

## 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

<b>TITLE (PROVISIONAL)</b>	45 Good Things: A Prospective Pilot Study of the Three Good Things Well-being Intervention in the US for Healthcare Worker Emotional Exhaustion, Depression, Work-life Balance, and Happiness
<b>AUTHORS</b>	Adair, Kathryn; Sexton, Bryan

### VERSION 1 - REVIEW

<b>REVIEWER</b>	Jesus Montero-Marin Red de Investigación en Atención Primaria (REDIAPP), Spain
<b>REVIEW RETURNED</b>	10-Apr-2018

<b>GENERAL COMMENTS</b>	<p><b>Abstract</b> Please, provide more information in the methods section (design, instruments, data analyses, etc).</p> <p><b>Introduction</b> Please, re-write the last paragraph clarifying the main and secondary aims of the study, with their corresponding hypothesis.</p> <p><b>Method</b> Any sample size or power estimates were developed prior to get the sample used? Please specify.</p> <p>What is the main outcome? All of the measures are main outcomes?</p> <p>The consideration of resilience as work-life balance and subjective happiness, etc needs to be justified. Please, explain in the introduction more deeply what resilience is, and justify why the indicators (variables) used in the present study might be reliable measures of it. Resilience appears as a different construct than burnout, depression, etc. The use of the construct in the text may drive to confusion.</p> <p><b>Results</b> Please, provide demographics for the whole group to facilitate a general view.</p> <p>Please, provide a measure of the programme compliance and analyse the possible determination of the level of compliance and improvements.</p> <p>Please, include the internal consistence of the metrics in the corresponding method section (it is not a study objective, isn't it?).</p>
-------------------------	--

	<p>Please, include standard errors in Figure 1 so that one can view possible differences along the time.</p> <p>Page 13, line 22-23, where it reads Table 1, it could be Table 2. Please revise it.</p> <p>Please, remove colours from tables to facilitate reading.</p> <p>How was managed missing data? A figure would clarify the flow of participants. Table 3 includes a very high n at baseline, and n is not the same across time points. Please, explain it in detail, because apparently you are providing a different number of participants than those really were included in the analyses.</p> <p>Were effect sizes corrected by correlated measures? Please, specify.</p> <p>I think it would be worth to analyse deeply the participant's experiences by systematic content analysis of submissions and comments, detailing the procedure and findings.</p> <p>The relevance of the changes in clinical terms would add important information. Please, analyse it. What is the risk-ratio decrement for each variable considering the "percent concerning"?</p> <p>Discussion Be careful when concluding about burnout, because only the exhaustion subscale was used.</p> <p>Please, include the effect size rules of thumb in the statistical analyses section. Page 19, line 27, please soften the statement "burnout..." (would be or could be ...) because it is hypothetical.</p> <p>References References number 40, and 47 are not complete.</p>
--	---

<b>REVIEWER</b>	Enriqueta Pujol-Ribera IDIAP JORDI GOL SPAIN
<b>REVIEW RETURNED</b>	27-Apr-2018

<b>GENERAL COMMENTS</b>	<p>I have assessed the manuscript "45 Good Things: A Prospective Pilot Study of the Three Good Things Intervention for Enduring Improvements in Healthcare Worker Resilience" (bmjopen-2018-022695).</p> <p>This is a relevant article about the effects of a low-cost web-based 15-day-long Intervention (Three Good Things) for enduring improvements in Healthcare Workers burnout, depression symptoms, and happiness at 1, 6, and 12-months, and in work-life balance at 1 and 6.</p> <p>In my opinion, this study has many strong points:</p> <ul style="list-style-type: none"> <li>• Brings a very new knowledge about the effects of a web-based 15-day-long low cost and simple Intervention (Three Good Things), cultivating positive cognition and emotions, in burnout, depression symptoms, work-life balance and happiness in healthcare workers</li> </ul>
-------------------------	--

	<ul style="list-style-type: none"> <li>• The cited literature is actual, relevant and relates to the objectives</li> <li>• The different sections of the manuscript are well developed</li> <li>• The sample included many different disciplines in healthcare workers</li> <li>• The instruments used to assess targeted variables are appropriate</li> <li>• Results are clearly described</li> <li>• Clarity and order of exposure</li> </ul> <p>Specific comments about some aspects that could be improved: I have raised specific comments to the authors with the aim to provide improvements in some issues of the manuscript. The TITLE should be more specific including more keywords such burnout, depression, work-life balance and happiness.</p> <p>I propose that all de manuscript should be consistent in the definition of the term “resilience” and the scales related to this term. In the measures section (page 7), authors mentioned that resilience is measured with 2 scales: “To assess changes in burnout and wellbeing hypothesized to result from 3GT, validated metrics included a burnout scale, a depression scale, and two measures of resilience (i.e., work life balance, and subjective happiness;....”.</p> <p>I recommend clarifying if the authors consider that resilience includes burnout, depression, work life balance, and subjective happiness or work life balance, and subjective happiness alone.</p> <p>The ABSTRACT provides a structured summary of the study. I recommend explaining the methods in-depth, especially considering that the journal admits up to 300 words. I suggest specifying the design (a repeated measures before-after study; the scales of measure of the variables and the assessment of effect size of the intervention methods).</p> <p><b>INTRODUCTION</b> The authors justify the need to do this study. They show in-depth knowledge about the scientific background and the current state of the subject and their explanation are well argued and rationale.</p> <p><b>METHODS</b> I suggest them the authors to include in the design “a non-randomized repeated measures before-after trial”. The setting, the eligibility criteria for participants, method of recruitment, and locations where data were collected are clear. Although it would be interesting to know the total number of people whom they proposed to participate and the number who have declined to be enrolled. Methods used to collect data and information about scales are well explained. In the statistical analysis section I propose adding how the size effect (Cohen D) has been calculated, and cut-offs his classification (small, moderate, and large).</p> <p><b>RESULTS</b> The results section presents a clear and concise description of the more important finding, without interpretation of them. The order is consistent with the methods section. There is not a figure with the participant’s flow to each phase. So that, the number of the participants of each measurement are recorded in the tables</p>
--	---

	<p>In the first paragraph, “Respondent Demographics” and in the discussion section, it seems me relevant to highlight that the majority of answers are from the professionals of the day shift. The “concerning group analysis” it's a very interesting subgroup analysis</p> <p><b>TABLES</b></p> <p>In Table 1, show respondents demographic characteristics in each assessment point, but there is no a legend explaining all abbreviations, In the title, I propose to avoid acronyms, and, in the second row headings, write “month” instead of “mo”. Moreover, usually we see the N of each evaluation in this row. It is an important information at the start of the table. Finally, I suggest adding a third row with n (%), removing all symbols %, except for reliability of the scales, and write correctly “Chrobach “.</p> <p>In Table 2, also there is no a legend and should be clearer and easier to read. I propose to indicate the complete name of the all measuring instrument, for example instead of only burnout “Maslach Burnout Inventory”... and indicate in the legend the range of values of the instruments and his sense, for example “Maslach Burnout Inventory of 22 items. Values range from 0 to 140. Higher scores indicate a higher degree of burnout or their subscales except for personal accomplishment”.</p> <p>The T value does not seem relevant, and the level of statistical significance of the test could be presented whit asterisks (*P&lt;0.05). Instead I miss a column whit the baseline values and I suggest to present the numerical values whit only one decimal place.</p> <p>In Table 3, take into account the recommendations on the previous tables. Table 3 informs us about the total sample and the effect size of the 3GT intervention without precision measures (such as 95% confidence interval).</p> <p><b>DISCUSSION</b></p> <p>The first sentences synthesize the main findings. The authors compare their results with those of other studies and they say that the positive effect size of the 3GT intervention are impressive and surprising taking in a count how quick and simple is to participate in this intervention. Their interpretation is consistent with results, and honestly recognizes their limitations (especially no randomization, no control group, attrition rate and the possibility the possibility that the effects observed are not the results of the intervention alone). Authors emphasize the relevance of the intervention given the need to reduce burnout through shorter, simple, low-cost and applicable interventions like 3GT.</p>
--	--

