00%	
Teacher Gender	Male	6	13.64%	$\chi^2 (1, 44) = .77, p = .38$
	Female	38	86.36%	
Teacher Race	Asian	1	2.27%	$\chi^2 (4, 44) = .46, p = .50$
	Black	3	6.82%	
	Latinx	5	11.37%	
	Mixed	3	6.82%	
	White	32	72.73%	
Teaching experience	<= 6 years	13	29.55%	$\chi^2 (2, 44) = 3.79, p = .15$
	7 - 12 years	14	31.82%	
	>= 13 years	17	38.64%	
Grade level	2	10	22.73%	$\chi^2 (3, 44) = 2.37, p = .50$
	3	12	27.27%	
	4	11	25.00%	
	5	11	25.00%	
Level 1: Students				
Study condition	Control	177	50.00%	-
	Treatme nt	177	50.00%	
Student Gender	Male	180	50.85%	$\chi^2 (1, 354) = .05, p = .83$
	Female	174	49.15%	
Student Race	White	211	59.60%	$\chi^2 (1, 354) = .03, p = .87$
	Non- White	143	40.40%	
FRPL	No	182	51.56%	$\chi^2 (1, 354) = .05, p = .83$
	Yes	171	48.44%	
				$\chi^2 (4, 354) = .01, p = .91$

Baseline Discipline	Never has been disciplin ed	259 95	73.16% 26.84%	$\chi^2 (1, 354) = .03, p = .87$
Baseline Outcomes	$M (SD)$ in treatment group	$M (SD)$ in control group	Baseline Equivalence (t-tests)	
Behavioral Expectation	2.67(0.64)	2.72(0.66)	$t (352) = 0.71, p = 0.48$	
Character Building	2.21(0.6)	2.31(0.58)	$t (352) = 1.59, p = 0.11$	
AET	75.23(12.93)	75.66(13.28)	$t (352) = 0.31, p = 0.76$	

Note. M=mean SD= standard deviation, FRPL= free/reduced price lunch, AET=academic engaged time. School demographics were not tested for baseline equivalence because in the multi-site cluster randomized trial design, study groups were randomized within each school.

Table 2*Correlation Among Key Variables at Student Level (n = 354)*

	1	2	3	4	5	6	7	8	9	10	11	12
1. Study condition	354	354	354	353	354	348	354	353	354	354	354	354
2. gender	.01 1	354	354	353	354	348	354	353	354	354	354	354
3. race	-.006	.08 4	354	353	354	348	354	353	354	354	354	354
4. FRPL	-.008	.06 0	.419 **	353	353	348	353	352	353	353	353	353
5. baseline behaviors	-.038	.08 5	-.23 9**	-.38 6**	354	348	354	353	354	354	354	354
6. Posttest behaviors	.175 **	.03 0	-.16 9**	-.31 2**	.776 **	348	348	347	348	348	348	348
7. baseline characters	-.084	.08 2	-.13 1*	-.12 2*	.565 **	.438 **	354	353	354	354	354	354
8. Posttest characters	.395 **	-.00 6	-.12 1*	-.13 2*	.452 **	.568 **	.636 **	353	353	353	353	353
9. baseline discipline	.00 6	-.09 8	.203 **	.262 **	-.66 0**	-.42 1**	-.41 6**	-.26 6**	354	354	354	354

10. Posttest discipline	-.19 4**	-.05 9	.02 0	.09 4	-.36 2**	-.38 5**	-.21 5**	-.26 8**	.481 **	354	354	3 5 4
11. baseline AET	-.017	.12 0*	-.20 5**	-.29 2**	.769 **	.549 **	.534 **	.415 **	-.76 3**	-.34 4**	354	3 5 4
12. Posttest AET	.243 **	.14 5**	-.19 8**	-.25 7**	.695 **	.653 **	.483 **	.545 **	-.61 2**	-.43 1**	.84 2**	3 5 4

Note. * $p < .05$; ** $p < .01$. The number above the diagonal line indicates the sample size used for calculation. All variables were raw, i.e., not centered. Point biserial correlation was used between binary and continuous variables. Pearson Product-Moment correlation was used between continuous variables.

