PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Exploring how ward staff engage with the implementation of a patient safety intervention: A UK based qualitative process evaluation
AUTHORS	Sheard, Laura; Marsh, Claire; O'Hara, Jane; Armitage, Gerry; Wright, John; Lawton, Rebecca

VERSION 1 - REVIEW

REVIEWER	Birgit Heckemann University of Maastricht, the Netherlands
REVIEW RETURNED	13-Nov-2016

GENERAL COMMENTS	**Comments to authors** Thank you for giving me the opportunity to review this paper reporting the process evaluation of the PRASE trial. The authors focus on how PRASE was implemented across the 17 participating intervention wards. **Overall impression** While this paper is interesting to read, particularly due to the novel methods the authors used for their data analysis and presentation, I have major concerns about the study's lack of theoretical grounding and scientific rigour. The discussion of results and the conclusions appear to be somewhat superficial. Furthermore I consider the paper poorly balanced. The description of the process evaluation, data collection and analysis is lacking and the results are not discussed in depth, while too much emphasis is given to the description of the actual intervention (which has been reported elsewhere). The rigour in reporting could be enhanced (by adhering to COREQ reporting criteria). **Detailed comments** • Title The title does not clearly reflect the topic under study: while the title implies that this study is about organizational context, the results presented are about ward/staff engagement and how this is

	influenced (or not) by facilitating processes and providing support at the level of healthcare trusts. Organizational context is a much more complex construct. It may be helpful for the authors to draw on the CFIR resource (Consolidated Framework for Implementation Research, http://www.cfirguide.org/constructs.html) to better define what they are actually looking at in their study. • Abstract The objectives of this study are not clearly stated, there is too much focus on the PRASE intervention rather than on the objectives of this process evaluation. Design: Lacks clarity Setting and participants: ok Data: the authors merely mention their data sources, not how these were analysed (pen portraits, adaptive theory). Findings: The authors present results about staff engagement and some influencing factors, but leave out further dimensions that the organizational context includes. Conclusion: These appear to be findings rather than conclusions. I was expecting some recommendations for process evaluation in general / to inform similar projects. Strengths and limitation are not compelling. I am not sure whether the novel approach taken (point 3) is a strength or a limitation. The same is true for point 4 which is speculation and point 5, engagement as a key factor: Engagement has been reported as a key factor in literature on implementation (see above, CFIR). • Background The background section lacks focus: line 4 to 16 are concerned with the PRASE trial and the importance of patient feedback. Instead, the authors could have focussed on introducing and discussing process evaluation in more depth. The manuscript could be enhanced by a brief mention of an overall guiding framework of the PRASE project, such as the MRC framework or a similar theoretical framework, to demonstrate how the process evaluation was planned from the outset and embedded within the overall trial. Furthermore, the researchers' underlying assumptions with regard to assumed specific barriers and facilitators to successful implementation are not discussed (these could have been explored systematically in the preparation of the trial by drawing on literature on implementation research). The background section therefore does not lead to specific, focussed research questions, but to the rather general question about how and why the intervention works with a focus on staff engagement.

	• Methods The methods section leaves me with many questions due to the poor description of how data were generated. For example, the authors state that they conducted 'short structured phone interviews' (p. 5, l. 32). I would have appreciated a short description of how these phone calls were structured. Was there a guide, maybe? And if so, how was this developed? Furthermore the description of how the data were analysed is scant. It is interesting and novel that the authors used a relatively new method, pen portraits, to synthesise their data. However, the authors' assertion that there is a lack of methodological literature pertaining to the construction of said pen portraits (compare p. 5 l. 50-52) does not justify the lack of rigour in the description of how the pen portraits were created in this particular study. The authors speak of a 'basic structure' that was created to write a linear account of how each ward engaged in the intervention. I do not get a clear idea on how data were synthesised from this description. General rules ensuring rigour in qualitative research can also be applied to novel methods (e.g. creating a guide for the creation of the pen portraits, creating an audit trail, member checking, etc.). The authors explain 'that they used techniques derived from adaptive theory' (p.6 l. 14/15) citing Layder 1998 [18] to further analyse the pen portraits. However, Layder (1998) states in his book: 'The generation of adaptive theory operates at each and every moment of the research from the preparation and planning of data collection (including choice of methods and techniques, problems of access and so on), through every phase of the actual collection and analysis of the data' (Layder, 1998, p. 174.) It is not clear how the authors ensured this kind of continuity throughout the process evaluation. The manuscript would be enhanced if the authors established a link between data sources, pen portraits and adaptive theory. • Findings The authors aim to describe 'context, circumstance and divergence' in the findings section. However, they actually present findings about staff engagement along with some influencing factors, such as trust level support and facilitative processes. While the findings regarding staff engagement are interesting, the number of influencing factors examined is limited and based on authors' assumptions, which are mentioned only in this section (e.g. p. 7, l. 37- 39, l. 42-43). These assumptions are essential and should be presented much earlier in the paper, for example in the background section. It would also be interesting to learn why these particular assumptions were made and considered most relevant with a view to existing implementation research (see CFIR, as noted above).

