PEER REVIEW HISTORY

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

ARTICLE DETAILS

TITLE (PROVISIONAL)	Success and limiting factors in health service innovation: a theory generating mixed-methods evaluation of UK projects
AUTHORS	Leedham-Green, Kathleen; Knight, Alec; Reedy, Gabriel

VERSION 1 – REVIEW

REVIEWER	Dewing , Jan Queen Margaret University Edinburgh
REVIEW RETURNED	23-Dec-2020

GENERAL COMMENTS	Thank you for your submission. It was an interesting read. You give minimal information about the projects (On p8). This may be entirely reasonable. When reading I was wondering about scale of projects - how many services, how many teams, length of time etc Would a summary table help readers? Its usual to have 6 keywords to assist with retrieval from searches - check what the journal requires. Abstract This is generally easy to read and sets out the paper. You would be better to reverse this statement; Exploratory analysis of a mixed methods. What about the method you used to generate the theoretical model. don't leave this out. Churn needs a more understandable term You can introduce churn in the main text. Background - This is concisely presented and flows well. P4 L27 I advise you to make more acknowledgment to the International Science Implementation Field here. L40-42 I suggest you reference this assertion. L44-45 You make an assumption that individual projects are heterogenous. This is not necessarily so. L50 As part of what larger programme? Methods P6 Great to see you have included your philosophical assumptions and identified the values and position of your research and how this then positions the research methodologically. What you haven't done is identify any theoretical framing. This is understandable given you are aiming to develop a theory yourselves. I advise you to change the tile of this section. Ethics I think you need to say clearly and simply that HIN funding and taking part in the project was not conditional to taking part in this research. L55 Please date the research approval P7 L4 the lack of public and patient involvement needs to be raised as a possible limitation of this research. As does the minimal validity

testing of the survey. At some point you should briefly mention what else is needed to improve the survey.

L10 the figure is essential to assist understanding the research design

L43 'non-responders' is labelling Please revise your terminology.

P8

L6 you may as well include the remaining number here

L18 'We' who is we please?

L24 Say more about the emergent framework please.

L27 what do you mean by the final framework?

L29-31 You are making an assumption here that meaning and value are found only in 'did it work?' who decided that the attribute 'did it work?' was to be used and on what evidence?

The second criteria -'did it embed/spread needs clarifying in terms of time and place(s)

And as for the previous one, who decided that the attribute was to be used and on what evidence?

L40 briefly justify to reader why you used this test and why it was a one way parametric analysis.

P9 I think you have connected the quantitative data with the qualitative nicely.

L39 You say it was interesting that no project related factors were critical to success - did you have some expectations here? Be careful not to move into theory testing space.

P15-16

All of the content here is appropriate . However, what is what I would expect is a more crafted discussion in the context of this research. At present the discussion is written more as general background. I really want to know if you do have a theory and what level or type of theory. What else needs doing to add to the credibility of your tentative theory?

REVIEWER	Avby, Gunilla Jönköping University, The Jönköping Academy for Improvement of Health and Welfare
REVIEW RETURNED	10-Feb-2021

GENERAL COMMENTS	The research ambition to explore factors that relate to an innovation success or failure is valuable (p.5). However, I am not convinced this is the aim according to the abstract? The abstract reads: to illuminate the factors that are associated with innovation success in healthcare, while the background also involves the aim of comparing successful projects with less successful projects and offering insights into the factors that differentiate between successful and unsuccessful innovations (also in the title and conclusions). In the abstract, number of participants is misleading, as these are stated as invited and not actual number. As the paper reads now, I find it to be an evaluation of projects financed by HIN. Design and Methods The methods section is undeveloped. I have a lot of questions, and have tried to illuminate a few here. I am not really familiar with the expression mixed methods survey (in the abstract)? Also the use of nested design survey is used. Do you mean mixed methods study? Would you say that you have conducted a scoping review (based on a literature review and interviews with practice representatives)? I do not find the method transparent, and believe

there is a need to include a paragraph that better describes the methodological steps and approach.

