PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (see an example) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below. Some articles will have been accepted based in part or entirely on reviews undertaken for other BMJ Group journals. These will be reproduced where possible.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Behavioral activation vs. Mindfulness-based guided self-help treatment administered through a smartphone application: a randomized controlled trial
AUTHORS	Ly, Kien Hoa; Trüschel, Anna; Jarl, Linnea; Magnusson, Susanna; Windahl, Tove; Johansson, Robert; Carlbring, Per; Andersson, Gerhard

VERSION 1 - REVIEW

REVIEWER	Prof. Dr. Heleen Riper VU University Amsterdam, Netherlands; Leuphana University, Lueneburg, Germany I declare no competing interests
REVIEW RETURNED	22-Jul-2013

THE STUDY	The abstract lacks the results of the non signifcant difference between the two groups on the primary outcome. Next, there are some language and grammar mistakes which can be checked by a native speaker.
RESULTS & CONCLUSIONS	The scope of the paper is clear. The main research question would benefit from a more detailed description. In the abstract the conclusion reads that ' indicate that this smartphone format works well for a depressed population'. I don't think this question has been answered in this study as both conditions made use of a smartphone and there was not a non-intervention control group. I have attached the pdf with my remarks. This may show my identity to the authors. If you want to have this feedback in another way, please let me know.
GENERAL COMMENTS	have read this paper with great pleasure. It is indeed one of the first papers to assess the effectiveness of two interventions that are delivered by smartphone. The scope of the paper is clear. The main research question would benefit from a more detailed description. In the abstract the conclusion reads that ' indicate that this smartphone format works well for a depressed population'. I don't think this question has been answered in this study as both conditions made use of a smartphone and there was not a non-intervention control group. It is unclear whether smartphone use was an explicit inclusion criterion. The main question was whether behavioral activation was more effective then MF delivered over the smartphone. This should be clear throughout the paper. Next the issue of power needs to be addressed more carefully.

	Next, there are some inacuracies in the text. I have marked these in yellow n the text. The authors could easily solve these (see also tables, figures and references).

REVIEWER	Dr. Bernet Kato Biostatistician National Heart and Lung Institute Imperial College London United Kingdom
REVIEW RETURNED	24-Jul-2013

GENERAL COMMENTS	This paper reports the results of a randomized trial on self- help treatment delivered through a smartphone application. The results are therefore of considerable interest for therapists working with people who have major depressive disorder. Using smartphone would be very handy as it would enable the therapist to reach their patients outside of the therapy room or when not sitting behind a computer. Minor comments Page 3 under conclusions: what is meant by "within group effects"? Page 10, first paragraph of results: what does the phrase "treatment credibility" mean? Page 11 under Credibility and therapist time: what is meant by a "C-scale"? Is this some measure that is used by therapists? --- Next I confine my remarks to statistical aspects of this paper. Methods I note that the study does not have a control group (a group that does not receive any treatment). Why was a control group not included? On page 3 it is mentioned that the efficacy of behavioural activation for treating major depressive disorder has been established in a number of studies over the past four decades. Does this imply that for this study the authors want to investigate if behavioural activation is still superior to mindfulness when both are delivered via smartphones? Since the authors are comparing two groups in which several measurements are taken on each participant they should use methods that are suited for analysing longitudinal data in order to answer the research questions e.g mixed effects models. Random effects models would make use of all the available data and would enable one to answer the relevant questions rather than computing Cohen's d for various pairs of outcome variables.