	• Discussion The authors discuss their findings in relation to patient safety. They state that 'the relationships between different parts and levels of the organization from senior management to ward teams to individuals were vital in achieving success. '(p. 10, l 43-45). This statement appears trivial, because it would be true for any complex intervention. Furthermore, the discussion lacks in-depth exploration about how interventions regarding patient safety differ from other complex interventions in clinical practice. • Limitations The first paragraph on the limitations section seems to be speculation. I do not really understand what the authors are implying here. The second limitation statement concerning the pen portrait methodology is plausible, but I think there is room for improvement in the rigour of reporting. • Conclusion While I understand what the findings mean in terms of the PRASE trial, the conclusion could be enhanced by an explanation of how the authors' findings could inform similar projects. • References Could be enhanced by including literature on implementation science. • Appendices Appendices 1,2 . These are difficult to read and should be, in my opinion, more concise. Appendix 3: This is a very nice illustration of the findings.

REVIEWER	Jamie Murdoch University of East Anglia, UK
REVIEW RETURNED	14-Nov-2016

GENERAL COMMENTS	I think this paper makes a good contribution to the process evaluation literature and is helpful for those planning similar initiatives to improve patient safety. There are clear explanations for the effects observed in the main study and the pen portrait method makes an interesting addition to existing PE methods. As it stands I

	think this paper warrants publication but I had thoughts about the study that might be useful for a more theoretical paper if one is planned. These are: 1. How do the authors define context and consequentially, how would they operationalise this definition within this process evaluation? This is important for the choice of method but also for how features of context are investigated. 2. Related to this point I think the pen portrait approach is interesting and helpful for examining engagement trajectories. However, as it stands, the method portrays a series of snapshot events. I wonder how this method could be developed so that explicit connections can be made across events so that one event might be empirically examined as consequential for the next. 3. The findings provide a descriptive account of how different wards did or did not engage with the PRASE intervention. However, I wonder if the authors are able to theorise the circumstances under which PRASE is likely to be successful? In other words, what are the particular contextual conditions, (from a broad macro to meso to micro level) that are likely to lead to engagement with PRASE which can then be transferred to other contexts? 4. What PE work could have been done prior to trial implementation? It seems that specifying some of the contextual conditions (e.g. staff mobility) prior to implementation helps understanding of the reasons for the subsequent engagement typologies, but could also assist the trial team by identifying wards that need additional support prior to trial delivery.

REVIEWER	Einar Hovlid University of Bergen, Norway
REVIEW RETURNED	15-Dec-2016

GENERAL COMMENTS	Thank you for the opportunity to review your interesting manuscript. I have some comments. Page 5, line 6. Your main research question is “where does the intervention work, how and why”. If I have understood your manuscript correctly the intervention did not have a significant effect on outcomes. Given the fact that the intervention did not work, your research question seems strange. What I as a reader and researcher would be interested to know more about is why the intervention did not have the intended effect. Was it because it was not implemented or did the intervention not work as intended? It can be summarized in the following questions: To what extend were the different key elements of the intervention implemented? (If not, why was it not implemented) Did the elements of the intervention have the intended effect when they were implemented? (If not, why did it not work) I suggest that you consider re-phrasing your research question. Methods P5 line 15.