More information is needed on how the systematic review was conducted as this is the basis for the survey. Also, who are the five respondents in the scoping interviews? How are they related to the study? And are they the same persons that were interviewed in the follow-up interviews? Are the 176 recipients the same as the number of projects financed of HIN? (some of which could have been run by the same person) Does this then mean that you landed at 56 projects (p. 8 …The innovation projects related to…)? How do you mean that the respondents scored according to a five-point Likert scale (research overview)? I would be good to see how the five interviews were able to identify a new theme – personal factors. As now, I cannot see the relevance and full value of the quite extensive use of quotes.
How was the post hoc analysis done?
Table 1. shading is difficult to see when printed in black & white. Consider patterns or other ways of illustrating results.

A theoretical framing is lacking.

Results and Analysis
The structure of this section is somewhat confusing. For example, I think that the first couple of paragraphs fit better under Methods. Another suggestion is to have a heading such as "Developing a framework of factors (categories) critical to success". Possibly by referring to Figure 3 (or Figure 4)? Maybe you could use some of the wording from the Discussion´s section, as an introductory text in the Results?
Also, I cannot find the six themes that are discussed and illuminated by quotes? The structure seems to better fit with the headings in Figure 2? But at the same time the six themes seem to be your main finding? I suggest that you either adjust the headings or use the same headings as in Figure e in the result´s section. Dependent on what is your main finding.

The development of categories of success is explained to be done by grouping similar projects together. Do you mean "projects with similar feature or factors"? I do not understand how these categories were developed? Were factors that had been rated as "agree" identified as success factors? Or were also factors "somewhat agree" included? If so, what is the basis for this? What does "lower" stand for when comparing innovations (p.8)?

When reading the results and analysis, I again question the value of the interviews, as FI2 is the main quoted respondent? How come a respondent, identified as R61, is included in the data, as only 56 respondents were included in the analysis?
On p. 8 you say that you adopted an exploratory approach to data analysis. Though this may be correct, I understand that the factors used were identified in the literature review?
Please observe that South West London is identified in a quote (p. 11).
I am not clear on the relationship between Figure 3 and 4? If Figure 3 is the main contribution, that is the framework, how does Figure 4 fit the picture?

Discussion

	The reflection on the study´s findings is shallow, the main part presents earlier research and models. As mentioned, a theoretical framing is lacking and most probably needed to enable a critical discussion of the study´s findings. The last couple of sentences in the discussion are important points!

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
Prof. Jan Dewing , Queen Margaret University Edinburgh

Comments to the Author:
Thank you for your submission. It was an interesting read.
You give minimal information about the projects (On p8). This may be entirely reasonable. When reading I was wondering about *scale* of projects - how many services, how many teams, length of time etc Would a summary table help readers?
Thank you for your helpful comments and your interest in the work.
We appreciate your desire to learn more about the projects and, as you note, we were unable to include too much information on the projects such that projects or individuals could be potentially identified. We have therefore enriched the description of the innovations' scope and scale. We have also referenced the range of projects using publicly available information to provide further context in the manuscript.

- "Projects were situated in secondary care (n=19); community care (n=14); academic sector (n=5); mental health sector (n=4); online (n=4); primary care (n=3); and the hospice sector (n=2), with the remainder working at the interfaces between services, or across sectors. Their scope ranged from national programmes at hundreds of sites, local programmes supporting tens of thousands of patients, to small intensive innovations working in new ways with a few dozen complex patients, and their duration ranged from 1-5 years. The innovation areas related to new ways of working in end-of-life care; disability enablement; support for complex or vulnerable patients; discharge support; pain management; patient safety innovations; recovery and rehabilitation; personalised care; chronic conditions; new models of integrated health and social care; health promotion; and novel simulation and workforce development strategies. Projects can be explored at the HIN website [19], however for reasons of confidentiality, we cannot specify which were included in this study."