	Outcomes measures It is mentioned that the primary outcome measures were the Beck Depression Inventory-II (BDI-II) and the 9-item Patient Health Questionnaire Depression Scale (PhQ-9) that were administered pre-treatment, at post-treatment and also six months after the treatment had ended. The PhQ-9 was also administered on a weekly basis during the entire treatment phase: *This suggests that each subject has three measurements on the outcome BD1-11 and at least 24 measurements on the outcome PhQ-9. The authors should clarify which measurements on the two outcome variables were used for the analyses.* *Were any analyses conducted on the secondary outcome measures (BAI, QOLI, TICP, and AAQ-II)?* Data analysis Independent t-tests and Chi-squared tests were used to test for group differences in demographics, pre-treatment data and in clinical significant improvement*: what is meant by "clinical significant improvement"?* Differences between the behavioral activation treatment and the mindfulness treatment were primarily investigated by modeling interaction effects of group and time. For the PHQ-9, where weekly measures were available, the continuous outcome variable was analyzed using mixed effects models, given their ability to handle missing data 41: *the authors state that mixed effects models were used for analyses, however they do not present any results (tables) that show the results from the mixed effects models. The results from the mixed models and their interpretations should be presented in the paper.* Random intercept models were selected. *Does this imply that all models used to analyse the data were random intercept models?* Also, several models were compared using availableinformation criteria, and the model with best fit was chosen. *The authors should mention what models were compared and what information criteria were used for the comparison.* Results Primary outcome measure *It is mentioned that "no significant interaction effects of group and time on the PHQ-9 and the BDI-II were found between the groups, …". The authors should show the results that were used to arrive at this conclusion.* Subgroup analyses *The main analyses included 81 participants randomised to two groups namely two smartphone-delivered treatments; one based on behavioural activation (40 participants) and the other based on*

	mindfulness (40 participants). It is not clear what the subgroup analyses are and how they differ from the main analyses – the authors should give more details about the subgroups analyses.

VERSION 1 – AUTHOR RESPONSE

Reviewer 1

Reviewer(s)' Comments to Author:

The abstract lacks the results of the non signifcant difference between the two groups on the primary outcome.
- This has now been added on page 3.

Next, there are some language and grammar mistakes which can be checked by a native speaker.
- We have now gone through the manuscript and corrected some grammatical errors.

The scope of the paper is clear. The main research question would benefit from a more detailed description.
- The main research question has been clarified in the Background section on page 5.

In the abstract the conclusion reads that ' indicate that this smartphone format works well for a depressed population'. I don't think this question has been answered in this study as both conditions made use of a smartphone and there was not a non-intervention control group.
- A new text has been added to the conclusion on page 3.

I have read this paper with great pleasure. It is indeed one of the first papers to assess the effectiveness of two interventions that are delivered by smartphone. The scope of the paper is clear. The main research question would benefit from a more detailed description.
- The main research question has been clarified in the Background section on page 5.

In the abstract the conclusion reads that ' indicate that this smartphone format works well for a depressed population'. I don't think this question has been answered in this study as both conditions made use of a smartphone and there was not a non-intervention control group. It is unclear whether smartphone use was an explicit inclusion criterion. The main question was whether behavioral activation was more effective then MF delivered over the smartphone. This should be clear throughout the paper. Next the issue of power needs to be addressed more carefully.
- The main research question has been clarified in the Background section on page 5, in the Article summary section on page 2, and in the Discussion section on page 12. Moreover, the conclusion that the study “indicates that this smartphone format works well for a depressed population” has either been removed or deemphasized.
We have also commented on the question of power on page 13.

Next, there are some inacuracies in the text. I have marked these in yellow n the text. The authors could easily solve these (see also tables, figures and references).
Thank you for spotting this. We have corrected the text according to the marks.

Reviewer 2

Reviewer(s)' Comments to Author:

Minor comments
Page 3 under conclusions: what is meant by “within group effects”?

- The conclusion has been changed and within-group effects are no longer mentioned in that section.

Page 10, first paragraph of results: what does the phrase “treatment credibility” mean?
- We have now clarified what treatment credibility is on page 7. This new text has replaced the older text. The older text has been moved to the Treatment credibility and therapist time section on page 11.

Page 11 under Credibility and therapist time: what is meant by a “C-scale”? Is this some measure that is used by therapists?
- This has now been clarified.

Methods

I note that the study does not have a control group (a group that does not receive any treatment). Why was a control group not included?
- It is correct that no wait list group was included in the study. Our main research question was to assess whether behavioral activation is more effective than mindfulness delivered over smartphone. Hence, we wanted to isolate all other components, such as the therapist support and the psychoeducation, and only investigate the two smartphone applications. This is now clarified in the Discussion section on page 13.

On page 3 it is mentioned that the efficacy of behavioural activation for treating major depressive disorder has been established in a number of studies over the past four decades. Does this imply that for this study the authors want to investigate if behavioural activation is still superior to mindfulness when both are delivered via smartphones?
- Yes, that is correct. Our aim was to investigate if the smartphone application build on behavioral activation mechanism that was developed for this study, was superior to the smartphone application build on mindfulness. This has now been clarified in the Background section on page 5.

