



OPEN ACCESS

EDITED BY

Andres Molano,
University of Los Andes, Colombia

REVIEWED BY

Garry Squires,
The University of Manchester, United Kingdom
Lin Lin,
Shanghai University of Finance and Economics,
China
Juliette Berg,
American Institutes for Research, United States
Rachel Abenavoli,
New York University, United States

*CORRESPONDENCE

Hui Zhao
✉ hzz23@psu.edu

RECEIVED 30 August 2022

ACCEPTED 05 April 2023

PUBLISHED 27 April 2023

CITATION

Zhao H, Lei P-W, Hart SC, DiPerna JC and Li X (2023) When is universal SEL effective under authentic conditions? Using LPA to examine program implementation in elementary classrooms.
Front. Educ. 8:1031516.
doi: 10.3389/educ.2023.1031516

COPYRIGHT

© 2023 Zhao, Lei, Hart, DiPerna and Li. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

When is universal SEL effective under authentic conditions? Using LPA to examine program implementation in elementary classrooms

Hui Zhao^{1*}, Pui-Wa Lei¹, Susan Crandall Hart¹, James Clyde DiPerna¹ and Xinyue Li²

¹Department of Educational Psychology, Counseling, and Special Education, The Pennsylvania State University, University Park, PA, United States, ²Cambium Assessment, Washington, DC, United States

As universal social-emotional learning (SEL) programs have become more common in K-12 schools, implementation practices have been found to affect program quality. However, research examining how multiple facets of program implementation interrelate and impact student outcomes, especially under routine conditions in schools, is still limited. As such, we used latent profile analysis (LPA) to examine implementation of a brief universal SEL program (Social Skills Improvement System SEL Classwide Intervention Program) in primary classrooms. Three latent profiles of implementation were identified based on dosage, adherence, quality of delivery, student engagement and teachers' impression of lessons. Although results suggested that classrooms with moderate- and high-level implementation practices generally showed higher gains in student outcomes than those with low-level implementation, these differences did not reach statistical significance except for academic motivation. Implications for school-based universal SEL program planning, implementation, and evaluation are discussed.

KEYWORDS

implementation fidelity, social-emotional learning (SEL), contextual factors, latent profile analysis, program quality, elementary classrooms, authentic conditions

1. Introduction

School-based universal social-emotional learning (SEL) aims to teach foundational social-emotional skills at the classroom or school level as a public health approach to improving student outcomes through prevention and promotion programming (Greenberg et al., 2017). Although studies demonstrate that SEL programs can increase student prosocial skills and attitudes, lower problem behaviors, and enhance academic performance (Durlak et al., 2011; Taylor et al., 2017), implementation is an important determinant for the success of such programs in practice (Durlak and DuPre, 2008; Low et al., 2016). When programs with prior evidence of efficacy are not implemented as intended in the real world, schools may lose time, money, and ultimately, the desired benefits for students (Sanetti and Collier-Meek, 2019).

Through implementation science, dimensions of program implementation have been identified, which can serve as indicators of how evidence-based interventions translate into real-world delivery systems and capture variation across teachers and contexts (Proctor et al.,

2011). Measuring and understanding these dimensions is a critical part of effectiveness trials in education, which allows decision-makers to evaluate the feasibility of intervention programs and the degree to which implementation practices affect program outcomes (O'Donnell, 2008). To date, there have been very few effectiveness trials of social-behavioral programs in authentic education settings (Chhin et al., 2018); therefore, little is known about how universal SEL programs are typically implemented under these conditions, including what contextual factors are associated with variation in implementation and what combination of approaches yield the most benefit to students in the real world. In addition to aiding educators in planning for resource allocation, training, and support for their local context, such information could also be helpful to intervention developers hoping to increase the feasibility and transportability of programs.

There is growing awareness that implementation is critical to understanding not just *if* universal SEL works for students and schools, but *how* and *under what conditions* it works (Jones et al., 2019). As implementation is widely considered a key influence on outcomes of prevention and intervention programs (Durlak and DuPre, 2008; Domitrovich et al., 2010; Durlak et al., 2011), SEL researchers are similarly moving away from examining *whether* implementation matters to focus on *what aspects* of implementation are most salient (Low et al., 2016). While implementation is multi-dimensional, implementation fidelity, which refers to the degree to which a program is delivered as intended by developers, has been studied most frequently in education science (Proctor et al., 2011; Durlak, 2016). Five aspects of implementation fidelity (i.e., dosage—amount delivered; adherence—components delivered as planned; implementation quality—competence in delivery; participant responsiveness—engagement/enthusiasm; and program differentiation—critical features that distinguish the program) have been identified as critical for achieving implementation that influences student outcomes (Dane and Schneider, 1998; Durlak and DuPre, 2008; Durlak, 2016).

A systematic review of 41 school-based mental health intervention studies found that aspects of implementation fidelity were positively associated with student outcomes 36% of the time; participant responsiveness most commonly (58%) related to intervention effects, with dosage (39%), adherence (28%), and quality (23%) observed less often (Rojas-Andrade and Bahamondes, 2019). These aspects are also thought to interact with each other within the context of systemic barriers and supports to promote benefits for students. For example, Carroll et al. (2007) proposed a conceptual model where interventions improve outcomes via adherence, with intervention complexity, implementation support, quality of delivery, and participant responsiveness moderating this relationship.

Individual efficacy studies of universal SEL programs have sought to identify and untangle the most important aspects of implementation for optimizing program outcomes, but findings have been mixed. For example, in one study of a preschool readiness program, student engagement and implementation quality were related to improvement in student outcomes, but program dosage was less salient (Domitrovich et al., 2010). Low et al. (2014) found that, while adherence did not influence program effects, higher classroom engagement in *Steps to Respect: A Bullying Prevention Program* (Committee for Children, 2005) was associated with more desirable student outcomes (e.g., student attitudes toward bullying, student climate and support).

Reyes et al. (2012) found no main effects for teacher training, dosage, and implementation quality on student outcomes of *RULER Approach* (Brackett et al., 2011), however, they did find an interaction effect. A combination of high attendance during training and high dosage had a positive impact on student social-emotional skills, but only in conjunction with above-average implementation quality; when implementation quality was low, high training attendance and dosage actually predicted negative student outcomes. In Humphrey et al.'s (2018) study of *Promoting Alternate Thinking Strategies* (PATHS; Greenberg et al., 1995), program fidelity and reach did not predict changes in student behaviors, but higher implementation quality and participant responsiveness were associated with lower reports of student externalizing behaviors. Surprisingly, however, higher dosage was associated with lower SEL skills and prosocial behaviors (Humphrey et al., 2018). Overall, these studies suggest a complex relationship between implementation and program outcomes that appears to differ by aspects of implementation, interventions, and outcomes assessed.

To date, most studies of universal SEL programs have examined independent variable-centered relationships between aspects of implementation and program outcomes (Durlak and DuPre, 2008; Low et al., 2016). As Durlak (2016) noted, though, implementation components can interact with one another, and despite evidence suggesting better implementation is associated with more promising program outcomes, there is a lack of systematic investigation of how patterns of implementation factors relate to different program outcomes. While traditional variable-centered analyses (e.g., Analysis of Variance, correlation, and regression) have been widely used in many implementation studies, these methods may be insufficient for examining inter-relationships among multiple implementation measures (Laursen and Hoff, 2006; Hennessey and Humphrey, 2020). Furthermore, variable-centered approaches assume that the population is homogeneous in terms of how predictors impact outcomes (Laursen and Hoff, 2006; Cheng et al., 2021). As such, studies that employ variable-centered analytic approaches may fail to account for the heterogeneous inter-relationships between implementation aspects and program outcomes within a population. Variable-centered approaches also might not be able to detect heterogeneity of small subpopulations when sample size is small (Muthén and Muthén, 2000). In contrast, person-centered analytic approaches, such as latent profile analysis (LPA) and latent class analysis (LCA), can be used to identify unobserved subgroups that share similar characteristics to examine population heterogeneity (Laursen and Hoff, 2006) without making assumptions (e.g., linear and homogenous relationship) that traditional variable-centered approach would (e.g., Magidson and Vermunt, 2005). Person-centered models are particularly well-suited to address questions regarding whether combination of multiple implementation variables have differential effects on program outcomes. Specifically, they are helpful when exploring combined effects of many predictors on outcomes given traditional moderation analyses have limited capability to test all possible interaction terms (Spurk et al., 2020).

Acknowledging the multi-faceted and multiplicative nature of implementation, the likely dynamic interplay between implementation aspects, and the lack of clarity about what matters most for school-based SEL, recent studies have identified person-centered profiles of implementation using multiple indicators and investigated their association with student outcomes. We located two studies that

explored how the dynamic interplay of multiple implementation measures affect outcomes of universal SEL programs. [Low et al. \(2016\)](#) examined how patterns of dosage, adherence, student engagement, and program differentiation of *Second Step* ([Committee for Children, 2012](#)) influenced student outcomes and found three latent classes: high-quality, low-engagement, and low-adherence. The analysis sample consisted of 160 teachers who implemented *Second Step* in kindergarten, first, or second grade. [Low et al. \(2016\)](#) used LPA to identify subgroups of implementation classrooms. All the implementation measures were based on teacher's weekly self-reports. Multilevel prediction models (i.e., students nested within classrooms) were adopted to examine the effect of profile membership on student gains in SEL competencies, functioning, and disruptive behavior while controlling for proactive classroom management and percentage of English Language Learners in the classrooms. Because only the low-engagement group showed associations with poorer student outcomes, Low and colleagues identified student engagement as the most pivotal aspect of implementation, though they acknowledged it was a necessary but not sufficient requirement for program success.