Its usual to have 6 keywords to assist with retrieval from searches - check what the journal requires.
Thank you for highlighting this. The journal's author guidelines required keywords to be taken from a pre-selected list, however we have asked the editorial team to consider adding the following non-listed words:

- Health services innovation; diffusion of innovation; sustainability of innovation; implementation science.

Abstract This is generally easy to read and sets out the paper.
You would be better to reverse this statement; Exploratory analysis of a mixed methods.
What about the method you used to generate the theoretical model. don't leave this out.
Churn needs a more understandable term You can introduce churn in the main text.
We are grateful for these suggestions, which we believe has improved the manuscript overall.

- We have clarified our mixed methods approach, which has both exploratory methods (answering the question, which factors are significant?) and more explanatory methods (answering the question, why are they significant?).

- We have added an extensive section on model development, and have added a section in our results where the mapping of factors onto our model is made explicit. The factors in the nested hierarchy can now be traced back to the underlying data.
- We have changed 'staff churn' to 'staff turnover', to avoid any misunderstanding about this concept.

Background - This is concisely presented and flows well.
Thank you.

P4 L27 I advise you to make more acknowledgment to the International Science Implementation Field here.
Thank you; we have inserted this into the background section.
- "At the vanguard are *implementation scientists* who work to translate research and innovation into clinical practice, navigating institutional, organisational, structural and cultural complexities to improve services [4]."

L40-42 I suggest you reference this assertion.
Thank you; we have softened this assertion and added a reference to Davidoff.
- The knowledge created through an individual innovation is likely to be complex and context-dependent, providing insights that *may not necessarily* be generalisable [6].

L44-45 You make an assumption that individual projects are heterogenous. This is not necessarily so.
Thank you, we have also softened this assertion.
- Meta-analyses are faced with the complexity of synthesising multiple project evaluations that *may* be reported in different ways.

L50 As part of what larger programme?
Agreed, this is confusing.
- We have deleted it from the abstract and clarified it in the main text.

Methods
P6
Great to see you have included your philosophical assumptions and identified the values and position of your research and how this then positions the research methodologically. What you haven't done is identify any theoretical framing. This is understandable given you are aiming to develop a theory yourselves. I advise you to change the tile of this section.
Thank you for this suggestion.
- We have changed the title of this section from "theoretical framing" to "methodological orientation and theory"

Ethics
I think you need to say clearly and simply that HIN funding and taking part in the project was not conditional to taking part in this research.
L55 Please date the research approval
Thank you, these have both been amended as follows.
- "HIN funding was not conditional on taking part in this research. Ethical approval was granted on 26-Mar-2019 by the Research Ethics Committee of Kings College London (LRS-18/19-10432)."

P7
L4 the lack of public and patient involvement needs to be raised as a possible limitation of this research. As does the minimal validity testing of the survey. At some point you should briefly mention what else is needed to improve the survey.

We have clarified that our primary stakeholders (innovators) were involved in the survey design and in checking back our interpretation. We have recognised the limited patient involvement and the limited validity testing of the survey in the discussion.

- "Stakeholder involvement: The primary stakeholders in this research were health service innovators. These stakeholders were involved in the survey design and in checking back our interpretation."
- "A limitation of this research is its highly contextual nature. Our results may not be generalisable to all contexts, however repeating these methods may produce locally relevant results. The analysis of variance depends on the universe of potential factors having been correctly identified and a large enough number of innovations to produce statistical significance. The research could be improved by more extensive validation of factors, patient and public involvement, further testing the directionality of tentative factors, a greater geographical spread, and a greater number of projects to allow for finer grading across the expansion axis."

L10 the figure is essential to assist understanding the research design
We agree and have tried to address this complexity.

- We have further enhanced the description of the research design, and all methods are now in a single section, rather than spaced sequentially.
- We have made minor changes and clarifications to figure 1

L43 'non-responders' is labelling Please revise your terminology.
Thank you.

