Prevention Science (2023) 24:663–675
https://doi.org/10.1007/s11121-022-01481-2

Implementing Adolescent Wellbeing and Health Programs in Schools: Insights from a Mixed Methods and Multiple Informant Study

Jacinda K. Dariotis[1] · Keren Mabisi[2] · Rachel Jackson-Gordon[3] · Nan Yang[2] · Emma Jane Rose[4] · Tamar Mendelson[5] · Diana H. Fishbein[6,7]

Accepted: 13 December 2022 / Published online: 11 January 2023

Abstract

Determining the factors that influence implementation of school-based wellbeing and health programs is essential for achieving desired program effects. Using a convergent mixed-methods, multiple informant design, this study considered factors that influence implementation of health programs for ninth grade students and in what ways implementation is differentially perceived by multiple informants (i.e., participants, instructors, and independent observers). Two types of programs—mindfulness and health education—were implemented with ninth graders ($N=70$) in three schools situated in low-resourced urban neighborhoods. Study outcomes were derived from four data sources: (1) focus group participants ($N=45$); (2) program instructor fidelity ratings; (3) independent observer fidelity ratings and notes; and (4) instructor open-ended session responses. Using thematic and mixed methods integration analyses, we identified themes related to implementation promoting or challenging factors. Theme names differed when data sources were separately analyzed by informant. Mixed methods integration analysis indicated that four themes were common across all informant groups: (1) competent, attentive, and engaging instructors are essential; (2) programs should involve interactive components (e.g., physical activities, applied learning opportunities); (3) adequate time for program delivery is key for student exposure and engagement; and (4) students' availability and preferences should guide program scheduling. A fifth theme, unique to instructor and observer perspectives, was that program implementation was negatively impacted by distractions from multiple sources, including instructors, students, and settings. Recommendations from students, instructors, and observers for implementation optimization are discussed.

Keywords Implementation · Fidelity · Mindfulness · Health education · Youth · School-based interventions

Introduction

Quality implementation is crucial for optimizing the effects of prevention programs (Durlak & DuPre, 2008; Meyers et al., 2012). Standards for developing and scaling prevention programs have emphasized the need to measure and report the quantity and quality of implementation (Gottfredson et al., 2015). To holistically understand program implementation, it is essential to assess the perspectives of multiple involved parties (e.g., implementers,

✉ Jacinda K. Dariotis
dariotis@illinois.edu

[1] Department of Human Development and Family Studies and Family Resiliency Center, College of Agricultural, Consumer and Environmental Sciences; Department of Biomedical and Translational Sciences, Carle Illinois College of Medicine, The University of Illinois at Urbana-Champaign, Urbana, IL, USA

[2] Evaluation Services Center, College of Education, Criminal Justice, & Human Services, University of Cincinnati, Cincinnati, OH, USA

[3] Charlotte Urban Institute, University of North Carolina, Charlotte, NC, USA

[4] The Child Study Center, Department of Psychology, College of Liberal Arts, The Pennsylvania State University, PA, University Park, USA

[5] Department of Mental Health, The Johns Hopkins University Bloomberg School of Public Health, Baltimore, MD, USA

[6] Frank Porter Graham Child Development Institute, The University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

[7] Department of Human Development and Family Studies, The Pennsylvania State University, University Park, PA, USA

recipients, program staff). However, intervention studies typically rely on only one type of participant perspective on intervention implementation rather than including various perspectives. The current study addresses this gap by integrating data on program implementation from three types of informants—student participants, instructors, and trained observers. Additionally, this study leverages multiple data sources, including focus group discussions, instructor logs, and session recordings to enhance our ability to detect implementation issues.

Implementation Science

Effective implementation strategies are necessary to produce program benefits (e.g., Fixsen et al., 2015), including for evidence-based programs. Through the examination and testing of factors facilitating and impeding the uptake, adoption, and integration of evidence-based practices or programs into real-world settings like schools, communities, and health care (Bauer & Kirchner, 2020), implementation science offers frameworks for understanding implementation processes and outcomes. Borrowing from Dane and Schneider (1998), Durlak and DuPre (2008) outlined eight key elements of program implementation: (1) fidelity, (2) dosage, (3) quality, (4) participant responsiveness, (5) program differentiation, (6) monitoring the control condition, (7) program reach, and (8) program modifications. In keeping with this framework, the current study utilized strategies to address fidelity, dosage, quality, and participant responsiveness.

Fidelity, one of the most frequently discussed implementation constructs, is concerned with adherence to a protocol and quality of delivery (Proctor et al., 2011). Research to gauge fidelity may focus on strategies to enhance participation (attendance and engagement), program adaptations, and implementation factors related to program outcomes (Gottfredson et al., 2015). Understanding program core components—or the active ingredients of a program (Blasé & Fixsen, 2013)—and adaptations or modifications are also helpful frameworks for studying fidelity in the context of programmatic success (Ferber et al., 2019). In the context of school-based programs, there are a variety of factors that may influence the quality of implementation. In a study of over 500 schools, Gottfredson and Gottfredson (2002) found that implementation quality is relatively low and could be improved by actions such as better integration with the school schedule and better training and planning overall. Additionally, in a study of school-based social and emotional learning programs, school staff using recommended practices produced better outcomes compared to staff who did not adopt the recommended practices (Durlak et al., 2011). Without monitoring implementation throughout a study—with attention to core components—it is challenging to understand program outcomes or replicate them in the future to produce intended results.

Methods to Uncover Implementation Drivers

Despite the importance of understanding program implementation, there are limitations in existing research methods. Across 542 studies reviewed by Durlak and DuPre (2008), self-report and observational methods were used most frequently when documenting implementation fidelity, dosage, and quality; there is limited use of validated quantitative methods. For example, Lewis et al. (2015) completed a systematic review of 104 implementation instruments related to mental and behavioral health based on implementation outcomes (Proctor et al., 2011) and found few with evidence of psychometric strength. Many implementation studies use qualitative methods to gather rich information about implementation but may only capture one type of perspective, such as from implementers or providers (e.g., Cutbush et al., 2017; Davidov et al., 2020; Koffel & Hagedorn, 2020). In contrast, studies that collect data from multiple informants have found diversity in perspectives to be a strength for optimizing implementation strategies (e.g., Deatrick et al., 2021). In summary, there are challenges to assessing the quality of program implementation for quantitative and qualitative studies.