Using hierarchical cluster analyses of observational data, [Hennessey and Humphrey \(2020\)](#) also identified four profiles of implementation for PATHS ([Greenberg and Kusche, 1993](#)) based on adherence, quality, dosage, reach, and engagement. The analysis sample was composed of 45 schools that implemented the program for children 9–11 years of age. Multilevel linear models (i.e., students nested within classrooms) were used to examine the association between clusters and student outcomes. Hennessey and Humphrey did not find evidence linking the PATHS implementation profiles and student academic outcomes (e.g., reading, writing, and math) controlling for student gender, percentage of free/reduced-price lunch eligibility, and academic outcomes at baseline, however. The authors noted that their profiles were primarily differentiated by dosage, as the other aspects of implementation were relatively high and stable across their sample.

Given the limited number of studies that have utilized person-centered methods to examine universal SEL implementation patterns, it is difficult to synthesize their results. The two aforementioned studies ([Low et al., 2016](#); [Hennessey and Humphrey, 2020](#)) differed in the methods of assessing implementation (teacher-report vs. observer-report), outcomes of the implementation (SEL competency vs. academic achievement), and covariates included in the prediction models. These factors may at least partially explain why the two studies yielded different profiles of implementation and patterns of association between profile membership and student outcomes. Both studies also treated implementation clusters or profiles as known groups and did not account for measurement error in classifications in their analysis of implementation outcomes. Notably, both studies were efficacy trials in which external (researcher-provided) support mechanisms such as training, coaching, and/or monitoring were in place to support implementation. To our knowledge, there currently are no published universal SEL studies that have used person-centered methods to examine implementation profiles using data from effectiveness trials (i.e., program conducted under routine conditions where implementation efforts are driven by schools and in accordance with their typical practices and available resources).

It is important to note that contextual factors at community, school and teacher levels also have been found to influence SEL program implementation ([Durlak and DuPre, 2008](#); [Anyon et al.,](#)

[2016](#); [Durlak, 2016](#)). [Domitrovich et al. \(2008\)](#) proposed a conceptual framework that synthesized macro-level (e.g., policies and funding), school-level (e.g., resources, school climate and culture), and individual-level (e.g., teachers' buy-in, confidence in delivery, professional training) determinants of school-based implementation. Similarly, studies have indicated a consistent positive association between teacher buy-in and aspects of implementation. For example, [Beets et al. \(2008\)](#) found that school climate affected teachers' beliefs and attitudes toward a school-based prevention program, *Positive Action* ([Flay et al., 2001](#)), which in turn impacted implementation dosage and adherence. [Anyon et al. \(2016\)](#) similarly found that teachers' lack of buy-in, affected by their principal's commitment or time pressure for academic-focused instruction, was a barrier to program delivery. Finally, [Domitrovich et al. \(2019\)](#) observed that teachers with positive attitudes toward the program tended to deliver SEL lessons more frequently, and their perceptions of SEL culture predicted material usage and quality of delivery.

Understanding the individual, school, and community factors associated with profiles of school-based universal SEL implementation can provide insights into what may facilitate or impede teacher practices. In [Low et al.'s \(2016\)](#) study, profile membership did not differ by individual-level factors such as gender, race, ethnicity, education, or grade taught; however, a larger number of older/more-experienced teachers were in the low-adherence class, while younger/less-experienced teachers tended to be in the low-engagement class. Nevertheless, their study did not examine other potential influences at the classroom- or school-level. Similarly, [Hennessey and Humphrey \(2020\)](#) did not examine any relationships between profile membership and contextual factors.

Implementation appears to play a critical role in accounting for the variability of school-based SEL program outcomes ([Reyes et al., 2012](#); [Low et al., 2014, 2016](#)); however, few studies have attempted to examine patterns of implementation that are associated with program outcomes and contextual factors under typical conditions in schools. The goal of the current study was to examine the implementation practices (e.g., adherence, dosage) of teachers implementing the Social Skills Improvement System SEL Classwide Intervention Program (SSIS SEL CIP; [Elliott and Gresham, 2017](#)), a universal program designed to improve students' prosocial skills and reduce problem behaviors ([DiPerna et al., 2015](#)). Using data from an effectiveness trial conducted under routine conditions, the primary aims of the study were to (a) examine if there are different patterns of program implementation and (b) determine if observed implementation patterns are associated with contextual factors and student outcomes. Our specific research questions included:

- (1) Are there different profiles of SSIS SEL CIP implementation classrooms based on dosage, adherence, quality of delivery, student engagement, and teachers' impression of lessons?
- (2) What are the contextual characteristics of the profiles of implementation?
- (3) Are profiles of implementation associated with student outcomes?

Results of this study can broaden our understanding of how evidence-based programs realistically translate into schools and potentially provide insight into "what works best" for developing and delivering school-based SEL programming.

2. Materials and methods

2.1. Participants

The analysis sample for this study consisted of 41 first- and second-grade classrooms from 13 elementary schools in the South Atlantic, East North and West North Central regions of the U.S. The number of participating classrooms within an individual school ranged from 1 to 6 (*median*=3). The racial composition of the analyzed student sample ($N=354$) was approximately 44.9% white, 30.2% Black/African American, 22.0% Hispanic/Latine, 4.5% other, 3.1% Asian, <1% American Indian/Alaskan Native, and <1% Native Hawaiian or Other Pacific Islander.¹ Most students (95.5%) spoke English as their primary language. At the time of baseline data collection, about 6.5% of students were receiving special education services, and 23.4% were receiving supplemental services (e.g., Title 1, reading support, tutoring). The analyzed teachers ($N=41$) were predominantly female (i.e., 90.2%), white (i.e., 78%), and native English speakers (i.e., 90.2%). About 9.8% of teachers were Hispanic/Latine, 4.9% were Black/African American, and 2.4% were Asian, other, or unknown. Approximately half (46.3%) were teaching Grade 1, with 63.4% of participating teachers having a Bachelor's degree and 36.6% having a Master's degree. The sample reported 14.39 years of teaching experience on average, and 34.3% had specialized teaching certificates in addition to regular education.

2.2. Measures

2.2.1. Aspects of implementation

To assess teachers' program delivery, data were collected regarding five aspects of implementation: dosage, adherence, quality of delivery, student engagement, and teacher's weekly impression of lessons. When possible, observer report was used (i.e., adherence, quality, student engagement); dosage, adherence, student engagement, and lesson impressions were assessed with teacher report.

Dosage was assessed via weekly and end-of-year survey responses. Specifically, teachers indicated the lesson(s) they taught each week and reported the completion of lessons again at the end of the implementation period (which corresponded with the end of the school year). These data sources were cross-referenced to create two dosage indicators: the number of lessons taught from the "core" SSIS SEL CIP units (out of 30 lessons across 10 core units) and the total number of lessons implemented across all units (out of 69 total lessons across 23 units).² The average number of core lessons implemented across classrooms was 23.23 ($SD=5.90$, range=7–30), and the average number of total lessons implemented was 26.20 ($SD=8.50$, range=7–46).

Adherence includes four indicators. *Self-reported adherence* was teachers' rating of their own adherence to the lesson scripts on a 5-point scale (1 = *Not at all* to 5 = *Completely*). The composite score was the average across weeks. *Observed steps* measured the completion

of steps described in the curriculum (i.e., tell, show/do, practice) via classroom observation. Each step was scored as 0 (*Non-occurrence*) or 1 (*Occurrence*), and the percentage of steps completed was averaged across lesson observations. *Observed adherence* reflected observers' ratings of the degree to which teachers followed the verbal script of the SSIS SEL CIP lessons using a 5-point scale (1 = 0–20% to 5 = 81–100%). *Observed level of implementation* assessed the extent to which teachers implemented the primary sections of each lesson (i.e., tell, show/do, practice, monitor progress, generalize) using a 5-point scale (1 = *Not at all* to 5 = *Completely*), and a composite score was averaged across sections. Interrater agreement was 92.15% for *observed steps*, 76.7% (90% with 1-point tolerance for disagreement) for *observed adherence*, and 62.92% (87.9% with 1-point tolerance) for *observed level of implementation*.

Observed quality of delivery was assessed during independent lesson observations. Specifically, observers completed five items that measured preparedness, enthusiasm, responsiveness to questions, clarity of presentation, and skill in facilitating activities using a 5-point scale (1 = *Not at all* to 5 = *Completely*). A composite score was computed for each lesson and then averaged across weeks. Interrater agreement was 72% (96.5% with 1-point tolerance).

Student engagement assesses students' active engagement, enthusiasm/interest, and understanding of lesson using teachers' and observers' report. Teachers responded to three questions regarding student engagement during weekly lessons using a 5-point scale (1 = *Very low* to 5 = *Very high*). Observers rated the same items after each lesson observation, and the interrater agreement on lesson observations conducted by two independent observers was 65% (95% with 1-point tolerance).

Weekly impression of lessons was measured via a single question using a 5-point scale (1 = *Poor*, 2 = *Fair*, 3 = *Acceptable*, 4 = *Very good*, 5 = *Excellent*) that teachers completed weekly throughout the implementation period. The item was "Overall, how would you rate the lessons you taught during the current week?" The composite score was averaged across weeks. This item was used to assess teachers' overall judgment of SSIS SEL CIP lessons that were taught during that week.

2.2.2. Student outcomes

Social skills and problem behaviors were rated by teachers using the social skills improvement system rating scales-teacher form (SSIS-RST; Gresham and Elliott, 2008). The *Social Skills* scale measures communication, cooperation, assertion, responsibility, empathy, social engagement, and self-control, whereas the *Problem Behavior* scale assesses externalizing, bullying, hyperactive-inattentive, internalizing, and autistic behaviors. Items in both scales use a 4-point format (0 = *Never* to 3 = *Almost always*). The SSIS-RST demonstrated sound psychometric evidence (e.g., $\alpha = 0.88-0.98$; DiPerna et al., 2018).

Approaches to learning were measured using teacher's ratings on the Academic Competence Evaluation Scales (ACES; DiPerna and Elliott, 2000). The *Motivation* scale assesses students' learning attitudes, persistence, and interest. The *Engagement* scale measures students' attention and participation in academic activities. Items on both scales were rated using a 5-point format (1 = *Never* to 5 = *Almost always*). Both subscales showed strong psychometric evidence (e.g., $\alpha = 0.95-0.98$; DiPerna et al., 2018). Composite scores for each scale were converted to item response theory (IRT) scores to ensure the equivalence of assessment at the pre-test and post-test.

