Child & Youth Care Forum
https://doi.org/10.1007/s10566-020-09558-1

ORIGINAL PAPER

Measuring the Fidelity of a School-Based Yoga and Mindfulness Curriculum for Youth: A Transdisciplinary Feasibility Study

Flint M. Espil[1] · John P. Rettger[1] · Carl F. Weems[2] · Erin L. Neill[2] · Victor G. Carrion[1]

Abstract

Background Although mindfulness and yoga interventions for youth continue to gain support, few studies have proposed a systematic way to study the implementation and fidelity of these approaches.

Objective The current study developed and tested the feasibility of measurement procedures for a yoga and mindfulness curriculum implemented in schools through a community partnership. This was accomplished in two studies.

Method In study 1 we created and analyzed the psychometric properties of a fidelity observation tool using data from 165 observations of students receiving a district-wide school yoga and mindfulness intervention. The findings were subsequently refined and implemented to collect 91 additional observations in study 2.

Results In study 1, a principle components analysis revealed a 3-factor solution across items, with 10 of 17 items reliably rated but only 4 of those 10 items rated as satisfactory or better fidelity. The fidelity observation tool was then revised and implemented in study 2, with 13 of 17 items reliably rated and 5 meeting satisfactory or better fidelity.

Conclusions The findings from both studies indicated that yoga and mindfulness interventions for youth can be feasibly implemented and reliably measured within school settings, although there were some limitations in measuring both the content and process of such interventions.

Keywords Yoga · Mindfulness · Youth · School · Fidelity · Wellness

Mindfulness and yoga practices have potential for the prevention of youth emotional and behavioral problems as well as for fostering wellness and building resilience (Greenberg

Dr. Greg Welk, the Special Issue Guest Editor, served as the Action Editor for this manuscript during the peer review process.

✉ Victor G. Carrion
vcarrion@stanford.edu

[1] Stanford Early Life Stress and Resilience Program, Stanford University School of Medicine, Stanford University, Stanford, CA, USA

[2] Department of Human Development and Family Studies, Iowa State University, Ames, IA, USA

Published online: 08 July 2020

and Harris 2012). This area of research is interdisciplinary by its very nature of integrating multiple intellectual traditions. These include perspectives, concepts, theories, and methods from domains including social emotional learning (e.g., Lawlor 2016; Shonert-Reichl et al. 2017), educational psychology (Roeser et al. 2013), and mindfulness research (Felver et al. 2016; Taylor et al. 2016). The confluence of these approaches into a universal set of practices has the potential to generate new models and methods not traditionally used in implementation research. Yoga and mindfulness practices may also have implications for stakeholder input and involvement. In this paper we draw from a transdisciplinary project involving direct collaborations and involvement of non-academic partners for enabling youth health and wellness in community and school settings.

Existing Evidence Base

A recent review (Carsley et al. 2018) of randomized controlled trials of yoga at elementary through high schools identified nine studies that met acceptable standards for inclusion. Three studies favored yoga (Hagins et al. 2013; Khalsa et al. 2012; Noggle et al. 2012), three favored control conditions (Haden et al. 2014; Sarokte and Rao 2013; Verma et al. 2014), and one study found no treatment effects in either condition (Telles et al. 2013). Reviews of yoga and mindfulness conclude that although findings are promising, the limited data on the fidelity of implementation preclude strong inferences regarding efficacy (Deitz and Rajan 2017; Greenberg, and Harris 2012; Kallapiran et al. 2015). For example, in a review of 48 mindfulness and/or yoga studies in North America, fewer than 10% of the studies outlined core program components or referenced a formal theory of action, 63% assessed at least one dimension of fidelity of implementation (FOI), 20% included an observational assessment, 10% monitored some dimension of dosage within a control/comparison condition, and 13% reported some FOI dimension (e.g. all studies reported on dosage) and outcomes (both significant and nonsignificant findings; Gould et al. 2016). The authors noted that more rigorous elements of FOI were rarely included, if at all. Examples of more rigorous FOI needed in studies of yoga and mindfulness include the articulation of a programmatic logic model, the division of structural and process dimensions of programming, an expanded assessment of other related variables such as program adherence, quality, participant responsiveness, the utilization of multiple data points for a single FOI variable, reporting on reliability and validity of FOI measures, a priori cutoffs for the quality of FOI, and reports of how program implementation was tailored during the delivery stage.

As mindfulness and yoga interventions continue to grow in popularity, systematic studies of fidelity of implementation are needed to ensure consistent and effective implementation of interventions (Gould et al. 2016). The goals of the current study were to review the literature on the FOI procedures across yoga and mindfulness interventions and then develop, test, and refine a systematic process to measure program implementation within a district-wide yoga curriculum for youth in schools.

Given the high level of variance across interventions, we limited our review to controlled studies conducted in U.S. schools utilizing yoga as the primary intervention for elementary and middle school students and uncontrolled yoga studies that included illustrative fidelity procedures, similar to that of Gould et al. (2016). Readers interested in a more comprehensive review should consult Khalsa and Butzer (2016). Of the 30 school-based yoga studies in the United States and the 15 in India since 2005, half of the curricula occurred in elementary schools. These curricula used either minimal or no FOI

procedures, minimal or no reporting of results, largely focused on giving general descriptions of implementation procedures, and included vague or absent descriptions of instructor training backgrounds (see Table 1). Yoga training in schools comprised teaching yoga postures, breathing techniques, relaxation training, meditation, and mindfulness didactics (Gould et al. 2016), yet there was considerable variability in how these components were delivered (Khalsa and Butzer 2016).