Table 3*Correlation Among Key Variables at Teacher Level (n = 44)*

	1	2	3	4	5	6	7	8
1. Study condition	44	44	44	44	44	44	44	44
2. Teacher Experience	-.027	44	44	44	44	44	44	44
3. Grade level	-.104	-.062	44	44	44	44	44	44
4. Teacher race	-.102	-.234	-.013	44	44	44	44	44
5. Teacher gender	-.132	-.001	-.052	.095	44	44	44	44
6. Class M of baseline behavioral expectation	-.072	-.025	-.565**	.053	.106	44	44	44
7. Class M of baseline Character	-.160	-.155	-.221	.093	-.005	.440**	44	44
8. Class M of baseline AET	-.051	.133	-.501**	.151	.062	.824**	.401**	44

Note. * $p < .05$; ** $p < .01$. M = mean. The number above the diagonal line indicates the sample size used for calculation. All variables were raw, i.e., not centered. Point biserial correlation was used between binary and continuous variables. Pearson Product-Moment correlation was used between continuous variables.

Correlation Among Key Variables at School Level ($n = 10$)

	1	2	3	4	5	6	7	8	9	10	11	12	13
1. %Reading proficiency	.10	.10	.10	.10	.10	.10	.10	.10	.10	.10	.10	.10	.10
2. %Non_white	-.071	.10	.10	.10	.10	.10	.10	.10	.10	.10	.10	.10	.10
3. %FRPL	-.248	.952*	.10	.10	.10	.10	.10	.10	.10	.10	.10	.10	.10
4. % Special education	-.167	.815*	.815*	.10	.10	.10	.10	.10	.10	.10	.10	.10	.10
5. %Drop-out	.058	.224	.123	.338	.10	.10	.10	.10	.10	.10	.10	.10	.10
6. School M of baseline behavior	.065	-.943**	-.903**	-.758*	-.226	.10	.10	.10	.10	.10	.10	.10	.10
7. School M of posttest behavior	-.075	-.942**	-.888**	-.708*	-.191	.981**	.10	.10	.10	.10	.10	.10	.10
8. School M of baseline character	-.148	-.412	-.503	-.265	.079	.284	.343	.10	.10	.10	.10	.10	.10
9. School M of posttest character	-.277	-.585	-.653*	-.372	.054	.525	.613	.906**	.10	.10	.10	.10	.10
10. School M of baseline AET	.122	-.892**	-.849**	-.663*	-.081	.948**	.926**	.323	.519	.10	.10	.10	.10
11. School M of posttest AET	.011	-.917**	-.884**	-.700*	-.007	.964**	.970**	.338	.588	.968**	.10	.10	.10
12. Total count of discipline at baseline	-.744*	.576	.656*	.401	-.037	-.639*	-.528	.023	-.046	-.665*	-.589	.10	.10
13. Total count of discipline at	-.831**	.403	.584	.424	-.04	-.36	-.25	-.07	-.07	-.31	-.29	.806**	.10

posttest	9	7	7	8	1	7	5
----------	---	---	---	---	---	---	---

Note. * $p < .05$; ** $p < .01$. M = mean. The number above the diagonal line indicates the sample size used for calculation. All variables were raw, i.e., not centered. Pearson Product-Moment correlation was used between continuous variables.

Table 5

Intra-Class Correlations (ICCs) and Variance Components of the Random Effects from All Unconditional Models

DV	Random Effects	Variance		Chi-		ICC
		Component	<i>df</i>	Square	<i>p</i> -value	
Behavior	Level 3 ()	0.090	9	57.892	< .001	0.241
Expectati	Level 2 ()	0.049	34	90.013	< .001	0.373
on	Level 1 ()	0.235	-	-	-	-
Characte	Level 3 ()	0.009	9	12.396	.191	0.027
r	Level 2 ()	0.096	34	156.422	< .001	0.323
Building	Level 1 ()	0.220	-	-	-	-
AET	Level 3 ()	12.888	9	33.363	< .001	0.099
	Level 2 ()	9.928	34	58.640	< .01	0.174
	Level 1 ()	108.017	-	-	-	-

Note. Level 1: $n_{student} = 354$; Level 2: $n_{teacher} = 44$, Level 3: $n_{school} = 10$.