1 Participants were allowed to endorse more than one category of racial/ethnic group.

2 See the section of procedures for a detailed description of the core units.

2.2.3. Multilevel contextual factors

School-, teacher-, and class-level demographic information were collected during the year of the study. School information included percent of children eligible for free/reduced-price lunch, percent of racial/ethnic minority children, school size, and location. Teachers provided information about their own demographic characteristics including gender, race/ethnicity, certification, educational level, primary language, and years of teaching experience.

The classroom assessment scoring system: kindergarten-third grade scale (CLASS K-3; [Pianta et al., 2008](#)) was used to evaluate the instructional climate of participating classrooms. Research staff observed the implementation classrooms and completed ratings in regard to emotional support, instructional support and classroom organization. Each item was rated by two observers on a 7-point scale (1 = *low* to 7 = *high*). Intraclass correlations for paired CLASS observations have been shown to be acceptable (0.65–0.76; [DiPerna et al., 2018](#)).

The *Teacher SEL Beliefs* scale includes four items reflecting teachers' comfort with teaching SEL, 4 items assessing commitment to SEL, and four items measuring perceived school culture relative to SEL ([Brackett et al., 2011](#)). The internal consistency for each subscale was good ($\alpha \geq .80$). The assumptions supporting social-emotional teaching (ASSET) scale assesses teachers' beliefs about the degree to which SEL skills are malleable, compatible, and influential ([Hart, 2021](#)). The internal consistency for ASSET subscales and total score were satisfactory ($\alpha \geq .87$).

Lastly, teachers were asked on an end-of-year questionnaire to indicate their opinion about the percent of school time should be allocated to facilitate academic or SEL skills. The question was "What percentage of early elementary students' (Grades 1–2) school time should be focused on each domain?" and included response options for academic subject areas (reading, math, etc.) as well as SEL. The total percentage across all the domains were required to sum to 100%.

2.3. Procedure

The larger effectiveness trial was approved by the university's Institutional Review Board. Consistent with the goal of testing the effectiveness of the SSIS SEL CIP, districts that were already considering adopting a universal SEL program in the early grades as part of their typical practice were recruited to participate. With the goal of reaching geographically diverse school sites, information about the trial was distributed through online platforms and national professional networks. School representatives who requested more details were provided with additional information through individual conversations. Prior to enrolling a school into the study, we sought and received permission to conduct the research with administration according to district guidelines. In addition, active teacher and parent consent was obtained prior to data collection. Schools were randomly assigned to treatment condition such that, within each school, the SSIS CIP SEL was taught in either first- or second-grade classrooms while the other grade levels maintained business-as-usual practices. Data from classrooms assigned to the treatment condition were used for this study.

The SSIS SEL CIP includes 10 core instructional units and 13 advanced units (3 lessons per unit) that focus on social and emotional

skills that a national sample of U.S. teachers identified as important to student success (e.g., listening to others, paying attention to your work, asking questions). Each SSIS SEL CIP lesson requires approximately 25–30 min and features multiple instructional phases (i.e., telling, showing, doing, practicing, monitoring progress, and generalizing) to promote skill acquisition and generalization. Materials include a teacher guide with scripted lesson plans, brief video clips that demonstrate examples and non-examples of social behaviors, scenarios describing common classroom scenarios for role plays, and cue card with emotion emojis. Because the goal of the larger project was to examine student outcomes under typical conditions (levels of support) and practices, teachers and schools decided how much, how often, and in what way to plan for and deliver SSIS SEL CIP units. Most teachers (79.5%) reported that they planned for implementation by preparing on their own, 38.5% reported planning with colleagues, and only about 12.8% attended a training conducted by their school or district.

Teachers' self-report and independent direct observations were completed to measure implementation fidelity of the SSIS SEL CIP. Teachers completed weekly surveys via online questionnaires to report the number of lessons completed, and rate the degree of adherence to the curriculum, quality of delivery, and student engagement. Teachers were compensated for their time spent completing questionnaires. Trained research assistants completed direct observations of implementation classrooms and rated teacher's adherence, quality of delivery, and student engagement. The average number of lessons observed per teacher was 5.40, with a minimum of three and a maximum of seven. Eighteen percent of observations were completed by paired observers. Student measures were completed immediately before and after program implementation.

The study duration was one school year. The SSIS SEL CIP materials were provided to the implementing grade levels prior to the baseline data collection window and then shared with the control classrooms at the end of post-test data collection.

2.4. Statistical analysis

The first step was to conduct a latent profile analysis (LPA) on implementation measures (i.e., dosage, adherence, quality, student engagement, and weekly impression of lessons) to explore if there were different patterns of classroom implementation of the SSIS SEL CIP. LPA is a statistical approach for identifying latent subgroups based on a set of observed indicators ([Ferguson et al., 2020](#)). LPA models can produce estimates of membership probability for each participant so that individuals sharing the same pattern of indicators are categorized into the same underlying class ([Spurk et al., 2020](#)). In this study, LPA was used as an exploratory tool to identify profiles in which classrooms shared similar patterns of implementation.

Multiple models were fitted to generate 1 to 4 latent profiles. Each model was compared against the previous one [i.e., ($k-1$) profiles] according to Akaike's Information Criterion (AIC), Bayesian Information Criterion (BIC), Sample-size Adjusted BIC (SABIC), entropy, LMR (Lo-Mendell-Rubin likelihood ratio test), and VLMR (Vuong-Lo-Mendell-Rubin likelihood ratio test). A smaller value of AIC, BIC, SABIC and a higher entropy indicate that a model fits better to the data ([Ferguson et al., 2020](#)). LMR and VLMR were used to test whether the model with [k] profiles fits better than the one with [$k-1$]

profiles. A non-significant test result suggests that the more parsimonious model is better fitting (Smith et al., 2021). We selected the final model based on all fit indices and interpretability of the profiles.

After the number of latent profiles were determined, the second step was to examine the association between the profile membership and contextual variables to identify ones that differed across profiles. Profile membership was determined based on the probability estimates produced by the solution of the chosen number of profiles (i.e., classrooms were assigned to the profile to which they had the highest probability of belonging). Contextual variables included school-, teacher-, and classroom-level demographic characteristics. Chi-square tests were conducted to examine whether profiles were associated with categorical factors (e.g., school location, teacher's certification, educational level). One-way ANOVA was used to test mean differences across profile for continuous variables (e.g., classroom environment, SEL belief, years of teaching). Due to the large number of contextual variables and the small classroom-level sample size by profile, judicious selection of contextual factors was necessary. We used this approach to identify statistically significant factors that should be considered for the next step of distal outcome analyses.

The third step was to explore the associations between implementation profile membership and outcomes of social skills, academic motivation, engagement, and problem behaviors controlling for contextual factors. We used auxiliary regression models for this purpose because they allowed auxiliary variables (contextual factors) to predict both profile membership and distal outcomes. Specifically, the manual BCH (Bolck et al., 2004) approach in Mplus was used because it could preserve profile membership and account for measurement error (Asparouhov and Muthén, 2014). Clustering by schools was accounted for by adjusting standard errors in Mplus using the cluster and "analysis=complex" commands. Robust maximum likelihood estimator (MLR) was used to estimate model parameters due to its robustness against violation of assumptions (e.g., normality) and ability to generate less biased estimates of parameters in comparison with the traditional maximum likelihood approach (Bakk and Vermunt, 2016).

A common-slope model was specified for each of the outcome variables to hold contextual factors constant across profiles. We fitted the model by constraining the regression slopes for covariates to be equal across profiles to estimate adjusted profile mean outcome differences (i.e., control for covariates). Contextual covariates included grade level, percentage of supplemental educational services, teaching experience, classroom organization, instructional support, and percentage of free/reduced-price lunch eligibility. These contextual variables were selected because they varied significantly across profiles and were included as covariates in prior implementation studies (Domitrovich et al., 2010; Reyes et al., 2012; Low et al., 2014, 2016; Cross et al., 2015; Humphrey et al., 2018). Even though percentage of students receiving supplemental services at classroom level was rarely examined in implementation studies, a previous efficacy trial of the SSIS-CIP (DiPerna et al., 2018) suggested that supplemental services had a negative effect on social skills and academic motivation. Therefore, it was included as a covariate in predicting outcome gains.

To account for baseline difference in student outcome variables, we calculated reliable change (RC) scores that reflect how much change occurred during the implementation period. The RC scores

were computed by subtracting the pre-test scores from the post-test and then dividing the difference scores by the standard error of the difference (Jacobson and Truax, 1991). If the RC score is larger than critical values (e.g., $z = \pm 1.96$ at α level of 0.05), the pre-post change is considered statistically reliable (Ferguson and Splaine, 2002). The sign of RC scores indicates the direction of change from pre-test to post-test. The RC score was used in the analysis because, first, we were interested in within-person change rather than relative change; second, it facilitated interpretation by accounting for standard error of measurement (i.e., whether the amount of pre-post change was reliable or due to random error) and was commonly used to decide clinical significance in mental and behavioral health (Ferguson and Splaine, 2002).

It is important to emphasize the exploratory nature of these analyses given the relatively small overall class-level sample size ($N = 41$) limits statistical power in detecting profile differences in both contextual factors and outcome gains. As such, the relevant results should be interpreted accordingly.

3. Results

3.1. Latent profiles of implementation classrooms

Descriptive statistics and intercorrelations for the implementation measures are presented in Table 1. There was a salient difference in variability of teacher's self-report and direct observation in terms of adherence and student engagement. It appears that teacher self-reported ratings of adherence were moderately associated with direct observations of adherence, with Pearson's r of 0.38. However, student engagement reported from teachers and their weekly impressions of lessons were weakly associated with all observer-reported implementation measures ($|$ Pearson's $r| < 0.10$).