Study sample sizes ranged from 30 to 209 students, 4–33 weeks in duration, and spanned anywhere between 8 and 100 sessions of 4–90 min in length. Studies also varied considerably in the style of yoga implemented. Although many used traditional approaches such as Ashtanga (Haden et al. 2014), Vinyasa (Hagins et al. 2013), and Hatha Yoga (Sarokte and Rao 2013; Telles et al. 2013) others implemented alternative curricula (Khalsa et al. 2012; Ramadoss and Bose 2010; White 2012). None of these studies examined how specific intervention and implementation factors might relate to outcomes.

Partnerships

The partnership for the current study began following initial collaborative work between the Stanford Early Life Stress and Resilience team and local schools in northern California. This weekly intervention gained the attention of the local district as well as Pure Edge, an organization dedicated to promoting youth health and wellness through the educational system. Pure Edge then invited both research and school administrators to view and assess their intervention as applied in other school districts, which garnered further interest from all partners. School administrators and principal investigators subsequently agreed upon an acceptable research design, including ways to measure fidelity and outcomes. Further, all parties agreed upon making implementation of the curriculum sustainable beyond the scope of the project should districts choose to maintain the intervention. To increase the successful implementation of this project, all parties agreed that community leaders within the school districts should be included to discuss the project, the curriculum, and purpose of the study. These informants were also involved in outreach efforts to ensure families were well informed prior to consenting. Finally, a multi-method evaluation was designed to study the impact of the curriculum implementation including the areas of cognitive, sleep, physiology (e.g. heart rate, etc.), brain imaging, and behavioral science.

The intervention selected for this study follows the pioneering FOI work by Gould et al. (2014) that demonstrated the importance of identifying "Core Activities and Core Processes" (p. 64). Core Activities are the curriculum "activities and practices" (p. 64) and Core Processes are "the manner in which the program is delivered" (p. 65). Examples of Core Activities identified in the Holistic Life Foundation, Inc. program they investigated included breathing, yoga, review of previous learning, discussion and assignments, and silent reflection. Core Processes included engaged learning through encouraging students to apply the material in their own lives, classroom management, modeling human compassion, and reinforcement of learning through continual review of concepts and poses. They assessed these utilizing three different methods: (1) session inventory forms completed by program instructors, (2) independent observations of videotaped sessions, and (3) focus groups with a "select group" (p. 69) of participants and school personnel. Results indicated acceptable participant engagement and program adherence, but differences between observer and instructor ratings on session delivery quality.

Table 1 Fidelity and curriculum assessment on current studies

References	Participants	TSS	Method	Curriculum	Setting	Weeks[a]	Sessions[a]	Minutes[a]	FOI measures and outcomes/relevant implementation information
White (2012)	ES	155	RCT	Mindful Awareness for Girls through Yoga	AS	8	8	60	Use of an intervention manual Instructor journal Intervention checklist Written instructions Homework with picture and audio instructions Feedback during the sessions
Fishbein et al. (2016)	HS/MS	85	RCT	Mindful yoga intervention	DS	7	20	50	Yoga Alliance registered instructors Classroom environment and equipment described
Mendelson et al. (2010)	ES	97	RCT	Holistic Life Foundation (HLF)	DS	12	48	45	HLF instructors of comparable racial and ethnic background to students; classroom environment noted (gym); Authors stated that fidelity not evaluated
Haden et al. (2014)	MS	30	RCT	Ashtanga-informed yoga[b]	DS	12	36	90	Teachers were 200-h yoga certified and with 3+ years' experience with curriculum
Hagins et al. (2013)	MS	30	RCT	Vinyasa Yoga[b]	DS	15	45	50	Teachers were 200-h yoga certified and 3+ years' experience and an additional 100-h of training specific to the yoga curriculum; Yoga curriculum complied with NY State Standards for Physical Education

Table 1 (continued)

References	Participants	TSS	Method	Curriculum	Setting	Weeks[a]	Sessions[a]	Minutes[a]	FOI measures and outcomes/relevant implementation information
Bergen-Cico et al. (2015)	MS	144	RCT	YogaKidz	DS	33	100	4	Teacher was a certified public-school teacher and completed both the 30-h YogaKids training program and was a 200-h Yoga Alliance registered yoga teacher with ongoing personal practice Use of yoga pose cards with instructions for each posture and the objective
Quach (2016)	MS	172	RCT	Shanti Generation Yoga	DS	4	8	45	Teachers were trained in child and adolescent yoga, 200-h certified, ongoing personal practice of 6–10 years and were teaching in private studios DVD provided for students and they were encouraged to practice daily for 15–30 min and to keep a home practice log but no analysis was conducted on logs

Table 1 (continued)

References	Participants	TSS	Method	Curriculum	Setting	Weeks[a]	Sessions[a]	Minutes[a]	FOI measures and outcomes/relevant implementation information
Frank et al. (2014)	HS	49	UCT	Transformative Life Skills (TLS)	DS	16	48	30	Certified yoga teachers with TLS training and certification, Fidelity was measured by regular supervision of instructors by program developers, regular review of implementation checklists, and observation. Instructor completed Fidelity checklist confirming lesson component delivery, overall student engagement rating, and a reflection on lesson delivery quality. After a unit was completed, lesson content with poor quality delivery or low engagement were repeated. Fidelity of implementation for all lessons was greater than 80%
Butzer et al. (2017)	MS	209	RCT	Kripalu Yoga In Schools (KYIS)	DS	24	32	35	Certified lead yoga teachers with advanced training in KYIS yoga (60-h), attendance at a research training session to promote fidelity and research protocol adherence and yoga certified at the 200-h level. Session logs and notes were kept by instructors, this data was not reported on; yoga curriculum integrated with PE classes