<b>REVIEWER</b>	Anthony Montgomery University of Macedonia, Thessaloniki, Greece
<b>REVIEW RETURNED</b>	05-Jun-2018

<b>GENERAL COMMENTS</b>	<p>The authors are to be commended for conducting an important pilot study of the 3GT method. The paper is severely weakened by the sampling and the fact that the authors did not include a control group. However, the paper is an important pilot study and I would recommend the paper for publication with the following minor revisions:</p> <ol style="list-style-type: none"> <li>1. The authors have not provided any meaningful information on the sample that participated in the research. The authors need to help the reader understand the context of research, in terms of the</li> </ol>
-------------------------	---

	<p>sample that participated. For example, the link to the survey was on their website. Do the authors have any idea of the composition of the HCWs who visit their website?</p> <p>2. The majority of respondents were women. Is this important or noteworthy?</p> <p>3. The suggestion that emotional exhaustion is a good measure of burnout is not consistent with the recent work by Leiter and Maslach (2016) suggesting that Emotional Exhaustion alone represents an overextended profile rather than burnout per se. In the research, the authors found that at the 1 month follow-up, baseline levels of burnout did not predict dropping out. This may be due to them using a less reliable definition of burnout? Reference: Leiter, M., &amp; Maslach, C. (2016). Latent burnout profiles: A new approach to understanding the burnout experience. <i>Burnout Research</i>, 3, 89-100.</p> <p>4. The category of 'present concerning' was an interesting idea for analysis, but the authors need to provide more information on the rationale for using this category. The cited paper (Lyubomirsky S, Lepper HS) does not discuss the category of 'present concerning'. It's difficult to assess how meaningful this category is.</p> <p>5. Additionally, how meaningful is the 'present concerning' category given the potential selectivity of the participants?</p> <p>6. Given that the sample was a convenience sample and pilot study, the 'updating' of the Job-Demands Resources theory is overextending the results of the research. I like the ideas that the authors suggest regarding the role of positive emotions, but their data do not warrant the conclusions they reach (at present). As already mentioned, they have failed to take account of the different profiles of burnout suggested by Leiter and Maslach (2016). I recommend that they delete this section.</p> <p>7. Finally, it would be interesting for the authors to provide information on what their participants were grateful about.</p>
--	---

<b>REVIEWER</b>	Michael K Howlett Department of Emergency Medicine, Dalhousie University, Canada
<b>REVIEW RETURNED</b>	08-Jun-2018

<b>GENERAL COMMENTS</b>	<p>Abstract: methods section vague. Results: I am not sure the significant claims of the article are warranted based on the methods used.</p> <p>Introduction: Research question p 6 lines 38-50: Study is a before-after prospective cohort survey style. Some methods are included in the paragraph, should go in next section. PICO is measuring that well being is changed. Was the sub group analysis declared prior to initiating the study?</p> <p>Not sure claims of outcome efficacy is being evaluated at a clinical level, especially when only portions of standardized scales are used and there is no specific diagnostic evaluation of study participants over time.</p> <p>Methods:</p> <p>Inclusion of diverse types of professionals that are not front line health care workers may significantly influence results. Front line MD and Nurse staff are a minority of the sample. They may have clinical burnout or depression but have increased exit rates, thus missed in the outcomes data.</p>
-------------------------	---

	<p>References for percent concerning thresholds seem to be from the authors' own previous work, use nontraditional methodology with reference to the tools, and have values arbitrarily assigned in a binary fashion for non binary data (ranges), leading to concerns about external validity and accuracy of the conclusions made. See below.</p> <p>I have a concern for example as to how Maslach Burnout Inventory scale is used, the tool I am most familiar with. A small subset of one parameter (emotional exhaustion) is used as a proxy for the MBI. The author's literature support for this methodology depends almost exclusively on a single group of authors' tool for assessing burnout in the context of safety culture measurements in the US. I question whether using an MBI subset in this manner, and using the arbitrary calculation of cutoffs for burnout in the quoted literature can be said to be valid, since this has not been externally validated nearly so rigorously as the MBI itself. Secondly the MBI does not set clear cutoffs for burnout vs nonburnout. I am not aware of a burnout scale sufficiently validated for burnout diagnosis at the individual level in the literature, and I am not sure the scales can be properly used in a binary fashion. Thirdly I believe that personal accomplishment, the third subscale in the MBI, is important as a mitigating factor in emergency medicine and nursing, and so I am very reluctant to throw it away as a factor in burnout. Finally, burnout measured solely as EE is insufficient, since burnout is present only when there is a resulting serious breakdown in the capacity to perform work (not measured). Suggesting the presence of the burnout syndrome based solely on these scales is a common misconception that leads to subjectivity and overcall. I do recognize the authors use the term "percent concerning" at a scale arbitrarily chosen as 50%, however the original MBI does not use this, and Christine Maslach and other collaborators e.g., Michael Leiter warn against over dependence on its own "low, medium and high" burnout allotments, that were at arbitrarily assigned 33% intervals.</p> <p>The CES-D10, as the author states, is a screening tool. It therefore should be used cautiously before drawing definitive conclusions.</p> <p>Work Life Balance questions did not include the whole Climate Scale, again leading to questions around validity since the original tool is not used, but a portion.</p> <p>Subjective Happiness: the benchmark used again was set based on "benchmarking data" but I was unable to trace this literature to understand how.</p> <p>Results: I am more familiar with the MBI so I will highlight my concerns with the outcomes claims of the study with reference to that particular tool.</p> <p>Front line workers such as nurses and MDs, where burnout syndrome tends to be greatest in health care formed a minority of the sample. Do we know if the improvement in scores was in the large (more than half) of study participants who were not front line caregivers?</p> <p>I question whether the statistically significant differences in the well being measures actually translates into a meaningful outcomes as is claimed. It is optimistic for example, to say that small changes in a subset of a subscale of the MBI using an arbitrary cutoff point actually means burnout, (i.e., effects of the workplace on work performance). Neither the MBI nor its subsets are valid for use in diagnosis nor does it have established binary</p>
--	---

	<p>cutoffs. Thus while p. 16 line 6,7 claims improvement in burnout I do not believe the stats actually prove this outcome.</p> <p>Line 10: it seems unlikely to me that 50% of the study population are clinically burned out. There seems to be a subtle shift in meaning during the article from risk of the syndrome to the existence and improvement of burnout as an outcome measure. Clinical burnout is simply not that common and I do not believe is measured here.</p> <p>The definition of burnout on page 18 lines 27 to 34 is insufficient in that burnout is differentiated by demonstrated effects on work performance, not just negative influence related to work on anxiety or stress. The subsequent hypothesis is interesting but the findings, dependent on partial measurements not externally validated in other populations and with significant adaptations of original scales, cannot be more than hypothesis generating. While the study may measure self reported subscale items, it in no way represents changes in levels of true burnout or depression. That is, in cannot be said from this data that these HCWs had improvements in diagnoses of burnout or depression, if that is how the author's wording is intended to be interpreted.</p> <p>Line 50 on outcomes. Following the above points, I do not see that this study shows improvement in outcomes. It measures improvements in subscale measures, not in actual diagnoses. The effect size relates to what was actually measured, not to diagnoses as well. This is not a clinical outcomes study.</p> <p>The outcomes effect size does not pass the sniff test. Knowing the difficulties of front line staff in the workplace, it is difficult to believe that true burnout or depression can be modified so easily, especially since there is no actual change in the work environment. The author acknowledges this limitation; in so doing I would suggest caution in over interpreting the results.</p> <p>We have found high attrition rates in our studies on burnout as well. It is possible that those more burned out exit the workplace and are not captured at later stages, thus biasing the sample toward those with greater resiliency. Burnout studies often fail to capture those with the most clinical concern.</p> <p>Conclusion:</p> <p>In my opinion the strength of the study is to present a tool that can be further studied. However, I feel this is in the range of "hypothesis generating" rather than representing a stable theory, and the methods used need further validation coupled with clinical outcomes data before one can say the intervention is effective at reducing burnout.</p> <p>In summary, in my opinion the methodology needs external validation, and claims of improvement in outcomes in this study are premature. The study is interesting from a hypothesis generating point of view. I would welcome arguments from the authors on the methods validity question.</p>
--	--