Since the authors are comparing two groups in which several measurements are taken on each participant they should use methods that are suited for analysing longitudinal data in order to answer the research questions e.g mixed effects models. Random effects models would make use of all the available data and would enable one to answer the relevant questions rather than computing Cohen’s d for various pairs of outcome variables.
- This is now corrected and clarified in the Data analysis section on page 9.

Outcomes measures

It is mentioned that the primary outcome measures were the Beck Depression Inventory-II (BDI-II) and the 9-item Patient Health Questionnaire Depression Scale (PhQ-9) that were administered pre-treatment, at post-treatment and also six months after the treatment had ended. The PhQ-9 was also administered on a weekly basis during the entire treatment phase: This suggests that each subject has three measurements on the outcome BD1-11 and at least 24 measurements on the outcome PhQ-9. The authors should clarify which measurements on the two outcome variables were used for the analyses.
- We have clarified how many measurements on the outcome BDI-II and PHQ-9 respectively in the Outcome measures section on page 6.

Were any analyses conducted on the secondary outcome measures (BAI, QOLI, TICP, and AAQ-II)?
- We have added that all outcome variables were analyzed using mixed effects models (including all secondary outcome measures (except from TiC-P)). This has been added in the Data analysis section on page 9.

Data analysis

Independent t-tests and Chi-squared tests were used to test for group differences in demographics, pre-treatment data and in clinical significant improvement: what is meant by “clinical significant improvement”?
- The word “clinical significant improvement” has been changed to “recovery rates” on page 9.

Differences between the behavioral activation treatment and the mindfulness treatment were primarily investigated by modeling interaction effects of group and time. For the PHQ-9, where weekly measures were available, the continuous outcome variable was analyzed using mixed effects models, given their ability to handle missing data 41: the authors state that mixed effects models were used for analyses, however they do not present any results (tables) that show the results from the mixed effects models. The results from the mixed models and their interpretations should be presented in the paper.
- All of the PHQ-9 results are results from the mixed models. For example on page 10, row 29-31 and row 36-38

Random intercept models were selected. Does this imply that all models used to analyse the data were random intercept models?
- Yes, Random intercept models were selected for all measures. This has now been clarified in in the Data analysis section on page 9.

Also, several models were compared using available information criteria, and the model with best fit was chosen. The authors should mention what models were compared and what information criteria were used for the comparison.
- A mistake was made and has now been corrected. This has been made and clarified in the Data analysis section page 9.

Results
Primary outcome measure
It is mentioned that “no significant interaction effects of group and time on the PHQ-9 and the BDI-II were found between the groups, …”. The authors should show the results that were used to arrive at this conclusion.
- We have added the Non-Significant results from pre-treatment to the 6-month follow up on page 11 (in the Primary outcome measure for BDI-II and PHQ-9, and in the Secondary outcome measure for BAI, AAQ-II, QOLI).

Subgroup analyses
The main analyses included 81 participants randomised to two groups namely two smartphone-delivered treatments; one based on behavioural activation (40 participants) and the other based on mindfulness (40 participants). It is not clear what the subgroup analyses are and how they differ from the main analyses – the authors should give more details about the subgroups analyses.
- This is now clarified in the Subgroup analyses section on page 10.

VERSION 2 – REVIEW

REVIEWER	Heleen Riper VU University Amsterdam, Dept. of Clinical Psychology
REVIEW RETURNED	30-Sep-2013

GENERAL COMMENTS	1) The paper needs to be worked out with more preciseness, all the ingredients are there but they miss a certain level of preciseness and detail (both in terms of building up the arguments, hypothesis and presenting of results and transition from one section to the other and references). 2) The conclusions need to be phrased in a much more modest manner given that: a) the study is it first in its kind so results need to be replicated; b) the study was underpowered; c) there was a lack of control group which doesn't rule out that the large within group results are due to natural recovery (although this is not expected given the severity of depression of study participants). 3) The paper could do with some more references to mobile health intervention studies, this would make the paper more interesting (even if these studies are not in the mental health domain or not rct's). 3) The good news is, that the above suggestions can be taken up rather easily as it is a question of fine-tuning the data and arguments.