- This has been changed to 'participants that had not responded"

P8 L6 you may as well include the remaining number here
Agreed, this has been changed as suggested.

L18 'We' who is we please?
Thank you.

- We have clarified specific members of the research team where relevant throughout.

L24 Say more about the emergent framework please.
We agree that this is an important aspect of the work and we appreciate the suggestion to add more detail.

- We have expanded the discussion to discuss the content and meaning of the framework.

L27 what do you mean by the final framework?
Thank you this was not clear.

- We have clarified the difference between the framework used to categorise projects, and the nested hierarchy of innovation needs which is our final framework throughout.

L29-31 You are making an assumption here that meaning and value are found only in 'did it work?' who decided that the attribute 'did it work?' was to be used and on what evidence? The second criteria -'did it embed/spread needs clarifying in terms of time and place(s) And as for the previous one, who decided that the attribute was to be used and on what evidence?
Thank you – this is important to clarify.

- We have clarified the term "did it work" throughout by changing this to "did it generate value for the intended beneficiaries" i.e. outcomes for patients/carers.
- We have also expanded the section that describes how we categorised projects both in the methods and in the results.

- Methods: "KLG and AK categorised projects into grades of success based on how the respondent self-categorised their project, triangulated against their qualitative survey responses. The categories of success were derived through an iterative process, involving both researchers agreeing a descriptive summary of the status of each project (e.g. scaled down despite achieving better than expected patient outcomes; scaled down because the intervention didn't achieve its aims). We grouped projects with similar project outcomes together, and through a process of constant comparison [14] constructed a categorisation framework that accounted for all the cases in the set."
- Results: "Our emergent framework categorised each project's success across two dimensions: the first relating to whether the innovation was reported as generating more or less than its anticipated value for patients/carers ('value creation axis'), the second according to whether the project became sustained or scaled up beyond the initial pilot, or whether it was scaled down or stopped ('expansion axis'). Innovations that were within the scope and intentions of the original pilot were positioned centrally. We initially scored projects into five categorisations across the expansion axis, as some projects expanded locally and some nationally, however there were not enough projects in each group and statistics became unreliable, so we made a pragmatic decision to adopt relative rather than absolute categories."

L40 briefly justify to reader why you used this test and why it was a one way parametric analysis. We have added the following justification:
- "The Kruskal-Wallis test does not assume a normal distribution in the data and can be used when the data is ordinal e.g. Likert scores. For asymmetric group sizes, the non-parametric Kruskal-Wallis test performs better than the parametric equivalent ANOVA method [16]."

P9 I think you have connected the quantitative data with the qualitative nicely.

Thank you.
- Our reviewer 2 could not see the relevance or value of the quotes, so we have moved quotes relating to success factors into table 2. We have however kept quotes in the limiting factor section, as this section is purely qualitative, and we feel that our interpretation needs to be appraised alongside our data.

L39 You say it was interesting that no project related factors were critical to success - did you have some expectations here? Be careful not to move into theory testing space.
Thank you. All factors were treated as potentially critical to success as they had been identified through the literature review and interviews as such. Our analysis of variance suggested however that they were not the underlying cause of the *relative* success of projects within this set.
- We have taken care to soften our wording and to make it clear that we are hypothesising about associations rather than proving associations throughout.
- We have also discussed this finding in relation of Dobson et al who also found that process and context were more important to success than what the project was trying to achieve.

P15-16
All of the content here is appropriate. However, what is what I would expect is a more crafted discussion in the context of this research. At present the discussion is written more as general background. I really want to know if you do have a theory and what level or type of theory. What else needs doing to add to the credibility of your tentative theory?

Thank you, we have re-written the discussion section.