Table based on Gould et al. (2014) and can be referred to for more information on these and other yoga studies more broadly

MS middle school, *RCT* randomized controlled trial, *NRT* non-randomized controlled trial, *UCT* uncontrolled trial, *TSS* total sample size of all groups, *DS* during school, *AS* after school

[a]Values may be estimates based on information in manuscripts

[b]Connotes a general style of yoga rather than a specific curriculum

Following our review of previous yoga and mindfulness FOI methods we aimed to (1) address the gaps in fidelity measurement, (2) begin the process of creating a standardized approach to measuring yoga program implementation, and (3) provide a framework for others aiming to conduct large-scale program implementation of mental and behavioral health curricula in schools. To this end, we examined the feasibility of developing, testing, and implementing fidelity monitoring practices in a large-scale, longitudinal trial of a yoga and mindfulness curriculum (Pure Power) in public schools. We accomplished this task in two phases. In study 1, we piloted a fidelity observation tool in a small sample of school-based classes. The results from study 1 were subsequently used to refine and implement the observation tool across another sample of school-based classes. In addition to establishing the feasibility and testing of a standardized fidelity measure, and consistent with previous work, we hypothesized that the components of the yoga and mindfulness curriculum would fall into two categories: measures of yoga process and measures of yoga content.

Study 1

The goal of study 1 was to develop a fidelity observation tool. As noted we partnered with a public-school district in Northern California. We hypothesized that most aspects of the Pure Power Curriculum could be observed and coded in a reliable manner, and that instructors would implement with fidelity curriculum requirements (i.e., average scores greater than 2 out of a possible 4) for each item deemed reliably observed.

Method

Pure Power

Pure Power was a 3-year longitudinal evaluation of a district-wide K-8 implementation of a yoga and mindfulness curriculum in a suburban public-school district serving approximately 4200 students in Northern California. Students in the district identify as Latino (79%), African American (10%), Pacific Islander (9%) and other (2%). The school district serves a historically underserved, low-SES, high stress, and trauma-impacted community. This study was approved by the principal investigator's Institutional Review Board and complies with the guidelines of the Helsinki Declaration, and study personnel did not have any conflicts of interest with Pure Power or its employees.

Participants

Pure Power Instructors. The manualized Pure Power curriculum (pureedgeinc.org/curriculum) was delivered to students twice weekly as an assigned class by yoga instructors (n = 21) employed by the school district. Yoga instructors were required to have a background in yoga practice, including a 200-hour Yoga Alliance accredited teaching certificate (Registered Yoga Teacher RYT® 200, Yoga Alliance) and previous experience working with youth. Experience working with youth was defined as any experience teaching yoga to youth within or outside of school settings. At the initiation of the program, each of the instructors completed a 5-day training followed by weekly professional development trainings to ensure ongoing quality of instruction and skill development. These trainings

occurred weekly for 18 months (with breaks during the summer when schools were not in session), at which point they were titrated to occur monthly for the remaining 2 years of the project. The initial training and subsequent professional development meetings were conducted by study personnel (JR, second author) and five Pure Power employees credentialed by their respective schools of yoga practices. Examples of topics addressed during these mentoring sessions include behavioral management, posture instruction, and time management. Instructors were also invited to regular district meetings for school teachers and administration to facilitate integration into the school culture. Report from weekly meetings and from schools indicated no problems with regard to instructor attendance.

Rating Team. The pilot research team included researchers within a child and adolescent psychiatry research program. The fidelity rating team was led by a psychologist (second author) who was also an Experienced Yoga Alliance Registered Yoga Teacher (E-RYT ® 200). The team leader attended the trainings on the Pure Power curriculum and provided training to the rating team through initial meetings and frequent team debriefing sessions.

Procedure

Class sessions were an average of 35 min (median) and were held in designated school classrooms with yoga mats. The number of classes provided daily by instructors varied between 4–8 over the course of the project. Fidelity measurements were taken from the moment the youth entered the class and continued until all youth finished the session and exited the classroom. Observation forms were then collected and stored in a secure location for data analysis.

Opening Breathing

Students were initially instructed in a brief breathing practice prescribed from the curriculum. Examples of breathing practices include anchor breathing (i.e., placing hands on a body part, such as the belly or chest, and breathing to feel the body move with the breath) and more active breathing practices (e.g., such as Ocean breathing, which involves slightly closing the back of the throat, breathing in through the nose, and breathing out through the nose with an audible exhale sounding like an ocean wave). Following opening breathing, students would either transition to movement, rest, or content instruction.

Content Instruction

The curriculum included brief content instruction of up to 5 min. Content themes included different "powers", such as the power to be calm, the taming of temper, focus, and cultivating a growth mindset among others. Neuroscience lessons were also embedded in these modules. The specific delivery of the content was flexible in that it could occur as an independent segment of the session or occur while students were engaged in yoga movement or rest. For the purposes of fidelity measurement, when content instruction occurred during movement or rest, content minutes were not logged but allocated to the other practice components (e.g., posture).

Posture/Movement Instruction

Most of the session time was allocated to movement and posture instruction. Specifically, instructors were given a teaching guideline of allocating approximately 70% of class time

to teaching breathing and posture sequences from the manual. Although yoga sequences varied according to requirements for a given session, the general sequence was as follows: warm-up postures, followed by standing or balancing poses, floor postures and backbends, and closing poses that ultimately led to a final relaxation posture.