VERSION 2 – AUTHOR RESPONSE

Reviewer 1

Reviewer's Comments to Author (the detailed feedback in the attached document, point-by-point):

ABSTRACT

• and comparing
- This has now been added.

• in presenting the results I would prefer another sequence corresponding the main question which was of a comparative nature, e.g. the difference between BA and Mindfulness by smartphone (corresponding design) and then go on to mention the within group effect sizes
- We agree with this and have now changed this section according to the suggestion. Also, the within-group effect sizes have been removed from the abstract.

• significance level and CI to be mention
- The sentence has been removed and therefore this is not applicable anymore.

• if PHQ 9 is also a primary outcome, the outcome should be mentioned here
- This has now been added.

• support with statistics

- This has now been added.

• see my previous remark, you should mention here the results of your main research question first
- This has now been clarified. Also, the conclusion about the within-group effect sizes have been removed.

BACKGROUND

• mention more then only BA, it is CBT
- The sentence has been removed.

• briefly explain van BA is
- Two short sentences about BA has now been added: BA is an established psychological treatment derived from learning theory. It is aimed at increasing adequate behaviors and learning about links between behavior and mood.

• link with previous sentence
- This has now been done.

• rct or meta-analysis?
- It has now been clarified that it was a RCT.

• with similar clinical outcome?
- This has now been added.

• for which disorders and eleborate slighlty on these studies and authors
- An additional reference has now been added. Also, a sentence clarifying that smartphone-administered CBT has been conducted for example in the treatment of MDD, has been added.

• negative? please explain
- The following sentence has been added: meaning that more mindfulness practice is associated with lower levels of depression.

• significant? CIs?
- This has now been added.

• other comorbidities?
- The following sentence has been added: as well as for comorbid disorders such as anxiety.

• keep the mobile text to gether, i.e. replace the section that starts with 'an important xxxxxx
- We agree with this and have now moved all sections that are targeting mobile use in CBT treatment to one place (page 5, right before the hypotheses and the aim is clarified).

• see previous remark and transition between sections
- We agree with this and have now moved all sections that are targeting mobile use in CBT treatment to one place (page 5, right before the hypotheses and the aim is clarified).

• ????
- This has now been removed.

• in what kind of direction? and this is not what we see with iCBT (see e.g. meta-analysis of Andrews et al.)
- The sentence has been removed.

• i would like to suggest to rewrite this section as the line of reasoning is not clear enough (the statements are correct though). W
- We agree with this and have now changed the last sentences to the following sentences: For example, Dimidjian et al. (2006) found that among more severely depressed patients, behavioral activation was as effective as antidepressant medication, and significantly outperformed cognitive therapy, whereas for the less severely depressed patients, no differential treatment effects were observed. However, in meta-analyses on BA versus cognitive therapy this has not been found.

• start with your hypothesis (and the section before should introduce this hypothesis thus describe not only BA but also Mindfullness and why you think that BA is better?) and then introduce your design
- We agree with this and have now moved the hypothesis to the beginning of this section.

• ????
- This has now been removed.

• cognitive therapy???
- This has now been removed.

• check nr of measurements and where is your subgroup hypothesis, see your abstract
- The subgroup hypothesis is in the beginning of this section.

METHODS

• is this one of the co-authors? mention
- This has now been added in the Authors' contributions section.

• please explain
- This has now been removed.

• reasons for exclusion after interest?
- This has now been clarified with the following sentences: The most common reason for exclusion was an ongoing psychological treatment. Other reasons for exclusion were wrong type of phone and score under 5 on the PHQ-9. 13 individuals were excluded after the diagnostic interview with the most common reason that the participants were judged to be in need of another kind of treatment.

• n = 81?
- This has now been clarified with the following sentence: Hence, 81 participants were finally included in the data analysis.

• why putting results in the method section?
- This has now been moved to the Results section.

• how, online? by smartphone?
- A new section (Administration format of self-report measures) has been added on page 7 to clarify this.

• why 10 measures in an 8 week treatment for the PHQ-9?
- Since PHQ-9 also included pre-treatment, post-treatment and six months follow up, there were a total of 10 measures. This has now been clarified with the following sentence: including pre-treatment, post-treatment and six months follow up.

• in your introduction and neither in your hypothesis you refer to Cost-Effectiveness???
- We used TIC-P as a measurement to able to evaluate the cost-effectiveness, but this was not included in this article because the data has not been analyzed yet. Therefore it has now been removed.