One strategy to develop a more comprehensive overview of implementation includes mixed methods designs, which combine qualitative and quantitative research strategies to "develop a science base for understanding and overcoming barriers to implementation" (Palinkas et al., 2011, p. 44). As such, a mixed methods approach may be an important strategy to understand program trial and implementation outcomes in real-world settings (Albright et al., 2013). Palinkas et al. (2019) described the advantages of mixed methods in three types of evaluations: effectiveness studies, implementation process studies, and effectiveness-implementation hybrid studies; the current study focuses on the implementation process. Others have employed mixed methods to study prevention programs, including Kozica et al. (2016) who utilized this approach in a hybrid study of the obesity prevention Healthy Lifestyle Program. Mixed methods data were complementary in determining overall program success; the program was effective and feasible to implement with quality, even in low-resource settings (Kozica et al., 2016). Similarly, Dobbie et al. (2019) studied the implementation of a smoking prevention school-based program through a process evaluation consisting of interviews, structured observations, and surveys to examine fidelity and acceptability, as well as the influence of context

on message diffusion. The investigators triangulated findings to produce overall implementation themes regarding program fidelity and acceptability, spread of messaging, and benefits. Similar to the Dobbie et al. (2019) study, the present study integrated findings from several data sources to gain insights and provide recommendations to improve program implementation.

Current Study

Incorporating multiple informants and methods provides an opportunity to understand program implementation more thoroughly and with greater rigor. Using a convergent mixed methods (Creswell & Plano Clark, 2018), multiple informant research design, the current study aimed to answer the following two research questions: which factors influenced implementation of two health programs administered to ninth grade students, and how was implementation differentially perceived by various reporters (e.g., student participants, instructors, observers)? Student participants provided qualitative focus group data, instructors provided qualitative and quantitative data from session ratings and open-ended comments, and observers provided qualitative and quantitative data through observation notes and session ratings. By leveraging data from multiple perspectives, this investigation has potential to clarify aspects that influence program implementation and inform future program delivery.

Methods

Using a mixed methods approach, quantitative and qualitative data were collected during the delivery of two health promoting programs to ninth grade students attending three urban public schools in low-income neighborhoods. A convergent mixed methods design was utilized, where quantitative and qualitative study components took place independently, and then results were compared and combined (Creswell & Plano Clark, 2018). The intent of the design was to bring together the quantitative and qualitative results for enhanced understanding of factors influencing implementation of the two programs. The programs were the Mind in Action (T-MIA) mindfulness program (Holistic Life Foundation, Inc.; www.hlfinc.org) and Healthy Topics (HT), a health education program (McGraw Hill Glencoe Health Curriculum) (see Dariotis et al., (under review) for detailed program descriptions). The programs were administered during the school day by instructors external to the school. This study was approved by the Institutional Review Boards of the school district and the universities affiliated with the research.

Participants

Schools serving predominantly low-income students, the majority of whom identified as Black, were recruited to partner with investigators to determine program effects on psychological, behavioral, and physiological indicators of coping in a high stress environment. Participants were ninth grade students ($N = 79$; 53% female, 84% Black) self-selected into the study at the participating schools. After obtaining signed student assent and parental permission, participants were randomized to the T-MIA or HT program. Key student demographic characteristics did not differ significantly between programs (Table 1). Of those consented, students who attended at least one program session were considered to comprise the "full sample" ($N = 70$). All 70 participants were invited to participate in focus groups, which were held within the next week after the last day of programming. To maintain consistency, focus groups were held the same day and time as program sessions and were conducted in the same location. A subset of these participants ($N = 45$) participated in one of the six focus groups (FG) after programming ended (size range: $n = 5$–11; mode = 6; $n = 21$ T-MIA; $n = 24$ HT). Background characteristics of this participant subset did not differ significantly from those of the full sample. FG participants, on average, attended 23 sessions (full sample: 20 T-MIA and 21 HT sessions).

Program Delivery

Programming was delivered by trained instructors external to the school during spring 2019 as part of a planned two-cohort parent study. Two Black male instructors delivered T-MIA, with one delivering the program at a single school and the other at the two remaining schools. The HT program was delivered in all schools by the same White male instructor. The T-MIA instructors were employees of the Holistic Life Foundation (HLF) and had received extensive training as yoga and meditation teachers through HLF. The HT instructor was a member of the study team who had been trained by one of the PIs in how to deliver HT. All instructors were paid for their time.

Although each program was planned to be delivered in 30-min sessions four times per week for 10 weeks, actual delivery time was often reduced somewhat below 30 min by school schedule changes and other interruptions. The average number of sessions ranged from 27 to 31 across programs. Prior to delivery, 25 T-MIA and 26 HT sessions were randomly selected for digital recording for fidelity coding by independent observers. Students received a $10 gift card each for completing each assessment (e.g., focus group,

Table 1 Demographic characteristics—full sample and focus group subsample

		Full sample ($N=70$)						Focus group subsample ($N=45$)					
		Total		T-MIA ($n=34$)		HT ($n=36$)		Total		T-MIA ($n=19$)		HT ($n=26$)	
Age (SD)	Mean	14.7 (0.83)		14.7 (0.99)		14.6 (0.65)		14.7 (0.94)		14.9 (1.20)		14.6 (0.70)	
		n	%	*n*	%	*n*	%	*n*	%	*N*	%	*n*	%
Sex	Female	37	52.9	17	50.0	20	55.6	23	51.1	9	47.4	14	53.8
	Male	32	45.7	17	50.0	15	41.7	22	48.9	10	52.6	12	46.2
Race/	Black	59	84.3	27	79.4	32	88.9	39	86.7	16	84.2	23	88.5
Ethnicity	Hispanic	3	4.3	2	5.9	1	2.8	1	2.2	0	0	1	3.8
	White	3	4.3	1	2.9	2	5.6	1	2.2	0	0	1	3.8
	Asian	2	2.9	1	2.9	1	2.8	2	4.4	1	5.3	1	3.8
	Other	2	2.9	2	5.9	0	0	1	2.2	1	5.3	0	0

Missing data for sex (1 full sample HT student); race/ethnicity (1 full and 1 subsample HLF student)

pre- and post-survey) and an additional $25 gift card for attending at least 80% of sessions.

Measures

Multiple measures were used to assess program implementation. After each session, implementers recorded students' attendance (dosage) and engagement and disruptiveness (student responsiveness) using a log and rated their adherence to the curriculum and quality of delivery during the session using a fidelity tool. Observers rated recorded sessions using the same log and fidelity tool and wrote observational notes. Students reported on their experiences during focus groups at the end of programming.

Attendance, Engagement, and Disruption Log

Instructors recorded each participant's attendance at each session (present/absent) as a measure of program dosage. The instructor also rated each participant's level of engagement and disruptiveness on 4-point scales (1 = not at all; 2 = a little; 3 = somewhat; 4 = a lot) as measures of student responsiveness. Quantitative methods were employed for these analyses.