To address the first research question, models with 1–4 latent profiles were fitted. Table 2 presents the fit statistics for each of the models. The 4-profile solution had a non-positive definite matrix and was difficult to interpret, therefore it was eliminated from further consideration. The entropies of the 2- and 3-profile solutions were equivalent (entropy = 0.974), which indicated a high classification certainty for both models (Ferguson et al., 2020). Smaller values of AIC, BIC and SABIC suggested the 3-profile solution was fitting better than the 2-profile. The difference in BIC (i.e., Δ BIC > 10) also provided strong evidence in support of the 3-profile model (Raferty, 1995). However, the non-significant results of VLMR and LMR tests suggested the 2-profile solution was better fitting ($p > 0.05$) than the 3-profile. Given the mixed results from multiple indices, we decided to retain the 3-profile solution because it provided more interpretive information.

Table 3 displays the descriptive data for the implementation measures by three latent profiles. Original scores from each measure were standardized to ease the interpretation of results across measures. One-way ANOVA and Tukey's HSD were conducted to compare the average variable scores across profile. Measures with asterisks indicate significant differences across profiles, and superscripts (i.e., a, b, and c) indicate significant pairwise differences. It suggested, for example, means of observer-reported adherence, quality of delivery and student engagement for Profile 1 were significantly lower than the other two

TABLE 1 Summary of descriptive statistics and intercorrelations for the implementation measures.

Measure	M (SD)	Min	Max	1	2	3	4	5	6	7	8	9
Direct observation												
1. Steps	0.74 (0.20)	0.16	0.99									
2. Adherence	3.89 (1.00)	1.00	5.00	0.90***	–							
3. Level of implementation	3.80 (0.88)	1.58	5.00	0.96***	0.86***	–						
4. Quality of delivery	4.26 (0.57)	2.25	5.00	0.82***	0.68***	0.87***	–					
5. Student engagement	4.29 (0.50)	3.22	5.00	0.58***	0.52***	0.65***	0.80***	–				
Teacher report												
6. Number of core lessons	23.27 (5.83)	7	30	0.45**	0.36*	0.45**	0.47**	0.40*	–			
7. Number of total lessons	26.20 (8.39)	7	46	0.35*	0.25	0.36*	0.42**	0.35*	0.77***	–		
8. Adherence	3.87 (0.42)	3.00	4.89	0.34*	0.38*	0.32*	0.18	0.02	0.19	0.30	–	
9. Student engagement	3.66 (0.40)	2.93	4.47	–0.02	–0.07	–0.03	–0.08	0.00	–0.16	–0.17	0.30	–
10. Impression of lessons	3.57 (0.45)	2.73	4.38	0.03	0.06	–0.05	–0.09	–0.01	–0.27	–0.18	0.19	0.72***

N=41. Scores of the implementation measures were on the original scale.
p*<0.05; *p*<0.01; ****p*<0.001.

TABLE 2 Latent profile analysis (LPA) model fit summary and profile proportions.

Model	LL	AIC	BIC	SABIC	Entropy	VLMR	LMR	Profile proportions
						<i>p</i> -value	<i>p</i> -value	
1	–576.70	1193.41	1227.68	1165.07				
2	–502.83	1067.65	1120.77	1023.73	0.974	0.0925	0.0971	0.29, 0.71
3	–462.92	1009.84	1081.81	950.33	0.974	0.3586	0.3664	0.12, 0.37, 0.51
4	–445.06	996.12	1086.94	921.03	0.968	0.7932	0.7975	0.24, 0.12, 0.17, 0.46

LL, log likelihood; VLMR, The Vuong-Lo-Mendell-Rubin likelihood ratio test between [*k*-1] profiles and [*k*] profiles; LMR, The Lo-Mendell-Rubin adjusted likelihood ratio test between [*k*-1] profiles and [*k*] profiles.

profiles, and Profiles 2 and 3 differed by one standard deviation on most of the measures. As such, the 3-profile solution provided more information of implementation patterns compared to the 2-profile solution which only identified High and Low profiles.

It is noticeable that the profile sizes found in the 3-profile solution were small, especially for the smallest profile (*N*=5). However, profile separation or the distance between profiles dictates the sample size required and hence power to detect the correct number of profiles in LPA (e.g., Tein et al., 2013; Ferguson et al., 2020). The high entropy of 0.974 for the 3-profile model suggests a high degree of profile separation certainty. Also as shown in Figure 1, the three profiles varied on the majority of the implementation measures, except for teachers’ report of student engagement and weekly impression of lessons.

Based on the 3-profile solution, five classrooms (12%) were classified to the first profile in which teachers completed fewer lessons with less adherence and quality, and students showed lower level of engagement. Therefore, we characterized this group of classrooms as

the “Low Implementation” Profile (Low IP). The second profile included 15 classrooms (37%) that reported lower dosage but moderate adherence, quality, and student engagement. Notably, the direct observation and teachers’ report yielded somewhat inconsistent ratings of adherence and engagement. Due to the small variability of teacher-reported ratings, we focused on the observational ratings and labeled this group as the “Moderate Implementation” Profile (Moderate IP). Twenty-one classrooms (51%) in the third profile completed the most lessons and had the highest ratings on all observational assessments, thus we labeled them collectively as the “High Implementation” Profile (High IP).

3.2. Contextual characteristics of the three implementation profiles

To address the second research question, we examined the contextual characteristics of the three profiles at the school, class,

TABLE 3 Descriptive statistics of the implementation scores by profile.

Measure	Mean (SD)		
	Profile 1	Profile 2	Profile 3
	Low IP (N=5)	Moderate IP (N=15)	High IP (N=21)
Direct observation			
Steps*	-1.92 (0.62)	-0.45 (0.40) ^a	0.78 (0.33) ^b
Adherence*	-2.03 (0.81)	-0.19 (0.53) ^a	0.62 (0.46) ^b
Level of implementation*	-1.89 (0.40)	-0.47 (0.39) ^a	0.78 (0.39) ^b
Quality of delivery*	-1.55 (1.26)	-0.45 (0.52) ^a	0.69 (0.48) ^b
Student engagement*	-1.16 (0.94)	-0.36 (0.86)	0.53 (0.76) ^b
Teacher report			
Number of core lessons*	-0.53 (1.13)	-0.63 (1.04)	0.57 (0.52) ^{bc}
Number of total lessons*	-0.64 (0.81)	-0.41 (1.06)	0.45 (0.80) ^b
Adherence*	-1.14 (0.84)	0.15 (1.04) ^a	0.16 (0.86) ^b
Student engagement	0.06 (1.10)	-0.06 (1.10)	0.03 (0.95)
Impression of lessons	-0.10 (0.82)	0.05 (1.00)	-0.01 (1.08)

N=41. Standard deviations are presented in the parenthesis. Scores of the implementation measures were on a z-score scale. Cell means on the same row were compared using One-way ANOVA and Tukey's HSD. ^aSignificant difference was found between Profile 1 (Low IP) and Profile 2 (Moderate IP).

^bSignificant difference was found between Profile 1 (Low IP) and Profile 3 (High IP).

^cSignificant difference was found between Profile 2 (Moderate IP) and Profile 3 (High IP).

*p < 0.05.

and teacher levels (Table 4). Results suggested that the 5 Low IP classrooms were all in the second grade and from two schools located in suburban districts (in the South Atlantic and East North Central regions of the United States). They had the smallest percentage of students receiving free/reduced lunch (37.80%), students of color (53.40% at school level; 38.25% at class level) and students receiving supplemental services (14.35%). Nevertheless, they had the highest percentage of students receiving special educational services (12%). Three teachers (60%) were female, two teachers (40%) received both regular and special education certificates, and one teacher (20%) was racial/ethnic minority. Teachers had the greatest amount of prior teaching experience (M = 20.60, SD = 9.34), and they tended to believe that more school time should be allocated to foster academic skills (55% of the school day) than social-emotional skills (11.25%).

The 15 Moderate IP classrooms were from eight schools distributed across rural, suburban, and urban districts. These classrooms were located in schools serving a significantly higher percent of students (86.79%) receiving free/reduced-price lunch and a marginally significantly higher percent of students of color (73.36% at school level; 72.24% at class level). These classrooms also were observed to have the lowest levels of instructional support and classroom organization. Only one teacher (7%) was male, four

teachers (27%) received regular and other credential, and four teachers (27%) were racial/ethnic minority.

