

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)	Developing Professional Caregivers' Empathy and Emotional Competencies Through Mindfulness-Based Stress Reduction (MBSR): Results of two Proof-of-Concept Studies
AUTHORS	Lamothe, Martin; McDuff, Pierre; Pastore, Yves; Duval, Michel; Sultan, Serge

VERSION 1 – REVIEW

REVIEWER	Sue Dean Faculty of Health, University of Technology Sydney
REVIEW RETURNED	03-Aug-2017

GENERAL COMMENTS	<p>This paper presents two well written pre-post proof -of -concept studies which provide important contributions to the growing literature on developing professional caregivers' empathy and emotional competencies through Mindfulness -Based Stress Reduction.</p> <p>In the introduction the authors draw attention to the lack of studies where mindfulness and empathy in healthcare have been selected as primary or secondary outcomes and I would like to draw the authors attention to a recent paper by Dean et al 2017, The effects of a structured mindfulness program on the development of empathy in healthcare students and I wonder why this paper wasn't cited.</p> <p>The authors have noted the limitation of the use of self report measures. The course would likely introduce response bias, when the students realise what is the wanted response.</p> <p>Further controlled studies should be carried out as was noted. The studies used validated questionnaires (MAAS, IRI, PEC,AAQ-II, ERS, GERT and GERT-S). Study 1 used the t-test statistic (2 measurement points) while study 2 used ANOVA (GLM) with post hoc Bonferroni due to the study having three measurement points. In addition demographic information was also obtained. The statistics are appropriate for the pre/post/follow-up design of the study and the posthoc tests identified where the significant differences occurred.</p> <p>The paper overall is well written and presented well and is a valuable contributor to the literature.</p>
-------------------------	---

REVIEWER	Gordon Kraft-Todd Yale University
-----------------	--------------------------------------

	USA
REVIEW RETURNED	18-Aug-2017

GENERAL COMMENTS	<p>In this paper, the authors present the results of two small-scale mindfulness-based stress reduction (MBSR) courses on longitudinal self-report outcomes related to empathy with a student (Study 1) and professional (Study 2) sample. They find that MBSR training significantly increases empathic abilities, and that this increase persists over a 3-month follow-up (Study 2). The studies were conducted well and the findings are interesting and (as the authors contend) provide evidence for the feasibility of larger-scale studies. I have concerns about four minor issues regarding a supplemental study, selection bias, discussion of the literature, and statistical reporting (as well as a few miscellaneous comments). Once these are adequately addressed, I believe this paper will be an important contribution to the literature worthy of publication in BMJ Open.</p> <p>Supplemental study</p> <ul style="list-style-type: none"> - Given the authors' concern with selection bias and their surprising null result for PEC, one minimally difficult study (data collection would take less than one day and likely cost less than \$300) they might run to supplement the current findings is the following: <ul style="list-style-type: none"> o Post a HIT on Amazon Mechanical Turk with the link to a survey (if unfamiliar with MTurk, see, e.g., citations below for an introduction). □ Horton, J. J., Rand, D. G., & Zeckhauser, R. J. (2011). The online laboratory: Conducting experiments in a real labor market. <i>Experimental Economics</i>, 14(3), 399-425. □ Arechar, A. A., Kraft-Todd, G. T., & Rand, D. G. (2017). Turking overtime: how participant characteristics and behavior vary over time and day on Amazon Mechanical Turk. <i>Journal of the Economic Science Association</i>, 3(1), 1-11. <ul style="list-style-type: none"> o First, ask about interest in an MBSR course (ideally using the same language as that used to recruit participants for the reported studies), as measured, perhaps, on a 1-7 likert scale (or via the method in reported studies) o Next, administer the same measures as in the reported studies o Such a study would accomplish two goals: <ul style="list-style-type: none"> □ By comparing your measures in this "representative" sample (not perfect, but better than college undergraduates) to the baseline measurements in the reported studies, you could provide evidence speaking to whether those who selected into the treatment have higher average baseline scores on these measures compared to the population □ If you do find a difference there, you could compare "highly motivated" subjects in this sample (as determined, e.g., by median-split on desire to participate in MBSR program) to the baseline measurements in the reported studies to provide evidence speaking to whether "high motivation to participate in an MBSR program" explains the difference between average baseline scores on these measures of those who selected into the treatment compared to the population <p>Selection bias</p> <ul style="list-style-type: none"> - One analysis the authors might consider adding which partially speaks to their selection bias concern would be to compare baseline measures of students versus professionals. Visually, there do not appear to be differences, and if this result is confirmed, it would be a piece of evidence against an interpretation that the professionals
-------------------------	--

are a “weird” population. There is still the concern, as the authors note, that all subjects opted into the treatment (though the supplemental study above can help speak to this concern).