Implementation Fidelity Tool

A fidelity rating tool was developed based on the CORE process model for assessing fidelity (Gould et al., 2014, 2016) of program components specified in manuals. After each session, using Likert-like scales, instructors rated their program delivery with respect to adherence to session core activities and processes and overall implementation quality. Means for each of the three fidelity components (core activities, core processes, overall quality) were computed by session. Instructors could also provide open-ended comments regarding challenges, well-received activities, and disruptions. The scale ratings were quantitative, and open-ended comments served as qualitative data.

Core activities assessed content described in the program manual to be covered in a particular session and were rated 0 (not covered at all), 1 (covered a little bit), 2 (covered most), and 3 (covered as outlined in the manual). These items varied by session and program.

Core processes included five items measuring teaching practices (maintain order and discipline; maintain student engagement and attention; closely follow curriculum outline; address emergent student needs; review concepts and skills to reinforce salience, learning, and mastery). Ratings were 0 (did not do at all), 1 (some of the time), 2 (most of the time), and 3 (did throughout session). *Overall quality* was assessed using two items—overall session quality (curriculum coverage, instructor delivery, and student engagement) and overall student responsiveness (to material, instruction, participation). Ratings were from 0 to 4 (0 = poor; 1 = fair; 2 = neutral; 3 = good; 4 = excellent). The same core processes and overall quality items were used to rate all sessions and both programs. Implementation characteristics (number of sessions, enrollment, attendance) are summarized in Table 2 by school and program for the full sample attending at least one session ($N=70$) and the focus group subsample ($N=45$).

Observer Fidelity Coding

Using the same implementation fidelity tool, three independent observers coded 29 randomly selected session recordings for fidelity of implementation. A total of 32 recorded

Table 2 Program implementation characteristics by school and program

	Full sample ($N=70$)								Focus group sample ($n=45$)		
	School A		School B		School C		Total				Total
Characteristic	T-MIA	HT	T-MIA	HT	T-MIA	HT	T-MIA	HT	T-MIA	HT	
Enrolled students (n)	11	12	11	13	12	11	34	36	19	26	45
Sessions offered (n)	27	27	30	31	31	31	29	30	29	30	29
Sessions attended (mean n)	23	19	13	20	24	25	20	21	23	23	23
Attendance (mean %)	85%	69%	44%	66%	77%	81%	69%	71%	80%	76%	78%

sessions (16 per program) had sufficient quality for coding (exclusion reasons were no audio, excessive noise, student face visibility, instructor not visible). Of these, three were missing instructor fidelity ratings, resulting in 29 recordings coded by observers. Observer scale rating data were quantitatively analyzed in the same way as instructor ratings.

Observers also took descriptive notes about fidelity-enhancing and detracting factors for quality improvement feedback. Prior to coding, the three observers attended two trainings, independently coded the same recording, and attended a quality assurance session during which coding results were discussed; interrater agreement was then established. Observers then reviewed program curricula and coded each session using the fidelity tool. They also documented other variables (e.g., set-up time, classroom management time) and provided comments about group session interactions (e.g., calm; talking over each other). These data are qualitative.

Focus Group Protocol

Focus group guides were adapted from protocols used in other mindfulness studies, (Dariotis et al., 2016, 2017) and wording was relevant to both programs. Questions specifically assessing program implementation included "How would you describe the program to one of your friends?" and a further probe included "What kinds of activities did you do?" Participants were asked about program weaknesses, strengths, changes they would make to the program, and whether they would join the program again and recommend it to others. At the end of the focus group, youth were able to offer additional information regarding their program experience. Data generated by focus groups were qualitative.

Procedure

Immediately after programs ended, six focus group discussions were conducted during the school day, one per program per school. All participants were invited; 65% attended, which is a large percent for a qualitative subset (Onwuegbuzie & Collins, 2007). Reasons for not attending included school absence, conflicting academic demands, and forgetting. Each focus group was facilitated by a team member while another team member served as note taker. These team members were not involved in program delivery or data analysis, reducing opportunities for bias. Three experienced team members were trained for these roles in two 2-h training sessions, as well as mock focus group discussions. Discussions lasted approximately 40 min and were recorded and transcribed.

Data Analysis

Session Fidelity and Quality—Mixed Methods

Ratings A total of 181 sessions (6 classes delivered 27–31 times) had attendance, engagement, and disruptiveness records and instructor self-reported fidelity ratings. Mean scores for each component were computed, coded as 0 (low fidelity) or 1 (high fidelity) based on the literature-suggested fidelity cut score (Hulleman & Cordray, 2009), and compared across schools, instructors, and programs using independent t tests and ANOVAs to identify fidelity patterns.

Observer and Instructor Comparisons Twenty-nine recorded sessions had instructor fidelity ratings, as well as sufficient visual and audio quality for observer coding. Fidelity rating differences between instructors and observers were tested using paired t tests for these sessions. Quantitative data were analyzed using IBM SPSS Statistics (Version 27) predictive analytics software and SAS (Version 9.4). Open-ended comments of instructors and observers and field notes by observers were summarized by common topics.

Focus Groups—Qualitative

Transcripts were analyzed using a codebook that was iteratively developed with an inductive approach (i.e., propose, discuss, identify, and refine; Braun & Clarke, 2006), which identifies patterns to enhance the meaning of the data.

Codebook consensus was established using two of the six transcripts, each coded and compared by two team members. Coding disparities were minimal, characterized by one coder using a subcode and the other using a higher order code; higher order codes were used in these instances. Using thematic analysis (Boyatzis, 1998), the coders independently reviewed data and codes, created and discussed categories, and refined these into themes. MAXQDA 2020 (VERBI Software, 2019) was used for data analysis. To ensure the reliability and validity of data interpretations, analyses were carried out independently by three researchers, with two coding the same transcript.

Mixed Methods Analysis—Triangulation

Triangulation is used in the study to "identify areas of corroboration and dissonance in the data by comparing quantitative and qualitative data" (Aschbrenner et al., 2022, p. 4). After data from each source were analyzed separately, the research team compared the findings across sources. As noted by Morgan (2019), triangulation can be a nebulous term with different purposes. Consistent with Morgan (2019), for this study convergence was the primary goal when comparing findings (Morgan, 2019). To identify themes across sources, researchers responsible for the analysis of each source met to discuss emergent ideas from different informants. Themes were revised iteratively, and original results were consulted for confirmation throughout the process. Ultimately, theme concordance (overlap) and uniqueness (non-overlap) across informants were agreed upon and themes finalized. Interpretation consistency and trustworthiness consistency across coders and data sources was ensured by reviewing and discussing coding and analytic approaches at each stage.