• how? by phone, face to face or telephone?
- It has now been clarified that it was by telephone.

• for patients or care givers???
- It has now been clarified that the C-scale measures the participants' perceived treatment credibility.

• how and by whom?
- It has now been clarified that the C-scale was administered via an online platform in the Administration format of self-report measures section.

• why this info here and not in the measurement section?
- This has now been moved to the Outcome measures section.

• how minimal?
- It has now been clarified that minimal refers to a maximum time of 20 minutes per participant and week.

• of the psycho-educational web module?
- It has now been clarified that the text refers to the web-based psychoeducation.

• who are these persons and what more content about their work
- It has now been added that text was inspired by the BA treatment manuals of the persons.

• ????
- This has now been explained by the following sentence: meaning that the application was coded in a specific programming language (Objective C).

• give an example
- The following example has now been added: Get ready in the morning.

• did they obtain messages from a real therapist or from an automated therapis?
- It has now been added that the therapists sending the messages are described later in the text, which clarifies that there were real therapists.

• what?
- This has now been removed.

• by smartphone?
- It has now been added that it was via e-mail.

• were

- This has now been changed.

• some content info is necessary here and about Williams
- It has now been added that the text was inspired by the self help book The Mindful Way Through Depression by Williams et al.

• was the amount of guidance similar for both interventions?
- The guidance refers to the mindfulness tracks. Later in the section, the guidance of the communication is explained to be similar (length and type of guided content in the feedback) for both interventions.

• such as xxx
- The following sentence has now been added: such as guided three minutes mindfulness exercise, which was one of the audio tracks in the application.

• why not?
- The mindfulness application did not have that function. This was an application sold on the open market, which means it was impossible for us to add that kind of function to the application.

• what was the content of the therapist in the M intervention?
- It has now been explained better that the content from the therapists in the mindfulness intervention was similar to the BA intervention, which means it was personal encouraging messages, as well as weekly general educational messages.

• and what about the MF training and supervision?
- It has now been added that the experienced psychotherapist has experience from mindfulness in depression treatment as well.

• check equal distribution
- The distribution between the two groups (high and low severity of depression) were not equal since it was not based on a median splitting.

• on tailed or two tailed?
- It has now been added that it refers to one tailed.

RESULTS

• table with measures
- The following sentences have now been added: See Table 3 for all outcome measurements at pre-treatment, post-treatment and at 6-month follow-up. Also, there was no significant difference in demographic characteristics between the groups according to chi-square analysis. See Table 1 for demographical data.

• 6 out of the 9? and dit 3 provide data at least?
- It has now been clarified that it was six out of the total (81) population.

• by imputation?
- In a ITT analysis.

• what is this telephone interview about? the MINI? to ask them again to fill in the quesiionnaire?
- It has now been clarified that it was to conduct the M.I.N.I. interview.

• difference in attrition between groups?
- The following sentences have now been added to make the distribution explicit: with a distribution of four participants from the BA group and five participants from the mindfulness group; and: , with a distribution of 35 participants from the BA group and 34 participants from the mindfulness group.

• significant differences?
- A Chi-2 analysis has now been added: No significant difference in adherence was found between the two groups ($\chi 2(N=81, df=1)=2.35$, $p=1.00$).

• or measures? see text before
- This has now been changed to measures.

• number of people in each of the groups?
- The number of people in each of the subgroups has now been added. High: total n=51, BA n=23, MF n=28; low: total n=30, BA n=17, MF n=13.

• post treatment?
- It has been added that no difference between the groups from pre-treatment to post-treatment was found.
• this is confusing, it is small or medium/moderate and why only from pre-treatment to 6 months and not pre-post treatment?
- This has now been changed to only small. No difference between the groups from pre-treatment to post-treatment was found, which has now been clarified.

• CEA???
- Everything that involved TIC-P have now been removed from this paper.

• n = ????
- It has been added that n=81.

• thus not a real significant difference between the two groups and what about the less depressed group?
- Yes, it is right that it was not a real significant difference, but rather a tendency. A section about the less severe depressed participants has now been added: Among the less severe depressed participants, 82.4 % (n=14) in the BA group recovered after treatment, compared to 92.3 % (n=12) in the mindfulness group ($\chi 2(N=30, df=1)=.63$, $p=.41$). At the 6-month follow-up, the number of participants from the BA group that had recovered remained the same as in the post-measurement (n=14). In the mindfulness group all participants (n=13) from the mindfulness group had recovered at the 6-month follow-up ($\chi 2(N=30, df=1)=2.549$, $p=.17$).

