PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (**http://bmjopen.bmj.com/site/about/resources/checklist.pdf**) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	The Moving 2 Mindful (M2M) Study Protocol: Testing a Mindfulness Group plus Ecological Momentary Intervention to Decrease Stress and Anxiety in Adolescents from High-Conflict Homes with a Mixed-Method Longitudinal Design
AUTHORS	Lucas-Thompson, Rachel; Seiter, Natasha; Broderick, Patricia; Coatsworth, J.; Henry, Kimberly L.; McKernan, Charlotte; Smyth, Joshua

VERSION 1 – REVIEW

REVIEWER	Michael Mullarkey University of Texas at Austin United States of America
REVIEW RETURNED	04-Jun-2019

GENERAL COMMENTS	Thank you for inviting me to review "The Moving 2 Mindful (M2M) Study Protocol: Testing a Mindfulness Group plus Ecological Momentary Intervention to Decrease Stress and Anxiety in Adolescents from High-Conflict Homes." This study primarily involves developing an add-on ecological momentary intervention to the Learning to Breathe intervention and testing its feasibility. This trial has the potential to do an excellent job of developing the ecological momentary intervention, but substantive edits, especially around the planned inferential testing, are necessary. My comments are as follows: Statistical Methods: The methods for Aims 1 and 2 seem to be sufficient for addressing their research questions. However, starting in Aim 2 the authors propose to use repeated measures ANOVAs to evaluate the liking ratings for the ecological momentary interventions differ based on methodology or dosage. The authors may wish to instead use linear mixed effect models, which can more effectively handle missing data. Proposed best practices for linear mixed effect models can be found here (doi: 10.31234/osf.io/h3duq), a brief tutorial on conducting them can be found here (https://qubeshub.org/resources/348), and a simulation based approach to understanding linear mixed effects can be found here (doi: 10.31234/osf.io/xp5cy). The authors may also wish to use these linear mixed effects models in Aim 4, though I am concerned about using inferential tests at all in a sample of this size, unless the expected differences are d = 0.45 or higher at 80% power for the primary tests (See: https://rpsychologist.com/d3/NHST/ for a useful resource for determining the smallest effect size that can be reliably detected

	between two groups at a given sample size, power, and alpha level). Even if the tests are exploratory, the authors should provide evidence that the tested effects are expected to be at least this large, perhaps by reference to the previous literature. If not, the inferential tests, even if exploratory, may not be appropriate. Also, please describe which specific analyses will be performed to examine the within-person links between mindfulness, stress, and anxiety as described on paragraph 3 of page 16. The current plan to test moderators if the effects are null, while common practice, may produce misleading results due to multiple testing. Further, this sample size is likely underpowered to detect moderators using traditional methods, as the sample size to detect an interaction effect must be four times larger than the sample size needed to detect a main effect (doi: 10.1080/10543401003618819). Examining moderators of the treatment effect using a machine learning approach such as Bayesian Causal Forests (https://arxiv.org/pdf/1706.09523.pdf) could allow the researchers to examine a wide variety of moderators while keeping the Type I error rate of the analyses low. This R package may be useful if the authors wish to implement Bayesian Causal Forests https://cran.r-project.org/web/packages/grf/index.html Other Minor Comments: Please further describe the blinding/masking procedure on page 14. How will blinding/masking be maintained? Under what specific circumstances would blinding/masking be broken? Please provide more details about the fidelity assessments of Learning to Breathe by the live coders and the principal investigator. What benchmark for fidelity would be considered a successful implementation of the intervention? Also, how will discrepancies between the principal investigator and live coders be resolved? In Table 1, it may be helpful to keep track of how long it takes to recruit the target N of 38 families.

REVIEWER	Tracy Pellatt-Higgins University of Kent, UK
REVIEW RETURNED	09-Jul-2019