- One limitation that requires more discussion is the lack of gender balance in the sample.

Given previous literature on gender differences in empathy (see citations below), there is reason to think that women may be more amenable to empathy-related training than men.

o Schmid Mast M, Hall JA, Klockner C, Choi E. Physician Gender Affects How

Physician Nonverbal Behavior Is Related to Patient Satisfaction. Medical Care.

2008; 46(12):1212-8

o Hall JA, Blanch-Hartigan D, Roter DL. Patients' Satisfaction With Male Versus

Female Physicians: A Meta-analysis. Medical Care. 2011; 49(7):611-7

- While the authors acknowledge the shortcoming of self-selection and lack of control

group, a partial remedy for future work would be to recruit potential participants who

expressed interested in the program but were not included in the treatment (e.g. those

who could not participate for scheduling reasons) to complete the same measures as those

in the treatment with the same timing (i.e. pre-, post-, and perhaps also at later followup).

This might be provided as a suggestion for future research.

Discussion of the literature

- Though the setting of Study 2 (and many of the citations) reference clinician empathy in

oncology, the study seems to have implications for and is consistent with findings from

clinician empathy interventions in a broader scope of medical specialties. I would

recommend that the authors either broaden the scope of their discussion (and citations) or

else make more explicit their treatment of the topic within the field of oncology. One

review which surveys a broader scope of outcomes to supplement the excellent Lelorain

et al. (2012) review of empathy and patient outcomes in cancer care, e.g.:

o Kelley, J. M., Kraft-Todd, G., Schapira, L., Kossowsky, J., & Riess, H. (2014).

The influence of the patient-clinician relationship on healthcare outcomes: a

systematic review and meta-analysis of randomized controlled trials. PloS one,

9(4), e94207.

- While there may not be many studies of MBSR, specifically, as a method for increasing

	<p>empathy, the authors might also discuss other empathy training programs in medical care, e.g.:</p> <ul style="list-style-type: none"> o Riess, H., Kelley, J. M., Bailey, R. W., Dunn, E. J., & Phillips, M. (2012). Empathy training for resident physicians: a randomized controlled trial of a neuroscience-informed curriculum. <i>Journal of general internal medicine</i>, 27(10), 1280-1286. <p>Statistical reporting</p> <ul style="list-style-type: none"> - More care should be taken to explicitly differentiate corrected from uncorrected significance thresholds. I.e. the authors explain in the “Statistical analysis” section that Bonferroni correction is used (good, though admittedly conservative), yet in the results it is unclear whether p-values are in terms of corrected or uncorrected values. E.g. the Bonferroni correction for 8 tests (the authors’ number of dependent measures) lowers the significance threshold per measure to $p=.00625$, yet when the authors report a result as “$p<.05$” is this the value obtained from the initial test or after correction? - Barring journalistic style (i.e. unless BMJ Open feels otherwise), the reporting of p-values should be precise to three decimal places (with the exception of $p<.001$). - Table 2 should include either exact p-values (as in SI Table) or stars (e.g. $*p<.1$, $**p<.05$, $***p<.01$). <p>Miscellaneous comments</p> <ul style="list-style-type: none"> - Regarding the replicability of the findings, it is unclear whether reference [20] has a more detailed account of the MBSR program methodology. If so, this should be made explicit. If not, more detail should be added in the supplement (e.g. a copy of text from the workbook and transcript from the audio recordings, or at least some example text of this nature if the complete text is prohibitively long and/or proprietary). - When referencing Figure 1, please specify which panel the reader should refer to in the discussion of the corresponding result (e.g. “Fig. 1A”). - The y-axis scaling of the figure is not very informative to the reader because it is in terms of the particular measure. I would propose the authors rescale by standardized scores
--	---

VERSION 1 – AUTHOR RESPONSE

REVIEWER 1

Reviewer Name: Sue Dean

Institution and Country: Faculty of Health, University of Technology Sydney

Reviewer's report:

This paper presents two well written pre-post proof -of -concept studies which provide important contributions to the growing literature on developing professional caregivers' empathy and emotional competencies through Mindfulness -Based Stress Reduction.

The authors have noted the limitation of the use of self report measures. The course would likely introduce response bias, when the students realise what is the wanted response. Further controlled studies should be carried out as was noted.

The studies used validated questionnaires (MAAS, IRI, PEC,AAQ-II, ERS, GERT and GERT-S).

