PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Using mixed methods to assess fidelity of delivery and its influencing factors in a complex self-management intervention for people with osteoarthritis and low back pain
AUTHORS	Toomey, Elaine; Matthews, James; Hurley, Deirdre

VERSION 1 - REVIEW

REVIEWER	Dr Fabiana Lorencatto City, University of London
REVIEW RETURNED	19-Dec-2016

GENERAL COMMENTS	Thank you for the opportunity to review this interesting manuscript, which aims to describe a mixed methods approach to assessing fidelity of intervention delivery. Fidelity is still a relatively new concept in the complex intervention literature, and it has been recognised that there is a need for further empirical examples of how different dimensions of fidelity have been operationalised, quantified and analysed. Therefore, I feel this study has the potential to make a significant contribution to the fidelity methodological literature. However, at this stage I have a few minor comments/ suggestions of points that could be further clarified in order to strengthen the manuscript: Title: Please clarify that this is a study assessing fidelity of delivery. Fidelity is multidimensional and encompasses more than just fidelity of delivery (i.e. fidelity of receipt, enactment, training etc)- so helpful to clarify the specific type of fidelity being investigated. Abstract: - Objectives/design: as per suggestion for title, please clarify this investigates fidelity of delivery. - Methods and outcomes: 'Quantitatively, fidelity was calculated using percentage, means, and standard deviations'- can the authors please clarify what %, means, and SD were calculated for? Similarly, 'qualitative data were analysed using thematic analysis'- it is unclear what data is being analysed for what purposes (i.e. what were summary statistics summarising, what is it % of intervention components delivered as intended? For qualitative analysis, what was the thematic analysis aiming to identify? Is this the 'reasons for findings' referred to in aims and objectives?). This detail is necessary to facilitate subsequent interpretation of presented results.

	- Results: 'Both qualitative and quantitative data found that physiotherapists' knowledge and previous experience were factors that influenced fidelity' – which data contributed to this from quantitative data? Was this a significant finding? If so, can the accompanying statistics be presented alongside a description of tests conducted in methods? - Results: If possible (I appreciate word count is often limited in abstracts), add example of 'participant' and 'programme' level factor that influenced fidelity. - Article summary- strengths and limitations: although a very robust assessment of fidelity of delivery, this study is not a 'comprehensive investigation of fidelity' in the broadest sense of the concept. Could the first bullet point therefore be rephrased slightly? (e.g. a mixed methods assessment of fidelity of delivery and its influencing factors provides valuable…..' Introduction: - Clear and succinctly written, citing relevant literature to contextualise the present study and the rationale for its importance. - Description of SOLAS intervention on p. 5: Is the intervention manualised? (i.e. is there an intervention manual specifying the components listed in Table 1, which the physiotherapists are expected to adhere to?)- would be helpful to clarify the nature of the intervention in this sense. - Objectives 1 and 2 are well supported rationale wise in the introduction, but some further rationale could be briefly provided in the intro (i.e. what are the types of potential factors that can influence fidelity results- is there any evidence in broader literature of what these may be? i.e. training, years experience etc) - Table 1- intervention components such as goal- setting, action planning, reviewing goals etc can be done in very different ways by different providers, to varying degrees of quality (which I appreciate is not the focus of this paper). Were there more expanded and formal criteria/definitions for what constituted goal setting/ goal review/ action planning etc? Against which actual delivery/performance during the intervention was assessed and compared? (i.e. BCT taxonomy definitions for these techniques?) - Table 1- might be helpful to separate the samples involved into a new table, as slightly confusing when presented alongside intervention content. Methods: - Quantitative phase; sample and procedure: I appreciate that the structure and feasibility of the checklists have been published elsewhere. However, without detailed description of these checklists, the subsequent results presented in this manuscript are difficult to interpret. It would be helpful to know what the response format was- is it a simple present/absent rating? Or present as intended; present but not as intended; absent but should be present; or not applicable (in line with BCC fidelity framework recommendations). Similarly 'checklists were structured into the SOLAS categories as detailed in Table 1 with components chosen based on the intended content'-