• what does this score mean? good, bad?
- This has now been clarified with the following sentences: participants in both groups rated their respective treatment as credible. Out of a possible total of 50, the average scores were…

• interpretation of the scale?
- Please see above.

DISCUSSION

• see remark about hypothesis

- It has now been added that the aim was to evaluate and compare the two interventions. Also, the hypotheses have now been highlighted with the following sentences: We hypothesized that BA treatment delivered over smartphone would be more effective than mindfulness treatment delivered over smartphone. We also expected that BA would be superior to mindfulness for participants suffering from more severe depression.

• replace
- The following sub-clause has been added to the section: In contrast to the meta-analysis by Cuijpers et al, Dimidjian et al found that BA was comparable in efficacy…

• chronic, reoccurent?
- It has now been clarified that it was recurrent depression.
• chronic, reoccurent?
- Please see above.

• what do you mean by that? and I think you should be very careful with making such a bold statement on the basis of initial results of this study
- This has now been removed.

• repetition
- This has now been removed.

• this comes slightly out of the blue, why do you think that? and if so provide some underpinning
- This has now been removed and replaced with the following sentences: A possible explanation to the results could be that there was a difference between the two treatment groups, although not significant, in the number of participants that were suffering from major depression. In the BA group 64.7% (n=11) were diagnosed with major depression in the initial screening, compared to 30.8 % (n=4) in the mindfulness group ($\chi 2(N=30, df=1)=3.39$, $p=.07$).

• again be careful with such a bold statement and as you did not have a control group these effects may also have been occured due to natural recovery
- The following sentence has now been added to make the statement less bold: . However, a replication with a waiting list group should be conducted to rule out that the effects occurred due to natural recovery.

• i find the rationale for doing so not very strong, as there are no intervention studies available yet, as you state several times, the lack of a control group makes it difficult to make such a strong statement/
- We agree with this and have now changed this to the following sentences: The first is that no wait list group was included. Even if our main research question was to assess whether behavioral activation is more effective than mindfulness delivered over smartphone a control group would have yielded a more clear result.

• this is not presented in the results
- We agree that power is a limitation but refrain from presenting the obtained power as a result, since power is something that should be calculated in advance and we did not expect the mindfulness treatment to be as effective as it was.

• if the study was underpowered then the subgroup analyses are even more underpowered
- We agree with this.

• describe in the method section how this was done?
- Since it was a post-hoc analysis, we are not able to describe a method for power calculation.

• what do you want to say here?
- The following sentence has now been added: meaning that the depression severity in this study was comparable to an outpatient psychiatric population.

• in which setting?
- The following sentence has now been added: i.e. an outpatient psychiatric population.

• what is your argument?
- The argument is that Vessey estimates that 50 % of patients seeking psychotherapy have some college education, and that and that educated patients may be more inclined to seek help for mental health problems in general.

• please explain already in the introduction
- This is explained in the Background on page 5.

• see my remarks above, you need to more cautious with your conclusions
- The following sentence has now been added: However, as mentioned above, a replication with a waiting list group should be conducted to rule out that the effects occurred due to natural recovery.

• see above
- This section has now been changed to the following: These results suggest that different treatments distributed digitally can target different subgroups of depression in terms of severity. However, more studies are needed to strengthen this hypothesis before a conclusion can be drawn.

• major
- This has now been added.

• significant levels?
- This has now been added in the table.

• thus non significant within? see discussion
- This has now been added and clarified in the table.

• non siginficant and large CI
- This has now been added and clarified in the table.

