12-20-2022

Extensions Beyond Program impacts: Conceptual and Methodological Considerations in studying the Implementation of a preschool social emotional Learning Program

Jaiya Rae Choles
Portland State University, jaiya.choles@gmail.com

Robert W. Roeser
The Pennsylvania State University

Andrew Mashburn
Portland State University, mashburn@pdx.edu

Citation Details
Choles, J. R., Roeser, R. W., & Mashburn, A. J. (2022, December). Extensions beyond program impacts: Conceptual and methodological considerations in studying the implementation of a preschool social emotional learning program. In Frontiers in Education (Vol. 7, p. 1035730). Frontiers Media SA.

frontiers | Frontiers in Education

Check for updates

Extensions beyond program impacts: Conceptual and methodological considerations in studying the implementation of a preschool social emotional learning program

Jaiya R. Choles[1]*, Robert W. Roeser[2] and Andrew J. Mashburn[1]

[1]Department of Psychology, Portland State University, Portland, OR, United States, [2]Department of Human Development and Family Studies, Pennsylvania State University, State College, PA, United States

Social−emotional learning (SEL) programs are frequently evaluated using randomized controlled trial (RCT) methodology as a means to assess program impacts. What is often missing in RCT studies is a robust parallel investigation of the multi-level implementation of the program. The field of implementation science bridges the gap between the RCT framework and understanding program impacts through the systematic data collection of program implementation components (e.g., adherence, quality, responsiveness). Data collected for these purposes can be used to answer questions regarding program impacts that matter to policy makers and practitioners in the field (e.g., Will the program work in practice? Under what conditions? For whom and why?). As such, the primary goal of this paper is to highlight the importance of studying implementation in the context of education RCTs, by sharing one example of a conceptualization and related set of implementation measures we created for a current study of ours testing the impacts of a SEL program for preschool children. Specifically, we describe the process we used to develop an implementation conceptual framework that highlights the importance of studying implementation at two levels: (1) the program implementation supports for teachers, and (2) teacher implementation of the curriculum in the classroom with students. We then discuss how we can use such multi-level implementation data to extend our understanding of program impacts to answer questions such as: "Why did the program work (or not work) to produce impacts?"; "What are the core components of the program?"; and "How can we improve the program in future implementations?"

Introduction

The recent wide-scale expansion of social emotional learning (SEL) programs in schools and classrooms has been informed by results of research studies that demonstrate, across age groups, SEL programs have positive impacts on students' academic success and well-being. More specifically, SEL programs have been found to produce demonstrably positive impacts on students' social and emotional skills (e.g., perspective taking, identifying emotions, interpersonal problem solving); attitudes toward self and others (e.g., self-esteem, self-efficacy); positive social behaviors (e.g., collaboration, cooperation); reduced conduct problems (e.g., class disruption, aggression); reduced emotional distress (e.g., depression and anxiety); and academic performance (e.g., standardized math and reading; Durlak et al., 2011). Positive impacts of SEL programs are also evident among preschool aged children; results from a meta-analysis of 39 SEL programs in early childhood education settings found small to medium effects (Hedge's g effect size estimates between 0.31 and 0.42) for improvements in children's social and emotional competencies, and reductions in their challenging behaviors (Luo et al., 2022). Results of studies of SEL program impacts on children's social and academic outcomes have gone on to be included in economic studies, and a recent estimate of a $11 return for every $1 invested in school-based SEL programs have compelled policy makers and program administrators nationwide to implement these programs at a wide scale (Belfield et al., 2015).

To date, studies of the impacts of SEL programs have mostly used field-based experimental designs (Boruch et al., 2002; Bickman and Reich, 2015). When applied to education contexts, these experimentally designed studies are commonly referred to as randomized controlled trials, or cluster randomized trials in the cases when the design considers the multiple levels of analyses that are familiar in school-based settings (e.g., classrooms clustered within schools, and children clustered within classrooms). The primary research question that is addressed by the experimental design is "What are the impacts of offering access to an SEL program on students' development of SEL competencies and well-being?." The experimental study then proceeds by randomly assigning schools, classrooms, or children (depending on whether the program is intended to be delivered school-wide, classroom-wide or to individual children) to either the "treatment" group that is offered the SEL program or to a comparison group that either may receive a different program (active control) or carry on with business as usual (control) and may receive the treatment at a later time (waitlist control).

The methodological premise of random assignment is that each school, classroom, or child has an equal chance of being assigned to the SEL program or the comparison group, and as such, there are no expected differences between these groups at the start of the study on any measurable or unmeasurable characteristics. Thus, any differences in outcomes between these two groups after the program concludes can be attributed to the one key difference between these groups—the SEL program itself

(e.g., Shadish et al., 2002). Given these methodological strengths, experimentally designed studies, when well implemented, may be more likely to achieve a label of "evidence-based," which is weighted heavily by policy makers and program administrators when considering which programs should be funded for expansion and wide scale use (Boruch, 2005; Mark and Lenz-Watson, 2011).

Despite these important strengths of the experimental design with regard to internal validity of the conclusions about program impacts, there are some notable weaknesses of the experimental design. Such weaknesses are related to the difficulty implementing field-based experimental studies in real world settings. For instance, the external validity of the results of an impact study—to whom the results of the study generalize—often is not clear. In addition, in the experimental study, the program is implemented under "ideal conditions" which may not reflect the actual conditions in which the program is implemented in the real world (e.g., more resources and/or implementation supports). And, despite using random assignment, there may be differences between the treatment and comparison groups at the start of the study, which would weaken the inference that the intervention caused the impacts.

Perhaps most notably, results of experimental studies leave a lot to be desired for researchers and program implementers who ask different questions about SEL programs that extend beyond the question about program impacts that the experimental study is primarily designed to address. This limitation of the experimental design in field-based settings (such as schools) becomes evident in studies where group assignment (treatment or comparison) does not reflect the experiences of those assigned to the condition. That is, random assignment determines that a school, classroom, or child was offered *access to the SEL program*. However, not all schools, classrooms or children that are offered access to the program will actually *participate in the program* at all, or in the ways that the program developers intended. In addition, those schools, classrooms, or children assigned to the comparison group may actually implement components of the SEL program that are intended to be accessed only by the treatment group. As such, the results of an experimental study provide estimated impacts of random assignment to the SEL program without regard to the actual experiences of those who participated in either study condition (Hollis and Campbell, 1999). Thus, this methodology ignores the often rich and meaningful variation in how the program was implemented and what the experiences were of those who participated in the treatment and comparison conditions, and how closely aligned these experiences were to how the program was intended to be implemented by those who developed it.

In sum, impact studies of SEL programs answer important policy questions, but they are limited in their capacity to answer questions that are relevant for education practitioners and non-policy-oriented research. We argue here, as others have (Fixsen and Blase, 2009; Moir, 2018) that studying the implementation of a program in the context of an impact study

creates an ideal context to understand program impacts and subsequently address many other important questions of interest to program developers and practitioners who directly implement these programs.

The field of implementation science offers rich frameworks for researchers to draw on that examine how variation in how the program was implemented is associated with the program's impacts. Carefully designed implementation studies can provide critical contextual information that helps researchers feel confident in answering the questions about impacts thoroughly, and they offer the opportunity to extend research questions to explore active ingredients in the program, explain why an intervention was not effective, and guide efforts to modify interventions that maximize their effectiveness in the future (Durlak and DuPre, 2008). For example, Low et al. (2016) used latent class analysis to study teachers' implementation within a RCT study of the Second Step® SEL curriculum in elementary schools. The authors incorporated multiple implementation measures into their latent structure specification, including adherence, dosage, generalization (i.e., application/integration in the classroom), and student engagement. The authors then used the determined latent structure to predict both teacher report and observed scores of student behavioral and academic outcomes and found a negative relationship between the low engagement latent class and student outcomes, when compared to the low adherence and high quality implementation latent classes.