	found this slightly confusing- were all the components in Table 1 added to the checklist? Or only a select few? And are there any other intervention components not present in Table 1 that originally constituted the SOLAS intervention but were not chosen as 'intended' or 'essential' components? (as alluded to in the discussion on p. 28 lines 22-25) - Quantitative phase- data analysis: not 100% clear what 'fidelity levels' being calculated are- can the authors please clarify what this % represents? (i.e. is this is the % of manual-specified components delivered as intended?) - Quantitative phase (general point applicable to manuscript as a whole)- there are a range of different terms being used throughout, 'fidelity of delivery' 'treatment delivery' 'treatment fidelity' 'session duration (dose),' 'fidelity of dose' 'adherence' 'fidelity of content' etc – this can get confusing at times, which hinders interpretation. There are also substantial issues in the wider fidelity literature around variable and inconsistent use of terminology. In order to avoid contributing to this variation, it would be helpful if the authors could try to limit and ensure consistency in the terms they are using to describe the type of fidelity being investigated, and to carefully define each new type of fidelity investigated. - Quantitative phase: analysis: lines 48-51 describing rationale for looking at these provider characteristics in relation to fidelity- could be moved to intro to address one of my earlier comments re. setting up rationale for objective 3. - Qualitative phase: more detail on the topic guide and nature of the interviews is required to facilitate interpretation. Not clear what exactly physiotherapists were asked. Were the physiotherapists provided with any feedback on their fidelity data, in order to guide the discussion on the factors that may have influenced their fidelity? If possible, it would be helpful to provide the topic guide as a supplementary file (along with checklist from phase 1). - Integration phase- very nicely and clearly described. Results: - Fidelity of content: would be helpful to present the mean % fidelity of delivery score (range: lowest- highest) for each assessment method in text. Current description on p. 12 lines 12-15 slightly confusing and unclear what referring to. - Fidelity of content: comment in text on what the significant differences were between physiotherapist scores and category scores. It is also unclear what 'category' refers to here. - Fidelity of dose: Is the data on observed intervention dose (session duration) presented anywhere? i.e. what the mean duration was (plus range), how this varied according to therapist/ session? Would be helpful to see actual data. - Group size- as per point for fidelity of dose, is this data presented in a table somewhere? Perhaps as supplementary file? - Qualitative findings are clearly presented, Table 4 in particular is

	very helpful. If possible, would be useful to know the frequency of each theme to get a sense for how prevalent that particular factor was across the sample. - Triangulation: Table 5 is particularly helpful. However, the in text summary of findings could be expanded upon slightly. It would be useful to also include a comment in text on the extent of convergence/ disagreement/ silence between data sources. Discussion: - Overall, the discussion is well written, providing a summary of the main findings, contextualising these in the broader fidelity literature and considering implications for wider research and practice. - However, it would be helpful to also acknowledge and discuss the earlier point I flagged, which is that this study only looks at a single dimension of fidelity- fidelity of delivery, and does not take into account the broader fidelity dimensions of design, training, receipt enactment- some of which could relate to the programme and participant level factors the authors identified through their qualitative work (particularly design for programme level, and receipt/enactment for participant factors). It is important to be mindful and cautious not to oversell this study's methods as a 'comprehensive fidelity assessment'- yes, this is an excellent example of a methodologically robust and comprehensive assessment of fidelity of delivery, but not fidelity in the broadest sense of its conceptualisation. Perhaps this could be considered under limitations? And implications for future research?

REVIEWER	Dawn Carnes University of Applied Sciences and The Arts Western Switzerland, Faculty of Health, Fribourg, Switzerland
REVIEW RETURNED	02-Jan-2017

GENERAL COMMENTS	This is an interesting article in an area of emerging research. The paper has two aims to assess fidelity (2&3) but also methodological approaches (1) (a sensitivity analyses ?). Mixed methods studies are particularly difficult to write, which makes the paper very 'busy' and sometimes difficult to follow. Abstract: The aims are slightly different to those in methods section as is the term concurrent/triangulation and then convergent / triangulation (in the manuscript methods section). In the methods section the sentence about the quantitative analysis is not clear. Consequently, the results are difficult to interpret. The conclusion doesn't tell the reader whether the treatment was delivered as intended or not. Introduction: Belg and Borelli describe the different components of fidelity, these might be useful to explain why the different methodological approaches were used. Methods: The methods section could be simplified, especially the flow chart. It would be useful to know what the criteria were used to establish % agreement (Table 1) for example presence or absence, fully delivered and or partially delivered The authors use the word 'dose' but this implies receipt: should this be exposure?