## VERSION 1 – AUTHOR RESPONSE

1	<p>Reviewer(s)' Comments to Author:</p> <p>Reviewer: 1  Reviewer Name: Jesus Montero-Marin  Institution and Country: Red de Investigación en Atención Primaria (REDIAPP), Spain Please state any competing interests or state 'None declared': None declared</p>
---	--

		Please leave your comments for the authors below
1	1	Abstract Please, provide more information in the methods section (design, instruments, data analyses, etc.
		Thank you, we have revised this section accordingly.
1	2	Introduction Please, re-write the last paragraph clarifying the main and secondary aims of the study, with their corresponding hypothesis.
		We have clarified our main and secondary aims of the study with corresponding hypotheses, where applicable.
1	3	Method Any sample size or power estimates were developed prior to get the sample used? Please specify.
		We have included information about power in our statistical analysis section.
1	4	What is the main outcome? All of the measures are main outcomes?
		Our measure of emotional exhaustion was our main outcome. We used a derivative of the Maslach Burnout Inventory's Emotion Exhaustion Subscale. We have clarified that it is the primary outcome throughout the paper. Thank you for this clarifying question.
1	5	The consideration of resilience as work-life balance and subjective happiness, etc needs to be justified. Please, explain in the introduction more deeply what resilience is, and justify why the indicators (variables) used in the present study might be reliable measures of it. Resilience appears as a different construct than burnout, depression, etc. The use of the construct in the text may drive to confusion.
		We agree that the term "resilience" is interpreted differently, and thus could lead to confusion. We have decided to use more widely shared language to describe our set of metrics by changing our wording to "well-being metrics".
1	6	Results Please, provide demographics for the whole group to facilitate a general view.
		The current demographics table outlines various variables (role, gender, shift type, years of experience, years in current position) at day 1 (N = 228), and all follow-up assessments. We believe this offers readers both a general view of the sample, and the ability to see if demographics vary widely at different time points (and in large part they do not).
1	7	Please, provide a measure of the programme compliance and analyse the possible determination of the level of compliance and improvements.
		This is a good point. We have conducted additional analyses to probe the question of compliance. Ostensibly the number of days a participant completes the 3GT activity reflects his/her compliance. To determine associations between compliance and improvements, we ran spearman correlations between wellbeing change scores and the number of days participant completed three good things. We did not find any correlations between number of days participants completed the activity and changes in wellbeing. These results can be found on pg. 13-14.

1	8	Please, include the internal consistence of the metrics in the corresponding method section (it is not a study objective, isn't it?).
		We have included the internal consistency for each measure in the method section and in Table 1.
1	9	Please, include standard errors in Figure 1 so that one can view possible differences along the time.
		We have added the standard error bars to Figure 1, thank you.
1	10	Page 13, line 22-23, where it reads Table 1, it could be Table 2. Please revise it.
		Revised, thank you.
1	12	Please, remove colours from tables to facilitate reading.
		We have removed grey shading from the tables.
1	13	How was managed missing data? A figure would clarify the flow of participants. Table 3 includes a very high n at baseline, and n is not the same across time points. Please, explain it in detail, because apparently you are providing a different number of participants than those really were included in the analyses.
		Our analyses utilized list-wise deletion for missing data. Sample sizes at each assessment point were clarified in the respondent demographic section of the results, page 13. We agree that the baseline column in Table 3 was confusing, so to avoid confusion, Table 3 now only contains results from participants with data from the paired t-tests.
1	14	Were effect sizes corrected by correlated measures? Please, specify.
		Yes, effect sizes were computed using correlated (or dependent) samples ( $d_{av}$ ). We have clarified this in the abstract and statistical analyses sections.
1	14	I think it would be worth to analyse deeply the participant's experiences by systematic content analysis of submissions and comments, detailing the procedure and findings.
		We agree that an in depth and systematic content analysis of participants' submissions and comments would be interesting, however that level of analysis is outside the scope of the current pilot study. Rippstein—Leuenberger and colleagues have an article on a thematic coding of 3GT submissions (citation below). To help readers understand the types of submissions participants made, and their comments about the 3GT experience in general, we have included a representative sample of both in the appendix. Our future research will include a systematic assessment of submissions and comments. Rippstein-Leuenberger K, Mauthner, Oliver, Sexton, J. Bryan, Schwendimann, Rene. Three Good Themes: a qualitative analysis of the Three Good Things intervention in health care workers. Rev. 2017.
1	15	The relevance of the changes in clinical terms would add important information. Please, analyse it. What is the risk-ratio decrement for each variable considering the "percent concerning"?
		The current study assessed changes in wellbeing across a set of 4 validated metrics. Although we also use thresholds to examine "concerning" at

		baseline as an exploratory subgroup, the current metrics are not to be used to diagnose clinical levels of mental health concerns. Reviewer 4 expressed that we should be careful to convey that our measures are not used to clinically diagnose. We have added text to reflect this notion, pg. 8, "This threshold, and all others, should not be considered clinically diagnostic, but rather identify those whose scores are concerning"
1	16	Discussion Be careful when concluding about burnout, because only the exhaustion subscale was used.
		This is an excellent point. Since we use a derivative of the emotional exhaustion subscale we have revised the paper throughout to more precisely describe our findings.
1	17	Please, include the effect size rules of thumb in the statistical analyses section.
		Thank you, we have moved this information to this section.
1	18	Page 19, line 27, please soften the statement "burnout..." (would be or could be ...) because it is hypothetical.
		Thank you, we have changed and softened the wording in this section.
1	19	References References number 40, and 47 are not complete.
		We have updated these references.
2		Reviewer: 2 Reviewer Name: Enriqueta Pujol-Ribera Institution and Country: IDIAP JORDI GOL, SPAIN Please state any competing interests or state 'None declared': NONE DECLARED  Please leave your comments for the authors below I have assessed the manuscript "45 Good Things: A Prospective Pilot Study of the Three Good Things Intervention for Enduring Improvements in Healthcare Worker Resilience" (bmjopen-2018-022695).
2	1	This is a relevant article about the effects of a low-cost web-based 15-day-long Intervention (Three Good Things) for enduring improvements in Healthcare Workers burnout, depression symptoms, and happiness at 1, 6, and 12-months, and in work-life balance at 1 and 6. In my opinion, this study has many strong points: <ul style="list-style-type: none"> <li>• Brings a very new knowledge about the effects of a web-based 15-day-long low cost and simple Intervention (Three Good Things), cultivating positive cognition and emotions, in burnout, depression symptoms, work-life balance and happiness in healthcare workers</li> <li>• The cited literature is actual, relevant and relates to the objectives</li> <li>• The different sections of the manuscript are well developed</li> <li>• The sample included many different disciplines in healthcare workers</li> <li>• The instruments used to assess targeted variables are appropriate</li> <li>• Results are clearly described</li> <li>• Clarity and order of exposure</li> </ul>
		Thank you for this positive feedback.
2	2	Specific comments about some aspects that could be improved: I have raised specific comments to the authors with the aim to provide improvements in some issues of the manuscript.