Other studies have examined moderation of program impacts with factors that may influence implementation. For example, McCormick et al. (2016) examined the role of parent participation in moderating program impacts of the INSIGHTS into Children's Temperament program in the subset of kindergarten and first grade participants of a larger experimental study. Results indicated program impacts for students (reading, math, and adaptive behaviors) were stronger for children of parents categorized in the low participation group. Similarly, Sandilos et al. (2022) explored whether implementing the Social Skills Improvement System SEL Classwide Intervention Program buffered the negative effect of low teacher well-being on the quality of teacher-child interactions in the classroom.

Despite this small but growing body of evidence in SEL programs, implementation is still often not studied in the context of testing impacts of SEL programs (Domitrovich et al., 2012) and mindfulness-based SEL programs (Roeser et al., 2020). Furthermore, these studies of implementation tend not to address the "complex" structure of the intervention design in education settings. That is, the focus of program implementation of SEL programs has been on the teachers (as intervention agents), rather than also considering whether the implementation of the training and supports offered to teachers (by implementation agents) have been effective (Fixsen et al., 2005); and, little consideration is given to the multitude of coordinating pieces required to implement this multi-level structure across diverse settings (Bryk, 2016). Additionally, with the increase in popularity of SEL curricula,

monitoring control group practices (i.e., describing business-as-usual practices) has become increasingly important.

In the following sections, we provide a conceptual framework for conducting an implementation study within the context of an experimental study of program impacts. We then apply this framework to a mindfulness-based SEL program for preschoolers and describe our process of creating a robust set of measures of program implementation that we include in our impact study. We then demonstrate how this framework and measures can be used to address interesting and important questions about the SEL program that extend beyond the question of program impacts, such as, "Why did the program work (or not work) to produce impacts?"; "What are the core components of the program?"; and "How can we improve the program in future implementations?." This paper will explore these questions using examples of our measures and offer suggestions to others implementing studies like ours.

Implementation study design and measurement framework

In this section, we highlight the importance of attending to the interconnections between an impact study and a study of program implementation. To do this, we provide an example of our own work assessing program implementation of a preschool mindfulness-based SEL program. We begin by briefly describing the SEL program and the design of the study testing the impacts of the program on children's outcomes. Then, we provide an overview of specific aspects of fidelity of implementation (FOI) that are commonly considered in implementation studies. Lastly, we present the key components of program implementation that are the focus of our implementation study and describe how they are incorporated into our multi-level FOI conceptual framework. In this last section, we also discuss the process of mapping measures onto our conceptual framework, piloting those measures, refining them, and integrating them into the design of our impact study.

Social emotional learning program description

The mindfulness-based SEL program for preschoolers under investigation in this study (MindUP™ PreK—The Goldie Hawn Foundation) consists of four main elements: the mindfulness-based SEL preschool classroom curriculum for students, the curriculum training for teachers, and two additional implementation supports for teachers—monthly community of practice meetings and coaching sessions. A cluster randomized trial testing the impacts of the MindUP™ program on children's social, emotional, and academic outcomes is the context for this implementation study. Specifically, the theory of change of this study hypothesizes that through implementation of the MindUP™

program, preschool children will develop key social and emotional skills (i.e., attentional, social, and emotional) and academic skills (i.e., early literacy and math). We also hypothesize that the impact of MindUP™ may be stronger for children who begin preschool with fewer SEL skills and/or who are in classrooms where students experience lower quality interactions with their teachers. Exploratory follow-up to these impacts will also examine which aspects of implementation of the program are positively associated with children's development of SEL and academic outcomes and whether aspects of FOI vary based on teacher and classroom characteristics measured at baseline.

The impact study consists of three sequential and independent cohorts of preschool classrooms that will be randomly assigned to either participate in the MindUP™ program or in a waitlist control group for one year. We began the MindUP™ trial in fall 2019 and successfully recruited our first cohort of 38 teachers from a range of preschool programs (e.g., private for profit, community-based organizations, Head Start) serving four-year-old children. These teachers were randomized and half were offered access to the MindUP™ training program and supports. However, the trial was interrupted in Spring of 2020 due to the COVID-19 pandemic and was then paused for two subsequent years. In fall of 2022, the trial was re-started with a new first cohort of teachers and minor adaptations to the implementation of the program (i.e., remote training rather than face-to-face training) and the research protocol (i.e., reduced in-person assessments for children).

Mindfulness-based social emotional learning curriculum

The MindUP™ curriculum for preschool students consists of 15 mindfulness and SEL-based lessons, and a daily core practice called the "brain break." Each weekly lesson comprises several related activities that are estimated to take 15 minutes to complete. For this study, teachers were instructed to implement two activities within each lesson per week in their classroom. Additionally, the curriculum included supplementary activities and instructions on ways to integrate the lessons into other classroom experiences. The 15 curriculum lessons are organized into three units. The first unit is *Mindful Me*, in which children learn about the structure and function of the brain and are introduced to the brain break. An example of an activity from this unit is *My Feelings,* in which children build emotion literacy skills by learning to name different emotions and identify how they feel when they experience them. The second unit is *Mindful Senses*, in which children focus on the relationship between their senses, their bodies, and how they think. An example activity from the second unit is *Mindful Touch*, in which children are instructed to be open and curious about touching mystery objects hidden in closed containers. Removing sight from the activity allows children to investigate the objects using only their tactile sense. The third unit is *Mindful Me in the World*, where children learn about mindsets, such as gratitude and perspective-taking, and how to apply mindful behaviors through interactions with the community and world. An example activity from unit three is *The Gratitude Tree*, in which students are

instructed to draw something or someone they are grateful for on a paper shaped like a leaf and then the children's work is displayed on a tree visual in the classroom.

The core practice of the MindUP™ curriculum is called the "brain break"—a brief focused attention activity that teachers implement in their classroom with their students four times daily—usually at the start of the day, after recess, after lunch, and at the end of the day. The brain break is initiated with a chime sound that children focus on and listen to in order to settle into their bodies. Children are then instructed to focus on the natural rhythm of their breath. For preschoolers, the brain break initially requires various scaffolds to help focus attention, which can include, for example, a "breathing ball" (e.g., Hoberman Sphere) that expands and contracts to simulate the inhale and exhale of breathing; or the placement of stuffed animals on the diaphragm to help children focus on the rising and falling of their breath. After repeating these mindful breaths two to three more times, the chime is rung a second time to conclude the brain break. Children are instructed to listen for as long as they can hear the chime and then provided time to bring their awareness back to the classroom. Conceptually, the brain break can be considered a focused attention practice (Maloney et al., 2016).

MindUP™ curriculum training

Before implementing the MindUP™ curriculum in the classroom, teachers attend a single day six-hour curriculum training led by a certified trainer (in this study, so named the implementation director). During the training, teachers learn foundational scientific research underpinning the curriculum, review the curriculum book, discuss implementation with other attendees, and practice lesson planning. During this training, teachers are also given a comprehensive materials kit to fully implement the curriculum in their classroom.

Community of practice

Based on our previous research on MindUP™ in the early years (Braun et al., 2018), we developed a new set of implementation supports for the purpose of this study. This included a community of practice for all teachers implementing the MindUP™ curriculum in their classrooms (see Mac Donald and Shirley, 2009). The community of practice component consists of monthly hour-long, face-to-face, small group meetings between participating teachers and the implementation director. At the start of each meeting, teachers are led through a mindfulness activity as a way to center the group and to provide teachers with their own opportunities to focus their attention. These meetings are facilitated by the implementation director and are used as time for teachers to discuss their progress in implementing the curriculum. More specifically, teachers are asked to reflect and share with the group regarding what went well or what was challenging implementing the curriculum activities since the last community of practice meeting. At this time, the implementation director, and other teachers, can offer feedback to support implementation improvement in the future. During these

meetings teachers are also provided with the opportunity to discuss their plan for upcoming lessons and to discuss as a group the purpose of the upcoming curriculum activities. During this segment of the meeting, teachers can also ask the implementation director for support in how to successfully implement the upcoming activities, or to address any aspects of the upcoming implementation that remain unclear.