Table 6*Fixed Effect Estimates of ML-ANCOVAs and GLMM Controlling for Baseline Outcome and Demographics*

Fixed effects	Behavioral Expectation			Character Building			AET			Discipline (Yes vs. No)		
	<i>b (SE)</i>	Partial Cohen' s d	<i>q</i>	<i>b (SE)</i>	Partial Cohen' s d	<i>q</i>	<i>b (SE)</i>	Partial Cohen' s d	<i>q</i>	<i>OR</i>	95% CI	<i>q</i>
School: Intercept	1.4 (0.21)	6.65	.0 0	1.46 (0.23)	6.31	.0 0	26.05 (3.02)	8.63	.0 0	0.01	(0, 4.78)	.01
School: %read	-0.33 (0.25)	-1.34	.0 2	-0.47 (0.28)	-1.69	.0 2	-4.48 (3.32)	-1.35	.0 2	0.60	(0, 2.85*10 ³)	.05
School: %non-white	-0.22 (0.39)	-0.56	.0 4	0.65 (0.43)	1.50	.0 2	-1.47 (5.2)	-0.28	.0 4	0.07	(0, 8.17*10 ³)	.03
School: %FRPL	-0.38 (0.36)	-1.05	.0 3	-0.86 (0.4)	-2.14	.0 2	-5.82 (4.8)	-1.21	.0 3	5.67	(0, 2.95*10 ⁵)	.03
School: %SpEd	1.09 (1.08)	1.01	.0 3	0.99 (1.19)	0.83	.0 3	-7.16 (14.41)	-0.50	.0 3	4348.5 8	(0, 1.42*10 ¹⁸)	.02
School: %Drop-out	-0.37 (1.25)	-0.29	.0 5	-0.78 (1.37)	-0.57	.0 4	17.38 (16.76)	1.04	.0 3	1.13	(0, 1*10 ²¹)	.05
Teacher: Treatment	0.25 (0.04)	1.93	.0 1	0.51 (0.05)	3.59	.0 1	5.9 (0.56)	3.42	.0 1	0.14	(0.047, 0.395)	.01
Teacher: Experience	0 (0)	0.33	.0 2	0 (0)	0.40	.0 3	0.12 (0.04)	0.92	.0 1	1.03	(0.957, 1.105)	.02
Teacher: Grade	-0.03 (0.04)	-0.24	.0 4	-0.03 (0.04)	-0.23	.0 4	0.05 (0.51)	0.03	.0 5	1.10	(0.46, 2.614)	.04
Teacher: Race	0.02 (0.05)	0.11	.0 5	0.05 (0.05)	0.30	.0 3	0.23 (0.64)	0.12	.0 4	0.89	(0.299, 2.679)	.04
Teacher: Sex	0.04 (0.06)	0.20	.0 4	0.02 (0.07)	0.08	.0 5	-0.04 (0.84)	-0.01	.0 5	0.61	(0.155, 2.373)	.02
Student: Sex	-0.04 (0.04)	-0.11	.0 2	-0.07 (0.04)	-0.21	.0 1	1.08 (0.54)	0.22	.0 2	0.84	(0.345, 2.067)	.03
Student: Race	0.07 (0.04)	0.18	.0 1	0 (0.04)	-0.01	.0 5	0.18 (0.62)	0.03	.0 4	0.56	(0.207, 1.522)	.01
Student: FRPL	0.04 (0.04)	0.10	.0 3	-0.02 (0.04)	-0.05	.0 4	1.18 (0.66)	0.20	.0 2	0.90	(0.317, 2.57)	.04
Student: Baseline	0.66 (0.03)	2.25	.0 1	0.61 (0.03)	2.08	.0 1	0.71 (0.02)	3.51	.0 1	51.31	(15.559, 1.69*10 ²)	.00

Note. Discipline is binary (1 = disciplined in 4 months, 0 = never disciplined in 4 months). Treatment = *PurposeFull People* versus business-as-usual; *b* = unstandardized fixed effect coefficient; *SE* = standard error; *q* value = FDR-adjusted *p*-value for a hypothesis test. Bolded & underlined *q* = statistical significance after FDR adjustment. OR = odds ratio, 95%CI = 95% confidence interval of OR.