GENERAL COMMENTS	1) what is the rational for not including members of the public in the design and implementation? 2) The sample size of 30 participants for aim 1 should be justified - why 30 participants? This seems a lot for the initial development phase. Can you justify why 30 are needed? 3) please can you clarify in the text whether you intend to include all 30 participants in Aim 1 in one focus group (this might not be conducive to open dialogue and discussion), or is the plan to have several smaller focus groups? 4) How many participants will be included in Aim 2? Include number and justification of sample size for this phase. 5) Aim 4 'examining the extent to which mindfulness reduces dysregulated stress and anxiety' can only really be done in an RCT, I think this aim should be more in line with objectives of pilot

	studies, to test study processes and implementation and estimate variability to inform sample size calculations for a definitive trial. 6) please justify the sample size for Aim 3 and 4, what level of precision in terms of the width of the confidence intervals will the sample size provide for the feasibility outcomes and expected primary outcome in the follow-on RCT? This will give you an idea of how well you may be able to estimate potential effect. 7) Have you defined stop/go criteria at the end of the feasibility and pilot phase. In what circumstances will you continue to a full RCT and in what circumstances will you discontinue? 8) what is the rational for collecting so many daily ecological momentary assessments from participants - will this not over-burden participants? 9) what is the justification of using graduate students to deliver the intervention.- will this not dilute the effect? Evidence suggests that training by experienced practitioners is more effective - please justify 10) What does extensive training mean? can you provide more detail on the training that will be provided to graduate students to deliver the intervention.

VERSION 1 – AUTHOR RESPONSE

Reviewer 1:

Thank you for the helpful suggestions and resources regarding linear mixed models and Bayesian Causal Forests! Again, we apologize that we do not have the leeway to propose these changes to NCCIH given their approval process and the need to maintain the clinical trial on schedule. Additional, specific changes we've made to address your feedback are detailed below.

- Also, please describe which specific analyses will be performed to examine the within-person links between mindfulness, stress, and anxiety as described on paragraph 3 of **page 16.**

We have added to the following paragraph to pg. 19 to clarify this element of the analytic plan: "**We will also conduct inferential statistics to examine within-person links between mindfulness, stress, and anxiety, as well as observed or experienced family conflict. More specifically, we will conduct both mixed-effects regression models to examine the extent to which on days that adolescents observe/experience family conflict they are less mindful, and similarly whether less mindful days are also days in which adolescents report greater stress and anxiety. Given the richness of the longitudinal data (i.e., EMA collected at least once day across the 8-week intervention period), we will also conduct heterogeneous mixed-effects models [48], which model not just random intercepts (as with mixed-effects regression models) but also random variances, allowing us to model and predict issues of varying erraticism within people (e.g., are adolescents more variable in their mindfulness on days that they observe/experience family conflict). We will also be able to model whether mindfulness becomes less erratic over the course of the intervention, as adolescents develop a personal mindfulness practice."**

- Please further describe the blinding/masking procedure on page 14. How will blinding/masking be maintained? Under what specific circumstances would blinding/masking be broken?

We have provided additional detail about maintaining and breaking the blind on pgs. 14-15. This revised section now reads is: "**The principal investigator, as well as individuals involved in outcome assessments and data entry/analyses, will be blind to participant condition. To maintain the blind, research team members who are not the PI or involved in outcome assessment or data entry/analyses will implement the randomization procedure and prepare unblinded reports. Intervention facilitators will necessarily not be blinded to study condition, but will not be involved with outcome assessment or data entry/analyses. The most likely condition in which the blind will be broken is that participants disclose their group assignment to someone on the assessment team. We will engage in training procedures so that assessment team members are well-prepared to handle this breaking of the blind. We will put**

procedures in place so that such a breaking of the blind would not threaten the integrity of the data. Data will be entered into REDCap either by participants, or by team members who are specifically trained in data entry and are not involved in assessment. *The blind will be broken only by study team members under critical incident conditions. Research team members responsible for randomization and preparing unblinded reports, as well as intervention facilitators* will be authorized to break the blind if one of the participants has a very serious adverse reaction or event. Then, the PI will be informed to help assure that the procedures of the DSMP are undertaken. There will be no reason that the blind will be broken with other study team members. Also, the PI will not be involved in assessment."

- Please provide more details about the fidelity assessments of Learning to Breathe by the live coders and the principal investigator. What benchmark for fidelity would be considered a successful implementation of the intervention? Also, how will discrepancies between the principal investigator and live coders be resolved?

We have added additional details to pg. 16 and Table 1 about these fidelity assessments. As noted in Table 1, live coders will evaluate curriculum adherence, group participation, facilitator delivery, and facilitation processes; the program developer will use the Teaching Mindfulness in Education Observation Scale (Broderick et al., 2019) to evaluate fidelity and quality of implementation. We have also clarified that both live coder and program developer ratings of fidelity will be reported, so that discrepancies will be clearly discussed (as well as reasons for them, if applicable), and that evidence of successful implementation will be if both forms of fidelity assessments indicate fidelity of ≥ 80%.