	Results: Because the aims of the SOLAS programme are not fully detailed it is difficult to interpret the results. For example: what were the a priori goals and or expectations of the SOLAS programme? For example how many groups were planned? How many physiotherapists did you hope to train and to what level? Did you assess their training? What were the criteria used to assess their competencies and or adherence to the programme? How many people did you hope would attend each group? What behaviour change techniques were you expecting to be delivered and how did you 'observe' these? How were the category sub components measured? Page 10 'Data analysis' section explains high, moderate and low fidelity adherence but to what? Discussion and conclusions: I agree with the authors that fidelity is an under-explored area and more needs to be published in this field so that we can learn from each other.

VERSION 1 – AUTHOR RESPONSE

Reviewer 1 comments

Title: Please clarify that this is a study assessing fidelity of delivery. Fidelity is multidimensional and encompasses more than just fidelity of delivery (i.e. fidelity of receipt, enactment, training etc)- so helpful to clarify the specific type of fidelity being investigated.
Response: Title has been amended to clarify fidelity of treatment delivery to 'Using mixed methods to assess fidelity of delivery and its influencing factors in a complex self-management intervention'

Abstract:
- Objectives/design: as per suggestion for title, please clarify this investigates fidelity of delivery.
Response: Clarified as suggested (page 2, lines 28 and 31; page 3, lines 61 and 67

Abstract:
Methods and outcomes: 'Quantitatively, fidelity was calculated using percentage, means, and standard deviations'- can the authors please clarify what %, means, and SD were calculated for? Similarly, 'qualitative data were analysed using thematic analysis'- it is unclear what data is being analysed for what purposes (i.e. what were summary statistics summarising, what is it % of intervention components delivered as intended? For qualitative analysis, what was the thematic analysis aiming to identify? Is this the 'reasons for findings' referred to in aims and objectives?). This detail is necessary to facilitate subsequent interpretation of presented results.
Response: Quantitative data has been clarified – i.e. percentage means and SD has been clarified within text as 'of intervention components delivered' (page 2, lines 39 and 43). Purpose of qualitative data analysis has been clarified as 'to explore potential reasons for fidelity scores', and has been clarified as data from the interviews (page 2, lines 44-45).

Abstract:
- Results: 'Both qualitative and quantitative data found that physiotherapists' knowledge and previous experience were factors that influenced fidelity' – which data contributed to this from quantitative data? Was this a significant finding? If so, can the accompanying statistics be presented alongside a description of tests conducted in methods? If possible (I appreciate word count is often limited in abstracts), add example of 'participant' and 'programme' level factor that influenced fidelity.
Response: It has been clarified in methods and results that associations between fidelity scores and physiotherapist variables were conducted using Spearman's correlations (page 2, lines 43-44), and that these variables are the source of the knowledge and experience quantitative data (page 2, line 52). P-values from Spearman's test for knowledge and experience have also been provided (page 2, lines 53-54). Have provided example of participant (e.g. individual needs) and programme level factor

(e.g. resources) (page 2, line 55).

- Article summary- strengths and limitations: although a very robust assessment of fidelity of delivery, this study is not a 'comprehensive investigation of fidelity' in the broadest sense of the concept. Could the first bullet point therefore be rephrased slightly? (e.g. a mixed methods assessment of fidelity of delivery and its influencing factors provides valuable…..'
Response: Amended as suggested to 'This mixed methods investigation of fidelity of delivery and its influencing factors provides valuable information on fidelity assessment methods and factors to be considered in developing and evaluating complex behaviour change interventions' (page 3, line 67)

Introduction:
- Description of SOLAS intervention on p. 5: Is the intervention manualised? (i.e. is there an intervention manual specifying the components listed in Table 1, which the physiotherapists are expected to adhere to?)- would be helpful to clarify the nature of the intervention in this sense.
Response: Intervention is manualised but at present this is not publicly available, however the intervention and its components are summarised in Table 1 and described in depth in the referenced protocol and intervention development papers. This has been better clarified on page 5, lines 137-140.