Results

To address the first research question on what factors influenced implementation, thematic findings that address implementation are presented, based on qualitative analyses. To address the second research question, instructor and observer quantitative comparisons and triangulation with qualitative data across all informants are presented. These results reveal differential and shared perceptions about implementation across students, instructors, and observers, relevant to the second research question.

Dosage and Student Responsiveness

Instructors reported the number of sessions held and student attendance by session using the attendance log. Study programs were originally intended to be delivered in 40 sessions, but the session number was later reduced to 32 sessions due to a few-week delay in study start limiting implementation to 8 rather than 10 weeks. Shown in Table 2, session offerings ranged from 27 to 31 for the full sample and 29–30 for the focus group subset. The average number of sessions students attended ranged from 13 to 25 across schools and programs.

Student responsiveness was assessed using engagement and disruptiveness items collected on attendance logs. The full and FG group samples had comparable engagement (mean (SD): 3.7 (0.28) FG sample vs. 3.6 (0.30) full sample; 4 = a lot) and disruptiveness (mean (SD): 1.1 (0.22) FG sample vs. 1.2 (0.25) full sample; 1 = not at all). Student focus group responses as well as instructor open-ended comments and observer field notes provide insights into factors impacting dosage and student responsiveness as described in the themes below.

Factors Influencing Implementation

Instructor and Observer Qualitative Perspectives on Factors Influencing Implementation

Data on implementation practices derived from two sources: instructor comments on session logs and observer notes on recorded sessions. Five themes emerged from both sources. Two of these themes addressed what worked well: (1) engaging and attentive instruction and (2) active student learning activities. Three themes related to implementation challenges: (3) insufficient time to cover material; (4) behavior management demands; and (5) environmental disruptions. Instructor and observer recommendations for improvement are included within relevant themes.

Engaging and Attentive Instruction Qualities of effective instruction included (a) engaging students through familiarity with student interests (rapport); (b) being aware and attending to student needs (attunement); (c) actively seeking student participation (active instruction); and (d) participating in the program with students (modeling). Some instructors brought students' focus back to the lesson using encouragement rather than threat of a penalty. Instructors spent time building rapport with students, which helped them find ways to pique their interest. One effective strategy for keeping students engaged was to review an agenda for the session to establish shared expectations. Furthermore, explaining the purpose of activities helped students remember the lesson. Strategies like rapport-building, attentiveness and attunement, active engagement with the material, and participation with students should be covered with specific examples during future instructor trainings, and programs should employ instructors displaying these skills.

Active Student Learning Activities Student-led session activities promoted higher student engagement. For example, in one observed T-MIA class, the instructor encouraged students seated in a circle to take turns leading breathing practices. As one instructor described, "Excellent session, [student] has taken ownership and likes to lead the meditation piece during our sessions." Furthermore, the HT instructor engaged students by asking questions—akin to the Socratic method—such that students actively generated answers rather than passively absorbing information. These active student learning approaches were effective in promoting participation and on-task time. Integrating more active learning activities is recommended as it increases student engagement and active learning and promotes a learning community.

Insufficient Time to Cover Material Instructors could not cover all the session material within 30 min due to a variety of constraints (e.g., time needed for transition to classes, check ins, distribution of homework sheets, redirecting). One instructor reported, "Student[s] were late coming in. I wasn't able to cover all the topics today. It feels like the time can be short to cover all new topics." They adjusted by decreasing the number of activities or time spent per activity (e.g., reducing a breath practice from 360 to 60 s). Activities requiring more time or student participation were often first to be eliminated. Some instructors tried to "catch up" by covering missed material in the next session, further reducing time for new material in that session. The impact of time constraint-induced adaptations on program effectiveness is unknown and warrants future study, specifically how much dosage is needed to observe hypothesized changes. To promote implementation fidelity, recommendations included (a) allocation of longer time for sessions; (b) revision of manuals to reduce the amount of content per session; and (c) helping instructors with time management to make start-up and ending time more efficient.

Behavior Management Demands One major challenge to implementing programming as intended was disruptive student behavior, such as off-task behaviors during instruction (e.g., playing on phones), interrupting or disrespecting the instructor (e.g., cursing, gossiping), and arriving late or leaving early. Some instructors were more adept at classroom management (e.g., patience, redirecting) than others. Those with better classroom management skills had good rapport with students and earned student respect. Those who struggled with classroom management promoted distraction via unclear expectations, starting late, blurring instructor and student roles, or asking tangential questions. Training future implementers in behavior management techniques may reduce off-task time and the need to redirect students. Instructors may benefit from reviewing session recordings and reflect on sources of disruption (including their role) immediately following a session.

Environmental Disruptions Environmental sources of disruption and distraction also posed challenges. For example, one instructor reported a session was unexpectedly cut short for a school event. Another reported that students missed sessions because teachers encouraged them to engage in other activities. School testing interrupted programming for a week. Changing rooms to accommodate testing or events confuses students and instructors, creates delays, and alters the program context. Securing a space with minimal distractions fosters greater implementation fidelity and student engagement. Loud, communal settings like gyms, cafeterias, and auditoriums should be avoided. Furthermore, offering snacks during programs took time and was distracting; ideally, healthy snacks should be offered as students transition to their next class.

Student Qualitative Perspectives on Factors Influencing Implementation

Four themes related to improving program implementation emerged from student focus groups. First, instructor qualities that promote student engagement and receptivity included personality and relatability. Second, more and varied activities are needed to maintain attention and interest, including active learning opportunities. Third, extending and expanding programming through longer sessions across more weeks, extending programs across multiple grades, and allowing students to attend both programs was anticipated to optimize program impacts. Last, to promote participation, program scheduling should be determined by students' availability and priorities.

Instructor Relational Qualities Are Valuable for Student Engagement Students emphasized the significance of instructor qualities for increasing their receptivity to program content. Desirable instructor characteristics included an engaging personality and ability to build relationships with students. Many students commented on instructors being skilled in communication, relationship building, and group management, being personable and fun, respecting student opinions, and creating a safe space where youth felt comfortable sharing. One HT student described how the instructor motivated the class to listen by demonstrating many of these skills and qualities:

> He's fun and at the same time he gets serious. He knows when to tell us when we need to be quiet cause we've been talking too much... And he's not mean, he don't get smart…

Although the majority of comments about positive instructor experiences were from HT students, some T-MIA students described how their instructor made them feel "more comfortable" and eased them into trying new breath techniques through the use of humor and willingness to participate in the same activities alongside students.