		The TITLE should be more specific including more keywords such burnout, depression, work-life balance and happiness.
		We agree and have revised the title accordingly.
2	3	<p>I propose that all de manuscript should be consistent in the definition of the term “resilience” and the scales related to this term. In the measures section (page 7), authors mentioned that resilience is measured with 2 scales: “To assess changes in burnout and wellbeing hypothesized to result from 3GT, validated metrics included a burnout scale, a depression scale, and two measures of resilience (i.e., work life balance, and subjective happiness;....”.</p> <p>I recommend clarifying if the authors consider that resilience includes burnout, depression, work life balance, and subjective happiness or work life balance, and subjective happiness alone.</p>
		This is an excellent point. We agree that all 4 of our metrics could be considered “resilience metrics”, however, since views on the definition of “resilience” differ more so than the definition of “well-being” we have decided to call our set of measures “well-being metrics”
2	4	The ABSTRACT provides a structured summary of the study. I recommend explaining the methods in-depth, especially considering that the journal admits up to 300 words. I suggest specifying the design (a repeated measures before-after study; the scales of measure of the variables and the assessment of effect size of the intervention methods).
		Thank you, we have changed the abstract accordingly.
2	5	<p><b>INTRODUCTION</b></p> <p>The authors justify the need to do this study. They show in-depth knowledge about the scientific background and the current state of the subject and their explanation are well argued and rationale.</p>
		Thank you.
2	6	<p><b>METHODS</b></p> <p>I suggest them the authors to include in the design “a non-randomized repeated measures before-after trial”.</p>
		We have made this change, thank you.
2	6	<p>The setting, the eligibility criteria for participants, method of recruitment, and locations where data were collected are clear. Although it would be interesting to know the total number of people whom they proposed to participate and the number who have declined to be enrolled.</p> <p>Methods used to collect data and information about scales are well explained.</p> <p>In the statistical analysis section I propose adding how the size effect (Cohen D) has been calculated, and cut-offs his classification (small, moderate, and large).</p>
		Thank you. We have made changes regarding the calculation of Cohen's d and the cut-off classifications.
2	7	<p><b>RESULTS</b></p> <p>The results section presents a clear and concise description of the more important finding, without interpretation of them. The order is consistent with the methods section. There is not a figure with the participant's flow to each phase. So that, the number of the participants of each measurement are recorded in the tables In the first paragraph, “Respondent Demographics” and in the discussion section, it seems me relevant to highlight that the majority of answers are from the professionals of the day shift.</p>

		We have added to the respondent demographics section that 76.3% of those who completed day 1 were day shift workers. Thank you.
2	8	The “concerning group analysis” it's a very interesting subgroup analysis TABLES In Table 1, show respondents demographic characteristics in each assessment point, but there is no a legend explaining all abbreviations, In the title, I propose to avoid acronyms, and, in the second row headings, write “month” instead of “mo”. Moreover, usually we see the N of each evaluation in this row. It is an important information at the start of the table. Finally, I suggest adding a third row with n (%), removing all symbols %, except for reliability of the scales, and write correctly “Chrobach “.
		These are excellent suggestions and we have made all of these changes with one exception. Rather than include a legend we have simply removed abbreviations.
2	9	In Table 2, also there is no a legend and should be clearer and easier to read. I propose to indicate the complete name of the all measuring instrument, for example instead of only burnout "Maslach Burnout Inventory"... and indicate in the legend the range of values of the instruments and his sense, for example “Maslach Burnout Inventory of 22 items. Values range from 0 to 140. Higher scores indicate a higher degree of burnout or their subscales except for personal accomplishment”.
		We have made these changes to table 1 and 2, however we have included the scoring and range information in the measures section.
2	10	The T value does not seem relevant, and the level of statistical significance of the test could be presented whit asterisks (*P<0.05). Instead I miss a column whit the baseline values and I suggest to present the numerical values whit only one decimal place.
		We appreciate these suggestions. We have replaced p-values with 95% CIs in Table 2 and 3, and designated levels of significance with asterisks. To preserve specificity, we have retained decimals to the hundredth place. Finally, since baseline values are reflected in table 3, we did not include them in table 2.
2	11	In Table 3, take into account the recommendations on the previous tables. Table 3 informs us about the total sample and the effect size of the 3GT intervention without precision measures (such as 95% confidence interval).
		We have added 95% CIs in Table 3 and have removed p-values and replaced them with asterisks. Thank you.
2	12	DISCUSSION The first sentences synthesize the main findings. The authors compare their results with those of other studies and they say that the positive effect size of the 3GT intervention are impressive and surprising taking in a count how quick and simple is to participate in this intervention. Their interpretation is consistent with results, and honestly recognizes their limitations (especially no randomization, no control group, attrition rate and the possibility the possibility that the effects observed are not the results of the intervention alone). Authors emphasize the relevance of the intervention given the need to reduce burnout through shorter, simple, low-cost and applicable interventions like 3GT.
		Thank you.
3		Reviewer: 3 Reviewer Name: Anthony Montgomery

		<p>Institution and Country: University of Macedonia, Thessaloniki, Greece Please state any competing interests or state 'None declared': None declared</p> <p>Please leave your comments for the authors below</p>
3	1	<p>The authors are to be commended for conducting an important pilot study of the 3GT method. The paper is severely weakened by the sampling and the fact that the authors did not include a control group. However, the paper is an important pilot study and I would recommend the paper for publication with the following minor revisions:</p> <p>1. The authors have not provided any meaningful information on the sample that participated in the research. The authors need to help the reader understand the context of research, in terms of the sample that participated. For example, the link to the survey was on their website. Do the authors have any idea of the composition of the HCWs who visit their website?</p>
		<p>Thank you for your feedback. Unfortunately we do not have a way to assess the composition of the HCWs who visited the website, but we inserted this sentence into the design and patient population section:</p> <p>"Generally, people who seek the content on our website have a background in patient safety, quality improvement or well-being."</p> <p>Also, we did ask a fairly comprehensive demographics questionnaire for participants in the study. The breakdown of position/roles, gender, shift type, years of experience, and years in current experience, can be found in Table 1.</p>
3	2	<p>2. The majority of respondents were women. Is this important or noteworthy?</p>
		<p>We have added the following to the discussion section: "We note also that our sample is comprised largely of women (81% of participants at baseline), however this is generally in line with base rates of women working in healthcare in the US (in 2017, 75% of healthcare workers were female).<sup>51</sup> Future research should oversample for men and examine whether 3GT is equally effective across genders."</p>
3	3	<p>3. The suggestion that emotional exhaustion is a good measure of burnout is not consistent with the recent work by Leiter and Maslach (2016) suggesting that Emotional Exhaustion alone represents an overextended profile rather than burnout per se. In the research, the authors found that at the 1 month follow-up, baseline levels of burnout did not predict dropping out. This may be due to them using a less reliable definition of burnout?</p> <p>Reference: Leiter, M., &amp; Maslach, C. (2016). Latent burnout profiles: A new approach to understanding the burnout experience. <i>Burnout Research</i>, 3, 89-100.</p>
		<p>This is an important point. We agree that emotional exhaustion is not the same thing as burnout, and therefore have changed our use of "burnout" to "emotional exhaustion" to precisely describe our primary measure of interest. We used 4 metrics of well-being, one of them was emotional exhaustion. The quartet that we used, EE, Depression, Happiness and Work-Life Balance provide a set of well-being indices, of which the burnout was represented by EE. We see EE as one facet of well-being, not that it represents all of burnout perfectly well by itself. Meta-analysis has revealed that of the three sub-scales (emotional exhaustion, depersonalization and personal accomplishment), emotional exhaustion consistently produces the largest and most consistent coefficient alpha estimates, while depersonalization and personal accomplishment were both lower and less consistent than emotional exhaustion. We added this under Measures, on page 8. Also, as we identified in the original manuscript draft, EE can be used to discriminate between burned out and non-burned out employees according to ICD-10 criteria. On a personal note, we agree with</p>