- We have focused this section more clearly, including what level of theory we think we are creating (using Nielsen's framework we would put this at the level of a meta-analysis or review as so many innovations were included), and how it differs to other theories (which tend to focus on scalability and sustainability, and lack evidence from innovations that do not reach literature, and therefore may not provide insights into a fundamental precedent to expansion: value creation for the intended beneficiaries). This dimension appears to be significantly associated with participatory practices, which is an important validation of patient, public and practitioner involvement in health service innovation.
- We have also added a section on strengths and limitations (which had been moved to five bullet points at the start in line with author guidelines, but we agree these need to be addressed in the discussion).

Reviewer: 2
Dr. Gunilla Avby, Jönköping University, Stockholm University

Comments to the Author:
The research ambition to explore factors that relate to an innovation success or failure is valuable (p.5). However, I am not convinced this is the aim according to the abstract? The abstract reads: to illuminate the factors that are associated with innovation success in healthcare, while the background also involves the aim of comparing successful projects with less successful projects and offering insights into the factors that differentiate between successful and unsuccessful innovations (also in the title and conclusions).
Thank you.
- We have clarified the exploratory nature of the quantitative side of this research (exploring which factors are critical to success) and the explanatory nature of the qualitative side (explaining why they are critical to success).
- We have further refined the abstract and methods accordingly.

In the abstract, number of participants is misleading, as these are stated as invited and not actual number.
Thank you.
- We have removed the number of invited participants from the abstract and stated the number of analysed projects.

As the paper reads now, I find it to be an evaluation of projects financed by HIN.
- We have refined the introduction and abstract, as well as other parts of the manuscript, so as to more fully explain the context, aims, and approach.

Design and Methods
The methods section is undeveloped. I have a lot of questions, and have tried to illuminate a few here. I am not really familiar with the expression mixed methods survey (in the abstract)? Also the use of nested design survey is used. Do you mean mixed methods study? Would you say that you have conducted a scoping review (based on a literature review and interviews with practice representatives)? I do not find the method transparent, and believe there is a need to include a paragraph that better describes the methodological steps and approach.
Thank you for highlighting that further clarification of the methods is needed. We have added the following statement on mixed methods designs to make our methods more transparent, and we have also made refinements to figure 1 to give a methodological overview.
- "We have adopted what Creswell refers to as a sequential mixed-methods design 10. According to Creswell, insight can emerge from exploring first through qualitative methods (in our case a published qualitative review and interviews) the types of factors that might be

important, and then designing questionnaires to explore to their salience to a population (called an 'exploratory sequential design'). Insight can also emerge from collecting survey data initially and then following up with interviews to help explain the survey results in more detail: an 'explanatory sequential design'. Where both qualitative and quantitative data are collected simultaneously, one set of data can be used to triangulate the other (for example, where the meaning of one is unclear), or they can be used in complementary ways to illuminate each other (for example, one determining which factors are important, the other illuminating why that might be). Our research process involves both exploratory and explanatory aspects as well as triangulation and illumination. It is summarised in Figure 1.."

More information is needed on how the systematic review was conducted as this is the basis for the survey. Also, who are the five respondents in the scoping interviews? How are they related to the study? And are they the same persons that were interviewed in the follow-up interviews? Are the 176 recipients the same as the number of projects financed of HIN? (some of which could have been run by the same person) Does this then mean that you landed at 56 projects (p. 8 …The innovation projects related to…)?
Again, thank you.

- We have added more detail on this review in the 'data generation' section (see below).
- More detail on the stakeholders involved in supporting the survey design and the follow-up interviews under the section "stakeholder interviews".
- We have clarified that 176 projects was the total number of projects that were supported by HIN in 2014-18. This is in the section on 'participants and sampling' and in the section on 'survey distribution'.
- The section on 'survey distribution' states what we did where more than one project was run by a single innovator, and where more than one grant was given to a single project.
- We have made it clearer in the 'descriptive summary of results' that we analysed data relating to 56 separate projects.

How do you mean that the respondents scored according to a five-point Likert scale (research overview)? I would be good to see how the five interviews were able to identify a new theme – personal factors.
Thank you.