- Objectives 1 and 2 are well supported rationale wise in the introduction, but some further rationale could be briefly provided in the intro (i.e. what are the types of potential factors that can influence fidelity results- is there any evidence in broader literature of what these may be? i.e. training, years experience etc)
Response: Further rationale and references to broader literature looking at factors influencing fidelity results have been added to the text (page 4, lines 102-104)

- Table 1- intervention components such as goal- setting, action planning, reviewing goals etc can be done in very different ways by different providers, to varying degrees of quality (which I appreciate is not the focus of this paper). Were there more expanded and formal criteria/definitions for what constituted goal setting/ goal review/ action planning etc? Against which actual delivery/performance during the intervention was assessed and compared? (i.e. BCT taxonomy definitions for these techniques?)
Response: Fidelity of delivery of SOLAS BCTs (including more specific detail on goal-setting/action-planning and their definitions) will be presented in another publication by other members of the trial team, and was beyond the scope of this paper. This has been clarified better within this text (p 5-6, lines 143-145).

- Table 1- might be helpful to separate the samples involved into a new table, as slightly confusing when presented alongside intervention content.
Response: Although it would be preferable to put the samples into a new table, unfortunately, the limit of tables allowed for the publication is five, and the authors feel that the remaining tables are integral to the paper. We also feel that the sample information should be presented in the tables within the main paper, and not as a supplementary table, however if it is felt by the reviewers that it would be better to present the samples in a supplementary table, we are happy to do this.

Methods:
- Quantitative phase; sample and procedure: I appreciate that the structure and feasibility of the checklists have been published elsewhere. However, without detailed description of these checklists, the subsequent results presented in this manuscript are difficult to interpret. It would be helpful to know what the response format was- is it a simple present/absent rating? Or present as intended; present but not as intended; absent but should be present; or not applicable (in line with BCC fidelity framework recommendations). Similarly 'checklists were structured into the SOLAS categories as

detailed in Table 1 with components chosen based on the intended content'- found this slightly confusing- were all the components in Table 1 added to the checklist? Or only a select few? And are there any other intervention components not present in Table 1 that originally constituted the SOLAS intervention but were not chosen as 'intended' or 'essential' components? (as alluded to in the discussion on p. 28 lines 22-25)
Response: Further information on the detail of the checklist content and components, and the scoring system (i.e. response format) has been provided within text (page 11, lines 183-188) as follows 'Checklists consisted of approximately 25 components for each session, were structured into according to the SOLAS categories as detailed in Table 1. Components for each session were chosen to address each element specified in the SOLAS intervention manual (summarised in Table 1) 25 ,26 to be delivered during that session. Each component was rated as 'Yes/Present' equating to a score of two points, 'No/Absent' (zero points), or 'Attempted' (one point)'.
Additionally, the checklists themselves have been provided as an additional supplementary file (Supplementary File 1) which we believe should provide enough detail regarding the queried aspects.

- Quantitative phase- data analysis: not 100% clear what 'fidelity levels' being calculated are- can the authors please clarify what this % represents? (i.e. is this is the % of manual-specified components delivered as intended?)
Response: It has been further clarified within text that fidelity levels are the percentage of manual-specified components delivered as intended (page 11, lines 194-195).

- Quantitative phase (general point applicable to manuscript as a whole)- there are a range of different terms being used throughout, 'fidelity of delivery' 'treatment delivery' 'treatment fidelity' 'session duration (dose),' 'fidelity of dose' 'adherence' 'fidelity of content' etc – this can get confusing at times, which hinders interpretation. There are also substantial issues in the wider fidelity literature around variable and inconsistent use of terminology. In order to avoid contributing to this variation, it would be helpful if the authors could try to limit and ensure consistency in the terms they are using to describe the type of fidelity being investigated, and to carefully define each new type of fidelity investigated.
Response: Terms have been edited throughout the manuscript to ensure that 'fidelity of delivery' is used, with reference to treatment fidelity or treatment delivery removed. At the start of the methods, it has been clarified that fidelity of delivery in this study includes both evaluation of fidelity of delivery of session duration, i.e. that providers deliver the session as long as intended (fidelity of duration) and fidelity of delivery of session content, i.e. that providers deliver the session content as intended (fidelity of content) (page 11, lines 173-177).