Rest

During the rest period students were instructed to lie down on their backs, close their eyes, and focus on their breathing and sensations. The target time guideline in the curriculum for rest in a 30-min session was 4 min. Classroom management was emphasized through structured entrance and exit routines and leaving the room in an orderly manner.

Observation Procedure

Observers arrived prior to sessions and sat in a non-intrusive corner of the room. Once the class began the rater would proceed to complete their observation form and remain in the classroom until the class ended and the students had exited. Observations were conducted both individually and in groups with other raters to examine inter-rater agreement and reliability.

Materials

Observation Form

We implemented a modified version of an existing fidelity observation form (FOF; Corke 2017) that was initially developed by Hagins and Rundle (2016) based on Gould et al. (2014) findings and recommendations. Specifically, the FOF included the assessment of clearly defined core content and process components, dosage, and student engagement. Content fidelity comprised class session structural and curriculum elements, including classroom routines and procedures, orderliness of classroom, entry/exit procedures, content instruction, yoga movement, and rest. Content fidelity was measured on a binary, true/false scale. Process fidelity included measures of compassion, classroom management, developmentally appropriate instructions, teaching clarity, cueing of breath and movement, professionalism, and engagement of students. The process items were measured using Likert scale items ranging from "Strongly Disagree" (a score of 0) to "Strongly Agree" (a score of 4).

Data Analysis

To examine the underlying structure of the FOF, exploratory principal components analysis was conducted for all observations (n = 165). Follow-up intraclass correlations using group ratings (n = 25) were calculated for all items that successfully loaded onto components. To examine fidelity, dependent sample *t*-tests were then conducted to determine whether instructors satisfied minimum curriculum requirements (i.e., average scores greater than 2 out of a possible 4) for each item deemed reliably observed via intraclass correlations.

Table 2 Study 1 descriptive statistics and intraclass correlation coefficients

	Mean/% (n = 165)	SD/frequency	ICC (n = 25)
Entering procedure	.86	141	.258
Exiting procedures	.91	145	**.631*****
Visual outline	.82	135	.121
Learning environment	.78	125	.303
Introduction content	.67	107	**.950*****
Inclusion content	.86	138	**.633*****
Movement included	.90	144	.345
Movement minutes	18.24	7.22	.068
Rest instruction	.92	144	1.0
Rest minutes	5.21	3.09	**.829*****
Models human compassion[a]	3.64	.617	**.635*****
Maintains order[a]	3.39	.800	**.840*****
Introduces/discusses topics appropriately	3.65	.617	.497**
Prepared and supportive[a]	3.68	.598	**.646*****
Instructs skills and cues appropriately[a]	3.62	.745	**.875*****
Avoids inappropriate behaviors	3.86	.474	.042
Students engaged in lesson	.77	.422	.580**

Bolded items indicate significantly reliable ratings after adjusting for family-wise error rates using the Bonferroni correction

[a]Denotes components rated reliably and significantly ($p < 0.05$) above required cutoffs for meeting fidelity

**$p < 0.01$, ***$p < 0.001$

Results and Discussion

A total of 165 observations were analyzed for study 1. An observation was defined as a fidelity form being completed by a rater. Observations were conducted based on feedback in weekly meetings. Descriptive statistics were calculated for items on the FOF, and intraclass correlations (ICCs) were calculated using a two-way mixed model for all raters conducting paired observations (Table 2). Kappa coefficients were calculated for all binary response items.

The factor structure of the FOF was examined using exploratory procedures. First, parallel analysis was conducted using monte carlo simulations to identify significant factors from the raw data. Eigenvalues greater than the 95th percentile (using 1000 parallel datasets) were considered significant. Results from the parallel analysis indicated three factors were present in the data. A subsequent principal components analysis confirmed a 3-factor structure of the FOF. Examination of the scree plot, Eigenvalues greater than 1, and of the amount of variance explained by the factors and salient factor loadings were employed. Loadings of .4 or greater were considered salient. A first analysis was conducted using principal components analysis and Varimax rotation. This was supplemented with principal factors extraction (i.e., Principal Axis Factoring in SPSS; Tabachnick and Fidell 2000) with direct oblimin rotation with Kaiser normalization (i.e., given the theoretical correlation among factors and to try to simplify factors by minimizing cross products of loadings; Tabachnick and Fidell 2000). Results of the principal components analysis with Varimax

Table 3 Study 1 fidelity observation form principal component analysis

	Component (n = 165)		
	1	2	3
Instructs skills clearly and cues appropriately	**.766**	.139	.018
Maintains order	**.756**	−.028	−.183
Students engaged in lesson	**.755**	−.138	−.292
Rest instruction	**.739**	−.080	−.212
Introduces and discusses topics appropriately	**.692**	.158	−.126
Movement included	**.659**	−.001	.123
Models human compassion	**.607**	.304	.161
Prepared and supportive	**.583**	.492	.058
Learning environment	**.537**	−.132	.221
Visual outline	.026	**.739**	.246
Inclusion content	.126	**.725**	−.126
Introduction content	−.124	**.468**	−.227
Entering procedure	.190	−.023	−.726
Exiting procedures	**.466**	.056	−.657
Avoids inappropriate behaviors	.371	−.106	**.520**

Rotations sum of square loadings: Factor 1 (4.814), Factor 2 (1.984), Factor 3 (1.704). Rotation converged in 16 iterations. Bolded items indicate a factor loading of .40 or greater on a given component

rotation indicated that there were three factors with Eigenvalues over 1 and they accounted for a total of 54.2% of the variance. Examination of the scree plot also corroborated the adequacy of a three-factor solution. Use of other extraction and rotation methods (e.g., principal factors extraction with direct oblimin rotation) did not change interpretation of the number of factors or salient loadings (Table 3).