Coaching

In addition to the community of practice, we also included a coaching program as an additional implementation support. Developed by our implementation director, the coaching program comprises monthly 30-min one-on-one check-in calls between the implementation director and each teacher participating in the MindUP™ program. The coaching model used was adapted from the National Center on Quality Teaching and Learning's *Practice Based Coaching* framework (for example see Snyder et al., 2015) to specifically align with the MindUP™ curriculum. Prior to each check-in call, the implementation director would review the teacher's most recent implementation log information to understand the teacher's implementation progress. During the coaching sessions, the implementation director and each teacher would discuss specific challenges or questions related to their curriculum implementation and brainstorm strategies for improving implementation in the classroom. The implementation director's agenda for each of these coaching sessions drew on the *Practice Based Coaching* framework and included time for discussing shared goals and action planning, teacher self-monitoring, and reflection and feedback related to the program.

Implementation of the MindUP™ program

In this section, we define components of FOI and apply them to the MindUP™ SEL program in particular, to create a conceptual framework for assessing fidelity in our study of the implementation of MindUP™. In our effort to develop a conceptual and assessment strategy for studying FOI in the context of this program, we drew on extant work on FOI. Broadly, implementation can be defined as the study of a program and its components, and how it is delivered in a specific context to optimize program outcomes (Durlak and DuPre, 2008). Variation exists among aspects of FOI terminology, however, common components often include the following: (1) *Dosage*, which describes the strength or quantity (in hours, sessions, etc.) of the program; (2) *Responsiveness*, which consists of the extent to which the program is engaging, interesting, and relevant to participants; (3) *Adherence*, which measures the extent to which the program is implemented as designed or planned; (4) *Quality* describes how well program components are delivered (e.g., clarity, organization); and, (5) *Differentiation* examines the degree to which the program under investigation is similar or different to others like it (Dane and Schneider, 1998; Durlak and DuPre, 2008). More recently, the field of implementation science has identified additional FOI components that should be measured, including (6) Program *adaptations*, which capture modifications (changes, omissions, additions) made to program components; (7) *Program reach*, which describes the generalizability or representativeness of program participants to the broader population of interest for the program; and, more specifically to experimental studies, (8) *Monitoring control group practices*, which seeks to measure the extent to which intervention activities or "intervention-like" activities are conducted by the comparison group (Durlak and DuPre, 2008).

Together, measures of FOI components help to capture the multiple elements of implementation that factor into program success; for instance, *dosage* data that measures the frequency with which participants attended program sessions can be used to calculate the percentage of the total program each participant actually received (i.e., their individual "dose"). However, it is often insufficient to simply measure attendance—research shows participants learn when content is engaging, relevant, and interesting—so measuring *responsiveness* becomes imperative to contextualize dosage. From the implementation support side, *adherence* and *quality* FOI components are similarly interrelated: *adherence* measures can assess the degree to which the program was implemented as planned, but in order to evaluate whether the program was engaging and relevant (i.e., participant responsive), assessing if the program components were delivered with a high level of *quality* is needed as well. Finally, *adaptations* are important to measure due to their potential influence on each of the other components described above. Adaptations may impact implementation in a complex way, such that they could increase the quality of delivery and therefore participant responsiveness, while simultaneously reducing adherence to the program as prescribed. Moving outside of the group receiving the intervention, *monitoring control group practices* is also important in order to fully understand the impact of an intervention in impact studies. If this FOI component is not considered, the true impact of the program on outcomes cannot be determined if the experiences of the control group are not known (Fixsen et al., 2005). In sum, each of these FOI components affect the conclusions that can be drawn about the impacts of the SEL program (Durlak and DuPre, 2008).

Other FOI components—differentiation and program reach—concern the external validity of results of a program and are particularly important to evaluate when comparing across different "evidence-based" programs (Durlak and DuPre, 2008). For instance, *differentiation* data can help answer questions that seek to determine why a specific program should be chosen over another (what makes a program unique?). Clearly outlined information regarding dosage, program components, and resources needed, as well as their associated costs, can help define a program's uniqueness and help determine which may be best used for a particular purpose or community. Similarly, paying attention to *program reach* can help program adopters estimate the extent previous data regarding a program's effectiveness will

generalize when implemented in communities the program was developed for (Durlak and DuPre, 2008). Without this broader lens, researchers remain in the dark on the extent to which a program will be successful when implemented outside of a highly controlled "ideal conditions" scenario of an impacts study.

In education interventions, measuring each of these aspects of implementation becomes increasingly complex due to the fact that program components are diverse and span across multiple levels (e.g., classrooms, teacher professional development, whole schools). In addition, the coordination of these components within and across levels, in a manner that is responsive to a wide range of local contextual and organizational conditions, is key to successful program implementation. Due to this multi-level, ecological complexity surrounding program implementation, Bryk (2016) describes FOI in education contexts as "adaptive integration" to emphasize that as much as program implementation involves adherence and compliance, it also fundamentally involves responsivity and adaptation across levels. Thus, although we adopt the term FOI here, we use this terminology as a broader term that acknowledges and incorporates this systems perspective of "adaptive integration" into our work.

Below we present our implementation study conceptual model in Figure 1, which includes the FOI components we believe are most important to measure during the implementation study that is part of an impacts study (see Bywater, 2012). However, it is important to note that there are numerous other factors that must be considered and coordinated prior to this stage that also influence implementation success. For example, researchers must obtain buy-in from interested parties and ensure schools, or in this case preschool centers, are ready and receptive to implementation

and change through program adoption (Fixsen et al., 2005; Moir, 2018). It is also important to ensure motivations are shared between leadership (often the gatekeepers for what programs are considered) and teachers (who often implement the program with children in classrooms; Fixsen et al., 2005). Intervention programs should also be piloted prior to testing their efficacy to determine if the program is accepted by participants and feasible, and to identify any areas that need to be improved for future use (Bywater, 2012; Bryk, 2016). Similarly, aspects of FOI focused on the external validity and generalizability of programs are connected to this work, but remain outside the focus of this paper.

In our implementation study, we believe it is very important to attend to the multi-level nature of the program under study (Fixsen et al., 2005). Specifically, the first level of the SEL program focuses on measuring the transfer of program knowledge from the project implementation team members (in our study, the implementation director) to participating teachers who will be implementing the curriculum in their classrooms with their students. As seen in our conceptual framework displayed in Figure 1, we call this level of measurement "teacher implementation supports." Specifically, this level comprises measuring FOI of the MindUP™ curriculum training, community of practice meetings and coaching sessions.