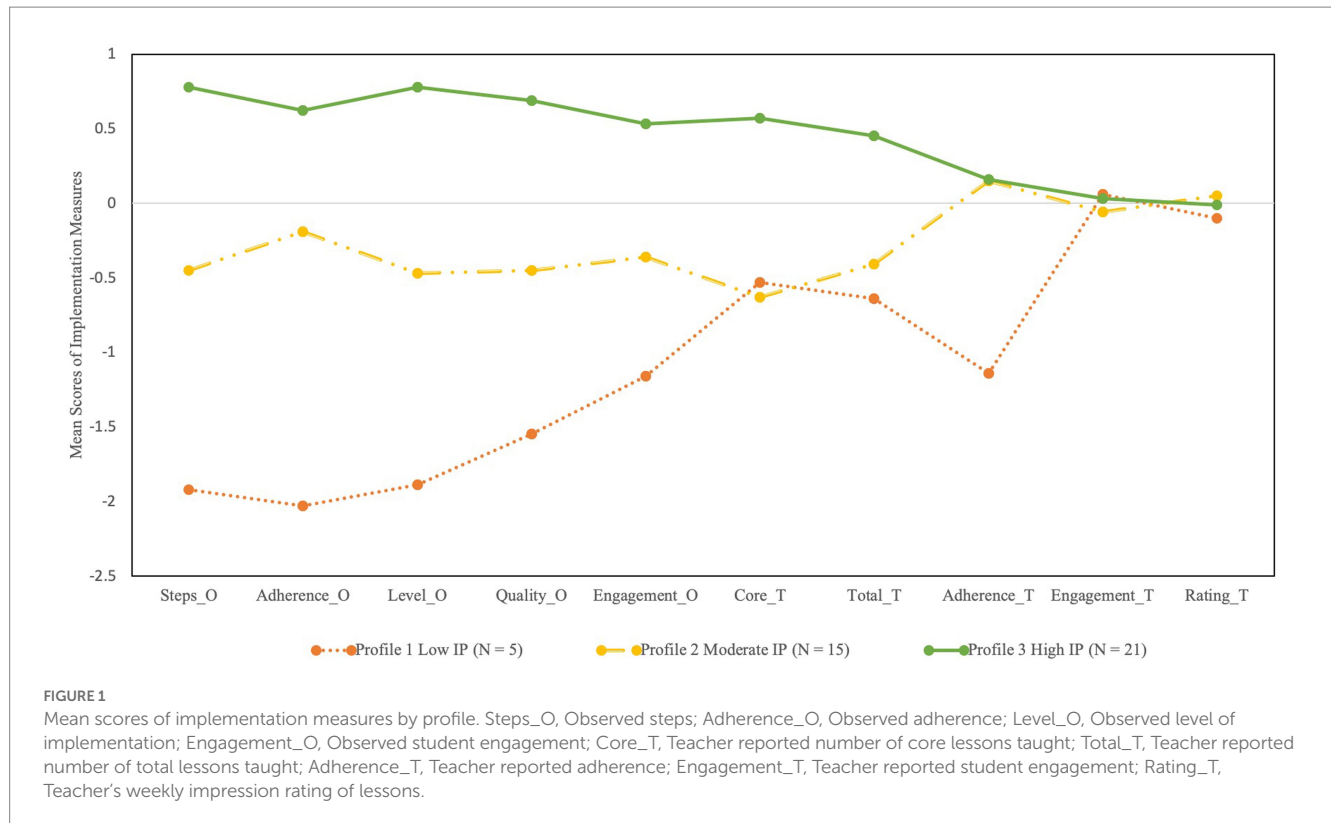
The 21 High IP classrooms were from 10 schools with diverse student populations (53.76% Students of color at school level; 51.64% at class level) where nearly 6 in 10 students who were eligible for free/reduced lunch (58.71%). One teacher (5%) was male, three teachers (14%) were racial/ethnic minority and eight teachers (38%) received other credential along with regular educational certificate. The average years of teaching experience was 13.86, which was close to the Moderate IP but much lower than the Low IP classrooms. These classrooms demonstrated significantly higher levels of instructional support and classroom organization relative to the Moderate IP classrooms.

Significant associations were found between profiles and grade ($\chi^2 = 9.29, df = 2, p = .01$), school location ($\chi^2 = 18.31, df = 6, p = .01$) and teachers' primary language ($\chi^2 = 7.68, df = 2, p = .02$). However, results must be interpreted with caution because there were a few zero counts in the frequency tables. Significant mean differences were only detected for percentage of students receiving free/reduced-price lunch, classroom organization, and instructional support. We also examined the distribution of classrooms by schools across profiles and found that each profile was represented by more than one school (2, 8, and 10 schools in Profiles 1, 2, and 3, respectively). Selected covariates were included to predict profile membership using the BCH approach. Profile classification generated from the BCH approach was consistent with the results from one-step LPA (i.e., 3-profile solution). Auxiliary regression analysis was adopted to examine the association between covariates and profile. Results indicated that only percentage of free/reduced-price lunch eligibility was marginally associated with profile membership. Classrooms with higher percentage of free/reduced-price lunch eligibility were more likely to be assigned to Moderate or High IPs compared to Low IP.

3.3. Association between profile membership and student gains

We used the probability of profile membership to predict student gains in social skills, problem behaviors, academic motivation, and engagement after controlling for grade level, teaching experience, percentage of students receiving supplemental services, percentage of free/reduced-price lunch eligibility, classroom organization, and instructional support. School-, teacher- and class-level covariates were included in the auxiliary regression models to account for the nonequivalence across profiles and their potential influence on student gains suggested by previous studies (e.g., Domitrovich et al., 2010; Cross et al., 2015).

The auxiliary regression model results are shown in Table 5 by outcome variables. Intercepts 1, 2, and 3 represent the adjusted mean of gains for Profiles 1, 2, and 3, respectively, when covariates were kept constant. In examining the direction and relative magnitude of gains shown in Table 5, the Moderate IP classroom mean gains appeared highest in social skills, academic motivation, and academic engagement, followed by the High IP classrooms. The Low IP classrooms appeared to show reductions in social skills, academic motivation, and academic engagement over time. High IP classrooms demonstrated a reduction in problem behaviors, while both Moderate



and Low IP classrooms showed an increase. It is important to note that these observations about the means are descriptive; we discuss statistically significant differences among profiles using confidence intervals in the next section.

The estimates of the mean gains for each outcome with 95% confidence intervals are shown in Figure 2. Intervals for adjusted means that did not overlap indicated significant difference between the adjusted means. There were no statistically significant differences in adjusted gains of social skills, problem behaviors, or academic engagement across profiles. However, the Moderate [95% CI = (0.20, 0.82)] and High IP classrooms [95% CI = (-0.25, 0.53)] showed significantly higher gains in academic motivation compared to Low IP [95% CI = (-1.55, -0.41)], even though no significant difference was found between the Moderate and High IP classrooms. Across profiles, second grade classrooms had a significantly higher gain in social skills ($b = 0.38$, $SE = 0.13$, $p = 0.003$), and classrooms with higher classroom organization tended to gain more in social skills ($b = 0.35$, $SE = 0.11$, $p = 0.001$). Percentage of free/reduced-price lunch eligibility appeared to be negatively associated with gains in academic motivation ($b = -0.86$, $SE = 0.37$, $p = 0.022$) and engagement ($b = -0.89$, $SE = 0.39$, $p = 0.021$). Results also suggested that instructional support was negatively associated with student gains in academic engagement ($b = -0.32$, $SE = 0.12$, $p = 0.009$).

4. Discussion

The study identified three latent profiles of implementation of the SSIS SEL CIP under routine implementation conditions based on dosage, adherence, quality of delivery, student engagement, and weekly impression of lessons. The Low IP included about 12% of the

classrooms and was characterized by low dosage, low fidelity to instructional scripts, less quality of implementation, and lower student engagement in the program lessons. The Moderate IP was comprised of about 37% of the classrooms; these classrooms demonstrated lower dosage but average adherence, quality of delivery, and engagement. About half of the classrooms fell into the High IP in which teachers delivered a greater number of lessons with high quality, the program was implemented with high adherence, and students appeared to be engaged to a high degree in the lessons.

It is important to note that the three profiles were labeled primarily based on observational data instead of teachers' self-report. Data collected directly from teachers (e.g., student engagement, impression rating) demonstrated limited variability. Social desirability may have played a role in the limited variability; however, one additional explanation is that the question used to solicit teachers' overall reflection on the lessons taught during that week was fairly broad and could have been interpreted differently across teachers, with some potentially rating the quality of curriculum and others potentially rating the quality of delivery (or some combination of the two). Student engagement reported by teachers and their impression rating of the lessons were also weakly associated with observer-reported implementation measures. A possible explanation could be that teachers and observers had different perceptions of implementation, particularly how students were engaged in SEL class activities. Teachers provided their perceptions of implementation practice retrospectively at the end of each week, whereas observers provided their ratings on individual lessons they observed in real-time. In their seminal meta-analysis, Durlak et al. (2011) noted that observational data are more objective and appear to be more correlated with program outcomes than teacher report data, and therefore recommended to use direct observation for implementation analysis

TABLE 4 Contextual characteristics of the three latent profiles.

Measure	Profile 1	Profile 2	Profile 3
	Low IP	Moderate IP	High IP
	(N =5)	(N =15)	(N =21)
School-level demographic (%)			
Free/reduced-price lunch*	37.80 (39)	86.79 (19) ^a	58.71 (30) ^b
Ethnic/racial minority	53.40 (36)	73.36 (30)	53.76 (34)
School location (%)**			
Rural (mid-size)	0	20.00	33.33
Suburban (large)	100.00	26.67	52.38
Urban (mid-size)	0	13.33	14.29
Urban (large)	0	40.00	0
Geographic region (%)			
South Atlantic	60.00	53.33	28.57
East North Central	40.00	26.67	38.10
West North Central	0	20.00	33.33
Teacher-level demographic (%)			
Female	60.00	93.30	95.20
Ethnic/racial minority	20.00	26.67	19.00
Bachelor's degree	80.00	73.33	52.38
Master's degree	20.00	26.67	47.62
English as primary language*	100.00	73.33	100.00
General education teacher	100.00	80.00	95.24
General & Special education teacher	0	13.33	0
Teaching experience (yrs.)	20.60 (9.34)	13.07 (10.69)	13.86 (9.09)
Classroom-level demographic (%)			
Grade 1*	0	73.33	38.10
Grade 2*	100	26.67	61.90
Special educational services	12.00 (15.54)	9.36 (9.33)	8.25 (9.17)
Supplemental educational services	14.35 (22.04)	39.81 (39.25)	18.13 (15.60)
English language learners	6.33 (11.30)	6.02 (9.44)	11.41 (13.25)
Ethnic/Racial minority	38.25 (36.66)	72.24 (34.01)	51.64 (32.66)
Classroom environment			
Emotional support	5.40 (0.66)	5.23 (0.98)	5.33 (1.21)
Class organization*	5.10 (0.64)	4.82 (1.02)	5.72 (1.05) ^b
Instructional support*	2.00 (0.90)	1.73 (0.75)	2.85 (1.11) ^b
Teacher's belief in SEL			
SEL comfort	4.05 (0.60)	4.05 (0.54)	3.89 (0.72)
SEL commitment	4.15 (0.49)	4.37 (0.55)	4.20 (0.56)
SEL culture	3.67 (0.94)	4.00 (0.65)	3.76 (0.86)
SEL malleable	4.46 (0.41)	4.32 (0.37)	4.49 (0.36)
SEL compatible	3.88 (0.97)	4.26 (0.67)	4.19 (0.53)
SEL influential	4.50 (0.46)	4.50 (0.47)	4.54 (0.49)
School time allocated (%)			
Social-emotional skills	11.25 (2.50)	21.79 (7.49)	18.33 (8.99)
Academic skills	55.00 (16.83)	45.00 (10.19)	45.24 (15.85)

N= 41. Standard deviations are presented in parenthesis. Cell means on the same row are compared using One-way ANOVA and Tukey's HSD.