Students also recognized that instructors cared about them and acknowledged their feelings. Some students reported arriving at their program having had a difficult day or expressing negative attitudes, and the instructor tried to improve their mood with humor or inquired how they could help. Students were particularly receptive to positive instructor–student interactions marked by acceptance, ease of discussion, and encouragement to be authentic. As one HT student described, "He just understood us. And he was really like cool and laid back." Other students mentioned non-judgment regarding how students spoke: "He would just be okay with the way we talk."

Students appreciated when instructors promoted safe and comfortable environments for sharing information. A few students noted how instructors attempted to build relationships with students through engaging them in conversations. In contrast, a few students found this bothersome, describing their instructor as too talkative, annoying, and "too close." Most students, however, provided positive feedback on instructors. Their comments highlighted that having an instructor who related well to students improved engagement and participation.

Integrating New Physical and Applied Activities to Supplement Didactic Components Students in both programs mentioned that they enjoyed the program activities but had a desire to incorporate new activities. Students recommended that more physical activities, field trips, presenters, and other similar engaging activities be added to the curricula. Students wanted programs to incorporate more movement-based activities to increase activity during the sessions. One HT student remarked, "Not so much sitting. We should at least get up and do something." Another student described:

> The program was cool for real. Everything we did was a lot of fun stuff. But, I feel like we stay in one spot most of the time. We just sit and listen... I feel like it should be more than we just sat there. We could clean the garden or something... Like you can go outside.

Some T-MIA students suggested including yoga to supplement the breathing techniques they learned. The suggestion to include field trips to visit farms, hospitals, or pharmacies was based on HT students' desire to gain real world experiences related to session content.

Extended and Expanded Programming May Optimize Program Impact Students were enthusiastic about the programs and suggested extending the number of sessions and time allotted for both programs, as well as expanding the programs to other grades. Some students noted "Like more time, we really didn't have enough," and "More sessions, just like longer class periods." Longer time in the program would give them greater familiarity, practice, and comfort with the techniques. As one T-MIA student described: "you get into it more, be more comfortable doing it." Similarly, another student noted how moving quickly through the session activity was not beneficial: "We would have to rush sometimes…It really wasn't too much effective." T-MIA students reported issues with timing as some sessions were shorter than others, and this could have an impact on what they learned.

Participants suggested continuing the programs next year and delivering the programs to different grades each semester. The students acknowledged that higher grades included more academic responsibilities and increased pressures; continuing the programs in subsequent years could be helpful for student wellbeing. As one T-MIA student noted, delivering the program in later grades could help with stress: "12th graders because they get more work and they feel stressed out…more essays…that's why you should include all grades."

Some students recommended the option to switch between T-MIA and HT in mid-semester or every other week, so students could experience both. Participants were curious about what was taught in the other program and were interested in learning something new. For example, a T-MIA student said:

> I think we should switch groups at one point of the time, so both of the groups can know what [the other] was doing. Cause they would get what we've been doing here and we could get what they were doing there.

Program crossover could expose students to more comprehensive knowledge and skill development.

Program Scheduling: Prioritize Students' Availability Participants in both programs expressed a desire to change the scheduling and timing to accommodate their priorities. Weekly program timing varied with some students participating during lunchtime while others attended during resource periods. This situation created competing demands, interfered with attentiveness, and ultimately resulted in suboptimal participation. As one HT student described:

> We had other activities. Sometimes we had to leave early. Sometimes we started missing work and stuff. Groups [were] going over and we had to leave class early. So I will be telling [instructor], "I can't come today." [He's say] "you're important here." I be like, "Well my class is more important than here."

Some T-MIA students recommended programming during morning hours to energize students for the rest of their

day. One student noted, "I feel like it would have been better if we did it in the morning because people when they come in, in the morning they're grumpy. Because it's early in the morning and everybody wants to go to sleep." Having programming in the morning would enable students to apply what they learned during their school day. In contrast, other T-MIA participants proposed end-of-the-day timing to foster sharing feelings and sustain calmness as they transitioned into afterschool activities. When offered earlier in the day, some students explained that stressful high school experiences and demands during the school day shortened program effects, making them temporary. As one T-MIA student explained:

> I feel like we would do it and then it would be like we all calm and collected now. Then we go right back outside and it's like 'oh yeah we back in high school.' … you're right back into everybody yelling in the hallway… It felt like it was temporary, like it didn't last that long because your mind is getting distracted by school right afterwards.

Differential Implementation Perceptions Among Instructors, Observers, and Students

Fidelity and—Instructor and Observer Ratings and Observer Notes Triangulation

Mean instructor-reported scores were positively skewed for core activities ($M = 2.5$; $SD = 0.67$), processes ($M = 2.7$; $SD = 0.62$), and overall quality ($M = 3.6$; $SD = 0.56$) with overall quality significantly differing between programs whereby T-MIA instructors rated themselves higher than the HT instructor ($M = 3.8$ vs. 3.59, $SD = 0.33$ vs. 0.56; $t(26) = -2.11$, $p = 0.044$). When comparing total instructor to observer scores, paired-samples t tests revealed instructors scored themselves higher on activities, processes, and quality (Table 3), consistent with other studies that found implementers rated their fidelity higher than observers (Gould et al., 2014). This demonstrates an instructor scoring effect. Observer scores, however, did not significantly differ across programs or schools for core activities, core processes, or overall quality. Using the gold standard of ratings by observers revealed no program implementation differences despite instructor perceived differences.

Regarding fidelity, observer notes revealed that some instructors were more proficient at class management, which left more time to cover curriculum material. These instructors were also more adept at adapting material to be more engaging (e.g., active student learning activities, participating with students) in ways that diverged from the manual. These two aspects may have differing associations with fidelity; more time on material (greater fidelity) coupled with more adaptations (less fidelity) may explain why observer quantitative ratings did not result in significant fidelity differences across instructors. This suggests fidelity rating measures could or should be designed to be sensitive to adaptations and instructor qualities.

Integrated Findings Across Instructors, Observers, and Students

Aligning session observations and instructor reports with focus group data provides a richer understanding of what promotes and hinders implementation. Theme alignment across data sources and informants is presented in an integration matrix (Table 4). Four themes aligned across instructor session notes, observer session notes, and student focus group data. First, qualities reported to promote greater student engagement and effective instruction are consistent and complementary across informants. Students described these qualities as personality attributes, whereas instructors and observers described them as instruction qualities, but the key characteristics that emerged overlapped across informants. Second, students were receptive to active learning opportunities and recommended future programs provide more of an integration. Observers and instructors also reported the highest student engagement during active student learning activities.