		Maslach that there may be a better way to conceptualize burnout than the seemingly relentless focus on EE, but EE outperforms the other two domains, consistently, across samples and cultures and translations, in terms of psychometrics.
3	4	4. The category of 'present concerning' was an interesting idea for analysis, but the authors need to provide more information on the rationale for using this category. The cited paper (Lyubomirsky S, Lepper HS) does not discuss the category of 'present concerning'. It's difficult to assess how meaningful this category is.
		<p>We are glad you find the sub-sample analyses of interest. We believe that it can be fruitful to assess efficacy of 3GT for those who are particularly struggling. The cut-offs for the "percent concerning" groups vary in how much prior research exists for them. None of the cut-offs are used here as clinically diagnostic thresholds.</p> <p>Threshold descriptions of results have been used by our team in debriefings of results to great effect because relative to showing means, SDs and distributions, threshold results can be grasped quickly. Thresholds provide an anchor for interpretation, offering a way of communicating something about the distribution of the data with a single number. We are not suggesting this is the only way to look at well-being metrics, but we are suggesting that this is one way that resonates with individuals and with groups.</p> <p>In the measures section we describe how each threshold was established. For the cited paper Lyubomirsky S, Lepper HS, table 1 outlines mean scores for a variety of samples, including 8 US college student samples (means between 4.79 and 4.99; SDs between 1.04 and 1.72), US adult community sample (mean = 5.62, SD = .96), and a US adult female community sample (mean = 4.80, SD = 1.12). Given that the means flank 5, we used that as the threshold. Lyubomirsky and others have published extensively on the links between Subjective Happiness and physical and mental health (including mortality studies) but we are using the scores less than 5 as an indicator of percent concerning because subjective well being is not normally distributed, with most people rating themselves in the top three points of the 7 point scale.</p> <p>For EE, we used a percent concerning threshold of mid-point or higher (50 or higher, out of 100) to represent folks who are struggling as they are not disagreeing with emotional exhaustion statements. This corresponds with the original scaling and designation of EE as mild, while scoring 100 would be considered severe. For practical purposes, reporting the percent concerning as the percent that are at least mildly emotionally exhausted is something that resonates with managers and with frontline staff. When we discuss EE over time and report the "rate of emotionally exhausted individuals increased from 32% to 51%," that is something that resonates much more with lay audiences than saying the mean EE score changed from 3.1 to 3.6 on a 5-point scale. Again, these thresholds are meant to convey a sense of struggling vs not struggling using relatively little information.</p> <p>We added the following clarification to the paper on page 9, just before the section labeled Depressive Symptoms:  These "percent concerning" thresholds should not be considered clinically diagnostic, but rather identify those whose scores represent more of a struggle on a given metric. The percent concerning threshold can be grasped quickly, providing an anchor for interpretation, and offering a way of communicating something about the distribution of the data within a single number.</p>
3	5	5. Additionally, how meaningful is the 'present concerning' category given the potential selectivity of the participants?

		Please see above response. We are regularly asked whether those who are suffering at greater levels are more or less likely to drop out of burnout research, as well as whether they are more or less likely to benefit from tools like Three Good Things. We therefore believe that the subscale analysis is useful and of interest. It is true that the generalizability is limited to only those who agree to participate, however it is noteworthy that we had a range of wellbeing scores throughout the study.
3	6	6. Given that the sample was a convenience sample and pilot study, the 'updating' of the Job-Demands Resources theory is overextending the results of the research. I like the ideas that the authors suggest regarding the role of positive emotions, but their data do not warrant the conclusions they reach (at present). As already mentioned, they have failed to take account of the different profiles of burnout suggested by Leiter and Maslach (2016). I recommend that they delete this section.
		We agree that the findings of the current study do not warrant statements regarding the updating of burnout theory as they were submitted. We have retained some of these general ideas, and removed sentences that overextend the current findings. Using suggestions from other reviewers, we have modified the language by softening it.
3	7	7. Finally, it would be interesting for the authors to provide information on what their participants were grateful about.
		We have included a representative sample of the 3GT submissions made by participants in the appendix. It is fascinating and uplifting to read through their logs, as both of us and a research assistant can wholeheartedly attest to – people have shared their tender moments and it is a privilege to witness their efforts.
4		Reviewer: 4 Reviewer Name: Michael K Howlett Institution and Country: Department of Emergency Medicine, Dalhousie University, Canada Please state any competing interests or state 'None declared': None declared  Please leave your comments for the authors below see enclosed file
4	1	Abstract: methods section vague.
		We have made several additions and clarifications to the methods section of the abstract.
4	2	Introduction: Research question p 6 lines 38-50: Study is a before-after prospective cohort survey style. Some methods are included in the paragraph, should go in next section. PICO is measuring that well being is changed. Was the sub group analysis declared prior to initiating the study? Not sure claims of outcome efficacy is being evaluated at a clinical level, especially when only portions of standardized scales are used and there is no specific diagnostic evaluation of study participants over time.
		Thank you, we have revised the research aims paragraph at the end of the introduction in line with your recommendations. The sub-group analysis was hypothesized prior to the study. The study is not examining clinically diagnostic levels for any of the measures. We have made note of this in the measures section, pg. 9.
4	3	Methods: Inclusion of diverse types of professionals that are not front line health care workers may significantly influence results. Front line MD and Nurse staff are a minority of the sample. They may have clinical burnout or depression but have increased exit rates, thus missed in the outcomes data. References for percent concerning thresholds seem to be from the authors' own previous work, use nontraditional methodology with reference to the tools, and have values arbitrarily assigned in a binary fashion for non binary data (ranges), leading to concerns about external validity and accuracy of the conclusions made. See below.
		We appreciate this reviewer's comments on these topics.

		<p>Traditional frontline clinical workers that comprised larger categories were represented in Table 1. Reviewer 4 may be expressing concern about the category of “other” in table 1 (24.6%) as being comprised by non clinical workers. To reduce the size of the already large table, smaller categories of participants, including some that were clinical (e.g., “pharmacists” “occupational therapists”) were grouped into the “other”. We have added a sentence to our participants section accordingly:</p> <p>Participants in the “other” role category most commonly identified as working in roles with clinical exposure not captured elsewhere (e.g., occupational therapist, physical therapist, pharmacist, long term care), patient safety or quality, administration, counseling and psychological services, and patient revenue management.”</p> <p>Regarding the second concern about the threshold analyses: We were curious to see whether 3GT had a greater (or lesser) impact on participants who began with more concerning levels of well-being. We decided to conduct exploratory analyses on this group, and the group was identified by having scores that exceeded thresholds that were identified in prior research – CESD-10 threshold was identified by Bjorgvinsson et al., 2013; Subjective Happiness score means fall around 5 across samples (Lyubomirsky &amp; Lepper, 1999), thus that threshold was used. Burnout and WLB thresholds were identified in our own previous work. We clarified how we established the burnout and WLB thresholds in the measures section.</p> <p>We added the following clarification to the paper on page 9, just before the section labeled Depressive Symptoms:</p> <p>These “percent concerning” thresholds should not be considered clinically diagnostic, but rather identify those whose scores represent more of a struggle on a given metric. The percent concerning threshold can be grasped quickly, providing an anchor for interpretation, and offering a way of communicating something about the distribution of the data within a single number.</p> <p>Although we use a cut off to identify this subsample, we did not “have values arbitrarily assigned in a binary fashion for non binary data” – all scores on the measures were retained, no replacement occurred, and we did not use binary data. Moreover, we reported the results two ways, first using the full range of scores converted to a 100 point scale, and second as a reflection of the percent concerning.</p>
4	4	Results: I am not sure the significant claims of the article are warranted based on the methods used.
		In the results section we made an effort to not make any claims, but to simply describe the results of the analyses. In the discussion section we have toned down our claims regarding the impact of our findings and their implication for theory, etc. We agree that that our study is limited by its method, and have outlined the extent of these limitations at the end of the discussion section.
4	5	<p>I have a concern for example as to how Maslach Burnout Inventory scale is used, the tool I am most familiar with.</p> <p>A small subset of one parameter (emotional exhaustion) is used as a proxy for the MBI. The author’s literature support for this methodology depends almost exclusively on a single group of authors’ tool for assessing burnout in the context of safety culture measurements in the US. I question whether using an MBI subset in this manner, and using the arbitrary calculation of cutoffs for burnout in the quoted literature can be said to be valid, since this has not been externally validated nearly so rigorously as the MBI itself.</p> <p>Secondly the MBI does not set clear cutoffs for burnout vs non-burnout. I am</p>