^aSignificant difference was found between Profiles 1 (Low IP) and Profile 2 (Moderate IP).

^bSignificant difference was found between Profiles 2 (Moderate IP) and Profile 3 (High IP).

*p < 0.05; **p < 0.01.

TABLE 5 Common-slope model estimates by outcome variables.

Effect	Social skills			Problem behaviors			Motivation			Engagement		
	B	SE	95% CI	B	SE	95% CI	B	SE	95% CI	B	SE	95% CI
Intercept 1	-0.58	0.56	[-1.67, 0.52]	0.13	0.72	[-1.28, 1.54]	-0.98	0.29	[-1.55, -0.41]	-0.56	0.47	[-1.48, 0.36]
Intercept 2	0.57	0.12	[0.33, 0.81]	0.27	0.26	[-0.24, 0.78]	0.51	0.16	[0.20, 0.82]	0.33	0.08	[0.17, 0.49]
Intercept 3	0.19	0.19	[-0.18, 0.56]	-0.18	0.14	[-0.45, 0.09]	0.14	0.20	[-0.25, 0.53]	0.22	0.17	[-0.11, 0.55]
Grade	0.38**	0.13		0.49	0.26		0.25	0.24		0.06	0.13	
Experience	0.00	0.01		0.01	0.02		0.01	0.01		-0.004	0.01	
% SUPED	-0.20	0.16		0.65	0.39		0.02	0.22		-0.45	0.24	
CO	0.35**	0.11		0.03	0.20		0.09	0.09		0.15	0.12	
IS	-0.13	0.24		-0.08	0.20		-0.17	0.09		-0.32**	0.12	
% SFR	-0.02	0.81		-0.36	0.86		-0.86*	0.37		-0.89*	0.39	
AIC	161.28			169.99			151.03			153.31		
BIC	205.19			213.91			194.94			197.22		
SABIC	123.83			132.54			113.58			115.86		

N=40. Intercept 1 = adjusted means of outcomes for the Low IP classrooms; Intercept 2 = adjusted means of outcomes for the Moderate IP classrooms; Intercept 3 = adjusted means of outcomes for the High IP classrooms. Mean outcomes are adjusted for grade, teaching experience (yrs), percentage of students eligible for free/reduced-price lunch (% SFR), classroom organization (CO), instructional support (IS), and percentage of students receiving supplemental education service (% SUPED).

*p < 0.05; **p < 0.01.

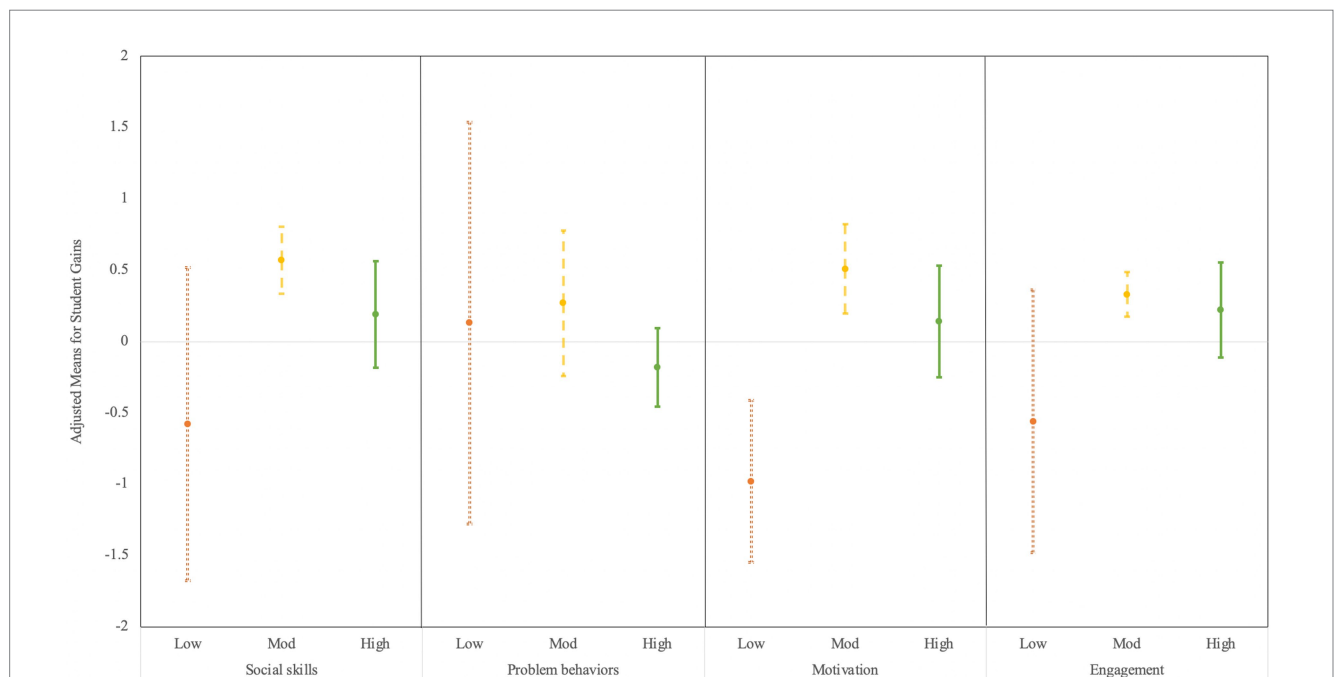


FIGURE 2

Auxiliary regression estimates of student gains with 95% CIs by Profile. Low, Low implementation profile; Mod, moderate implementation profile; High, high implementation profile. Mean outcomes are adjusted for grade, teaching experience (yrs), percentage of students eligible for free/reduced-price lunch (% SFR), classroom organization (CO), instructional support (IS), and percentage of students receiving supplemental education service (% SUPED).

if it is possible. One noticeable difference between our study and prior profile research is that we used different methods to assess implementation practices. Low et al. (2016) used teacher-reported rating, and their profile classification gave more weight to engagement

and adherence than dosage and generalization. Hennessey and Humphrey (2020) adopted observational data, and the clusters they identified were primarily based on dosage. Our study used both teacher- and observer-reported rating of implementation and results

suggested that adherence, quality of delivery, and student engagement may play an integral role in differentiating profiles. Given the interpretation of program outcomes relies upon accurate assessment of implementation, there is a need for future research that incorporates and evaluates multiple methods to measure program implementation.

Regarding contextual factors, we found that the Moderate IP classrooms were comprised of a significantly higher percentage of students with free/reduced-price lunch eligibility. In addition, these classrooms were observed to have significantly lower instructional support and classroom organization than the High IP classrooms. Teachers also reported putting more daily instructional emphasis on social-emotional skills (21.79%) in the Moderate profile than the others. This suggests that teachers' emphasis on social-emotional skills across and throughout the school day and their interactions with students may facilitate the development of these skills even in the context of lower dosage of a formal SEL program. The Low IP teachers had the greatest amount of teaching experience and tended to place more daily instructional emphasis on academic skills (55%) than social-emotional skills (11.25%). Prioritizing academic instruction is perhaps the most likely reason why Low IP teachers completed fewer lessons with fidelity. This finding also supports the argument that time pressure for academic instruction may be a barrier to SEL program delivery in some schools and classrooms (Anyon et al., 2016). Even though prior literature suggests school climate and teacher's SEL beliefs may affect implementation (Durlak and DuPre, 2008; Durlak, 2016), our current study did not yield evidence to support this hypothesis. There are at least two possible explanations. First, the small sample size and resulting lower levels of statistical power may have been insufficient to detect significant differences; second, teacher's self-reported ratings may have been affected by social desirability, resulting in response patterns that were more positive than reality (Holden and Passey, 2009).

In investigating the association between profile membership and student gains, we found only one statistically significant difference. Moderate and High IP classrooms demonstrated higher gains in students' academic motivation compared to Low IP classrooms and controlling for contextual factors. Although not reaching a threshold of statistical significance, based on the adjusted means, the Moderate and High IP classrooms appeared to have more positive gains in social skills and academic engagement, with Moderate IP showing the most benefit. The Low IP classrooms, however, had negative gains in all of the outcomes. One noteworthy consideration is that increasing dosage beyond a certain threshold may not necessarily relate to improved student outcomes resulting from implementing universal SEL, which is consistent with and extends prior efficacy research (Domitrovich et al., 2010; Reyes et al., 2012; Humphrey et al., 2018). For example, Hennessey and Humphrey (2020) found their profiles, which were primarily differentiated by dosage, did not appear related to students' academic outcomes and Low et al. (2016) suggested that aspects of implementation delivery that are harder to manualize (e.g., implementer competency and student engagement) are necessary for maximal program benefit. In this sample, with the same amount of lessons completed, the Moderate IP classrooms showed higher gains in academic motivation compared to the Low IP classrooms; however, teachers in the Moderate IP classrooms delivered the lessons with higher adherence, quality, and student engagement.