	Did you transcribe the taped discussions between ward staff? Did you have any predefined guide or structure for what you should observe and document in the field notes? What kind of "context knowledge" do the observers have? Are they familiar with clinical work? Did you transcribe the telephone interviews? Did you develop an interview guide for the telephone interviews, how was it developed, did you use theory? P 5 line 54 We created a basic structure for the pen portraits…. How did you develop this basic structure? Did you use existing theory, models, frameworks? Example: (Damschroder, Aron et al. 2009, Kaplan, Brady et al. 2010, Kaplan, Provost et al. 2012) What did you include in the portrait and what did you exclude? P6, line 14 Could you please provide more detail on how you did the analysis? What guided the analysis, did you use theory? How did you identify the second and third themes in the findings section? Findings P 6 line 21 "We now set out to understand context, circumstance and divergence in the ways in which the 17 intervention wards engaged with the intervention." I like your 5 typologies, but I suggest that you define or clarify what you mean by "engage". Does it refer to engaging people in implanting the intervention or does it also include to what extent the intervention actually was implemented/put into use. A key question is to what extent the intervention was implemented/put into use. In my mind "implementation of the actual intervention" is in not necessarily the same as engaging in the process of implementing the intervention. Example: Damschroder et al. (Damschroder, Aron et al. 2009) define engaging as: Attracting and involving appropriate individuals in the implementation and use of the intervention through a combined strategy of social marketing, education, role modeling, training, and other similar activities. If you use a definition like the one above you should distinguish more clearly between who is involved in the implementation process, and to what extent are key elements of the intervention used, and by whom? When you are more specific in theme 1 of your findings section, I think you can use theme 2 and 3 to support/deepen our understanding of the implantation trajectories described in theme 1. P 7 line 47 As I understand it, your findings are drawn from the pen portraits which are written on a ward level. Where do the findings about the Trust come from? There are parts of your findings section where it seems like you start the discussion of your findings, example p 8 line 16-20. Discussion General comment:

	I would like to see a more thorough discussion that more specifically addresses your main findings. Was the intervention implemented? Did the intervention work? You should also consider relating your findings more specifically to previous research and theory/frameworks. P 10 line 20 Our process evaluation found that context was so varied within the intervention group that this led to a general 'dilution' of intervention implementation. Can you be more specific, how did context vary and what do you mean by "dilution". Was the intervention not implemented, or were only some elements of the intervention implemented, or was it the interventions itself that did not work? I think that you can develop the link between the implementation trajectories and theme 2 and 3 when explaining the dilution effect. P11 line 16 We believe this a simplistic view which does not take into account the wealth of positive benefits which patients and staff gained. Your intervention did not have a significant outcome. I think you should consider developing the argument for why the examples you mention are positive benefits. Conclusion P 12 line 6: Again, what do you mean by a dilution effect of the intervention. Line: 6 and 7. …ward interacting with the intervention in highly divergent manner.. What do you mean by interacting with the intervention. Was it implemented, was part of it implemented, did it not work as intended, why did it not work as intended? Line 8 and 9: How was the facilitative process inadequate to fully "embed" the intervention? What to you mean by embed, is it different from implement? Line 10 and 11: … how ward staff on the ground engaged.. again what do you mean by engage? General comment What I miss in your manuscript is a better and more precise description of the following: To what extend were the different key elements of the intervention implemented/put into use? What factors contribute to explain implementation/lack of implementation? Did the intervention work as intended where it was implemented? Why did it work/not work? Given the answers to the above questions; how do you explain that the intervention had no significant effects on the outcomes? It seems to me like a major problem was that the action plans that

	the wards developed based on feedback from the patients not necessarily were implemented. Maybe you should consider developing this finding more in your manuscript. Examples of relevant references Damschroder, L., D. Aron, R. Keith, S. Kirsh, J. Alexander and J. Lowery (2009). "Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science." Implementation Science 4(1): 50. Kaplan, H. C., P. W. Brady, M. C. Dritz, D. K. Hooper, W. M. Linam, C. M. Froehle and P. Margolis (2010). "The Influence of Context on Quality Improvement Success in Health Care: A Systematic Review of the Literature." The Milbank Quarterly 88(4): 500-560. Kaplan, H. C., L. P. Provost, C. M. Froehle and P. A. Margolis (2012). "The Model for Understanding Success in Quality (MUSIQ): building a theory of context in healthcare quality improvement." BMJ Quality & Safety 21(1): 13-20.

REVIEWER	Tara Lamont Deputy Director NIHR Dissemination Centre University of Southampton Alpha House Venture Road SO16 7NS
REVIEW RETURNED	21-Dec-2016