VERSION 3 - REVIEW

REVIEWER	Heleen Riper VU University Amsterdam, Dept. of Clinical Psychology
REVIEW RETURNED	21-Nov-2013

GENERAL COMMENTS	The authors have answered most questions in their reply to the reviewer and the paper has improved considerably. There are however still a number of issues to be addressed in the paper in more detail. I have added my questions/remarks in the attached pdf file. One of the main issues is the rationale for the hypothesis and the subsequently chosen (one tailed-test). Please describe (with references) why it was expected that the BA group would lead to better outcomes compared to Mindfulness training (and therefore a one tailed test was chosen and power calculated). Next, how was recovery measured? The paper would improve with some specific attention to the English used and careful reading/checking. The discussion section could given some more specific attention to other work (both regarding the therapies and results of other smartphone studies on depression). As indicated before, I do know and collaborate with these authors, apologizes for the delay in reviewing, I am incredibly busy these days

VERSION 3 – AUTHOR RESPONSE

Reviewer's Comments to Author (the detailed feedback in the attached document, point-by-point):

• introduce abreviation
- Abbreviation for behavioral activation (BA) has now been added.

• on which is the d = 0.50 based? and ref for cohen's d
- This has now been clarified with the following text:
A meta-analysis by Mazzucchelli, Kane, and Rees detected a significant moderate pooled effect size of Hedges's g=0.33 (Cohen's d=0.31) when comparing BA with other psychological interventions, such as psychoeducation about depression, problem solving, assertiveness training and brief interventions.

Also reference for Cohen's d has now been added.

• see previous statement in your text about these authors, should be consistent with each other, otherwise please explain
- This has now been clarified.

• informed consent?
- It is now clarified that written informed consent was obtained from all participants by surface mail before the study started.

• please explain

- Not owning a smartphone has been added as the reason for exclusion because of wrong type of phone.

• on a pc by internet, including the smartphone data? security of email exchange?
- It has now been clarified that the back-end system was accessible from a website. Also, it has been clarified that all internet (including the therapists' back-end system) and smartphone activities (including the participants' mobile application) were secured, with SSL-encrypted information.

• could this have influenced the results, in a way that MF was almost unguided?
- Even if the therapists could not give specific feedback on activities or exercises that the participants had done, the communication was guided with a lot of suggestions on how to work with mindfulness activities based on the reflection that was sent every week.

• one tailed because ba was expected to far better?
- We have added an explanation why it was expected that the BA intervention would lead to better outcomes compared to mindfulness intervention (and therefore a one tailed test was chosen and power calculated).

• post treatment?
- It is clarified that Cohen's d was done both from pre-measurements to post-measurements, and from pre-measurements to the 6-month follow up data.

• start with table 1
- This has now been changed so table 1 is presented first.

• homework assignments why chosen for reflection?
- This has now been clarified with the following text:
In a study by Andersson et al, the number of postings in a discussion group was used as a process factor. Therefore, we defined adherence to treatment as the number of weekly reflections the participants sent to their therapist.

• natural recovery?
- There is of course a risk for natural recovery since we did not include a wait list group in the study. However, when bench marking the within-group effect sizes with other studies, they are comparable to other depression treatments. Reference for these studies are now added.

• sleepers effect?
- This is a good point and it is possible that the results are due to the sleepers effect. However, since the participants with less severe depression showed the same pattern but in favor for the mindfulness intervention, we believe that this might not have to do with the sleepers effect.

• which groups are these and 0.47 - moderate?
- This has now been changed and clarified.

• .
- Done

• between groups
- This has now been added.

• how measured?
- This has now been clarified with the following sentence:

Recovery rates were defined as no longer fulfilling the criteria for depression according to M.I.N.I.

• check how recovery was measured
- This has now been clarified with the following sentence:
Recovery rates were defined as no longer fulfilling the criteria for depression according to M.I.N.I.

• at post treatment not at 6 months
- This has now been added.

• non significant
- This has now been clarified.

• statistics how measured?
- This has now been clarified with the following sentence:
The therapists reported that the time they spent did not differ between the two treatment groups.

• but what abou the comparison between midfullness?
- This has now been clarified with the following sentences:
Since it is known that depressed individuals in greater extent have concentration difficulties, distractibility and problems in engaging in effortful cognitive processes, Beck and colleagues have long suggested that therapists should focus on behavioral strategies early in treatment when patients are more depressed and return to that emphasis later if patients start to worsen. We expected that the BA intervention would be more suitable for the more severely depressed participants since mindfulness require more cognitive functioning in initial stages, such as the ability to control attention in order to focus on the present moment.

• reference?
- This has now been added.

• or no intervention control group does not have to be only a waitlist
- This has now been clarified.

• power limitations as well
- This has now been added.