- In Table 1, it may be helpful to keep track of how long it takes to recruit the target N of 38 families.

We have added that we will track this helpful information.

Reviewer 2:

- What is the rational for not including members of the public in the design and implementation?

We have clarified that members of the patients/members of the public did inform the design and implementation in that adolescents provided extensive feedback on the nature of the multi-method adaptive intervention supplement to L2B, as well as delivery dosage, but did not otherwise contribute to other elements of study design or implementation.

- The sample size of 30 participants for aim 1 should be justified - why 30 participants? This seems a lot for the initial development phase. Can you justify why 30 are needed?

We have clarified on pg. 8 that this number was selected on the basis of providing facilitator training and certification, and participants were also asked to contribute to text-message development.

- Please can you clarify in the text whether you intend to include all 30 participants in Aim 1 in one focus group (this might not be conducive to open dialogue and discussion), or is the plan to have several smaller focus groups?

We have clarified that multiple cohorts of 4-10 adolescents participated in Learning 2 BREATHE, and, as a result, the focus groups (see pg. 8).

- How many participants will be included in Aim 2? Include number and justification of sample size for this phase.

***We have corrected this oversight, and added (see pg. 11) that 10 adolescents will participate in the procedures relevant to aim 2. We have also added our justification: "*This sample size was selected to allow us to get feedback on the refinement of the EMI supplement and delivery plan from two cohorts of adolescents (with an estimate of 5 teens per cohort)."**

- Aim 4 'examining the extent to which mindfulness reduces dysregulated stress and anxiety' can only really be done in an RCT, I think this aim should be more in line with objectives of pilot studies, to test study processes and implementation and estimate variability to inform sample size calculations for a definitive trial.

We have revised the specific phrasing of the aim in line with this feedback, as well as provided more detail (see response to reviewer #1's feedback) about the within-person analyses that will contribute to this part of aim #4. However, this aim was reviewed and approved by NCCIH and so cannot be substantially altered at this point.

- Please justify the sample size for Aim 3 and 4, what level of precision in terms of the width of the confidence intervals will the sample size provide for the feasibility outcomes and expected primary outcome in the follow-on RCT? This will give you an idea of how well you may be able to estimate potential effect.

NCCIH has specifically asked us to avoid doing this at this stage of this work (i.e., as we are doing a "small-scale test of the methods and procedures to be used on a larger scale", https://nccih.nih.gov/grants/whatnccihfunds/pilot_studies). Rather, as noted (pg. 18), these analyses will be underpowered but will allow us to explore potential group differences that will explore a adequately powered trial in the future.

- Have you defined stop/go criteria at the end of the feasibility and pilot phase. In what circumstances will you continue to a full RCT and in what circumstances will you discontinue?

The full RCT discussed in aim #4 will be conducted with the same participants as aim #3 (i.e., at the same time) and so there are not stop/go criteria to determine whether we will move from feasibility/acceptability to the RCT.

- What is the rational for collecting so many daily ecological momentary assessments from participants - will this not over-burden participants?

First, we agree that the issue of burden is of course very important. The goal of aim #2 will be to determine the delivery rate of the text-messages but also EMA that will not over-burden participants, and so the specific delivery plan used for aims #3 and #4 will depend on the results of that first L2B Plus pilot. Second, the goal of the EMA is to provide intervention content to participants in times of need (in line with a "just-in-time" intervention approach, e.g., Smyth & Heron, 2016). In other words, EMA will be the tool that the research team uses to determine moments of need (in terms of elevated stress, anxiety, or conflict) in which participants may benefit from mindfulness content. More generally, this level of EMA is acceptable based on prior studies (Heron & Smyth, 2010).

- What is the justification of using graduate students to deliver the intervention.- will this not dilute the effect? Evidence suggests that training by experienced practitioners is more effective - please justify

Steps have been taken to assure that the graduate students who deliver the intervention will be experienced facilitators. We have provided additional detail about this on pgs. 15-16, which reads: **"Both group interventions will be administered by graduate students who will receive extensive training in implementation and supervised by the first author (RGLT) as well as program developer (PB). To be certified, potential facilitators need: 1) a foundation of mindfulness practice (e.g., participating in a mindfulness-based stress reduction course), 2) relevant professional expertise (e.g., teaching and/or counseling experience); 3) participation in 24-hours of training by a master training (including a foundational training about the background of L2B, plus an intensive training about facilitation involving demonstrations of L2B lessons and discussions, and opportunities to practice leading sessions with feedback from experienced trainers), and 4) individual supervision and coaching (i.e., practice leading full L2B sessions with feedback from experienced trainers). In the current study, minimum qualifications for facilitators will be a Bachelor's degree and specific and thorough training in program implementation. In addition, facilitators will be current or former students in Marriage and Family Therapy who have received extensive clinical training and supervision".**

- What does extensive training mean? Can you provide more detail on the training that will be provided to graduate students to deliver the intervention.