The second and more frequently considered level of implementation is what we refer to as "curriculum implementation"—teachers' implementation (as intervention agents) of the curriculum with students in the classroom. In our study, this level includes measurement of the program curriculum activities as well as the daily brain break practice. We also posit that it is important to monitor these two levels in the control

FIGURE 1
Conceptual framework for fidelity of implementation of MindUP™.
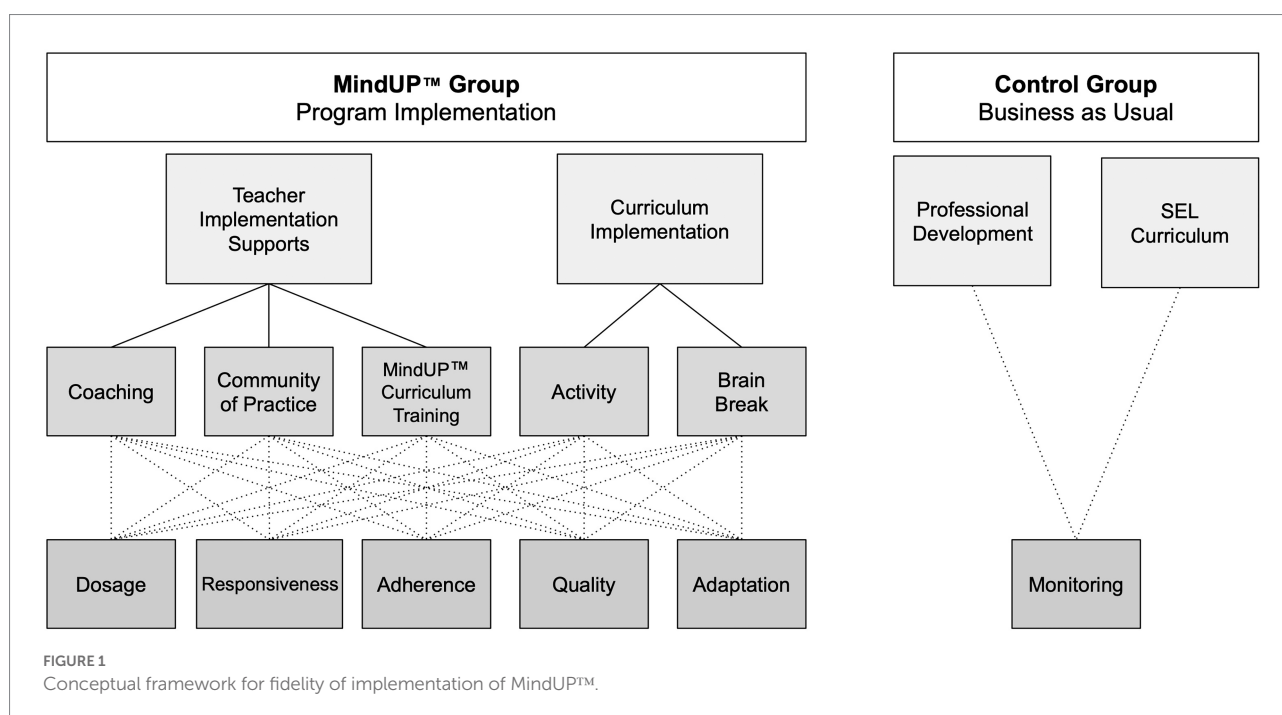
group as well. To operationalize similar activities that may take place as part of business-as-usual practices for the control group, we define these two levels more broadly as professional development for teachers and SEL activities for students for this group. The right side of Figure 1 visualizes this component of our framework and highlights the importance of attending to this aspect of FOI in the context of an impacts study.

The lower half of our conceptual model in Figure 1 identifies the comprehensive set of FOI components of the program we sought to measure in our implementation study. Specifically, we chose to measure dosage, responsiveness, adherence, quality, and adaptations for each of the teacher implementation supports (i.e., curriculum training, community of practice, and coaching) and each curriculum implementation component (curriculum activities and brain break). Similarly, we defined the FOI measurement at this level for the control group as "monitoring" of both control group teachers' professional development activities and SEL curricular activities with their students. In Table 1,

we provide a glossary that defines all of these terms. Next, we describe in detail how we compiled and developed measures, and when we determined it essential to have measures from multiple informants.

Mapping measures onto the implementation conceptual framework

With a clear conceptualization of FOI in general, and of the MindUP™ program in particular, we then began to develop specific measures of FOI that map onto our conceptualization. We began this mapping process by stating two explicit goals for our set of FOI measures: first, the set of measures should be comprehensive, such that there is at least one measure for each FOI component outlined in the framework for each program element within each program level. Secondly, that when feasible, we believe there should be more than one measure for an FOI element, and that these multiple measures should be from different reporters or sources (i.e., multi-informant). Table 2 presents the results of this process by summarizing the informant or informants for each FOI component measured. Specifically, it highlights whether the informant was the implementation agent (i.e., implementation director), the classroom teacher, a researcher who is a third-party observer, or a combination therein. In our study, it did not seem developmentally appropriate to have preschoolers report on their experiences participating in the curriculum implementation, however, older students can be included as an additional informant source when applicable.

As mentioned above, in order to develop a robust FOI measure set, when feasible and appropriate, we sought triangulation of perspectives and experiences through multi-informant measures (see Table 2). Thus, our study design included third-person observers, with a member of our research group attending and observing the MindUP™ curriculum training and community of practice sessions. For these program elements among those in the "treatment" group, the observer comprehensively rated all FOI components defined in our conceptual framework (i.e., dosage, responsiveness, adherence, quality, and adaptation). We supplemented this complete set of observer ratings with teacher self-report ratings of those FOI components that require the first-person experience of the participant (i.e., responsiveness, quality). We did not feel it was appropriate for a third-party observer to rate the one-on-one coaching sessions, so instead, for this program element, we had the implementation director provide ratings that measure dosage, responsiveness, adherence, and adaptation. Again, we supplemented these measures with responsiveness and quality ratings from the teacher.

For the curriculum implementation level, teachers were the primary source who reported on the complete set of FOI measures (except for ratings of quality) for their implementation in the classroom. This required teachers to fill out an implementation log on a weekly basis as they implemented each program activity. In

TABLE 1 Definitions of terms.

Construct	Definition
Program	Synonymous with intervention
Fidelity of implementation (FOI)	General term used to define the systematic study of implementation, includes complex interventions that have been defined as "adaptive integrations"[1]
FOI component	Aspects of implementation that can be measured, such as dosage, responsiveness, adherence, quality, or adaptations
Implementation level	Label to differentiate implementation supports to teachers by an implementation agent (first level) and curriculum implementation by the teacher to students in the classroom (second level)
Program element	Factors that make up the program as a whole, including both implementation supports (curriculum training, community of practice, and coaching) and curriculum implementation (curriculum activities and brain break practice)
Implementation agent	A person who provides professional development to the teacher to support their implementation of the program with students in the classroom
Intervention agent	The teacher or other person who delivers the program in the classroom to students

[1] Bryk (2016).

TABLE 2 Example of multi-informant measures of fidelity of implementation components across levels and groups.

| FOI component | Implementation supports (implementor to teacher) | | | | Curriculum implementation (teacher to student) | | |
	Curriculum training	Community of practice	Coaching	Control monitoring (PD)	Curriculum activity	Curriculum brain break	Control monitoring (SEL)
Dosage	OR	OR	IR	TR	OR; TR	OR; TR	TR
Responsiveness	OR; TR	OR; TR	IR; TR	TR	OR; TR	OR; TR	TR
Adherence	OR	OR	IR		OR; TR	OR; TR	
Quality	OR; TR	OR; TR	TR		OR	OR	
Adaptation	OR	OR	IR		OR; TR	OR; TR	

OR, Observer Report; TR, Teacher Report; IR, Implementor Report; PD, Professional development; SEL, Social emotional learning curriculum.

addition, teachers provided weekly data on the daily brain break practice as well. For triangulation of information at this level, researchers also comprehensively rated all FOI components measured in the "treatment" group, for both a curriculum activity and brain break practice, by visiting teachers in their classrooms on two separate observation occasions during the implementation phase. During these observation sessions, the researchers recorded the number of students who were present for the observation session (as a collective measure of dosage) for both the curriculum activity and the brain break. The observer then rated adherence, adaptations, student engagement, and total minutes spent conducting the activity (a specific measure of dosage) separately for the curriculum activity and brain break (see Figure 2 for an example of how these FOI components were scored by the observer during the Brain Break implementation). After observing both program components, the researcher scored the teacher's quality of delivery of both the curriculum activity and brain break (as a collective measure of quality).