Following the components analysis, intraclass correlations and kappa coefficients were conducted for those observations that were conducted by multiple raters (total of n = 25 paired observations). Significant correlations were found for 10 of the 17 items, with reliability ranging from poor to excellent for all 17: exiting procedures, introduction to content, inclusion of content, minutes of rest, modeling compassion, maintaining order, appropriate topic introduction/discussion, prepared and supportive, skill instruction and cueing, and student engagement. All items meeting moderate or better ($r > 0.6$) reliability were analyzed to determine instructor fidelity using dependent samples *t*-tests. Results from the tests indicated four items were significantly ($p < 0.05$) rated, on average, as being satisfactory or better: Models human compassion, maintains order, prepared and supportive, and skill instruction and cueing (Table 2).

Results from study 1 demonstrated the connection of the FOF items to curriculum content delivery (component 1), and the didactic content itself (component 2). Additionally, we discovered that a third component emerged that was tied to both session management and professional decorum. These results suggest the FOF items reflection of both session content and process. Furthermore, our results in study 1 suggest that raters were generally able to reliably rate sessions. The results of study 1 provided insight into areas where our procedures were performing as expected and areas in which our methods could be strengthened for study 2. These enhancements are described in study 2 methods.

Study 2

The results of study 1 provided insight into areas where our procedures were performing as expected and areas in which our methods could be strengthened for study 2. These enhancements are described in study 2 methods. Again, we hypothesized that most aspects of the Pure Power Curriculum could be observed and coded in a reliable manner, and that instructors would implement with fidelity curriculum requirements (i.e., average scores greater than 2 out of a possible 4) for each item deemed reliably observed.

Method

Participants

Pure Power Instructors. Study 2 continued with the same study 1 yoga instructors.

Rating Team. In study 2 the rating team included a psychologist (second author) who was also an Experienced Yoga Alliance Registered Yoga Teacher (E-RYT® 200) and two clinical psychology graduate students trained by the lead researcher.

Procedures

Implementation procedures continued from study 1 with the addition of weekly Session Plans (e.g., Carrion 2020a and a Pacing Guide (Fig. 1) developed by the Pure Power curriculum team for the instructors to promote greater standardization of content delivery and dosage prescriptions (e.g. specifically 21 min of posture; 4 min of rest, and 5 min of content in a 30-min session) in study 2.

Fig. 1 Opening sequence A

Materials

FOF Revisions

After piloting the form in study 1, informed revisions were made to the FOF prior to implementation in study 2. Specifically, entering procedures, visual outline, and learning environment were removed from the FOF. Items added included the following: total breathing prompts, total mindfulness prompts, and draws attention to breath and movement. Additionally, several items (i.e., minutes of rest/posture/instruction, exiting procedures, and introduces/discusses topics appropriately) were reworked and expanded to provide more detail. Finalized versions of these items included minutes of breath instruction, minutes of posture instruction, minutes of rest instruction, and minutes of content instruction. An electronic copy of the final version of the FOF is available online (Carrion 2020b).

Inter-rater Consensus

Inter-rater debriefing meetings following class observations were introduced to review ratings, achieve consensus, and prevent rater drift. Where clear agreement could be established, ratings were changed to reflect the mutually agreed upon score. When consensus was not reached, item scores were left in disagreement.

Data Analysis

Fidelity observations were collected over the course of 16 months. Data and collection analytic procedures were identical to those employed in study 1 for both total observations ($n = 91$) and group ratings ($n = 19$).

Results and Discussion

A total of 91 observation forms were conducted and analyzed for study 2. Observations were conducted based on feedback in mentoring and training meetings. Descriptive statistics were calculated for 17 items on the FOF. Results are presented in Table 4.

The factor structure of the FOF was examined using the same procedures employed in study 1. Results from the parallel analysis indicated three factors were again present in the data. A subsequent principal components analysis confirmed the 3-factor structure of the FOF, as did examination of the scree plot and salient (greater than 0.4) factor loadings. These results were consistent across extraction and rotation methods (Table 5).

Intraclass correlations and kappa coefficients were calculated for observations conducted by multiple raters (total of $n = 19$ paired observations). Significant correlations were found for 13 of the 17 items, with reliability ranging from poor to excellent for all 17: Content introduced in first 15 min, content discussed with students, content objectives included in the lesson, total breath instruction minutes, total posture instruction minutes, total rest instruction minutes, breathing prompts, modeling compassion, maintaining order, preparation, clear posture instruction, attention to breath and movement, and student engagement. All items meeting moderate or better ($r > 0.6$) reliability were analyzed to determine instructor fidelity using dependent samples t-tests. Results from the tests indicated five

Table 4 Study 2 fidelity observation form descriptive statistics and intraclass correlation coefficients