- Quantitative phase: analysis: lines 48-51 describing rationale for looking at these provider characteristics in relation to fidelity- could be moved to intro to address one of my earlier comments re. setting up rationale for objective 3.
Response: Rationale for looking at these provider characteristics in relation to fidelity- has been moved to introduction section as addressed earlier in the introduction comments (page 4, lines 102-104).

- Qualitative phase: more detail on the topic guide and nature of the interviews is required to facilitate interpretation. Not clear what exactly physiotherapists were asked. Were the physiotherapists provided with any feedback on their fidelity data, in order to guide the discussion on the factors that may have influenced their fidelity? If possible, it would be helpful to provide the topic guide as a supplementary file (along with checklist from phase 1).
Response: Topic guide has been provided as an additional supplementary file as requested (Supplementary File 3). Physiotherapists were not provided with feedback on fidelity as this data was not available at the time of the interviews. This was outlined previously in the study limitations but has been better clarified (page 29, lines 445-449).

Results:
- Fidelity of content: would be helpful to present the mean % fidelity of delivery score (range: lowest-highest) for each assessment method in text. Current description on p. 12 lines 12-15 slightly confusing and unclear what referring to.
Response: Mean % and ranges for each method have been provided in text (page 13, lines 255-258).

- Fidelity of content: comment in text on what the significant differences were between physiotherapist scores and category scores. It is also unclear what 'category' refers to here.
Response: Further clarification regarding significant differences and category has been added to text (page 13, lines 259-262).

- Fidelity of dose: Is the data on observed intervention dose (session duration) presented anywhere? i.e. what the mean duration was (plus range), how this varied according to therapist/ session? Would be helpful to see actual data.
Response: Because of limits on the number of tables included, the data for fidelity of duration was discussed within text – however, an additional table has been added as a supplementary file (5).

- Group size- as per point for fidelity of dose, is this data presented in a table somewhere? Perhaps as supplementary file?
Response: Group size data has been added as supplementary file 6.

- Qualitative findings are clearly presented, Table 4 in particular is very helpful. If possible, would be useful to know the frequency of each theme to get a sense for how prevalent that particular factor was across the sample.
Response: The frequency of the themes discussed (i.e. number of physiotherapists discussing theme) has been provided within text (page 18, line 293, 297; page 19 lines 311, 313, 320, 324; page 20, lines 328, 332, 334).

- Triangulation: Table 5 is particularly helpful. However, the in text summary of findings could be expanded upon slightly. It would be useful to also include a comment in text on the extent of convergence/ disagreement/ silence between data sources.
Response: Further detail on the triangulation findings has been provided within text (page 23, lines 352-356).

Discussion:
- Overall, the discussion is well written, providing a summary of the main findings, contextualising these in the broader fidelity literature and considering implications for wider research and practice.
- However, it would be helpful to also acknowledge and discuss the earlier point I flagged, which is that this study only looks at a single dimension of fidelity- fidelity of delivery, and does not take into account the broader fidelity dimensions of design, training, receipt enactment- some of which could relate to the programme and participant level factors the authors identified through their qualitative work (particularly design for programme level, and receipt/enactment for participant factors). It is important to be mindful and cautious not to oversell this study's methods as a 'comprehensive fidelity assessment'- yes, this is an excellent example of a methodologically robust and comprehensive assessment of fidelity of delivery, but not fidelity in the broadest sense of its conceptualisation. Perhaps this could be considered under limitations? And implications for future research?
Response: The authors thank the reviewer for her comments, and would like to confirm that the fidelity of training and receipt/enactment domains are being addressed in separate publications by the trial team. This has been further clarified in the study limitations section (page 29, lines 455-457)

Reviewer 2 comments
Abstract: The aims are slightly different to those in methods section as is the term

concurrent/triangulation and then convergent / triangulation (in the manuscript methods section). In the methods section the sentence about the quantitative analysis is not clear. Consequently, the results are difficult to interpret. The conclusion doesn't tell the reader whether the treatment was delivered as intended or not.
Response: The aims have been edited to ensure consistency (page 6, lines 147-149), as has the term convergent/triangulation. The quantitative analysis sentence has been edited to clarify further (page 2, lines 42-45). A brief overview of fidelity findings has been added to the conclusion (page 3, line 58).