Generally, across both levels of implementation data, the desire for a high degree of adherence to intervention elements needs to be balanced with flexibility and adaptation of the program to meet local school, community, and student needs. Known as the adherence/adaptation trade off, researchers must balance the desire for internal validity with an understanding that achieving high-quality implementation often requires the implementation agent and intervention agents to be afforded flexibility in meeting the needs of participants and contextual demands (Dane and Schneider, 1998). To contend with this issue, Fixsen et al. (2005) recommend requiring adherence to main intervention principles but allowing flexibility in how the principles are implemented (i.e., processes, strategies) in a manner that retains the objective or function of the component.

In our own work, we attempted to produce this balance when developing our measures of adherence and adaptation. First, the way the MindUP™ curriculum was developed helped us with this balance. The curriculum allows for flexibility in the implementation of the brain break practice, for instance, by including multiple scaffolded ways teachers can implement the brain break in the classroom. The manual also discusses the need for flexibility regarding the context in which the practice is implemented, such as the location/time of day (e.g., circle time) and structure (whole group or small group). This flexibility directly translates into the classroom implementation adherence measure we developed, where we capture teachers' use of different scaffolds when offering the brain break (e.g., Hoberman sphere, stuffed animals; see the "extra materials utilized" section of Figure 2) as well as the presence/absence of (what we believe are) core components and the extent to which the component's objective was met (see Jennings Brown et al., 2017; Doyle et al., 2019). For example, Figure 2's section on adherence depicts adherence expectations for teachers during the brain break practice, in which each element is considered essential for full implementation of this practice. Given teachers' diverse expertise and our desire to leave space for developmentally appropriate practice, we believe measuring adherence in multiple ways (e.g., presence/absence and degree objective was met) will capture more nuance and variability across adherence in classroom implementation. Additionally, we included space to qualitatively describe any adaptations to the brain break practice that move outside the scope of flexible options provided within the program.

To monitor control group practices, control group teachers were asked to describe their professional development and SEL curriculum activities as part of their spring (end of program) survey. In terms of FOI components, we focused on asking questions about program dose and responsiveness, and asked teachers to describe their experiences regarding curriculum trainings and their engagement in communities of practice, and/or coaching. Similarly, we asked them whether or not they implemented any SEL curricula and/or attentional practices with their students, and explicitly asked them whether they implemented any portion of the MindUP™ curriculum over the course of the year.

In sum, Table 3 provides a complete list of the comprehensive set of FOI measures developed for and used in this study. Table 4 provides example items drawn from these measures for each of the FOI components measured. Developing this set of measures was one of the main activities during the first year of the project, which was a planning year for the project intended for this purpose. After initial development, we conducted a small-scale pilot study

Brain Break
Indicate below if the brain break was implemented, the level of adherence, any extra materials utilized. Adaptations, and student engagement

Observed: ☐ Yes   ☐ No   Start Time: _____:_____   End Time: _____:_____

Were each of the following components included?

☐ Children are invited to participate

☐ Teacher tells children to find comfortable position

☐ Beginning Chime   ☐ Three breaths   ☐ Ending Chime

☐ Time to bring attention back to classroom (30-45 seconds)

Extra materials utilized:

☐ Flower/candle metaphor   ☐ Stuffies/duckies   ☐ Hoberman sphere   ☐ Pinwheels

☐ Other: _____

Notes:

Brain Break Adherence: How closely was the brain break implemented as prescribed? Note: use in conjunction with Activity Adherence Guide

☐ Very high: followed the directions as prescribed (80%-100%)

☐ Moderately high: Followed most of the activity directions (60%-80%)

☐ Average: Followed about half of the activity directions (40%-60%)

☐ Moderately low: Followed less than half of the activity directions (20%-40%)

☐ Very low: did not follow most of the activity directions (<20%)

Was the Brain Break Objective Met?   0 1 2 3 4

Brain Break Adaptations: e.g., were there any practices added from other SEL programs? Any other information added? Anything left out that should have been included? Any disruptions? Any accommodations made for diverse learners?

Student Engagement
For the questions below, please rate student engagement during each implementation component - from 0 (lowest) to 4 (highest)
0 = Very low 1 = Low 2 = Average 3 = High 4 = Very high

Brain Break   NA 0 1 2 3 4

Number of students who didn't accept invitation to participate: _____

Alternative activity offered: N/Y _____

FIGURE 2
Example of third-party observer ratings of the brain break practice: adherence, adaptations, and student responsiveness. Dosage was additionally measured by student attendance during the observation session. Quality of the teacher's implementation was rated once for the observation session, such that quality scores comprise both the curriculum activity and brain break. Neither of these measures are represented in this figure.

to test our set of FOI measures, which led to a cycle of revisions of the measures in preparation for their use within the context of the MindUP™ impact study. Also of note, the process of development required close collaboration between the implementation director and researchers, particularly in operationalizing adherence measures of the teacher implementation supports. We also drew on expertise from a FOI consultant and other researchers supporting our project, who had extensive knowledge and many

TABLE 3 Fidelity of implementation measures of program elements by frequency and components measured.

Program element	Measure	Report frequency	FOI components measured
Curriculum training	MindUP™ training teacher survey	Once	Quality, responsiveness
	MindUP™ training observation form	Once	Dosage, adherence, quality, responsiveness, adaptations
Community of practice (CoP)	CoP teacher survey	Monthly	Quality, responsiveness
	CoP observation form	Monthly	Dosage, adherence, quality, responsiveness, adaptations
Coaching	Implementation director coaching log	Monthly	Dosage, adherence, responsiveness, adaptations
	CoP teacher survey	Monthly	Quality, responsiveness
Curriculum activity	Teacher implementation calendar	Weekly	Dosage, adherence, responsiveness, adaptations
	Classroom observation form	Twice	Dosage*, adherence, quality*, responsiveness, adaptations
Brain break	Teacher implementation calendar	Weekly	Dosage, adherence, quality, responsiveness, adaptations
	Classroom observation form	Twice	Dosage*, adherence, quality*, responsiveness, adaptations
Control group practices	Teacher survey	Once	Dosage, responsiveness, descriptions of professional development experiences

*Quality of teacher delivery and dosage (attendance of students) were each rated once for the observation session, such that quality and dosage scores comprise both the MindUP™ Curriculum Activity and Brain Break. Dosage and responsiveness were measured in the control group for both implementation supports and curriculum implementation level program elements.

TABLE 4 Examples of fidelity of implementation items and informants.

FOI component	Measurement type	Sample item
Dosage	OR; IR	Total minutes of each session attended, additive to total minutes of program attended overall
	OR	Attendance (present, arrived late, left early)
Responsiveness	OR	Engagement (e.g., asking questions, active listening, making eye contact, lack of off task behavior)
	TR	Satisfaction with program element (e.g., "How satisfied are you with the training today?")
Quality	TR	Clarity of the [program elements] activities (e.g., "How clear was the content of this unit delivered during the training today?")
	OR	Observer ratings of implementor clarity of [program elements] activities (e.g., delivers instructions for activities, aids participants in completely activities; confusion is noticed and addressed; misinformation is corrected)
Adherence	OR	Presence or absence of each agenda item
	OR; IR	Degree objective was met for each agenda item—scale of 0 (did not meet any participant objective) to 4 (all participant objectives were met at an exemplary level)
Adaptations	OR; IR; TR	Qualitative notes on changes, additions, or omissions from planned program components
Control group monitoring	TR	Description of professional development opportunities (equivalent to implementation supports including curriculum trainings, community of practices, coaching)
	TR	Implementation of social emotional learning curriculum or practices, including MindUP™

Items in table represent a sample of constructs we measured and how they were operationalized. This list is not comprehensive for most fidelity of implementation components presented. OR, Observer Report; TR, Teacher Report; IR, Implementor Report.

example measures from previous studies examining FOI of the same curriculum that we were studying with different age groups (e.g., Schonert-Reichl et al., 2015). We also drew on an established set of FOI measures developed in the context of teacher-focused, mindfulness-based professional development programs (Doyle et al., 2019). Our final set of measures is the result of this joint effort among many researchers and we acknowledge their effort in this work. Those who are interested in additional detail regarding our comprehensive set of measures should contact the first author for more information.