		<p>not aware of a burnout scale sufficiently validated for burnout diagnosis at the individual level in the literature, and I am not sure the scales can be properly used in a binary fashion. Thirdly I believe that personal accomplishment, the third subscale in the MBI, is important as a mitigating factor in emergency medicine and nursing, and so I am very reluctant to throw it away as a factor in burnout. Finally, burnout measured solely as EE is insufficient, since burnout is present only when there is a resulting serious breakdown in the capacity to perform work (not measured). Suggesting the presence of the burnout syndrome based solely on these scales is a common misconception that leads to subjectivity and overcall. I do recognize the authors use the term “percent concerning” at a scale arbitrarily chosen as 50%, however the original MBI does not use this, and Christine Maslach and other collaborators e.g., Michael Leiter warn against over dependence on its own “low, medium and high” burnout allotments, that were at arbitrarily assigned 33% intervals.</p>
		<p>We believe we addressed some of these concerns in response to comment #3 above.</p> <p>You have expressed some strong feelings about this and we appreciate your passion on this topic, as we share it ourselves. You are absolutely correct in saying that we should not consider EE alone to be burnout. We have changed the descriptions from burnout to emotional exhaustion throughout, which is an important improvement to this paper. We refer to our results as emotional exhaustion results, not burnout results.</p> <p>To clarify why we use EE as part of our set of well-being metrics, at the bottom of page 8 you will find:  Meta-analysis has revealed that of the three sub-scales (emotional exhaustion, depersonalization and personal accomplishment), emotional exhaustion consistently produces the largest and most consistent coefficient alpha estimates, while depersonalization and personal accomplishment were both lower and less consistent than emotional exhaustion. We used a 5-item derivative (31) of the original 9-item emotional exhaustion scale. In addition to being more psychometrically robust, emotional exhaustion can be used to discriminate between burned out and non-burned out outpatients suffering from work-related neurasthenia (according to ICD-10 criteria).<sup>32</sup></p> <p>Now about the nature of our sub-group analyses as they pertain to aspects of the current comment. First and foremost, although we were interested in the effect of 3GT on a concerning subsample, we did not mean to convey that the threshold established a clinical criteria for dichotomizing “burned out” and “not burned out”. To identify the subgroup, we used a cut off, but we did not assign a binary code, e.g., burned out =1, not burned out = 0 – rather we simply examined changes in wellbeing scores across assessment points for those who scored above the thresholds at baseline.</p> <p>We focused on the emotional exhaustion (EE) component of burnout because it is incredibly psychometrically robust, and significantly more robust than DP and PA, as demonstrated by a meta-analysis of 98 samples that we included in the revisions. Second, our intervention is emotions-related; i.e., increasing positive emotions. This decision is strengthened by recent research showing that person-directed interventions tend to be more effective at reducing emotional exhaustion than organizational-interventions (Dreison et al., 2018). Moreover, the current study looks beyond just emotional exhaustion – but also to depressive symptoms, subjective happiness, and work-life balance. These factors are of considerable interest to the field as we understand the scope of what it means to be a “well” healthcare worker.</p>
4	6	<p>The CES-D10, as the author states, is a screening tool. It therefore should be used cautiously before drawing definitive conclusions.</p>

		<p>We agree and have added “those who met the threshold for a positive screen for depression” to the discussion section on the CESD findings. In addition, we believe that the overarching message of this manuscript is that we would like to greenlight the use of 3 Good Things for further study in healthcare workers. We believe this because our pilot project using repeated measures over 4 assessment periods, including 12 month follow-ups, suggest that further clinical trials would be a good next step.</p> <p>We have toned down the conclusion accordingly.</p>
4	7	<p>Work Life Balance questions did not include the whole Climate Scale, again leading to questions around validity since the original tool is not used, but a portion.</p>
		<p>The current study was conducted before an 8<sup>th</sup> item (“felt frustrated with technology”) was added to the Work-life Scale. However, Cronbach’s alphas for the score for the 7 –item version of the scale were good (see table 1): .81, .81, .79, .72. There are two publications of the WLB scale, one with 7 items and one with 8 items. Psychometrically, they are interchangeable but more to the point, we provided sufficient evidence across all 4 assessment time points in the current study for you to know that validity of this scale is not a significant concern.</p>
4	8	<p>Subjective Happiness: the benchmark used again was set based on “benchmarking data” but I was unable to trace this literature to understand how.</p>
		<p>Please see comments 3 and 5 above about the use of thresholds and their intended meaning in this context.</p> <p>Threshold descriptions of results have been used by our team in debriefings of results to great effect because relative to showing means, SDs and distributions, threshold results can be grasped quickly. Thresholds provide an anchor for interpretation, offering a way of communicating something about the distribution of the data with a single number. We are not suggesting this is the only way to look at well-being metrics, but we are suggesting that this is one way that resonates with individuals and with groups. Also, we report results on a 100 point scale AND as percent concerning, not one instead of the other, but together.</p>
4	9	<p>Results: I am more familiar with the MBI so I will highlight my concerns with the outcomes claims of the study with reference to that particular tool. Front line workers such as nurses and MDs, where burnout syndrome tends to be greatest in health care formed a minority of the sample. Do we know if the improvement in scores was in the large (more than half) of study participants who were not front line caregivers?</p>
		<p>The largest group in our sample is registered nurses (24%) followed by Other manager (22.8), Other (21.5%, which is comprised of clinical social workers, clinical support (CMA, Nurses aid, etc.), Dietician/Nutritionists, Occupational Therapist, Respiratory Therapists, and those who selected “other”) and then Physicians (13.6%). These percentages indicate that sizable portion of our sample have at least some if not considerable clinical exposure. Note that we have clarified the groups that comprise “other” below table 1, given your concern about frontline workers.</p> <p>In our ongoing studies, we have found that emotional exhaustion rates are very high throughout healthcare, whether you work in finance, IT, the OR, or in administrative roles. Rates are higher among MDs and RNs, but not dramatically different. We find that there are bigger differences in work-life balance than in emotional exhaustion.</p>
4	10	<p>I question whether the statistically significant differences in the well being measures actually translates into a meaningful outcomes as is claimed. It is optimistic for example, to say that small changes in a subset of a subscale of the MBI using an arbitrary cutoff point actually means burnout, (i.e., effects of the workplace on work performance). Neither the MBI nor its subsets are</p>