Study 1 used the t-test statistic (2 measurement points) while study 2 used ANOVA (GLM) with post hoc Bonferroni due to the study having three measurement points. In addition demographic information was also obtained. The statistics are appropriate for the pre/post/follow-up design of the study and the posthoc tests identified where the significant differences occurred. The paper overall is well written and presented well and is a valuable contributor to the literature.

1. In the introduction the authors draw attention to the lack of studies where mindfulness and empathy in healthcare have been selected as primary or secondary outcomes and I would like to draw the authors attention to a recent paper by Dean et al 2017, The effects of a structured mindfulness program on the development of empathy in healthcare students and I wonder why this paper wasn't cited.

Response 1: As suggested by the reviewer, we cited the article by Dean et al. 2017 (see reference below) in the introduction of our manuscript. As this article is quite recent, we just did not come across it when initially searching the literature. Please see reference in the text on page 5 and on page 27 in the reference section (reference #19). The subsequent reference numbers in the text have been adjusted accordingly to this change (page 7, 8, 9, 10, 11, 19, 20, 21, 22).

Dean S, et al. The effects of a structured mindfulness program on the development of empathy in healthcare students. *NursingPlus Open*, 2017; 3:1–5.

REVIEWER 2

Reviewer Name: Gordon Kraft-Todd

Institution and Country: Yale University, USA

Reviewer's report:

In this paper, the authors present the results of two small-scale mindfulness-based stress reduction (MBSR) courses on longitudinal self-report outcomes related to empathy with a student (Study 1) and professional (Study 2) sample. They find that MBSR training significantly increases empathic abilities, and that this increase persists over a 3-month follow-up (Study 2). The studies were conducted well and the findings are interesting and (as the authors contend) provide evidence for the feasibility of larger-scale studies. I have concerns about four minor issues regarding a supplemental study, selection bias, discussion of the literature, and statistical reporting (as well as a few miscellaneous comments). Once these are adequately addressed, I believe this paper will be an important contribution to the literature worthy of publication in *BMJ Open*.

Supplemental study

1. Given the authors' concern with selection bias and their surprising null result for PEC, one minimally difficult study (data collection would take less than one day and likely cost less than \$300) they might run to supplement the current findings is the following:

o Post a HIT on Amazon Mechanical Turk with the link to a survey (if unfamiliar with MTurk, see, e.g., citations below for an introduction). □ Horton, J. J., Rand, D. G., & Zeckhauser, R. J. (2011). The online laboratory: Conducting experiments in a real labor market. *Experimental Economics*, 14(3), 399-425. □ Arechar, A. A., Kraft-Todd, G. T., & Rand, D. G. (2017). Turking overtime: how participant characteristics and behavior vary over time and day on Amazon Mechanical Turk. *Journal of the Economic Science Association*, 3(1), 1-11.

o First, ask about interest in an MBSR course (ideally using the same language as that used to recruit participants for the reported studies), as measured, perhaps, on a 1-7 likert scale (or via the method in reported studies)

o Next, administer the same measures as in the reported studies

o Such a study would accomplish two goals:

- By comparing your measures in this “representative” sample (not perfect, but better than college undergraduates) to the baseline measurements in the reported studies, you could provide evidence speaking to whether those who selected into the treatment have higher average baseline scores on these measures compared to the population

- If you do find a difference there, you could compare “highly motivated” subjects in this sample (as determined, e.g., by median-split on desire to participate in MBSR program) to the baseline measurements in the reported studies to provide evidence speaking to whether “high motivation to participate in an MBSR program” explains the difference between average baseline scores on these measures of those who selected into the treatment compared to the population

Response 1: Having a comparison sample as you suggest is a wonderful idea. We also believe that comparing the baseline scores obtained on the PEC in our studies, especially the Identify Other's Emotions' subscale, to a “representative” sample could help determine if the participants selected in our studies had higher average baseline scores compared to a representative population. We realized we had a more powerful option for this comparison than the one you suggested (i.e. posting a hit on Turk), namely to use existing published norms available with 4000+ people to assess the normality of baseline scores of the PEC obtained in the two studies of the manuscript (Brasseur et al., 2013). Participants had very similar levels on the PEC's Identify Others' Emotions than the published norms. This suggests that the participants in our study were not different from others from the general population at identifying others' emotions. Thus, the absence of effect on the PEC's Identify Others' Emotions in both studies seems to be attributable to other reasons, including the fact that all formal meditation exercises learned during the MBSR program focused on being attentive to one's own experiences. We added the results of additional analyses and rephrased a portion of the discussion. See changes in the manuscript (underlined) on page 20.

Brasseur S, Gregoire J, Bourdu R, et al. The Profile of Emotional Competence (PEC): development and validation of a self-reported measure that fits dimensions of emotional competence theory. *PLoS One* 2013;8(5):e62635.