Discussion

Developing a robust, comprehensive, multi-informant set of FOI measures for our multi-level SEL program creates numerous

opportunities to use the data collected to extend beyond program impacts to address other questions about a program of interest to program administrators, researchers, and practitioners alike. In the following section, we discuss several of these additional avenues of inquiry that become answerable with implementation study data. These areas include (a) a richer potential explanation for why program impacts were found (or not found), (b) a refinement of our understanding of the core components or "active ingredients" of the program, and (c) how FOI data can be used to improve SEL programs over time.

Using implementation data to understand program impacts

In our own work, we view implementation data as critical for answering "*Why did the program work (or not work)?.*" Our implementation conceptual framework is meant to guide analyses that can assess both implementation and program outcomes, and therefore, how implementation affects intervention effectiveness. In our study, implementation outcomes are those that, when taken together, allow us as researchers to understand whether or not our implementation supports were successful in transferring the knowledge and skills needed to teachers (as intervention agents) to successfully implement the curriculum in their classroom, and whether or not teachers were successful in transferring the knowledge and skills of the curriculum to their students to promote positive outcomes (Fixsen et al., 2005; Dunst and Trivette, 2012).

Examples of how implementation outcomes are hypothesized to impact intervention outcomes in our study are displayed in Figure 3. Specifically, this figure depicts an example chain of hypotheses that link implementation supports for teachers (first level) to teachers' curriculum implementation in the classroom (second level) and then on to intervention outcomes for children, using the FOI components of quality, adherence, and responsiveness to demonstrate. To analyze these relations, first, we intend to use implementation data to understand the extent to which our program implementation was successful. Examples of implementation agent outcomes that would indicate success in this area are: that our implementation director was on time, organized, and prepared for the training, community of practice, and coaching sessions (i.e., prepared to provide a high quality session); that they were knowledgeable and implemented each session with fidelity (i.e., high adherence), and that they were also respectful to participants and responsive to their needs.

Second, we plan to examine relations between implementation agent outcomes and teacher implementation outcomes. As is displayed in the first example of Figure 3 (Implementation Supports Level), we hypothesize that the quality of the implementation director outcomes mentioned above will influence teachers' participation in the program. Examples of teacher outcomes that would indicate success in this area are: that all teachers attend the curriculum training (100%

attendance); that they find the training content useful and relevant, and are engaged in the training (high responsiveness); and that they leave the training feeling efficacious in implementing the curriculum with their students. After the training, we also desire to see additional implementation support outcomes that indicate success such as: that teachers attend the majority of ongoing implementation support sessions throughout the study (high dosage of additional supports); that teachers find the sessions engaging and supportive (high responsiveness); and that they will implement the curriculum to a higher level of fidelity through participation in these regular touchpoints and supports. As is displayed in the second example in Figure 3 (Implementation Supports to Curriculum Implementation), we will also be able to determine if greater attendance and engagement in these supports is associated with teachers' fidelity of implementation (adherence) of the curriculum in their classroom, such that teachers who attend support sessions more frequently and/or are more engaged in these sessions will implement the curriculum in their classroom to a higher degree of fidelity compared to teachers who do not attend as frequently, or are not as engaged.

Third, we also plan to examine whether teachers' implementation quality and adherence are associated with student responsiveness to the curriculum (see the third example of Figure 3, Curriculum Implementation Level). Specifically, we hypothesize that teachers who implement curriculum activities on schedule (i.e., high level of adherence), with the majority of students in attendance for each lesson and for each daily Brain Break offered (high dosage), and in a manner that is supportive of where students are developmentally, will lead to higher levels of student engagement and responsiveness overall.

Finally, we plan to test the relations between program implementation and child outcomes. In the final example of Figure 3 (Curriculum Implementation to Outcomes), we hypothesize that implementation support outcomes (indirectly) and curriculum implementation outcomes (directly) will influence child outcomes (impacts). Specifically, we hypothesize that teachers who implement curriculum activities on time, as designed, in a high quality and engaging manner will contribute to larger impacts for their students' social emotional and academic outcomes compared to other teachers who implement activities less frequently, at a lower level of quality, or with a lower degree of adherence. However, we only expect to see larger impacts for students if student attendance and engagement are high.

When displayed in the context of a logic model (such as in Figure 3), the interconnectedness between intervention implementation and outcomes becomes clear. In an ideal scenario, researchers could engage in analyses, such as the ones we have outlined above, to understand if the intervention was implemented as planned. This can help to ensure that the intervention "on paper" matches its effectiveness through its implementation "in practice", as well as any conclusions made about its effectiveness. We believe that it is only when the implementation is understood
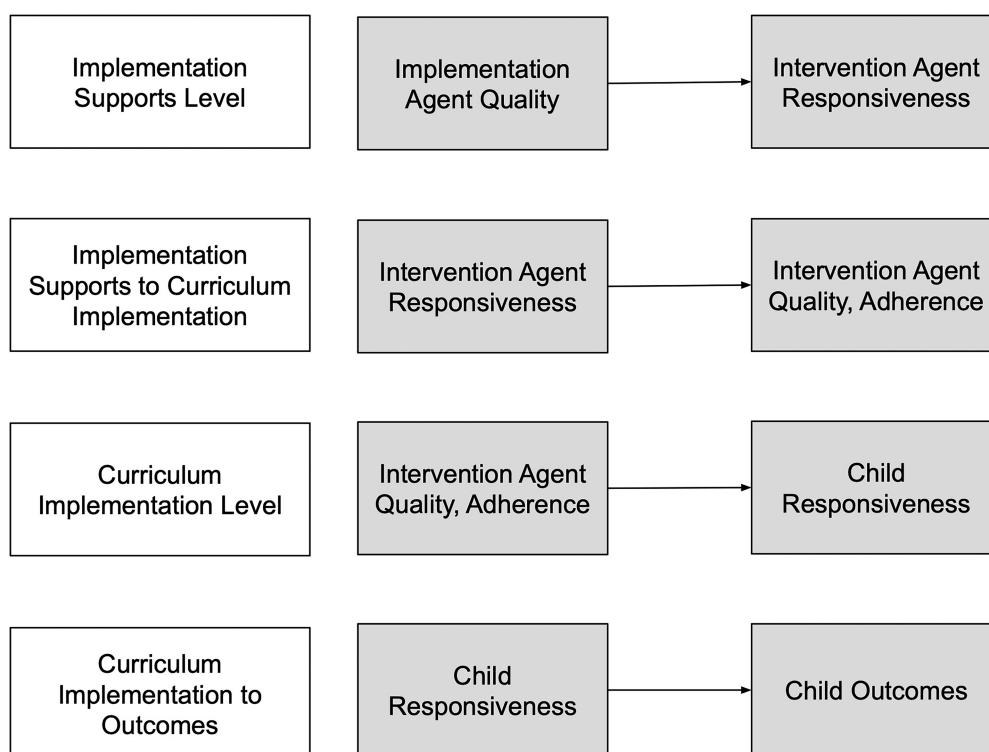
FIGURE 3
Example hypothesized relations between and within intervention implementation levels and outcomes. Outcomes refers to program impacts for children. In our study, implementation agent refers to our implementation director and intervention agent refers to preschool teachers implementing the MindUP™ curriculum in their classroom with students.

in these basic ways that one should move on to testing efficacy for the impact study. Furthermore, publishing implementation and impacts together may reduce the bias present in published findings and support more accurate metanalytic reviews (Torgerson and Torgerson, 2008).