Third, students, instructors, and observers reported that the session duration was not sufficient to cover program material in depth and that sessions were rushed. All informants agreed program benefits would be enhanced through longer sessions. Students also recommended extending programming across more sessions and other grades, as well as student participation in both programs (crossover). Fourth, student participation was disrupted by multiple environmental sources. To optimize participation and sustain program effects, students recommended prioritizing their scheduling preferences, whereas instructors and observers highlighted the impact of physical space constraints on implementation.

Table 3 Mean fidelity scores for activities, processes, and overall quality by rater

	Instructor		Observer				
Items	Mean	SD	Mean	SD	t value	df	p value
Core activities	2.5	0.67	2.0	0.73	0.18	28	0.015
Process	2.7	0.62	2.4	0.51	0.16	28	0.075
Quality	3.6	0.56	2.9	0.88	0.21	27	0.004

Table 4 Integration matrix of themes by informants

	Student themes	Instructor and observer themes	Alignment/integration
Implementation Promoting	*Instructor skills and personal qualities* • Communication • Relationship building • Group management • Personable, fun, respectful • Create safe space	*Engaging and attentive instruction* • Rapport • Attunement • Active instruction • Participation	Instructor and instruction qualities reported to promote greater student engagement and effective instruction are consistent and complementary. Students described these as personality or individual qualities whereas instructors and observers described them as instruction qualities.
	Integrating new physical and applied activities to supplement didactic components	*Active student learning activities*	Active learning opportunities were requested by students and were observed to promote student engagement during sessions
Implementation Challenging	*Extending and expanding programming*	*Insufficient time to cover material*	Session duration (25 min) was insufficient for delivering programming as intended. Sessions were rushed. More time for program delivery was recommended (e.g., longer sessions, more sessions, additional grade levels, multiple years, and program crossover)
	Program scheduling	*Environmental disruptions*	Student participation was disrupted by multiple environmental sources. To optimize student participation and sustain program effects, students focused on prioritizing their availability and timing preferences whereas instructors and observers detailed physical space and conflicting school events constraints on implementation
	Not mentioned	*Behavioral management demands*	Instructors and observers noted how behavioral management demands interfered with program implementation. Both student- and instructor-driven distractions were described

Last, instructors and observers described how time spent on behavior management disrupted program implementation. When comparing on- and off-task time across instructors, statistically significant differences were found ($F(2, 3) = 13.44$, $p = 0.03$) with ranges of 2.96 to 11.12 min off task. Student-driven distractions included gossiping, disrespect, interruption, and tardiness. Instructor-driven distractions included unclear expectations, late class start time, blurred roles, and off-topic discussions. Students described management skills as an effective instructor quality (Theme 1) even though student data did not explicitly address behavioral management demands. This may be due to the lack of explicit requests for this information, reluctance to discuss such challenges in a group setting for fear of stigma or retribution from peers, or lack of awareness.

Discussion

This study addressed two research questions and examined four of the eight key elements of program implementation previously identified—participant responsiveness, dosage, fidelity, and quality (Durlak & DuPre, 2008) in the context of health and wellness program delivery for ninth graders in urban schools serving low-income families. By documenting and triangulating the perspectives of instructors, independent observers, and students on school-based prevention program implementation, this study presents an in-depth depiction of factors that facilitate or impede effective program implementation. Study findings contribute several insights that may guide effective implementation.

First, to determine factors that influence implementation of these two programs, qualitative data were reported by and later integrated across informants, including instructor self-reported fidelity logs, observer-rated fidelity field notes, and student focus group discussion points. Our findings suggest specific instructor and program-related features exerted an impact on program implementation. Instructor-specific influences included skills, personal qualities, teaching strategies, and classroom management experiences that promote participant responsiveness. Furthermore, active learning opportunities were requested by students and were observed to promote student engagement during sessions.

Dosage and *fidelity* were impacted by program-related factors like scheduling, session numbers and length changes including unexpected session cancellation, and environmental disruptions. Ultimately, minimizing competing demands in terms of scheduling, distractions, frequency, and duration will promote utility, translation, and sustainability of skills

learned. Instructors, observers, and students alike reported insufficient time to implement, absorb, and sustain lessons and material throughout the day and as students transitioned into afterschool environments. Competing demands—such as school events or student field trips—placed pressure on instructors to make program modifications to fit sessions into condensed time periods or into fewer sessions. Behavioral management demands emerged as a factor impacting effective program implementation.

Second, to shed light on how implementation factors may be differentially perceived by various reporters (e.g., instructors, observers, students), data from multiple informants were triangulated and explored for commonalities and differences across informants. Our findings suggest a self-report instructor effect such that different instructors had varying perceptions about their instruction and the overall quality of their program sessions, whereas observer reports did not reveal instructor effects on activities, processes, and quality. This measurement effect has implications for results of studies that rely exclusively on instructor reports. Systemically integrating health education and mindfulness skills may reduce some implementation challenges reported by implementers, observers, and students, thereby optimizing contexts that support maximizing program effects.

Some of these findings are consistent with previous research. For instance, prior work has assessed the facilitating roles of quality instructors and importance of program integration with school schedules (Gottfredson & Gottfredson, 2002). Other findings have not been commonly addressed in the prevention science literature. For example, the desirability of learning promoted by physical activities and opportunities for real-world applications was supported in this study by both instructor measures and students' focus group discussions. Triangulation across the data sources served to validate this finding and minimize reporter biases. These findings suggest that knowledge-based programs may benefit from experiential and youth-led learning.

With respect to fidelity and quality of program implementation measures, this study highlights the complexity of fidelity self-ratings and suggests the importance of including independent observational ratings. Self-rated fidelity varied across instructors, with some under-reporting and others over-reporting adherence to program manuals, whereas, according to observer ratings, both programs were implemented with similar levels of fidelity and quality. This finding suggests an instructor or rater effect rather than an implementation effect. Reliance on instructor-rated fidelity to adjust quantitative models is particularly problematic when the instructor is held constant within the program, resulting in confounding of program and instructor effects. This finding further emphasizes the importance of collecting multiple measures of fidelity. Without comparison of instructor and observer fidelity ratings, confounding could not be identified, and quantitative adjustments based on these ratings would result in different interpretations.