Selection bias

2. One analysis the authors might consider adding which partially speaks to their selection bias concern would be to compare baseline measures of students versus professionals. Visually, there do not appear to be differences, and if this result is confirmed, it would be a piece of evidence against an interpretation that the professionals are a “weird” population. There is still the concern, as the authors

note, that all subjects opted into the treatment (though the supplemental study above can help speak to this concern).

Response 2: As advised, we compared baseline scores of Study 1 (students) versus those of Study 2 (professionals). Differences were small/negligible and none were statistically significant. We added the results of this comparison on page 15 and added a table in Supplementary File 2.

3. One limitation that requires more discussion is the lack of gender balance in the sample. Given previous literature on gender differences in empathy (see citations below), there is reason to think that women may be more amenable to empathy-related training than men.

o Schmid Mast M, Hall JA, Klockner C, Choi E. Physician Gender Affects How Physician Nonverbal Behavior Is Related to Patient Satisfaction. *Medical Care*. 2008; 46(12):1212-8

o Hall JA, Blanch-Hartigan D, Roter DL. Patients' Satisfaction With Male Versus Female Physicians: A Meta-analysis. *Medical Care*. 2011; 49(7):611-7

Response 3: As you advised, we included the lack of gender balance in both samples (students and professionals) as a limitation to our studies in the discussion section of the article. We also cited the meta-analysis by Hall et al (2011). Please see changes on page 23 of the manuscript.

4. While the authors acknowledge the shortcoming of self-selection and lack of control group, a partial remedy for future work would be to recruit potential participants who expressed interested in the program but were not included in the treatment (e.g. those who could not participate for scheduling reasons) to complete the same measures as those in the treatment with the same timing (i.e. pre-, post-, and perhaps also at later followup). This might be provided as a suggestion for future research.

Response 4: Indeed, there are ways to address the shortcoming of self-selection and lack of control groups. In fact, we mentioned in the discussion section of the manuscript that "The results of our studies justify progressing toward more rigorous testing with larger samples in randomized controlled trials" (p.23) which implies the inclusion of control groups in future research. The fact that the MBSR intervention was associated with clinical signals on measured outcomes provides adequate grounds for recommending the inclusion of control groups in future, controlled comparisons such as wait-list trials as you mentioned.

Discussion of the literature

5. Though the setting of Study 2 (and many of the citations) reference clinician empathy in oncology, the study seems to have implications for and is consistent with findings from clinician empathy interventions in a broader scope of medical specialties. I would recommend that the authors either broaden the scope of their discussion (and citations) or else make more explicit their treatment of the topic within the field of oncology. One review which surveys a broader scope of outcomes to supplement the excellent Lelorain et al. (2012) review of empathy and patient outcomes in cancer care, e.g.:

o Kelley, J. M., Kraft-Todd, G., Schapira, L., Kossowsky, J., & Riess, H. (2014). The influence of the patient-clinician relationship on healthcare outcomes: a systematic review and meta-analysis of randomized controlled trials. *PloS one*, 9(4), e94207.

Response 5: As recommended, we cited the review and meta-analysis by Kelley et al. (2014), which suggests that the clinician-patient relationship has a positive impact on patient health outcomes. We cited the review in the introduction of our article (page 4). The subsequent reference numbers in the text have been adjusted accordingly to this change (page 7, 8, 9, 10, 11, 19, 20, 21, 22). We also added in the discussion that the program could have interesting applications in the training of

professionals not only in pediatric oncology, but also in other emotionally challenging specialties (page 24).

6. While there may not be many studies of MBSR, specifically, as a method for increasing empathy, the authors might also discuss other empathy training programs in medical care, e.g.:

o Riess, H., Kelley, J. M., Bailey, R. W., Dunn, E. J., & Phillips, M. (2012). Empathy training for resident physicians: a randomized controlled trial of a neuroscience-informed curriculum. *Journal of general internal medicine*, 27(10), 1280-1286.

Response 6: As suggested, we cited the excellent article by Riess et al. (2012) on empathy training grounded in neuroscience (See page 24). We believe that mindfulness programs could be greatly improved by integrating other forms of evidence-based empathy training and mentioned it in our discussion (page 24). The subsequent reference numbers in the text have been adjusted accordingly to this change (page 7, 8, 9, 10, 11, 19, 20, 21, 22).