- We have included the full survey in the appendix and the statements that participants were invited to score in table 1.
- We have clarified how the interviews served to identify a new theme in the section on 'survey design'.
- "The survey design began with the extraction of potential success factors for health service innovation from a recently published qualitative systematic review [12]. This review aimed to identify all the factors and theories associated with sustainability and scale-up (capacity building) of innovations in health services research. KLG validated and expanded these factors through scoping interviews with five experienced health service innovators. The interviews started with an open exploration of what the innovator felt had impacted on the success of their project, followed by discussion on the factors identified through the literature. Personal factors were mentioned by all stakeholders in addition to the factors from the review, suggesting these may be under-reported. An additional theme (personal factors) with related sub-factors was therefore included, based on these interviews. Themes and factors are listed in Figure 2. These were used to create a mixed methods nested design survey [13] using Qualtrics software (full text in supplementary data).

The survey asked respondents to:

- o Categorise and describe their project's current status (no longer running / likely to finish soon / stable at the level of the original pilot / scaled-up beyond the original pilot / too early to say / other)
- o Score statements (listed in table 1) relating to the impacts of each factor on their project's outcomes, grouped into nine themes on a five-point disagree/agree Likert scale.
- o Describe the status of their project and provide qualitative insights into each of the nine themes.

Our five stakeholders helped to improve the clarity, acceptability and usability of the survey questions and instructions."

As now, I cannot see the relevance and full value of the quite extensive use of quotes.
Thank you for this feedback, which we have had to balance with feedback from reviewer 1.
- • We have removed the quotes from the section on success factors and put them in table 1 as these were illuminative rather than inductive evidence.
- • We have kept the quotes in the section on limiting factors, as the purpose of these quotes is provide the evidence for the claims we are making and therefore we feel they are needed.

How was the post hoc analysis done?
Thank you for highlighting this.
- • We have clarified that this is generated automatically by SPSS after a Kruskal Wallis test and consists of a box plot of the results in each category. We have explained this in the section 'Determination of salience of success factors'. We have also aligned our terminology to that used by SPSS and called this a 'secondary analysis'.

Table 1. shading is difficult to see when printed in black & white. Consider patterns or other ways of illustrating results.
Thank you.
- • We have further highlighted the significant results in bold, and collated them in table 2 to enhance the visual identification of significant factors for readers where shading and colour are not accessible.

A theoretical framing is lacking.
We agree with the reviewer about the centrality of theory and its importance in research. This project uses inductive exploratory methods to generate theory, rather than imposing an existing theoretical framework onto a phenomenon. We are working up from our data to create theory. We have made this clearer (including our philosophical assumptions) and cited (Varpio 2020) in the section on 'Methodological orientation and theory'.
- • "Adopting Varpio's terminology on the philosophy of research, we are taking an inductive approach that works towards a theoretical framework, rather than applying a pre-existing theoretical framework to this study [10]."

Results and Analysis
The structure of this section is somewhat confusing. For example, I think that the first couple of paragraphs fit better under Methods. Another suggestion is to have a heading such as "Developing a framework of factors (categories) critical to success". Possibly by referring to Figure 3 (or Figure 4)? Maybe you could use some of the wording from the Discussion´s section, as an introductory text in the Results?
We had attempted to explain the stages in turn, as the results from one section fed into the methods for the next. We appreciate that the multi-stage research design needed to be explained more clearly.
- • In response to similar comments from Reviewer 1, we have moved all our methods together into a single section.

- We have also made further modifications figure 1 that further explains the research.
- We have added sections in the methods on developing the framework for categorisation, and creation of our nested hierarchy of innovation needs.
- We have moved the model from the discussion into the results.

Also, I cannot find the six themes that are discussed and illuminated by quotes? The structure seems to better fit with the headings in Figure 2? But at the same time the six themes seem to be your main finding? I suggest that you either adjust the headings or use the same headings as in Figure e in the result´s section. Dependent on what is your main finding.
Thank you for pointing this out. Our results were presented under their original survey themes, and not the emergent success factor themes, which we have corrected. We had to keep them under the original themes as we needed to discuss negative findings as well as positive ones.
- We have made this clearer in the results section.