The difference between student outcomes in the Moderate and High IP classrooms was not statistically significant in any of the outcomes, suggesting that there may be a certain threshold of covered

content and implementation practices that result in the greatest benefit. Identifying such "core components" (Lawson et al., 2019; Wigelsworth et al., 2021) of SSIS SEL CIP implementation is an important direction for future research. In general, classrooms where lessons were delivered with higher levels of adherence and quality, as well as engaged students in class activities more often, were associated with higher gains in prosocial behavior, academic motivation, and engagement. This finding is consistent with results from a previous efficacy trial of an earlier edition of the SSIS CIP program (e.g., DiPerna et al., 2015, 2016, 2018), in which the program, implemented with high levels of fidelity, yielded positive effects for students. To date, the existing research on universal SEL implementation in schools has largely come from efficacy trials, during which a research team provides implementation supports in order to enhance internal validity when evaluating and isolating the causal impact of programs. However, studies conducted under authentic or routine conditions, in which educators independently determine implementation based on their needs, resources, and capacities, are few and far between. Such effectiveness research is critical for extending the external validity of efficacy trials—that is, results from studies conducted with minimal researcher oversight may better represent and generalize to the real-world context of typical schools.

As implementation strategies facilitated by school psychologists and other support personnel have been shown to improve delivery of evidence-based practices in schools (Merle et al., 2022), identifying teacher implementation profiles may have practical implications for those supporting educators delivering universal SEL programs in real-world conditions. Collier-Meek et al. (2017) discussed the need for feasible and flexible implementation supports to promote integrity of universal program delivery across teachers with varying levels of need. They found brief daily emails containing tips and reminders were sufficient for improving observer-rated adherence and quality for some, but not all, teachers, suggesting that teachers may need to be matched with differentiated implementation support much like student needs are matched with interventions of varying intensity in multi-tiered service delivery systems. For the profiles that emerged in our sample, teachers in the Low IP group may require more individualized support regarding program implementation (e.g., emailed performance feedback, in-person coaching, and/or modeling) than those in the other two groups). Rather than support focused on program implementation, teachers in the High IP group may benefit from guidance in how SEL skill development can be prioritized, integrated, and generalized into their instructional time and interactions with students. Knowledge of differing implementation profiles can ensure that scarce resources like time are used strategically and optimally to support teachers with effective program delivery (Fallon et al., 2018), and that teachers receive targeted support that meets their needs.

5. Limitations and conclusion

There are several limitations to the current study. First, even though there appeared to be high degree of separation among the three identified profiles, the small sample size did not provide sufficient power to detect smaller group differences or examine the differential effects of contextual factors on student outcomes for profiles of classrooms. As such, findings of this study should be viewed as preliminary and interpreted with caution. Additional

future studies with larger and more diverse samples are necessary to verify the number and patterns of implementation profiles in real-world conditions. Second, we only assessed implementation in terms of dosage, adherence, quality, student engagement and teacher's impression of lessons. Based on anecdotal data, some teachers made modifications to the lessons due to a variety of factors (e.g., lack of time, perceived student needs); however, these adaptations are not accounted for in the study. While adaptations may decrease adherence, thoughtful and intentional changes informed by accurate student data to better meet student needs may actually improve student outcomes within the context of a universal program, and this is an important area for future research (Hunter et al., 2022; Neugebauer et al., 2023). Third, as noted previously some of the questions on the teachers' weekly survey may have (unintentionally) been ambiguous or susceptible to social desirability. It is also important to note that the current study was correlational in nature. Besides implementation and contextual factors at the class and school levels, there could be other factors contributing to student outcomes (e.g., student-level demographic and behavioral characteristics) that we did not measure or control. As such, no causal inference should be made without further investigation.

Nonetheless, examining typical practices of teachers when delivering universal SEL and associated contextual factors provides insight regarding the role of aspects of implementation in facilitating program outcomes. Specifically, we examined the role of multiple facets of implementation and how they potentially associated with student's gains (or lack thereof) from the SSIS SEL CIP when implemented under routine conditions in elementary classrooms. Results suggest that considering a single component of implementation (e.g., dosage) is potentially insufficient to account for the overall quality of implementation. Also, given the small variability of teachers' self-report scores, direct observation may provide a more accurate evaluation of implementation quality. Findings also suggest the need to evaluate implementation via multiple dimensions and measures. In addition, results suggest that the relationship between implementation factors and student outcomes may be more nuanced than prior studies featuring individual indicators of program implementation.

Data availability statement

The data analyzed in this study is subject to the following licenses/restrictions: The datasets presented in this article are not readily available because the grant period for the larger effectiveness trial is still ongoing. Requests to access the datasets should be directed to PI of the larger effectiveness trial (jdiperna@psu.edu).

References

- Anyon, Y., Nicotera, N., and Veeh, C. A. (2016). Contextual influences on the implementation of a schoolwide intervention to promote students' social, emotional, and academic learning. *Child. Sch.* 38, 81–88. doi: 10.1093/cs/cdw008
- Asparouhov, T., and Muthén, B. (2014). Auxiliary variables in mixture modeling: three-step approaches using Mplus. *Struct. Equ. Model.* 21, 329–341. doi: 10.1080/10705511.2014.915181
- Bakk, Z., and Vermunt, J. K. (2016). Robustness of stepwise latent class modeling with continuous distal outcomes. *Struct. Equ. Model.* 23, 20–31. doi: 10.1080/10705511.2014.955104
- Beets, M. W., Flay, B. R., Vuchinich, S., Acock, A. C., Li, K. K., and Allred, C. (2008). School climate and teachers' beliefs and attitudes associated with implementation of the

Ethics statement

This study was reviewed and approved by the Institutional Review Board of the Pennsylvania State University. Written informed consent was provided by participating teachers and the parents/legal guardians of participating students. Verbal assent also was provided by participating students.

Author contributions

HZ contributed to the conceptual framework, conducted the statistical analysis, and wrote the draft of the manuscript. P-WL contributed to the conceptual framework and conducted the statistical analysis. SH contributed to the introduction and discussion sections. JD contributed to the conception and revised the manuscript. XL organized and prepared the data set. All authors contributed to the article and approved the submitted version.

Funding

This research was supported, in whole or in part, by the Institute of Education Sciences, US Department of Education, through Grant R305A170047 to The Pennsylvania State University.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Author disclaimer

The opinions expressed are those of the authors and do not represent the views of the Institute or the US Department of Education.

positive action program: a diffusion of innovations model. *Prev. Sci.* 9, 264–275. doi: 10.1007/s11121-008-0100-2

Bolck, A., Croon, M., and Hagenaars, J. (2004). Estimating latent structure models with categorical variables: one-step versus three-step estimators. *Polit. Anal.* 12, 3–27. doi: 10.1093/pan/mp001

Brackett, M. A., Reyes, M. R., Rivers, S. E., Elbertson, N. A., and Salovey, P. (2011). Assessing teachers' beliefs about social and emotional learning. *J. Psychoeduc. Assess.* 30, 219–236. doi: 10.1177/0734282911424879

Carroll, C., Patterson, M., Wood, S., Booth, A., Rick, J., and Balain, S. (2007). A conceptual framework for implementation fidelity. *Implement. Sci.* 2, 1–9. doi: 10.1186/1748-5908-2-40