Examining data from three perspectives—implementers, independent observers, and students— and leveraging mixed methods are study strengths. Limitations, however, warrant noting. First, students who participated in focus groups are a subset (64%) of the full sample of participants. A more ideal scenario, albeit impractical, would have been to include the entire sample in the focus group sessions to avoid criticism of selection bias. We are confident, however, that selection bias was minimized because the subset and full samples did not differ with respect to demographic characteristics or program attendance, engagement, or disruptiveness. Second, a random subset of 51 sessions were selected a priori to be video recorded. Unfortunately, only 32 had sufficient audio and visual quality for coding and, of these, 29 had implementer ratings, with each program having a comparable number of usable recordings. The reduced number of usable videos resulted primarily from technical problems and canceled sessions. Third, perspectives of administrators and teachers were not captured. These informants may have had limited capacity to report on daily implementation successes, challenges, and quality, although their input would have provided broader perspectives on program feasibility and acceptability. Previous research documents the importance of school-implementer-researcher partnerships for effective implementation (Mendelson et al., 2013). Effective approaches for gaining additional perspectives are needed. Last, the findings from this study are not generalizable or comparable to other samples and settings. However, our results pertain to very low income, urban, minoritized adolescents, a group often targeted by preventive interventions and thus may be useful for program developers and implementers to consider in future studies.

Conclusions and Implications

Rigorous measurement and reporting of fidelity of implementation remain limited (e.g., Lewis et al., 2015), despite consensus on its importance. One challenge is self-report bias, as reflected by the differences in ratings of instructors compared to observers in this and previous studies (Gould et al., 2014). Although fidelity of implementation measures has improved over time (Gould et al., 2016), additional training of instructors and normalizing expectations that fidelity levels will fall short of perfect may help in this regard. Adding ratings for adaptations and modifications, in addition to solely fidelity measures, could be reflective of the level of acceptance and the expectation that positive changes will occur. Furthermore, assessing instructor characteristics that predict reporting styles may be of importance. Last, ensuring programs have similar complexity and demands (e.g., set up, clean up, activity vs. didactic instruction) and amount of new material to cover per session will promote equitable

ratings. Fidelity measure development and instructor self-report training merit ongoing attention.

In addition to fidelity, findings suggest that effective program implementation would be optimized through adequate time for both delivering content and fully engaging students, consistent and minimal schedule disruptions, and desirable instructor qualities. These factors should be considered in the design of future school based-prevention programs and training of program facilitators in various skills, including behavior management and active listening techniques.

In this study, incorporating perspectives of multiple informants and using mixed methods were methodological strengths that increased confidence in identified themes. While these methods can be time-consuming and resource intensive, we argue that they are as important as program outcome measures and enhance both interpretability of outcome findings and potential for intervention sustainability and impact. We encourage other prevention researchers to collect and analyze both qualitative and quantitative data from multiple informants.

Acknowledgements We are very grateful for the support of participating schools and program implementers and students' time and willingness to participate in this study. We appreciate the high level of support from research associates and assistants: Jessica Stavig, Yuanfang Liu, Jeffrey Krick, Marcus Nole, Rachel Dows, Violet Odom, Alex Welna, Steven Sheridan, and Drs. Laura Clary and Qing Duan.

Funding This work was supported by the National Center for Complementary and Integrative Health (grant number: R61AT009856; MPIs: Fishbein, Mendelson).

Data Availability De-identified aggregate data analyzed for this paper are available from the corresponding author upon reasonable request.

Declarations

Conflict of Interest The authors declare no conflicts of interest.

Ethics Approval This study was performed in accordance with the ethical standards as laid down in the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards. Consenting and procedures were approved by the institutional review board at the Johns Hopkins Bloomberg School of Public Health, and reliance agreements were established with the Pennsylvania State University, the University of Cincinnati, and the University of Illinois.

Consent to Participate Parental written consent and student written assent was obtained.

Competing Interests The authors declare no competing interests.

Disclaimer The opinions and conclusions expressed are solely the authors' and should not be construed as representing the opinions of NIH or any agency of the Federal Government. NIH did not have a role in study design; data collection, analysis, and interpretation; and writing or submission for publication.

References

Aschbrenner, K. A., Kruse, G., Gallo, J. J., & Plano Clark, V. L. (2022). Applying mixed methods to pilot feasibility studies to inform intervention trials. *Pilot and Feasibility Studies, 8*(1), 1–13. https://doi.org/10.1186/s40814-022-01178-x

Albright, K., Gechter, K., & Kempe, A. (2013). Importance of mixed methods in pragmatic trials and dissemination and implementation research. *Academic Pediatrics, 13*(5), 400–407. https://doi.org/10.1016/j.acap.2013.06.010

Bauer, M. S., & Kirchner, J. (2020). Implementation science: What is it and why should I care? *Psychiatry Research, 283*, 112376. https://doi.org/10.1016/j.psychres.2019.04.025

Blase, K., & Fixsen, D. (2013). *Core intervention components: Identifying and operationalizing what makes programs work* (ASPE Research Brief). OHSP, US Department of Health and Human Services (USDHHS). Retrieved July 13, 2020, from http://aspe.hhs.gov/hsp/13/KeyIssuesforChildrenYouth/CoreIntervention/rbCoreIntervention.cfm

Boyatzis, R. E. (1998). *Transforming qualitative information: Thematic analysis and code development.* Sage.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3rd ed.). Sage Publications.

Cutbush, S., Gibbs, D., Krieger, K., Clinton-Sherrod, M., & Miller, S. (2017). Implementers' perspectives on fidelity of implementation: "Teach every single part" or "Be right with the curriculum"? *Health Promotion Practice, 18*(2), 275–282.

Dane, A. V., & Schneider, B. H. (1998). Program integrity in primary and early secondary prevention: Are implementation effects out of control. *Clinical Psychology Review, 18*, 23–45. https://doi.org/10.1016/S0272-7358(97)00043-3

Dariotis, J. K., Cluxton-Keller, F., Mirabal-Beltran, R., Gould, L. F., Greenberg, M. T., & Mendelson, T. (2016). "The program affects Me 'Cause it gives away stress": Urban students' qualitative perspectives on stress and a school-based mindful yoga intervention. *Explore, 12*(6), 443–450. https://doi.org/10.1016/j.explore.2016.08.002

Dariotis, J.K., Mabisi, K., Jackson-Gordon, R., Rose, E., Fishbein, D.H., & Mendelson, T. (under review). Perceived Benefits of Mindfulness and Health Education Programs for Minoritized Adolescents: A Qualitative Analysis. *Mindfulness*.