	Mean (n = 91)	SD	ICC (n = 17)
Content introduced within 15 min[a]	2.31	1.47	**.920*****
Content discussion	2.05	1.13	**.929*****
Content objectives included[a]	2.88	1.15	**.909*****
Posture objectives included	2.55	1.33	.003
Minutes of breath instruction	3.10	2.05	**.779*****
Minutes of posture instruction	17.95	5.90	**.831*****
Minutes of rest instruction	5.00	2.58	**.931*****
Minutes of content instruction	4.54	4.49	.093
Total breathing prompts	13.59	3.28	.429*
Total mindfulness prompts	7.88	5.18	.100
Models compassion[a]	3.54	.633	**.630****
Maintains order	3.24	.820	.526**
Topic introduction and discussion	3.41	.650	.241
Is prepared[a]	3.10	.915	**.640****
Teaches postures clearly	3.21	.960	.500*
Draws attention to breath and movement[a]	3.00	1.18	**.677*****
Avoids inappropriate behaviors	3.66	.745	.062
Students engaged in lesson	3.29	.585	.578**

Bolded items indicate significantly reliable ratings after adjusting for family-wise error rates using the Bonferroni correction

[a]Denotes components rated reliably and significantly ($p < 0.05$) above required cutoffs for meeting fidelity

*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$

items were significantly ($p < 0.05$) rated, on average, as being satisfactory or better: content introduced in first 15 min, content objectives included in the lesson, modeling compassion, preparation, and drawing attention to breath and movement (Table 4).

General Discussion

This study provides initial support for the feasibility of measuring yoga and mindfulness implementation through a Pure Power curriculum in large school settings. It further identifies items with appropriate fidelity of implementation. To date, most of the school-based mindfulness research has not focused on or included an evaluation of fidelity of implementation. The absence of fidelity research limits our understanding of what constitutes the essential ingredients in yoga curricula, their relative roles in driving program outcomes, and the ability of researchers to replicate implementation and evaluation efforts. Our findings, therefore, advance the field of yoga and mindfulness curricula implementation by describing specific methods to evaluate fidelity and offering insight into refining the process in future studies.

Consistent with our hypothesis, the FOF items in study 1 aligned with the content processes (component 1) and the didactic content (component 2), but a third component emerged with mixed items of session management and professional decorum. Further refinement of the FOF in study 2 led to most items loading on components 1 and 3, again

Table 5 Study 2 fidelity observation form principal component analysis

	Component (n = 91)		
	1	2	3
Draws attention to breath and movement	**.892**	−.084	−.036
Total breathing prompts	**.795**	−.045	−.041
Teaches postures clearly	**.409**	.199	.235
Content discussion	**.115**	.051	−.091
Content objectives included	**.168**	.065	.139
Content introduced within 15 min	−.262	−.096	**.173**
Minutes of content instruction	**.188**	−.191	−.104
Is prepared	.305	**.436**	.174
Minutes of breath instruction	**.308**	−.163	.102
Minutes of posture instruction	−.261	**.792**	.019
Posture objectives included	−.046	**.603**	.141
Minutes of rest instruction	−.131	−.589	**.158**
Maintains order	.138	.148	**.694**
Models compassion	.063	−.093	**.666**
Topic introduction and discussion	.175	.006	**.626**
Students engaged in lesson	−.082	.195	**.575**
Total mindfulness prompts	−.185	−.173	**.538**
Avoids inappropriate behaviors	.285	−.027	**.349**

Rotations sum of square loadings: Factor 1 (4.201), Factor 2 (1.674), Factor 3 (1.509). Rotation converged in 13 iterations. Bolded items indicate a factor loading of .40 or greater on a given component

predominantly reflecting content and process, respectively. Component 2 was again mixed, with items related to content, preparation, and process. In addition to our goal of developing a measure reflective of both content and process of yoga instruction, we found that raters were generally able to reliably rate these sessions across both studies, but reliability varied by item type.

Consistent with research on curriculum adherence (Gould et al. 2014; Metz et al. 2013; Schonert-Reichl and Lawlor 2010) in school-based mindfulness and yoga programs, the instructors effectively met expectations regarding the modeling of compassion, classroom management, preparedness, and skillful instruction and cueing in study 1. Similarly, in study 2 we found curriculum adherence on two content items (introduction of the content objective within the first 15 min of the session, inclusion of the content objective) and three process items (modeling of compassion, instructor preparedness, and the cueing of breath and movement). Consistent with White (2012), our FOF served as an "intervention checklist" in which raters were able to quantify content and process implementation of a manualized intervention. Our work also extends that of Metz et al. (2013) by including the measuring periods of content, posture, and rest. We were also able to conclude that instructors were prepared to teach the lessons and further the research (e.g., Frank et al. 2014) on assessing student engagement.

The training level of the instructors in the Pure Power program was comparable to other programs (see Table 1) in that nearly all of them were 200-h yoga certified, had specialized training in the Pure Power curriculum, and were under ongoing supervision

with program managers and regularly attended professional development trainings. We believe these practices also supported FOI in that instructors were able to receive ongoing mentoring and guidance on how to deliver the curriculum with fidelity.

In addition to demonstrating the feasibility of implementing a district-wide yoga and mindfulness curriculum, our study also created and maintained a lasting community partnership within the district. Specifically, most of the staff members trained and assigned to schools remained at these schools after the conclusion of the study. These positions were supported not only by students, but also by administration, teachers, and coaches alike. Informal reports from schools indicated overall favorable impressions of the partnership—which was also maintained after the study ended.

Investigations of this scale necessitate collaboration among a multidisciplinary team. Our approach demonstrates how academia, community, and private organizations can integrate efforts to provide interventions in a feasible and reliable manner. Outcomes resulting from multi-method evaluations (i.e., academic, sleep, imaging) will be presented in separate manuscripts.