We have included more detail about the extensive training that graduate students will undergo on pgs. 15-16, which is detailed above.

VERSION 2 – REVIEW

REVIEWER	Michael Mullarkey University of Texas at Austin USA I received a research sponsorship to conduct a clinical trial evaluating a single-session virtual reality intervention created by Limbix Health, Inc. I am a Co-Investigator on this project. All funds associated with this sponsorship are allocated directly to research costs (i.e., no salary support is included in the sponsorship.) I am under contract to

	author the following book, to be published by New Harbinger Publications upon completion in 2020: Schleider, J. L., Mullarkey, M. C., & Dobias, M. L. (under contract). The Growth Mindset Workbook for Teens: Say Yes to Challenges and Reach Your Full Potential. New Harbinger Publications.
REVIEW RETURNED	14-Oct-2019

GENERAL COMMENTS	I believe the authors have addressed my comments as thoroughly as possible without modifying their analytic plan. I still have some reservations about the plan to interpret moderators using traditional interactions at this sample size, especially given that these analyses will be conducted specifically if there is a null main effect. However, I do not believe these reservations warrant not accepting the protocol for publication.

REVIEWER	Tracy Pellatt-Higgins University of Kent, UK
REVIEW RETURNED	29-Oct-2019

GENERAL COMMENTS	I would like to thank the authors for their comprehensive response to the previous reviewers comments. I have two further comments requiring minor revision to the manuscript: 1) please can the authors provide more clarity on the timing of the daily ecological assessments and how these will be completed. The paper states that these will at multiple random intervals throughout the day. How many will be sent to each participant on each day? Are these entered electronically so the information is available in real time to the research team? The paper states that practices will be recommended based on responses to these assessments. How will this be done? Is this automated and based on algorithms? How will this work in a full RCT? The paper would benefit from more clarification around this. 2) In the proposed analysis under Aim 4, it is unlikely that there will be enough data to estimate random variances. The proposed model is quite complex given that there are only 38 participants in this phase and a large number are required to accurately estimate variances.

VERSION 2 – AUTHOR RESPONSE

Reviewer: 1

- Remaining reservations about the plan to interpret moderators, particularly because they will be conducted if there is a null main effect

Thank you for making it clear that this issue does not, in your opinion, warrant not publishing this protocol paper. However, we did want to address the concern. To do so, first, we have revised this section to indicate that we will conduct these moderator analyses whether or not the main effect is null; in addition, we indicate that these analyses (as are other inferential tests discussed) will be exploratory.

Reviewer: 2

- Provide more clarity about the timing of the daily ecological assessments; how many will be sent each day? Will information be automatically available to the research team?

We have revised this section to address these issues. We have also tried to clarify that the EMA can trigger the delivery of intervention content, but that this is separate from the sending of intervention content at random intervals throughout the day. We now also discuss that the

EMA will be completed three times a day (first thing in the morning, after school, and in the evening), whereas the dosage of the pre-programed EMI (text messages with intervention content) will vary across weeks (with orders counterbalanced across participants) of 1-5 messages a day. We also clarify that these EMA links will be texted to teens, who will complete them online, which allows us to send the text messages in moments of need automatically.

-Address concerns that there will not be enough data to estimate random variances (given there will only be 38 participants)
Because we will focus on the within and not between person modeling of random variances (i.e., we'll focus on consistency/erraticism within person), we hope that the small sample size will be offset by the large number of repeated measurements (as participants will be invited to provide measurements over at least 49 days). We have clarified this focus in the manuscript. In addition, we want to note that our sample size is comparable to the sample sizes in one of the studies included in the original paper documenting heterogeneous mixed effects models (n=46) in which both between and within person effects were successfully modeled (Hedeker & Mermelstein, 2007).

Thank you again for the opportunity to revise and resubmit this manuscript! We hope that this revised version is suitable for publication, but are happy to make additional changes.