Introduction: Belg and Borelli describe the different components of fidelity, these might be useful to explain why the different methodological approaches were used.
Response: The authors are aware of the Bellg and Borrelli work having used this work to guide much of the fidelity planning and evaluation as discussed and referenced in the introduction. We are unsure how best to further address this comment in the text as requested.

Methods: The methods section could be simplified, especially the flow chart. It would be useful to know what the criteria were used to establish % agreement (Table 1) for example presence or absence, fully delivered and or partially delivered
Response: The flow chart has been created following mixed methods guidance from Ivankova et al. 2006, and we are therefore unsure how best to further simplify this. Further information on the detail of the checklist content and components, and the scoring system (i.e. response format) has been provided within text (page 11, lines 183-188) as follows 'Checklists consisted of approximately 25 components for each session, were structured into according to the SOLAS categories as detailed in Table 1. Components for each session were chosen to address each element specified in the SOLAS intervention manual (summarised in Table 1) 25 ,26 to be delivered during that session. Each component was rated as 'Yes/Present' equating to a score of two points, 'No/Absent' (zero points), or 'Attempted' (one point)'.
Additionally, the checklists themselves have been provided as an additional supplementary file (Supplementary File 1) which we believe should provide further clarification and simplify the methods section.

The authors use the word 'dose' but this implies receipt: should this be exposure?
Response: Due to lack of consensus in the literature regarding terminology and definitions, the term dose has been changed to duration (e.g. fidelity of duration) to clarify which exact aspect this work is assessing.

Results: Because the aims of the SOLAS programme are not fully detailed it is difficult to interpret the results. For example: what were the a priori goals and or expectations of the SOLAS programme? For example how many groups were planned? How many physiotherapists did you hope to train and to what level? Did you assess their training? What were the criteria used to assess their competencies and or adherence to the programme? How many people did you hope would attend each group? What behaviour change techniques were you expecting to be delivered and how did you 'observe' these? How were the category sub components measured? Page 10 'Data analysis' section explains high, moderate and low fidelity adherence but to what?
Response: More detail regarding the SOLAS aims has been provided within text 'The trial aims to evaluate the feasibility of providing the SOLAS intervention (experimental group) to promote self-management for patients with osteoarthritis (OA) of the hip/ knee and/or chronic low back pain (CLBP) compared to usual physiotherapy, which will serve as the pragmatic control group in order to determine the feasibility of moving to a full scale trial by following the MRC guidelines' (page 5 lines 131-135). The intervention also intended to deliver the intervention to a group of six to eight people, which is clarified on page 5, lines 136.
The training of the physiotherapists (provider training), fidelity of this training and pre-post training competency assessment is being conducted in an additional publication, as is the assessment of

fidelity to behaviour change techniques. This information has been further clarified within the text (page 5-6, lines and page 29 lines 455-457).
More clarification regarding the categories and components measured and what fidelity scores relate to has been provided (page 11, lines 183-188), and additionally the checklists used have been provided as supplementary files to provide further clarification and information.

VERSION 2 – REVIEW

REVIEWER	Fabiana Lorencatto City University of London, United Kingdom
REVIEW RETURNED	21-Mar-2017

GENERAL COMMENTS	Thank you for the opportunity to review this revised manuscript. Thank you to the authors for thoughtfully and thoroughly addressing my previous comments. The manuscript is much improved and I am satisfied that most of my previous concerns and suggestions have been adequately addressed. There is, however, one point that I remain unclear on. This concerns the 'components' of the intervention sessions for which fidelity is being assessed. In their response to my previous comments the authors have clarified that the fidelity of delivery of specific behaviour change techniques (BCTs) will be reported as part of a separate publication. This is also highlighted in the article summary-strengths and limitations final bullet point (line 28-31, p. 37). However, in the methods on p. 45, lines 10-15 the authors now state that fidelity will be assessed in terms of the delivery 'of session content, i.e. providers deliver the session categories and components as intended as summarised in Table 1'. BCTs are frequently conceptualised as the components of interventions, so if these are being reported in a separate paper, it is unclear what 'components' mean in this manuscript/ fidelity analysis. Furthermore, Table 1 now has no reference to the word 'component' so it is unclear which cell in this table represents the intervention 'components' being investigated. Is it whether or not 'materials' 'introduction and review,' 'education,' 'review and planning,' and 'exercise' were delivered? Presumably each of these categories are made up of BCTs (i.e. components)? So to assess whether or not these categories were delivered involves an analysis of BCT delivery too? Could the authors clarify further, please? Alternatively, the authors may wish to consider whether the BCT fidelity of delivery analysis should be folded into this paper rather than reported separately. This would help substantially strengthen this paper.