Using implementation data to understand core program components

Once a program has been found to be efficacious, implementation data can also be used to explore why a program was effective and answer "*what are the core components of the program?*" (see Baelen et al., 2022). Core components can be defined as "the most essential and indispensable components of an intervention practice or program" (Fixsen et al., 2005, p. 24). The goal of this framework is to carve away everything nonessential to the program over time to ensure the intervention promotes valuable activities (i.e., "active ingredients") and only demands what is necessary from participants for positive outcomes (Moir, 2018).

Determining the core components of a program requires an iterative process of testing and refining an intervention or program over time (Fixsen et al., 2005). In the initial implementation of a

program, implementation data can be used to clarify and solidify the core components of the program being tested (see Baelen et al., 2022). Ideally, researchers will not stop at determining that an intervention is effective, but rather continue investing in a program with additional studies that test the accuracy and impact of specific core components. This can aid in refining the program as needed by adding or subtracting core components and/or delivery methods of program elements (Fixsen et al., 2005). This work between initial testing and future implementation efforts is imperative to engage in a process of continuous improvement, to reduce resources wasted on implementation components that are non-essential, and to explore context-specific considerations of the program (e.g., does it work for everyone? In all setting types? For all ages?). Additionally, the effectiveness of a program would be expected to improve by paring down nonessential elements, as participants may receive a higher dose of the "active ingredients" of the intervention, rather than a combination of inactive and active components (Fixsen et al., 2005).

More specifically, dosage data can be used to determine active ingredients of the intervention, by examining if specific activities or practices were drivers of intervention outcomes. These analyses examine the extent to which dose of the intervention component for each participant relates to their outcomes or benefits gained from the intervention. These explorations provide interesting insights by going beyond the primary question of the impact

study. However, they involve teachers in the treatment group only and rely on quasi-experimental methods (classrooms were not randomly assigned to different levels of implementation). Nonetheless, they can be valuable additional tests that explore questions about what program factors are related to positive intervention impacts (Baelen et al., 2022; Roeser et al., 2022).

In our study of MindUP™, we plan to examine the unique effect of the Brain Break practice, as it is likely a primary core component of this program and also a practice that differentiates this program from many other SEL curricula. We hypothesize that students who receive a "higher dose" of the Brain Break will demonstrate greater gains in cognitive and behavioral measures of self-regulation at the end of the year. Beyond simply examining dosage, we also hypothesize that the impact of the brain break will be greatest for children in classrooms where teachers implemented the Brain Break in a manner that is engaging for students. Thus, we plan to use a combination of the dosage (total number of Brain Breaks offered to the class), student attendance, and child engagement (as a measure of responsiveness) to examine whether the Brain Break is a core component of this program.

Using implementation data to improve the program over time

Analysis of implementation data can also support a process of continuous improvement in the program under evaluation in an impact study. By identifying and understanding the core or active ingredients of an intervention, researchers are in a better position to answer the question *"how do we make the program better?"* in the future. While there are numerous ways to approach program improvements, for this section, we focus on three areas in terms of program optimization, including improvements for teaching, enhancement of active ingredients, and considerations focused on equity.

Implementation data collected as part of an impact study can be used to support teachers' future implementation of a program. In general, FOI data can be synthesized to highlight strengths across the group of implementors, as well as to identify common challenges with implementation. Feedback on adherence can be particularly useful if teachers plan to continue implementing the program after the intervention is over to support high fidelity and counteract program drift over time (see Domitrovich et al., 2012). For example, in our own study, we plan to analyze teachers' use of materials to support and scaffold students in the Brain Break practice and provide teachers with a summary of the most frequently used supplementary materials and tools. This implementation summary resource may be helpful to teachers who found this aspect of implementation challenging.

Similarly, an analysis of FOI components collected regarding coaching may reveal certain aspects of implementation that were particularly challenging for teachers (intervention agents) and that could use careful review for potential revisions in future implementations by the curriculum developers. For example, our

coaching log asks teachers whether they would implement each activity in the curriculum again and if they report that they would not, asks them to elaborate on reasons why. We also ask teachers on the coaching log whether they have any feedback for the implementation team regarding each curriculum activity. Having the implementation director read the coaching logs throughout the program allows for a direct line between teachers and the implementation team in a manner that supports this work. Additionally, researchers can acknowledge and honor teachers' expertise by exploring adaptations they implemented, taking careful note of any that may be particularly useful for diverse learners and that could be incorporated into curriculum revisions to improve the program. These are just several examples that highlight how researchers can give back the data to those participating in ways that are useful to them and their profession.

Analyses examining dose–response relations, as described in the previous section, can also inform potential revisions to a program by helping to determine whether the "full-dose" was necessary to produce impacts or if a smaller dose is effective. Given the time demands associated with teaching, if a smaller dose is found to be effective, reducing program demands may actually increase teachers' ability to adhere to the program, and on a broader scale, may increase the total number of teachers who find the program feasible to implement. Furthermore, examining unique effects of each intervention element can establish if specific intervention elements are driving intervention outcomes, in which case these active ingredients could be amplified in future implementations to maximize positive outcomes for participants. We have also learned that incorporating a new FOI measure of generalization (for example see Low et al., 2016) into our future work will likely be important in this vein, to better understand the informal use of the program through integration and reinforcement of program components into daily classroom activities.

Finally, FOI data can also inform future changes to a program if null or negative impacts are found for specific subgroups of participants through moderated impacts analyses (Bywater, 2012). From an equity perspective, this area is essential to ensure an intervention touted as universal does in fact equally benefit individuals from diverse backgrounds. To highlight this issue, Rowe and Trickett (2018) conducted a meta-analytic review examining various diversity characteristics and how they moderated impacts of SEL interventions for students. One prominent finding of this study was the general lack of attention on this issue overall, in which most studies included in the review did not test for differential impacts. For those that did, the authors found mixed program impacts dependent on the diversity characteristic examined. These results speak to the need for examining these issues, particularly in the context of interventions considered appropriate for students of all cultures and backgrounds. Furthermore, we argue these analyses should be examined even if a positive impact of the intervention is found overall for participants, to determine if specific subgroups did not benefit, or alternatively, if certain subgroups benefited significantly

more than others. In the latter case, these analyses can be used to further explore active ingredients of the intervention and potentially inform program changes to increase the impact across all participants in the future.

Regardless of whether we find an overall positive effect on children's outcomes in our study, we plan to examine whether effectiveness of the program varied for different subgroups of children. In particular, we are interested in understanding whether the effectiveness of the program is the same or different based on the classroom setting and/or type of program. As described in our theory of change, we hypothesize that the effects of this program will be stronger in programs in which the quality of interactions between teachers and students is lower at baseline. Furthermore, it will be important to examine whether the program is effective in a variety of program types (e.g., community-based program, Head Start, for-profit) to better understand whether the program context impacts the effectiveness of the program. Lastly, if we find that the impact of the program varies for subgroups of children, it will be important to examine teachers' adherence and quality data, as well as students' dosage and responsiveness data, to understand whether implementation differed across subgroups. If differences in implementation by teachers are not found, but student engagement differs, it may indicate that the program was less relevant for some children. Each of these examinations of differential impacts can inform program changes in the future to enhance the program's effectiveness.

Conclusion

The main purpose of this paper was to demonstrate the interconnected nature between implementation and impacts in the context of studying complex, multi-level education interventions generally, and SEL programs more specifically. We believe, as others do, that robust FOI data provide essential contextual information about an intervention's inner workings. In this vein, these data inform both the internal validity of the conclusions from our impacts study and provide valuable information that can be used to support effective application in real-world contexts. We aimed to illustrate this key point by drawing on our own learnings developing the conceptual framework and associated FOI measures for our ongoing evaluation study testing a mindfulness-based SEL program for preschoolers, and by exploring numerous ways these data can extend beyond simple tests of intervention effectiveness and be used to describe, explain, and optimize education programs.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

Funding

Acknowledgments

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 U.S. Department of Education.