The development of categories of success is explained to be done by grouping similar projects together. Do you mean "projects with similar feature or factors"? I do not understand how these categories were developed? Were factors that had been rated as "agree" identified as success factors? Or were also factors "somewhat agree" included? If so, what is the basis for this? What does "lower" stand for when comparing innovations (p.8)?
Thank you.
- We have made our process of categorisation of success clearer, including being explicit about how we graded projects along each axis.
- We have also made it clearer that the survey included categorical questions (not agree disagree likert scores) in relation to whether the project had scaled up or down since its pilot.
Categories relating to the value creation for patients were made from qualitative responses, and this was not difficult as we collected a qualitative description of the status each project and their insights into the nine themes were suffused with talk of their innovation's value for patients in relation to whether it had become scale up or down.
- This inductive process is described more clearly in the methods section 'development of categories of success' and in the results section 'categories of success'.

When reading the results and analysis, I again question the value of the interviews, as FI2 is the main quoted respondent? How come a respondent, identified as R61, is included in the data, as only 56 respondents were included in the analysis?
Thank you. The follow-up interviews were primarily to check back our findings and emergent framework. We have stated that no new themes arose from these interviews. Checking back findings with stakeholders is a standard part of qualitative research and included in the SRQR reporting guidelines. For most themes, the survey responses were used. Quotes from follow-up interviews were only used where they were clearer or more succinct.
- We have made it clearer that we had to exclude 7 incomplete or duplicate results in the section 'descriptive summary' hence numbering went up to R63.
- We have expanded and clarified the section on the purpose of the 'stakeholder follow-up interviews' in the methods.
- We have attached the SRQR guidelines which include the pages where we have addressed each of its quality criteria.
- "KLG checked back our results and interpretation with five stakeholders identified by HIN as experienced innovators, one of whom was also involved in the original scoping interviews. Interviews lasted 30 - 45 minutes and transcription was facilitated by automated software (otter.ai). These stakeholders helped to refine the model and confirmed its the applicability and utility in their context. No new themes arose however quotes were used to enrich our survey data."

On p. 8 you say that you adopted an exploratory approach to data analysis. Though this may be correct, I understand that the factors used were identified in the literature review?
Thank you, yes, qualitative review followed by quantitative survey would be an exploratory approach.

- We have made it clearer that this research involves both an exploratory and explanatory aspects. The whole research process is made clearer in figure 1 and we have included an explanation of the terminology of sequential mixed methods research in the section on 'data generation'.

Please observe that South West London is identified in a quote (p. 11).
Thank you.

- We have removed this geographical identifier.

I am not clear on the relationship between Figure 3 and 4? If Figure 3 is the main contribution, that is the framework, how does Figure 4 fit the picture?
Thank you for identifying the potential lack of clarity some readers may have experienced.

- We have revised much of the text to make the process more explicit, and we believe this has helped to improve the manuscript.
- Our intention is that Figure 3 (how we categorised success) illustrates a step on the way to our final framework (now represented in figure 5).

Discussion
The reflection on the study´s findings is shallow, the main part presents earlier research and models. As mentioned, a theoretical framing is lacking and most probably needed to enable a critical discussion of the study´s findings. The last couple of sentences in the discussion are important points!
Thank you for highlighting what we agree are important aspects of our findings. We are not asking or testing a research question that is embedded in a pre-existing theory, but we argue that this work is purposefully generating theory.

- We have expanded and focused our discussion to include more on the meaning of the resulting model.
- We have added a comment on this, and included a reference to Varpio's work on the philosophy of research.
- We have emphasised throughout that we are theorising and not proving.

Reviewer: 1
Competing interests of Reviewer: None to declare

Reviewer: 2
Competing interests of Reviewer: None declared