REVIEWER	Dawn Carnes University of Applied Sciences and The Arts, Western Switzerland.
REVIEW RETURNED	28-Feb-2017

GENERAL COMMENTS	The reviewer completed the checklist but made no further comments.

VERSION 2 – AUTHOR RESPONSE

Reviewer comment:
Thank you for the opportunity to review this revised manuscript. Thank you to the authors for thoughtfully and thoroughly addressing my previous comments. The manuscript is much improved and I am satisfied that most of my previous concerns and suggestions have been adequately addressed.

There is, however, one point that I remain unclear on. This concerns the 'components' of the intervention sessions for which fidelity is being assessed. In their response to my previous comments the authors have clarified that the fidelity of delivery of specific behaviour change techniques (BCTs) will be reported as part of a separate publication. This is also highlighted in the article summary-strengths and limitations final bullet point (line 28-31, p. 37).

However, in the methods on p. 45, lines 10-15 the authors now state that fidelity will be assessed in terms of the delivery 'of session content, i.e. providers deliver the session categories and components as intended as summarised in Table 1'. BCTs are frequently conceptualised as the components of interventions, so if these are being reported in a separate paper, it is unclear what 'components' mean in this manuscript/ fidelity analysis. Furthermore, Table 1 now has no reference to the word 'component' so it is unclear which cell in this table represents the intervention 'components' being investigated. Is it whether or not 'materials' 'introduction and review,' 'education,' 'review and planning,' and 'exercise' were delivered? Presumably each of these categories are made up of BCTs (i.e. components)? So to assess whether or not these categories were delivered involves an analysis of BCT delivery too? Could the authors clarify further, please?

Alternatively, the authors may wish to consider whether the BCT fidelity of delivery analysis should be folded into this paper rather than reported separately. This would help substantially strengthen this paper.

Response:
The authors would like to thank the reviewer for her comments and the opportunity to provide further clarity on this issue. This paper is assessing the fidelity of delivery of intended session content and duration. By this we mean that each session was intended to consist of an education section (to last approximately 45 minutes in duration) and an exercise section (to last approximately 45 minutes). Each education section for each session comprised four categories – Materials, Introduction and Review, Education and Review and Planning. For each category, a number of components (i.e. specific elements or activities) were intended to be delivered or addressed during the session by physiotherapists, which varied according to the session. The Materials category typically consisted of components such as provision of participant handbooks, use of Powerpoint slides etc., materials that were intended to enhance the intervention. Introduction and Review consisted of components such as introducing the aims of the programme or reviewing the previous week's activities. Education consisted of components such as provision of information on pacing or balanced weight, and Review and Planning consisted of components such as reviewing the session content and participant's action plans. Table 1 has been amended to provide examples of components for each category, with all components listed in the provided fidelity checklists (supplementary files) and in the referenced trial protocol paper (Hurley et al 2016a) and intervention development paper (Hurley et al 2016b). Further information has been provided on page 5 and 6 which we hope will provide sufficient clarity.

While there may be some degree of overlap with a small number of the components assessed in this study and the BCT analysis (i.e. in the Review and Planning category), the two papers overall address different aspects of fidelity of delivery and the components covered by this paper enabled a self-report analysis by physiotherapists and direct-observations by researchers. The separate BCT paper (being led by another researcher) is assessing the fidelity of delivery of specific pre-specified intended BCTs, e.g. BCT 8.7 Graded tasks or 15.1 Verbal persuasion, using the exact definitions

specified by the BCT taxonomy from audio-recorded session transcripts, and therefore is too detailed for inclusion within this paper, nor was it feasible for physiotherapists to complete self-report analyses or for researchers to complete direct observations of specific BCT delivery. However, future work aims to look at effectiveness outcomes in relation to the relative fidelity for both of these aspects, which we hope will enable a better understanding of the essential elements of SOLAS (i.e. specific BCTs or specific intervention materials and components (as defined in this paper)). This future plan has also been highlighted on page 30.