		valid for use in diagnosis nor does it have established binary cutoffs. Thus while p. 16 line 6,7 claims improvement in burnout I do not believe the stats actually prove this outcome.
		<p>We understand this concern. First, we do not propose that our thresholds are to be used for clinical diagnosis, but rather to help readers understand the impact of 3GT on those reporting higher levels of emotional exhaustion (and our other measures). See pg. 8., “ This threshold, and all others, should not be considered clinically diagnostic, but rather identify those whose scores are concerning.” Second, we report effect sizes (Cohen’s d) to transparently communicate the sizes of the effects in our data, rather than simple levels of significance. We provide effect sizes for the entire sample, as well as the “concerning” subsample, in table 3. Although the effect sizes for the subsample are larger, we do not mean to overshadow the effect sizes found for the entire sample (in which, of course, no cut-off was used).</p> <p>In terms of “meaningfulness”: The effect sizes for EE in the current study are lower than those from considerably more intensive interventions. Our effect sizes for emotional exhaustion for the overall sample were .20-.34 (and .61 to .77 for the percent concerning subsample). We transparently communicate this in Table 3. More intensive interventions (e.g., mindfulness training) find effect sizes similar to those in our percent concerning subsample. Across our whole sample, we find the largest effect sizes (.41 to .52) for Depression symptoms. This indicates that 3GT may be particularly useful for those struggling with depression.</p>
4	11	Line 10: it seems unlikely to me that 50% of the study population are clinically burned out. There seems to be a subtle shift in meaning during the article from risk of the syndrome to the existence and improvement of burnout as an outcome measure. Clinical burnout is simply not that common and I do not believe is measured here.
		Our threshold for the concerning subscale includes participants who, on average, do not disagree with the emotional exhaustion items (i.e., mean scores of 3 or higher). We have revised our description of scoring above our concerning threshold as reflecting “at least mild emotional exhaustion”; pg. 18, “ Similar to previous work, half of our sample reported at least emotional exhaustion at baseline”. We agree that we are not capturing “clinical” burnout, rather, we are identifying a subgroup that is “concerning” in their reporting of ee symptoms.
4	12	The definition of burnout on page 18 lines 27 to 34 is insufficient in that burnout is differentiated by demonstrated effects on work performance, not just negative influence related to work on anxiety or stress. The subsequent hypothesis is interesting but the findings, dependent on partial measurements not externally validated in other populations and with significant adaptations of original scales, cannot be more than hypothesis generating. While the study may measure self reported subscale items, it in no way represents changes in levels of true burnout or depression. That is, in cannot be said from this data that these HCWs had improvements in diagnoses of burnout or depression, if that is how the author’s wording is intended to be interpreted.
		We have removed the definition of burnout on pg 21, based on recommendations from reviewers, which we agree with. We agree that we are not reporting changes in diagnosis in depression or burnout. However we are reporting changes in reported levels of emotional exhaustion and depression symptoms, in so far as respondents are honest and accurate in their responses. A wide literature in psychology relies on respondent reports of symptom level, without using “actual diagnosis”.
4	13	Line 50 on outcomes. Following the above points, I do not see that this study shows improvement in outcomes. It measures improvements in subscale measures, not in actual diagnoses. The effect size relates to what was actually measured, not to diagnoses as well. This is not a clinical outcomes study.

		<p>The outcomes effect size does not pass the sniff test. Knowing the difficulties of front line staff in the workplace, it is difficult to believe that true burnout or depression can be modified so easily, especially since there is no actual change in the work environment. The author acknowledges this limitation; in so doing I would suggest caution in over interpreting the results.</p>
		<p>We disagree that this study does not show improvement in outcomes. However this is likely due to a difference in how we view the term “outcomes”. We do not use diagnosis of “burnout” or not “burned out” as our outcome, as described in the paper. Instead, like other researchers, we used reduction in EE scores an indication of improvement in the EE domain of burnout.</p> <p>We agree with this reviewer that the effect sizes are notable given the ease of the intervention. Martin Seligman, the well-known psychologist who created the Three Good Things intervention, was also impressed with the effect of Three Good things (which he had participants do for 1 week). In an RCT comparing 3GT to a placebo, he found significant effects on happiness, and depression symptoms (also using CESD).</p> <p>A straightforward and simple tool to improve well-being is desperately needed by many healthcare workers right now. This pilot study indicates that 3GT may offer significant relief. Yet we agree that there are a number of limitations to the study and we should not over interpret the results. We have thoroughly outlined the limitations to the study in the discussions sections, and have reviewed our language throughout to reduce the appearance of over interpretation.</p>
4	14	<p>We have found high attrition rates in our studies on burnout as well. It is possible that those more burned out exit the workplace and are not captured at later stages, thus biasing the sample toward those with greater resiliency. Burnout studies often fail to capture those with the most clinical concern.</p>
		<p>We agree that those who are high in burnout are likely more vulnerable to not completing studies, and may simply have exited the workplace. Interestingly, we did not find that those higher in burnout at baseline were significantly more likely to drop out (<math>p = .12</math>); however we did find that those higher in depression, and lower in happiness and work-life balance, at baseline were more likely to drop out at 1-month (Table 2).</p> <p>Since research tends to find that those struggling have more to benefit from interventions (a finding we replicate in our current study), we do not believe that our sample was biased in a way that skewed for resiliency in our results. Our simple paired-t-tests used list-wise deletion, meaning that those who dropped out after baseline (potentially due to higher burnout or depression) were not captured in our analyses. Taking this together, this suggests that if those who dropped out actually stayed in the study, the effects of 3GT might have been even stronger.</p>
4	15	<p>Conclusion:</p> <p>In my opinion the strength of the study is to present a tool that can be further studied. However, I feel this is in the range of “hypothesis generating” rather than representing a stable theory, and the methods used need further validation coupled with clinical outcomes data before one can say the intervention is effective at reducing burnout. In summary, in my opinion the methodology needs external validation, and claims of improvement in outcomes in this study are premature. The study is interesting from a hypothesis generating point of view. I would welcome arguments from the authors on the methods validity question.</p>
		<p>We agree with this reviewer on a couple points. We agree that the central aim of this pilot study is to present a tool that can be further examined for improvements in wellbeing for workforce which direly needs wellbeing tools. We also agree that this is a hypothesis generating paper. We disagree that we do not demonstrate improvements in outcomes; however, as discussed above, this may be due to differences in the term “outcomes”. We use the term to capture improvement in a continuous measures of well-being. WE</p>

		agree that future research should use randomized samples to further test the efficacy of 3GTs. Considering the urgency for brief tools for well-being in healthcare, the demonstrated success of 3GTs in the psychology literature (Seligman et al., 2005), and the results of the current work, we believe this pilot study of the tool should be shared within the healthcare field for further examination.
--	--	--

STROBE Statement—checklist of items that should be included in reports of observational studies

	Item No	Recommendation	Checklist (including pg #s)
Title and abstract	1	(a) Indicate the study's design with a commonly used term in the title or the abstract	We have included the terms "prospective pilot study" and "Prospective repeated measures" in the title and abstract, respectively
		(b) Provide in the abstract an informative and balanced summary of what was done and what was found	We believe we have achieved this aim.
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	We believe we have achieved this aim; pgs. 4-7
Objectives	3	State specific objectives, including any prespecified hypotheses	Pg.6., "We hypothesized that 3GT participants would report reductions in burnout and depressive symptoms, as well as improvements in subjective happiness and work-life balance, between baseline assessments and all three follow-up assessments."
Methods			
Study design	4	Present key elements of study design early in the paper	We convey the study design in the first sentence of the methods section, p. 7. "This is a non-randomized repeated measures before-after trial of 3GT..."
Setting	5	Describe the setting, locations, and relevant dates, including periods of	Pg. 7, "data collected between February 2014 and March of

		recruitment, exposure, follow-up, and data collection	2015. HCWs enrolled over 4 weeks through a link on our center's website, labeled "Three Good Things February 2014." "Participants completed assessments at baseline, at 1-month (i.e., 1 month after the last day of 3GT), 6-months, and 12 months follow-up."
Participants	6	<p>(a) Cohort study—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up</p> <p>Case-control study—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls</p> <p>Cross-sectional study—Give the eligibility criteria, and the sources and methods of selection of participants</p> <p>(b) Cohort study—For matched studies, give matching criteria and number of exposed and unexposed</p> <p>Case-control study—For matched studies, give matching criteria and the number of controls per case</p>	<p>Pg. 7, "all HCWs (clinical and non-clinical) 18 and older were eligible to participate." "HCWs enrolled over 4 weeks through a link on our center's website, labeled "Three Good Things February 2014." "Links to online assessments were sent to participants via email."</p>
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	Outcomes are described in the measures section (burnout, depression symptoms, subjective happiness, work-life balance). Exposures, pg. 10, "we sent daily email reminders to enter "What went well today, and what was your role in making it happen?" in three text boxes."
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment	Sources of data and details of methods assessment are