Dariotis, J. K., Mirabal-Beltran, R., Cluxton-Keller, F., Feagans Gould, L., Greenberg, M. T., & Mendelson, T. (2017). A qualitative exploration of implementation factors in a school-based mindfulness and yoga program: Lessons learned from students and teachers. *Psychology in the Schools, 54*(1), 53–69. https://doi.org/10.1002/pits.21979

Davidov, D. M., Hill, K., Bush, H. M., & Coker, A. L. (2020). The green light for Green Dot: A qualitative study of factors influencing adoption of an efficacious violence prevention program in high school settings. *Violence against Women, 26*(12–13), 1701–1726.

Deatrick, J. A., Kazak, A. E., Madden, R. E., McDonnell, G. A., Okonak, K., Scialla, M. A., & Barakat, L. P. (2021). Using qualitative and participatory methods to refine implementation strategies: Universal family psychosocial screening in pediatric cancer. *Implementation Science Communications, 2*(62), 1–10. https://doi.org/10.1177/1077801219886377

Dobbie, F., Purves, R., McKell, J., Dougall, N., Campbell, R., White, J., Amos, A., Moore, L., & Bauld, L. (2019). Implementation of a peer-led school based smoking prevention programme: A mixed

methods process evaluation. *BMC Public Health, 19*(1), 1–9. https://doi.org/10.1186/s12889-019-7112-7

Durlak, J. A., & DuPre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology, 41*(3–4), 327–350. https://doi.org/10.1007/s10464-008-9165-0

Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. B. (2011). The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions. *Child Development, 82*, 405–432. https://doi.org/10.1111/j.1467-8624.2010.01564.x

Ferber, T., Sileo, A., & Wiggins, M. E. (2019). *Advancing the use of core components of effective programs*. Forum for Youth Investment. Retrieved July 13, 2020, from https://forumfyi.org/knowledge-center/advancing-core-components/

Fixsen, D., Blase, K., Metz, A., & Van Dyke, M. (2015). Implementation Science. In J.D. Wright (Ed.), *International encyclopedia of the social and behavioral sciences* (2nd ed.). Elsevier.

Gottfredson, D. C., Cook, T. D., Gardner, F. E. M., Gorman-Smith, D., Howe, G. W., Sandler, I. N., & Zafft, K. M. (2015). Standards of evidence for efficacy, effectiveness, and scale-up research in prevention science: Next generation. *Prevention Science, 16*(7), 893–926. https://doi.org/10.1007/s11121-015-0555-x

Gottfredson, D. C., & Gottfredson, G. D. (2002). Quality of school-based prevention programs: Results from a national survey. *Journal of Research in Crime and Delinquency, 39*(1), 3–35. https://doi.org/10.1177/002242780203900101

Gould, L. F., Dariotis, J. K., Greenberg, M. T., & Mendelson, T. (2016). Assessing fidelity of implementation (FOI) for school-based mindfulness and yoga interventions: A systematic review. *Mindfulness, 7*(1), 5–33. https://doi.org/10.1007/s12671-015-0395-6

Gould, L. F., Mendelson, T., Dariotis, J. K., Ancona, M., Smith, A. S. R., Gonzalez, A. A., Smith, A. A., & Greenberg, M. T. (2014). Assessing fidelity of core components in a mindfulness and yoga intervention for urban youth: Applying the CORE process. *New Directions for Youth Development, 2014*(142), 59–81. https://doi.org/10.1002/yd.20097

Hulleman, C. S., & Cordray, D. S. (2009). Moving from the lab to the field: The role of fidelity and achieved relative intervention strength. *Journal of Research on Educational Effectiveness, 2*(1), 88–110. https://doi.org/10.1080/19345740802539325

Koffel, E. E., & Hagedorn, H. H. (2020). Provider perspectives of implementation of an evidence-based insomnia treatment in veterans affairs (VA) primary care: Barriers, existing strategies, and future directions. *Implementation Science Communications, 1*(1), 1–10. https://doi.org/10.1186/s43058-020-00096-4

Kozica, S. L., Lombard, C. B., Harrison, C. L., & Teede, H. J. (2016). Evaluation of a large healthy lifestyle program: Informing program implementation and scale-up in the prevention of obesity. *Implementation Science, 11*(1), 1–13. https://doi.org/10.1186/s13012-016-0521-4

Lewis, C. C., Fischer, S., Weiner, B. J., Stanick, C., Kim, M., & Martinez, R. G. (2015). Outcomes for implementation science: An enhanced systematic review of instruments using evidence-based rating criteria. *Implementation Science, 10*(1), 155–172. https://doi.org/10.1186/s13012-015-0342-x

Mendelson, T., Dariotis, J. K., Gould, L. F., Smith, A. S., Smith, A. A., Gonzalez, A. A., & Greenberg, M. T. (2013). Implementing mindfulness and yoga in urban schools: A community-academic partnership. *Journal of Children's Services, 8*, 276–291. https://doi.org/10.1108/JCS-07-2013-0024

Meyers, D. C., Durlak, J. A., & Wandersman, A. (2012). The quality implementation framework: A synthesis of critical steps in the implementation process. *American Journal of Community Psychology, 50*(3), 462–480. https://doi.org/10.1007/s10464-012-9522-x

Morgan, D. L. (2019). Commentary—After Triangulation, What Next? *Journal of Mixed Methods Research, 13*(1), 6–11. https://doi.org/10.1177/1558689818780596

Onwuegbuzie, A. J., & Collins, K. M. (2007). A typology of mixed methods sampling designs in social science research. *Qualitative Report, 12*(2), 281–316. https://doi.org/10.46743/2160-3715/2007.1638

Palinkas, L. A., Aarons, G. A., Horwitz, S., Chamberlain, P., Hurlburt, M., & Landsverk, J. (2011). Mixed method designs in implementation research. *Administration and Policy in Mental Health and Mental Health Services Research, 38*(1), 44–53. https://doi.org/10.1007/s10488-010-0314-z

Palinkas, L. A., Mendon, S. J., & Hamilton, A. B. (2019). Innovations in mixed methods evaluations. *Annual Review of Public Health, 40*(1), 423–442. https://doi.org/10.1146/annurev-publhealth-040218-044215

Proctor, E., Silmere, H., Silmere, H., Raghavan, R., Raghavan, R., Hovmand, P., Aarons, G., Bunger, A., Griffey, R., & Hensley, M. (2011). Outcomes for implementation research: Conceptual distinctions, measurement challenges, and research agenda. *Administration and Policy in Mental Health and Mental Health Services Research, 38*(2), 65–76. https://doi.org/10.1007/s10488-010-0319-7

VERBI Software. (2021). MAXQDA 2022 [computer software]. Berlin, Germany: VERBI Software. Available from maxqda.com.