Statistical reporting

7. More care should be taken to explicitly differentiate corrected from uncorrected significance thresholds. I.e. the authors explain in the “Statistical analysis” section that Bonferroni correction is used (good, though admittedly conservative), yet in the results it is unclear whether p-values are in terms of corrected or uncorrected values. E.g. the Bonferroni correction for 8 tests (the authors’ number of dependent measures) lowers the significance threshold per measure to $p=.00625$, yet when the authors report a result as “ $p<.05$ ” is this the value obtained from the initial test or after correction?

Response 7: As the reviewer mentioned, a large number of independent variables lowers the significance threshold. P values here are obtained from initial testing. As mentioned in the manuscript, both studies were designed as Phase IIa Proof-of-concept studies. According to the developmental guidelines that we followed (Czajkowski et al., 2015), one of the goals of Phase IIa Proof-of-concept studies is to assess whether clinically significant improvements can be reached following an intervention. At this stage, “the sample size can be small because clinical, not statistical, is sought” (Czajkowski et al., 2015, p. 7). Thus, we did not want to emphasize statistical significance in this article, but rather determine whether the intervention could reach a clinical signal to inform future controlled trials.

Czajkowski SM, Powell LH, Adler N, et al. From ideas to efficacy: The ORBIT model for developing behavioral treatments for chronic diseases. *Health Psychol* 2015;34(10):971–82.

8. Barring journalistic style (i.e. unless BMJ Open feels otherwise), the reporting of p values should be precise to three decimal places (with the exception of $p<.001$).

Response 8: The BMJ Open guidelines have been followed in the reporting of p-values.

9. Table 2 should include either exact p-values (as in SI Table) or stars (e.g. $*p<.1$, $**p<.05$, $***p<.01$).

Response 9: As recommended, the p-values were added in Table 2. See page 17.

10. Miscellaneous comments

- Regarding the replicability of the findings, it is unclear whether reference [20] has a more detailed account of the MBSR program methodology. If so, this should be made explicit. If not, more detail should be added in the supplement (e.g. a copy of text from the workbook and transcript from the

audio recordings, or at least some example text of this nature if the complete text is prohibitively long and/or proprietary).

Response 10: Following your comment, we attached the transcript from one of the audio recordings that were created for this program as supplementary material. The transcript was translated into English, as the original audio recordings created for this program were in French. Both English and French versions are now made available to readers. See page 4 and Supplementary File 1.

11. When referencing Figure 1, please specify which panel the reader should refer to in the discussion of the corresponding result (e.g. "Fig. 1A").

Response 11: As requested, we specified the exact panel to which the reader should refer in the result section of the manuscript. (Please see underlined changes on page 15,16 and 18).

12. The y-axis scaling of the figure is not very informative to the reader because it is in terms of the particular measure. I would propose the authors rescale by standardized scores.

Response 12: We understand that some readers might not be familiar with some of the measures reported in the article. However, we believe that by replacing the actual questionnaire scores by standardized scores we would lose valuable information. The y-axis values could help readers better understand the results reported in the article. For example, someone might want to compare the results reported here with those reported in previous research. For example, Cohen-Katz et al. (2005), Figure 3. page 30 or Horner et al. (2014), Table 1, page 2. For these reasons, we decided to leave the y-axis scaling unchanged in Figure 1.

Cohen-Katz J, Wiley SD, Capuano T, et al. The effects of mindfulness-based stress reduction on nurse stress and burnout, part II: a quantitative and qualitative study. *Holist Nurs Pract*, 2005;19(1):26–35.

Horner JK, Piercy BS, Eure L, et al. A pilot study to evaluate mindfulness as a strategy to improve inpatient nurse and patient experiences. *Appl Nurs Res*, 2014;27(3):198–201.

- Please ensure that your DATA SHARING STATEMENT in your main document and Scholar One submission system are the same.

Response: As requested, we ensured that both Data Sharing Statements are the same.

- You have cited 'Figure 1B' rights after 'Figure 1H' which makes your citations incorrect. Please review again your main document and ensure that all figures will be cited and will appear in numerical order.

Response: In the text, we have reported the effect sizes from the largest to the smallest, so it is normal that "Figure 1B" appears after "Figure 1H"

- Please review your main document reference number #30 is missing. References must be numbered consecutively in the order in which they are mentioned in the MAIN TEXT. Please take note reference citations must begin in the introduction of the main text and not in the abstract section of the paper.

Response: Following your comment, we have removed Reference #30 in the Reference section and modified the reference numbers in the text and the Reference section accordingly.

VERSION 2 – REVIEW

REVIEWER	Gordon Kraft-Todd Yale University USA
REVIEW RETURNED	01-Oct-2017
GENERAL COMMENTS	The authors have done an excellent and thorough job addressing reviewer concerns. This paper clearly communicates the results of interesting and important studies. I look forward to seeing it in print!