- Cheng, W., Lei, P. W., and Diperna, J. C. (2021). Using a person-centered analysis to examine second grade outcomes of a universal SEL program. *Sch. Psychol. Forum*. doi: 10.1037/spq0000539
- Chhin, C. S., Taylor, K. A., and Wei, W. S. (2018). Supporting a culture of replication: an examination of education and special education research grants funded by the Institute of Education Sciences. *Educ. Res.* 47, 594–605. doi: 10.3102/0013189X18788047
- Collier-Meek, M. A., Fallon, L. M., and DeFouw, E. R. (2017). Toward feasible implementation support: E-mailed prompts to promote teachers' treatment integrity. *Sch. Psychol. Rev.* 46, 379–394. doi: 10.17105/SPR-2017-0028.V46-4
- Committee for Children (2005). *Steps to Respect: A Bullying Prevention Program*. Seattle, WA.
- Committee for Children. (2012). *Second Step Early Learning*. Seattle, WA.
- Cross, W., West, J., Wyman, P. A., Schmeelk-Cone, K., Xia, Y., Tu, X., et al. (2015). Observational measures of implementer fidelity for a school-based preventive intervention: development, reliability, and validity. *Prev. Sci.* 16, 122–132. doi: 10.1007/s11211-014-0488-9
- Dane, A. V., and Schneider, B. H. (1998). Program integrity in primary and early secondary prevention: are implementation effects out of control? *Clin. Psychol. Rev.* 18, 23–45. doi: 10.1016/S0272-7358(97)00043-3
- DiPerna, J. C., and Elliott, S. N. (2000). *Academic Competence Evaluation Scales*. San Antonio, TX: Psychological Corporation.
- DiPerna, J. C., Lei, P., Bellinger, J., and Cheng, W. (2015). Efficacy of the social skills improvement system Classwide intervention program (SSIS-CIP) primary version. *Sch. Psychol. Q.* 30, 123–141. doi: 10.1037/spq0000079
- DiPerna, J. C., Lei, P., Bellinger, J., and Cheng, W. (2016). Effects of a universal classroom behavior program on student learning. *Psychol. Sch.* 53, 189–203. doi: 10.1002/pits.21891
- DiPerna, J. C., Lei, P., Cheng, W., Hart, S. C., and Bellinger, J. (2018). A cluster randomized trial of the social skills improvement system—Classwide intervention program (SSIS-CIP) in first grade. *J. Educ. Psychol.* 110, 1–16. doi: 10.1037/edu0000191
- Domitrovich, C. E., Bradshaw, C. P., Poduska, J. M., Hoagwood, K., Buckley, J. A., Olin, S., et al. (2008). Maximizing the implementation quality of evidence-based preventive interventions in schools: a conceptual framework. *Adv. School Ment. Health Promot.* 1, 6–28. doi: 10.1080/1754730X.2008.9715730
- Domitrovich, C. E., Gest, S. D., Jones, D., Gill, S., and DeRousie, R. M. S. (2010). Implementation quality: lessons learned in the context of the head start REDI trial. *Early Child. Res. Q.* 25, 284–298. doi: 10.1016/j.ecresq.2010.04.001
- Domitrovich, C. E., Li, Y., Mathis, E. T., and Greenberg, M. T. (2019). Individual and organizational factors associated with teacher self-reported implementation of the PATHS curriculum. *J. Sch. Psychol.* 76, 168–185. doi: 10.1016/j.jsp.2019.07.015
- Durlak, J. A. (2016). Programme implementation in social and emotional learning: basic issues and research findings. *Camb. J. Educ.* 46, 333–345. doi: 10.1080/0305764X.2016.1142504
- Durlak, J. A., and DuPre, E. P. (2008). Implementation matters: a review of research on the influence of implementation on program outcomes and the factors affecting implementation. *Am. J. Community Psychol.* 41, 327–350. doi: 10.1007/s10464-008-9165-0
- Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., and Schellinger, K. B. (2011). The impact of enhancing students' social and emotional learning: a meta-analysis of school-based universal interventions. *Child Dev.* 82, 405–432. doi: 10.1111/j.1467-8624.2010.01564.x
- Elliott, S. N., and Gresham, F. M. (2017). *SSIS SEL Edition Classwide Intervention Program Manual*. Bloomington, MN: Pearson.
- Fallon, L. M., Collier-Meek, M. A., Kurtz, K. D., and DeFouw, E. R. (2018). Emailed implementation supports to promote treatment integrity: comparing the effectiveness and acceptability of prompts and performance feedback. *J. Sch. Psychol.* 68, 113–128. doi: 10.1016/j.jsp.2018.03.001
- Ferguson, S. L., Moore, E. W., and Hull, D. M. (2020). Finding latent groups in observed data: a primer on latent profile analysis in Mplus for applied researchers. *Int. J. Behav. Dev.* 44, 458–468. doi: 10.1177/0165025419881721
- Ferguson, R., and Splaine, M. (2002). Use of the reliable change index to evaluate clinical significance in SF-36 outcomes. *Qual. Life Res.* 11, 509–516. doi: 10.1023/A
- Flay, B. R., Allred, C. G., and Ordway, N. (2001). Effects of the positive action program on achievement and discipline: two matched-control comparisons. *Prev. Sci.* 2, 71–89. doi: 10.1023/A:1011591613728
- Greenberg, M. T., Domitrovich, C. E., Weissberg, R. P., and Durlak, J. A. (2017). Social and emotional learning as a public health approach to education. *Futur. Child.* 27, 13–32. doi: 10.1353/foc.2017.0001
- Greenberg, M. T., and Kusche, C. A. (1993). *Promoting Social and Emotional Skills in Deaf Children: The PATHS Project*. Seattle, WA, US: University of Washington Press.
- Greenberg, M. T., Kusche, C. A., Cook, E. T., and Quamma, J. P. (1995). Promoting emotional competence in school-aged children: the effects of the PATHS curriculum. *Dev. Psychopathol.* 7, 117–136. doi: 10.1017/S0954579400006374
- Gresham, F., and Elliott, S. N. (2008). *Social Skills Improvement System Rating Scales*. Minneapolis, MN: Pearson Assessments.
- Hart, S. (2021). Beliefs about social-emotional skills: Development of the assumptions supporting social-emotional teaching (ASSET) scale [Unpublished doctoral dissertation]. The Pennsylvania State University.
- Hennessey, A., and Humphrey, N. (2020). Can social and emotional learning improve children's academic progress? Findings from a randomised controlled trial of the promoting alternative thinking strategies (PATHS) curriculum. *Eur. J. Psychol. Educ.* 35, 751–774. doi: 10.1007/s10212-019-00452-6
- Holden, R. R., and Passey, J. (2009). "Social desirability" in *Handbook of individual differences in social behavior*. eds. M. R. Leary and R. H. Hoyle (New York: The Guilford Press), 441–454.
- Humphrey, N., Barlow, A., and Lendrum, A. (2018). Quality matters: implementation moderates student outcomes in the PATHS curriculum. *Prev. Sci.* 19, 197–208. doi: 10.1007/s11211-017-0802-4
- Hunter, L. J., DiPerna, J. C., Hart, S. C., Neugebauer, S., and Lei, P. (2022). Examining teacher approaches to implementation of a classwide SEL program. *School. Psychology* 37, 285–297. Advance online publication. doi: 10.1037/spq0000502
- Jacobson, N. S., and Truax, P. (1991). Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *J. Consult. Clin. Psychol.* 59, 12–19. doi: 10.1037/0022-006X.59.1.12
- Jones, S. M., Farrington, C. A., Jagers, R., Brackett, M., and Kahn, J. (2019). National Commission on social, emotional, and academic development: a research agenda for the next generation. *The Aspen Institute*. Available online at: <https://www.aspeninstitute.org/publications/a-research-agenda-for-the-next-generation/> (accessed March 13, 2023).
- Laursen, B., and Hoff, E. (2006). Person-centered and variable-centered approaches to longitudinal data. *Merrill-Palmer Q.* 52, 377–389. doi: 10.1353/mpq.2006.0029
- Lawson, G. M., McKenzie, M. E., Becker, K. D., Selby, L., and Hoover, S. A. (2019). The core components of evidence-based social emotional learning programs. *Prev. Sci.* 20, 457–467. doi: 10.1007/s11211-018-0953-y
- Low, S., Smolkowski, K., and Cook, C. (2016). What constitutes high-quality implementation of SEL programs? A latent class analysis of second step[®] implementation. *Prev. Sci.* 17, 981–991. doi: 10.1007/s11211-016-0670-3
- Low, S., Van Ryzin, M. J., Brown, E. C., Smith, B. H., and Haggerty, K. P. (2014). Engagement matters: lessons from assessing classroom implementation of steps to respect: a bullying prevention program over a one-year period. *Prev. Sci.* 15, 165–176. doi: 10.1007/s11211-012-0359-1
- Magidson, J., and Vermut, J. K. (2005). *Technical Guide for Latent GOLD 4.0: Basic and Advanced*. Belmont, MA: Statistical Innovations Inc.
- Merle, J. L., Thayer, A. J., Larson, M. F., Pauling, S., Cook, C. R., Rios, J. A., et al. (2022). Investigating strategies to increase general education teachers' adherence to evidence-based social-emotional behavior practices: a meta-analysis of the single-case literature. *J. Sch. Psychol.* 91, 1–26. doi: 10.1016/j.jsp.2021.11.005
- Muthén, B., and Muthén, L. K. (2000). Integrating person-centered and variable-centered analyses: growth mixture modeling with latent trajectory classes. *Alcohol. Clin. Exp. Res.* 24, 882–891. doi: 10.1111/j.1530-0277.2000.tb02070.x
- Neugebauer, S. R., Sandilos, L. E., DiPerna, J. C., Hunter, L. J., and Hart, S. C. (2023). 41 teachers, 41 different ways: exploring teacher implementation of a universal social-emotional learning program under routine conditions. *Elem. Sch. J.*
- O'Donnell, C. L. (2008). Defining, conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in K-12 curriculum intervention research. *Rev. Educ. Res.* 78, 33–84. doi: 10.3102/0034654307313793
- Pianta, R. C., La Paro, K. M., and Hamre, B. K. (2008). *Classroom Assessment Scoring System™: Manual K-3* Baltimore, MD: Paul H Brookes Publishing.
- Proctor, E., Silmere, H., Raghavan, R., Hovmand, P., Aarons, G., Bunger, A., et al. (2011). Outcomes for implementation research: conceptual distinctions, measurement challenges, and research agenda. *Adm. Policy Ment. Health Serv. Res.* 38, 65–76. doi: 10.1007/s10488-010-0319-7
- Raferty, A. E. (1995). Bayesian model selection in social research. *Sociol. Methodol.* 25, 111–163. doi: 10.2307/271063
- Reyes, M. R., Brackett, M. A., Rivers, S. E., Elbertson, N. A., and Salovey, P. (2012). The interaction effects of program training, dosage, and implementation quality on targeted student outcomes for the RULER approach to social and emotional learning. *Sch. Psychol. Rev.* 41, 82–99. doi: 10.1080/02796015.2012.12087377
- Rojas-Andrade, R., and Bahamondes, L. L. (2019). Is implementation fidelity important? A systematic review on school-based mental health programs. *Contemporary school. Psychology* 23, 339–350. doi: 10.1007/s40688-018-0175-0
- Sanetti, L. M., and Collier-Meek, M. A. (2019). Increasing implementation science literacy to address the research-to-practice gap in school psychology. *J. Sch. Psychol.* 76, 33–47. doi: 10.1016/j.jsp.2019.07.008
- Smith, P. E., Witherspoon, D. P., and Lei, P. W. (2021). The "haves, have some, and have Nots": a latent profile analysis of capacity, quality, and implementation in community-based afterschool programs. *Prev. Sci.* 22, 971–985. doi: 10.1007/s11211-021-01258-z

Spurk, D., Hirschi, A., Wang, M., Valero, D., and Kauffeld, S. (2020). Latent profile analysis: a review and “how to” guide of its application within vocational behavior research. *J. Vocat. Behav.* 120:103445. doi: 10.1016/j.jvb.2020.103445

Taylor, R. D., Oberle, E., Durlak, J. A., and Weissberg, R. P. (2017). Promoting positive youth development through school-based social and emotional learning interventions: a meta-analysis of follow-up effects. *Child Dev.* 88, 1156–1171. doi: 10.1111/cdev.12864

Tein, J. Y., Coxe, S., and Cham, H. (2013). Statistical power to detect the correct number of classes in latent profile analysis. *Struct. Equ. Model.* 20, 640–657. doi: 10.1080/10705511.2013.824781

Wigelsworth, M., Verity, L., Mason, C., Qualter, P., and Humphrey, N. (2021). Social and emotional learning in primary schools: a review of the current state of evidence. *Br. J. Educ. Psychol.* 92, 898–924. doi: 10.1111/bjep.12480