Table 7

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

	Behavioral Expectation			Character Building			AET			Discipline (Yes vs. No)		
	<i>b</i> (<i>SE</i>)	Partial Cohen's <i>d</i>	<i>q</i>	<i>b</i> (<i>SE</i>)	Partial Cohen's <i>d</i>	<i>q</i>	<i>b</i> (<i>SE</i>)	Partial Cohen's <i>d</i>	<i>q</i>	<i>OR</i>	95% CI	<i>q</i>
Fixed effects												
School: Intercept	1.27 (0.24)	5.27	.0 0	1.26 (0.25)	5.14	.0 1	19.04 (3.3)	5.77	.0 0	0.01	(0, 4.71)	.0 1
School: %read	-0.31 (0.26)	-1.17	.0 1	-0.43 (0.28)	-1.55	.0 2	-3.47 (3.23)	-1.07	.0 3	0.55	(0, 2.7*10 ³)	.0 4
School: %non-white	-0.23 (0.41)	-0.56	.0 1	0.59 (0.44)	1.34	.0 3	-1.06 (5.05)	-0.21	.0 5	0.07	(0, 9.34*10 ³)	.0 3
School: %FRPL	-0.37 (0.38)	-0.98	.0 1	-0.81 (0.4)	-2.01	.0 2	-5.56 (4.66)	-1.19	.0 3	5.24	(0, 3.11*10 ⁵)	.0 3
School: %SpEd	1.16 (1.14)	1.01	.0 2	1.01 (1.21)	0.84	.0 3	-5.5 (14)	-0.39	.0 4	3881	(0, 1.89*10 ¹⁸)	.0 3
School: %Drop-out	-0.4 (1.32)	-0.30	.0 2	-0.66 (1.39)	-0.47	.0 4	16.66 (16.28)	1.02	.0 3	1.35	(0, 1.77*10 ²¹)	.0 5
Teacher: Treatment	0.46 (0.16)	0.91	.0 2	0.89 (0.15)	1.97	.0 0	20.09 (3.1)	2.10	.0 1	0.31	(0.03, 3.34)	.0 1
Teacher: Experience	0 (0)	0.31	.0 3	0 (0)	0.37	.0 3	0.11 (0.04)	0.87	.0 2	1.03	(0.96, 1.11)	.0 2
Teacher: Grade	-0.03 (0.04)	-0.24	.0 3	-0.03 (0.04)	-0.25	.0 4	-0.21 (0.5)	-0.13	.0 3	1.11	(0.47, 2.67)	.0 4
Teacher: Race	0.02 (0.05)	0.11	.0 3	0.04 (0.05)	0.25	.0 3	0.22 (0.62)	0.11	.0 4	0.87	(0.29, 2.65)	.0 4
Teacher: Sex	0.04 (0.07)	0.21	.0 3	0.01 (0.07)	0.06	.0 5	0.23 (0.82)	0.09	.0 5	0.60	(0.15, 2.36)	.0 2
Student: Sex	-0.04 (0.04)	-0.11	.0 4	-0.07 (0.04)	-0.22	.0 2	1.03 (0.53)	0.21	.0 2	0.85	(0.34, 2.09)	.0 3
Student: Race	0.06 (0.04)	0.17	.0 4	-0.01 (0.04)	-0.03	.0 5	0.23 (0.6)	0.04	.0 4	0.56	(0.21, 1.54)	.0 1
Student: FRPL	0.03 (0.04)	0.08	.0 4	-0.02 (0.04)	-0.05	.0 4	0.96 (0.64)	0.16	.0 2	0.92	(0.32, 2.66)	.0 5

Student:	0.69	1.82	<u>.0</u>	0.7 (0.04)	1.70	<u>.0</u>	0.8 (0.03)	3.03	<u>.0</u>	62.41	(16.11, 241.79)	<u>.0</u>
Baseline	(0.04)		<u>5</u>			<u>1</u>			<u>1</u>			<u>0</u>
cross-level:	-0.08	-0.14	.0	-0.17	-0.30	<u>.0</u>	-0.19 (0.04)	-0.51	<u>.0</u>	0.39	(0.03, 4.82)	.0
Baseline X	(0.06)		5	(0.06)		<u>1</u>			<u>1</u>			2
Treatment												

Note. Discipline is binary (1= disciplined in 4 months, 0 = never disciplined in 4 months). Treatment = *PurposeFull People* versus business-as-usual; b = unstandardized fixed effect coefficient; *SE* = standard error; *q* value = FDR-adjusted *p*-value for a hypothesis test. Bolded & underlined *q* = statistical significance after FDR adjustment. OR= odds ratio, 95%CI = 95% confidence interval of OR.