		(measurement). Describe comparability of assessment methods if there is more than one group	outlined in the design and measures sections, pgs. 7-10.
Bias	9	Describe any efforts to address potential sources of bias	We attempted to address bias by informing participants that study responses were confidential. By including a variety of ages (18+) and healthcare worker roles, we reduced the likelihood that any effects were biased by sample characteristics.
Study size	10	Explain how the study size was arrived at	This pilot study allowed for open enrollment to a 3GT cohort. Due to concerns about attrition, a sample larger than 200 was sought. A power analysis indicated that our sample size of 228 provided high power $(1-\beta) > .9$ to detect relatively small effect sizes (Cohen's $d = .3$ ).
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	We used four quantitative survey measures of wellbeing. At each assessment participants completed survey measures, which were scored and analysed.
Statistical methods	12	(a) Describe all statistical methods, including those used to control for confounding	This non-randomized repeated measures cohort study employed paired t-tests to examine changes in survey outcomes between baseline assessments and follow-up assessments.
		(b) Describe any methods used to examine subgroups and interactions	Exploratory subgroup analyses tested for changes in wellbeing

	across the study for those who had particularly concerning wellbeing scores at the baseline assessment. This group was identified by scored above thresholds on the scores identified by previous research.
(c) Explain how missing data were addressed	Our paired t-tests used list-wise deletion. Pg. 11
(d) Cohort study—If applicable, explain how loss to follow-up was addressed  Case-control study—If applicable, explain how matching of cases and controls was addressed  Cross-sectional study—If applicable, describe analytical methods taking account of sampling strategy	All enrolled participants received links to complete all follow-up measures. Participants who were lost to follow up were still included in any analysis for which they had provided data. Pg. 11
(e) Describe any sensitivity analyses	No sensitivity analyses were conducted.

Continued on next page

Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	Numbers of participants at each assessment time point can be found in table 1., pg. 13. Given the extremely broad eligibility characteristics, the true number of eligible participants is unknown.
		(b) Give reasons for non-participation at each stage	The main reason given for non-participation include 'not having time to complete surveys'.
		(c) Consider use of a flow diagram	Since this a non-randomized cohort study, and participant numbers are available in Table 1, we did not feel that a flow diagram would be useful.
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	Demographic information is outlined in Table 1, and described on pg. 11. Exposures are described in the measures section, "3 Good Things Intervention"
		(b) Indicate number of participants with missing data for each variable of interest	Sample sizes of data available for each measure can be found in Table 3, pg. 15. All available data were utilized.
		(c) Cohort study—Summarise follow-up time (eg, average and total amount)	This can be found on page 7.
Outcome data	15*	Cohort study—Report numbers of outcome events or summary measures over time	Numbers of outcome events can be found in Table 3, pg. 15. Results are described
		Case-control study—Report numbers in each exposure category, or summary measures of exposure	
		Cross-sectional study—Report numbers of outcome events or summary measures	
Main results	16	(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	Unadjusted variable scores are displayed in Table 3, pg. 15. 95% CIs are provided. Given the nature of the study we did

			not adjust for confounders.
		(b) Report category boundaries when continuous variables were categorized	All measures were used according to the scoring and use recommendations by the scales' authors. All measures of interest produced continuous outcomes which were used in simple paired t-tests.
		(c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	N/a, this is not a risk study
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	We conducted exploratory analyses on participants whose scores were concerning at baseline on our wellbeing metrics. The results of these analyses can be found in Table 3 and are briefly described in the results section, pg. 14.
Discussion			
Key results	18	Summarise key results with reference to study objectives	Achieved on pg. 17
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	Limitations are listed in detail on pgs. 21-22.
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	We carefully describe and interpret the results of the study in the discussion section, particularly since this was a non-randomized pilot study. We were mindful not to overstate the impact of these findings, and rather present them as a promising first step in evaluating 3GT with healthcare workers.
Generalisability	21	Discuss the generalisability (external validity) of the study results	Generalisability is conveyed in the end of the discussion, and the conclusion section.
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	This information is included in the "funding" section, pg. 24.

\*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at <http://www.plosmedicine.org/>, Annals of Internal Medicine at <http://www.annals.org/>, and Epidemiology at <http://www.epidem.com/>). Information on the STROBE Initiative is available at [www.strobe-statement.org](http://www.strobe-statement.org).

## VERSION 2 – REVIEW

<b>REVIEWER</b>	Jesus Montero-Marin Primary Care Prevention and Health Promotion Research Network (RedIAPP), Zaragoza, Spain
<b>REVIEW RETURNED</b>	17-Oct-2018

<b>GENERAL COMMENTS</b>	<p>The authors have improved substantially the manuscript. However, there are five points more that, in my opinion, would enrich the draft:</p> <ul style="list-style-type: none"> <li>-Please, could you introduce a flow diagram of participants along the study?</li> <li>-Although less robust than exhaustion, ¿why did not you used the other two dimensions of burnout? If you have measures of them, they should be included to gain a greater overview.</li> <li>-You have enough statistical power to develop multivariate analyses. Why did not you use them, to control for some sociodemographic or baseline variables?</li> <li>-In my opinion, the qualitative part of the study (3GT submissions, and comments of experiences) might be enriched by using a content analyses, pointing the topics emerged and even the percent of each of them, making a mixed-analysis.</li> <li>-Finally, I think that the relationships of baseline levels of some outcomes and drop-outs at the first time point should be treated in more detail because it is an important limitation of the study (those who need more the intervention were lost at the first follow up). Maybe some suggestions on how to treat this problem should be added to the discussion section.</li> </ul> <p>In general, I would like to congratulate authors because the proposed program seems to have great beneficials with very low costs.</p>
-------------------------	---

<b>REVIEWER</b>	Anthony Montgomery University of Macedonia Thessaloniki Greece
<b>REVIEW RETURNED</b>	29-Oct-2018

<b>GENERAL COMMENTS</b>	<p>The authors have done a comprehensive job in addressing the concerns of the reviewers.</p> <p>I have one minor (pedantic) revision.</p> <p>In the text the authors state; 'A meta-analysis has revealed that of the three sub-scales (emotional exhaustion, depersonalization, and personal accomplishment), emotional exhaustion consistently produces the largest and most consistent coefficient alpha estimates, while depersonalization and personal accomplishment were both lower and less consistent than emotional exhaustion'. However, the cited study (no 31) is not a meta-analysis?</p> <p>I have no problem with them using EE as an outcome, but they should be aware of the 2016 paper by Maslach &amp; Leiter which suggests that cynicism (not exhaustion) comes closer to the negative endpoint of burnout.</p> <p>Reference Maslach C, Leiter MP. Understanding the burnout experience: recent research and its implications for psychiatry. World Psych 2016; 15: 103-111.</p>
-------------------------	---

<b>REVIEWER</b>	Michael K Howlett Dalhousie University Department of Emergency Medicine Saint John Regional Hospital Saint John NB Canada
<b>REVIEW RETURNED</b>	24-Oct-2018

<b>GENERAL COMMENTS</b>	<p>I think the authors have done an admirable job of very carefully addressing all of the issues. I think it is now time to put it out to the wider readership for their consideration. I would find this article useful in some of my own work since several of its issues resonate with a lot that is happening in Emergency Medicine these days. My only question is regarding length, i.e., word count. I do see some opportunities for making it more concise. I admit some of the length appears to be due to responses to reviewer comments. Having seen all their explanations, i'd be satisfied with some of it edited down. I leave that up to the decision editor and authors to discuss.</p>
-------------------------	--

### VERSION 3 – REVIEW

<b>REVIEWER</b>	Jesus Montero-Marin University of Zaragoza
<b>REVIEW RETURNED</b>	30-Dec-2018

<b>GENERAL COMMENTS</b>	In my opinion the latest version of the manuscript is publishable
-------------------------	---