Although we were able to test the feasibility of measuring yoga curriculum fidelity on a large scale, there were some limitations in our findings. The most notable is the number of items on the FOF where instructors failed to meet acceptable levels of fidelity. For example, in study 1 six items were rated on average as being acceptable (i.e., "agree" or "strongly agree" that objective was met), four of which were reliably rated. In study 2 there were again 6 items meeting standards of acceptability, five of which were reliably rated. Acceptable items across both studies generally loaded on the respective process components, indicating such elements of yoga instruction may be easier to measure, to implement, or both. The remaining content items, however, were reliably scored across raters but consistently did not meet acceptable levels of implementation by instructors. This finding highlights an area where training of instructors could be improved. Alternatively, it may also suggest the need to further revise the FOF, specifically on items related to content. This might be achieved by reducing the number of items, keeping the rating system consistent across items, and simplifying items. Several items on the FOF also include multiple criteria. For example, process item number six assesses four separate criteria: Drawing attention to breath and movement, transitions between poses, holding poses, and transition into rest. Assessing multiple criteria across one item may introduce variability across raters.

Another limitation was the relatively low number of observations conducted and inability to collect observations based on an a priori designed schedule. Classes were held daily throughout the academic year and throughout the district, but fidelity monitoring was limited for several reasons. These include limited research staff availability and the significant time commitment required to travel among the district campuses. Research staff also experienced numerous failed rating attempts due to school field trips, assemblies, classroom maintenance issues, student health issues, and instructor absences. Schedule disruptions also occurred due to other school teachers being absent and yoga instructors assisting with school coverage needs. One of the strengths of naturalistic designs involving community settings is the identification of such issues. Knowledge of these is vital for proper interpretation and generalization of results. These systemic issues emphasize the need to intensify flexibility when tailoring methods to evaluate curricula for schools. For example, utilizing video recordings of classes would maximize observation and fidelity rating opportunities and reduce time commitments related to logistics and unexpected disruptions.

This report is one of the first of its kind to systematically assess both content and process measures of yoga instruction taking place within a large-scale trial. A multidisciplinary

approach was critical in the formation, execution, and completion of our design. Not only were we able to feasibly assess content and process, but we were also able to create a standardized measurement system that we, and hopefully others, can continue to refine and implement in future trials. Given the high variability and lack of standardization across studies of school-based yoga and mindfulness, this study represents a crucial step in the larger body of research examining the effectiveness of such curricula.

Given this protocol is freely available to others interested in implementing and/or studying the impact of yoga and mindfulness curricula, future studies should be aware of and address these limitations. Nevertheless, our study provides an initial investigation of a yoga curriculum and systematic way to evaluate that curriculum. It may also be a useful starting point for others interested in creating or refining yoga and mindfulness protocols, or alternative treatments more generally.

Acknowledgements We are grateful for the contributions of community members of the Ravenswood City School District.

Funding This project was supported by the Lucile Packard Foundation for Children's Health and NIH award UL1 TR001085.

Compliance with Ethical Standards

Conflict of interest None to declare.

References

Bergen-Cico, D., Razza, R., & Timmins, A. (2015). Fostering self-regulation through curriculum infusion of mindful yoga: A pilot study of efficacy and feasibility. *Journal of Child and Family Studies, 24*(11), 3448–3461.

Butzer, B., LoRusso, A., Shin, S. H., & Khalsa, S. B. S. (2017). Evaluation of yoga for preventing adolescent substance use risk factors in a middle school setting: A preliminary group-randomized controlled trial. *Journal of Youth and Adolescence, 46*(3), 603–632.

Carrion, V. (2020a). *Example session plan from pure power curriculum; unit 4, lesson 1, session 1* [Online]. Retrieved April 30, 2020, from https://www.researchgate.net/publication/341685678_Example_Session_Plan_from_Pure_Power_Curriculum_Unit_4_Lesson_1_Session_1. https://doi.org/10.13140/rg.2.2.31281.20324.

Carrion, V. (2020b). *Fidelity observation form* [Online]. Retrieved April 30, 2020, from https://www.researchgate.net/publication/341685781_Fidelity_Observation_Form. https://doi.org/10.13140/rg.2.2.24570.31680.

Carsley, D., Khoury, B., & Heath, N. L. (2018). Effectiveness of mindfulness interventions for mental health in schools: A comprehensive meta-analysis. *Mindfulness, 9*(3), 693–707.

Corke, M. (2017). Fidelity-Observation-Validity-Report-2017.pdf, *Pure edge fidelity observation form: Exploring its reliability and validity* [Online]. Retrieved August 10, 2017, from http://pureedgeinc.org/wp-content/uploads/2017/05/Fidelity-Observation-Validity-Report-2017.pdf.

Deitz, R., & Rajan, S. (2017). Access and barriers to implementing yoga-based interventions for at-risk adolescent youth. *International Journal of Health Promotion and Education, 55*(1), 30–34.

Felver, J. C., Celis-de Hoyos, C. E., Tezanos, K., & Singh, N. N. (2016). A systematic review of mindfulness-based interventions for youth in school settings. *Mindfulness, 7*(1), 34–45.

Fishbein, D., Miller, S., Herman-Stahl, M., Williams, J., Lavery, B., Markovitz, L., et al. (2016). Behavioral and psychophysiological effects of a yoga intervention on high-risk adolescents: A randomized control trial. *Journal of Child and Family Studies, 25*(2), 518–529.

Frank, J. L., Bose, B., & Schrobenhauser-Clonan, A. (2014). Effectiveness of a school-based yoga program on adolescent mental health, stress coping strategies, and attitudes toward violence: Findings from a high-risk sample. *Journal of Applied School Psychology, 30*(1), 29–49.