References

Baelen, R. N., Gould, L. F., Felver, J. C., Schussler, D. L., and Greenberg, M. T. (2022). Implementation reporting recommendations for school based mindfulness programs. *Mindfulness*. doi: 10.1007/s12671-022-01997-2

Belfield, C., Bowden, A. B., Klapp, A., Levin, H., Shand, R., and Zander, S. (2015). The economic value of social and emotional learning. *J. Benefit Cost Analys.* 6, 508–544. doi: 10.1017/bca.2015.55

Bickman, L., and Reich, S. M. (2015). *"Randomized controlled trials: A gold standard or gold plated?" in credible and actionable evidence: The Foundation for Rigorous and Influential Evaluations.* Thousand Oaks, CA: SAGE Publications, Inc., 83–113.

Boruch, R. (2005). Better evaluation for evidence-based policy: place randomized trials in education, criminology, welfare, and health. *Ann. Am. Acad. Pol. Soc. Sci.* 599, 6–18. doi: 10.1177/0002716205275610

Boruch, R., de Moya, D., and Snyder, B. (2002). "The importance of randomized field trials in education and related areas" in *Evidence matters: Randomized trials in education research.* eds. F. Mosteller and R. Boruch (Washington, DC: Brookings Institution Press), 50–79.

Braun, S. S., Roeser, R. W., Mashburn, A. J., and Skinner, E. (2018). Middle school teachers' mindfulness, occupational health and well-being, and the quality of teacher-student interactions. *Mindfulness* 10, 245–255. doi: 10.1007/S12671-018-0968-2

Bryk, A. S. (2016). Fidelity of implementation: is it the right concept? Carnegie commons blog. Available at: https://www.carnegiefoundation.org/blog/fidelity-of-implementation-is-it-the-right-concept/ (Accessed 14 July 2022).

Bywater, T. (2012). "Developing rigorous programme evaluation" in *Handbook of implementation science for psychology in education.* eds. B. Kelly and D. F. Perkins (New York, NY: Cambridge University Press), 37–53.

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

Domitrovich, C. E., Moore, J. E., and Greenberg, M. T. (2012). "Maximizing the effectiveness of social-emotional interventions for young children through high quality implementation of evidence-based interventions" in *Handbook of implementation science for psychology in education.* eds. B. Kelly and D. F. Perkins (New York, NY: Cambridge University Press), 207–229.

Doyle, S. L., Jennings, P. A., Brown, J. L., Rasheed, D., DeWeese, A., Frank, J. L., et al. (2019). Exploring relationships between CARE program fidelity, quality, participant responsiveness, and uptake of mindful practices. *Mindfulness* 10, 841–853. doi: 10.1007/s12671-018-1034-9

Dunst, C. J., and Trivette, C. M. (2012). "Meta-analysis of implementation practice research" in *Handbook of implementation science for psychology in education.* eds. B. Kelly and D. F. Perkins (New York, NY: Cambridge University Press), 68–91.

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

Fixsen, D. L., and Blase, L. A. (2009). *Implementation: The missing link between research and practice* No. 1; NIRN Implementation Brief; The National Implementation Research Network. Chapel Hill, NC: University of North Carolina at Chapel Hill.

Fixsen, D. L., Naoom, S. F., Blasé, K. A., Friedman, R. M., and Wallace, F. (2005). *Implementation research: A synthesis of the literature* FMHI Publication #231; The National Implementation Research Network. Tampa, FL: University of South Florida.

Hollis, S., and Campbell, F. (1999). What is meant by intention to treat analysis? Survey of published randomised, controlled trials. *Br. Med. J.* 319, 670–674. doi: 10.1136/bmj.319.7211.670

Jennings Brown, J. L., Frank, J. L., Doyle, S., Oh, Y., Davis, R., Rasheed, D., et al. (2017). Impacts of the CARE for teachers program on teachers' social and emotional competence and classroom interactions. *J. Educ. Psychol.* 109:1010. doi: 10.1037/edu0000187, 1028

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/s11121-016-0670-3

Luo, L., Reichow, B., Snyder, P., Harrington, J., and Polignano, J. (2022). Systematic review and meta-analysis of classroom-wide social–emotional interventions for preschool children. *Top. Early Child. Spec. Educ.* 42, 4–19. doi: 10.1177/0271121420935579

Mac Donald, E., and Shirley, D. (2009). *The mindful teacher.* New York: Teachers College Press.

Maloney, J. E., Lawlor, M. S., Schonert-Reichl, K. A., and Whitehead, J. (2016). *A mindfulness-based social and emotional learning curriculum for school-aged children.* New York, NY: Handbook of mindfulness in education, Springer, 313–334.

Mark, M. M., and Lenz-Watson, A. L. (2011). "Ethics and the conduct of randomized experiments and quasi-experiments in field settings" in *Handbook of ethics in quantitative methodology.* eds. A. T. Panter and S. K. Sterba (New York, NY: Routledge), 185–209.

McCormick, M. P., Cappella, E., O'Connor, E., Hill, J. L., and McClowry, S. (2016). Do effects of social-emotional learning programs vary by level of parent participation? Evidence from the randomized trial of INSIGHTS. *J. Res. Educ. Effect.* 9, 364–394. doi: 10.1080/19345747.2015.1105892

Moir, T. (2018). Why is implementation science important for intervention design and evaluation within educational settings? *Front. Educ.* 3, 1–9. doi: 10.3389/feduc.2018.00061

Roeser, R. W., Galla, B. M., and Baelen, R. N. (2020). Mindfulness in schools: Evidence on the impacts of school-based mindfulness programs on student outcomes in P-12 educational settings. Available at: www.prevention.psu.edu/sel

Roeser, R. W., Greenberg, M., Frazier, T., Galla, B. M., Semenov, A., and Warren, M. T. (2022). Envisioning the next generation of science on mindfulness and compassion in schools. *Mindfulness.* 114, 408–425. doi: 10.1007/s12671-022-02017-z

Rowe, H. L., and Trickett, E. J. (2018). Student diversity representation and reporting in universal school-based social and emotional learning programs: implications for generalizability. *Educ. Psychol. Rev.* 30, 559–583. doi: 10.1007/s10648-017-9425-3

Sandilos, L. E., Neugebauer, S. R., DiPerna, J. C., Hart, S. C., and Lei, P. (2022). Social–emotional learning for whom? Implications of a universal SEL program and teacher well-being for teachers' interactions with students. *Sch. Ment. Heal.* doi: 10.1007/s12310-022-09543-0, 1, 12

Schonert-Reichl, K. A., Oberle, E., Lawlor, M. S., Abbott, D., Thomson, K., Oberlander, T. F., et al. (2015). Enhancing cognitive and social–emotional development through a simple-to-administer mindfulness-based school program for elementary school children: a randomized controlled trial. *Dev. Psychol.* 51, 52–66. doi: 10.1037/a0038454

Shadish, W. R., Cook, T. D., and Campbell, D. T. (2002). *Experimental and quasi-experimental designs for general causal inference.* Boston: Houghton Mifflin.

Snyder, P. A., Hemmeter, M. L., and Fox, L. (2015). Supporting implementation of evidence-based practices through practice-based coaching. *Top. Early Child. Spec. Educ.* 35, 133–143. doi: 10.1177/0271121415594925

Torgerson, D. J., and Torgerson, C. J. (2008). *Designing randomised trials in health, education and the social sciences.* London, UK: Palgrave Macmillan.