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.

Gould, L. F., Mendelson, T., Dariotis, J. K., Ancona, M., Smith, A. S., Gonzalez, A. A., et al. (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.

Greenberg, M. T., & Harris, A. R. (2012). Nurturing mindfulness in children and youth: Current state of research. *Child Development Perspectives, 6*(2), 161–166. https://doi.org/10.1111/j.1750-8606.2011.00215.x.

Haden, S. C., Daly, L., & Hagins, M. (2014). A randomised controlled trial comparing the impact of yoga and physical education on the emotional and behavioural functioning of middle school children. *Focus on Alternative and Complementary Therapies, 19*(3), 148–155.

Hagins, M., Haden, S. C., & Daly, L. A. (2013). A randomized controlled trial on the effects of yoga on stress reactivity in 6th grade students. *Evidence-Based Complementary and Alternative Medicine, 2013,* 607134.

Hagins, M., & Rundle, A. (2016). Yoga improves academic performance in urban high school students compared to physical education: A randomized controlled trial. *Mind, Brain, and Education, 10*(2), 105–116.

Kallapiran, K., Koo, S., Kirubakaran, R., & Hancock, K. (2015). Effectiveness of mindfulness in improving mental health symptoms of children and adolescents: A meta-analysis. *Child and Adolescent Mental Health, 20*(4), 182–194.

Khalsa, S. B. S., & Butzer, B. (2016). Yoga in school settings: A research review. *Annals of the New York Academy of Sciences, 1373*(1), 45–55.

Khalsa, S. B. S., Hickey-Schultz, L., Cohen, D., Steiner, N., & Cope, S. (2012). Evaluation of the mental health benefits of yoga in a secondary school: A preliminary randomized controlled trial. *The Journal of Behavioral Health Services & Research, 39*(1), 80–90.

Lawlor, M. S. (2016). Mindfulness and social emotional learning (SEL): A conceptual framework. In K. Schonert-Reichl & R. Roeser (Eds.), *Handbook of mindfulness in education. Mindfulness in Behavioral Health*. New York, NY: Springer.

Mendelson, T., Greenberg, M. T., Dariotis, J. K., Gould, L. F., Rhoades, B. L., & Leaf, P. J. (2010). Feasibility and preliminary outcomes of a school-based mindfulness intervention for urban youth. *Journal of Abnormal Child Psychology, 38*(7), 985–994.

Metz, S. M., Frank, J. L., Reibel, D., Cantrell, T., Sanders, R., & Broderick, P. C. (2013). The effectiveness of the learning to BREATHE program on adolescent emotion regulation. *Research in Human Development, 10*(3), 252–272.

Noggle, J. J., Steiner, N. J., Minami, T., & Khalsa, S. B. S. (2012). Benefits of yoga for psychosocial well-being in a US high school curriculum: a preliminary randomized controlled trial. *Journal of Developmental and Behavioral Pediatrics, 33*(3), 193–201.

Quach, D., Mano, K. E. J., & Alexander, K. (2016). A randomized controlled trial examining the effect of mindfulness meditation on working memory capacity in adolescents. *Journal of Adolescent Health, 58*(5), 489–496.

Ramadoss, R., & Bose, B. (2010). Transformative life skills: Pilot study of a yoga model for reduced stress and improving self-control in vulnerable youth. *International Journal of Yoga Therapy, 20*(1), 73–78.

Roeser, R. W., Schonert-Reichl, K. A., Jha, A., Cullen, M., Wallace, L., Wilensky, R., et al. (2013). Mindfulness training and reductions in teacher stress and burnout: Results from two randomized, waitlist-control field trials. *Journal of Educational Psychology, 105*(3), 787.

Sarokte, A. S., & Rao, M. V. (2013). Effects of Medhya Rasayana and Yogic practices in improvement of short-term memory among school-going children. *Ayu, 34*(4), 383.

Schonert-Reichl, K. A., & Lawlor, M. S. (2010). The effects of a mindfulness-based education program on pre-and early adolescents' well-being and social and emotional competence. *Mindfulness, 1*(3), 137–151.

Shonert-Reichl, K. A., Kitil, M. J., & Hanson-Peterson, J. (2017). *To reach students, teach the teachers: A national scan of teacher preparation and social and emotional learning*. A report preapared for the Collaborative for Academic, Social, and Emotional Learning (CASEL). Vancouver, B.C.: University of British Columbia.

Tabachnick, B. G., & Fidell, L. S. (2000). *Computer-assisted research design and analysis*. Boston: Allyn & Bacon Inc.

Taylor, C., Harrison, J., Haimovitz, K., Oberle, E., Thomson, K., Schonert-Reichl, K., & Roeser, R. W. (2016). Examining ways that a mindfulness-based intervention reduces stress in public school teachers: A mixed-methods study: Erratum.

Telles, S., Singh, N., Bhardwaj, A. K., Kumar, A., & Balkrishna, A. (2013). Effect of yoga or physical exercise on physical, cognitive and emotional measures in children: a randomized controlled trial. *Child and Adolescent Psychiatry and Mental Health, 7*(1), 37.

Verma, A., Shete, S. U., & Singh Thaku, G. (2014). The effect of yoga practices on cognitive development in rural residential school children in India. *Memory, 6*(2.80), 6–24.

White, L. S. (2012). Reducing stress in school-age girls through mindful yoga. *Journal of Pediatric Health Care, 26*(1), 45–56.