GENERAL COMMENTS	This is a thorough piece of qualitative research which adds to our understanding of how and why patient safety and service improvement initiatives take hold in service settings. It is clear, well-written and is appropriately grounded in relevant literature. I just have a few observations/comments about how this could be strengthened, but no substantive criticisms. 1) The description of approach to analysing the qualitative data using adaptive theory is rather vague and generalised. This could be strengthened, with some examples of emerging themes and how team members worked to challenge or test emerging framework with the data. 2) The study is rather atheoretical and could explain how our knowledge of how change and improvements spread (or don't) by use of theoretical frameworks like NPT and how this study is positioned in relation to other work in the field of implementation science. 3) The engagement taxonomy is interesting - again, as above (2) would help to place this in relation to other frameworks and categories for judging levels of implementation or embedding. I think the temporal dimension is interesting in terms of engagement or disengagement over time - which was not immediately obvious from the labels of `upward' and `downward' engagement. 4) The discussion section is helpful but I would have liked to see slightly more high-level analysis of what this adds to existing

	knowledge. I thought the findings were very interesting around the lack of fit between trust-level support and `good' ward engagement. It would be good to ground this in related literature on eg high-performing wards and clinical microsystems. Similarly, although some of Mary Dixon-Wood's later work is cited, I think directly relevant is her original work in Michigan looking at how units adopted improvements http://onlinelibrary.wiley.com/doi/10.1111/j.1468-0009.2011.00625.x/abstract. It would be helpful to relate the findings here to this seminal work on patient safety improvement.

VERSION 1 – AUTHOR RESPONSE

Reviewer and comment ID	Comment text	Response from author team
Jamie Murdoch – 1	I think this paper makes a good contribution to the process evaluation literature and is helpful for those planning similar initiatives to improve patient safety. There are clear explanations for the effects observed in the main study and the pen portrait method makes an interesting addition to existing PE methods. As it stands I think this paper warrants publication but I had thoughts about the study that might be useful for a more theoretical paper if one is planned.	Thank you. We have recently published a theoretical paper based on our study in the journal Social Science and Medicine. The reference is: Sheard L, Marsh C, O'Hara J et al (2017) The Patient Feedback Response Framework – Understanding why UK hospital staff find it difficult to make improvements based on patient feedback: A qualitative study http://www.sciencedirect.com/science/article/pii/S0277953617300850 In this paper, we develop a conceptual framework to explore the circumstances under which staff were able and were not able to act on patient feedback to bring about change. We found that, firstly, staff must exhibit 'normative legitimacy' which is the belief that listening to patients is a worthwhile exercise. Secondly, structural legitimacy must be in place – adequate autonomy, ownership and resource to solve problems. Thirdly, organisational readiness to change needs to exist at the level of the hospital if improvement is to occur. Interestingly, some ward teams are able to make improvements when the elements they want to change are within their immediate control and within the boundary of their ward environment. Inter-departmental collaboration and high level assistance is often not forthcoming and therefore prevents structural level changes being implemented.
Jamie Murdoch – 2	How do the authors define context and consequentially, how would they operationalise this definition within this process evaluation? This is important for the choice of method but also for how features of context are investigated.	Given comments from other reviewers, we have moved away from the overt use of the term "context" and have sought to focus the paper more fully on the concept of "engagement". Further detail on how we have operationalised this is given in the response to comment 10.

Jamie Murdoch – 3	Related to this point I think the pen portrait approach is interesting and helpful for examining engagement trajectories. However, as it stands, the method portrays a series of snapshot events. I wonder how this method could be developed so that explicit connections can be made across events so that one event might be empirically examined as consequential for the next.	The pen portrait approach we used is relatively novel and we agree with this reviewer that the method could be developed further in the manner suggested. We may consider this approach for a future methodological paper
Jamie Murdoch – 4	The findings provide a descriptive account of how different wards did or did not engage with the PRASE intervention. However, I wonder if the authors are able to theorise the circumstances under which PRASE is likely to be successful? In other words, what are the particular contextual conditions, (from a broad macro to meso to micro level) that are likely to lead to engagement with PRASE which can then be transferred to other contexts?	In the recently published Social Science and Medicine paper (which we discuss in response to comment 1), we theorise more broadly under what circumstances ward staff are able to successfully respond to patient feedback in general. Whilst our empirical case study is the PRASE intervention, we look wider to demonstrate how our conceptual model could be applied to other forms of patient feedback, be that safety, quality or experience. Incorporated in this is significant comment on the macro, meso and micro levels conditions which enable or hinder improvement work by staff.
Jamie Murdoch – 5	What PE work could have been done prior to trial implementation? It seems that specifying some of the contextual conditions (e.g. staff mobility) prior to implementation helps understanding of the reasons for the subsequent engagement typologies, but could also assist the trial team by identifying wards that need additional support prior to trial delivery.	The research team undertook a feasibility and acceptability study prior to the main RCT. The reference is: O'Hara J, Lawton R, Armitage G et al. (2016). The patient reporting and action for a safe environment (PRASE) intervention: a feasibility study *BMC Health Services Research* 16:676 https://bmchealthservres.biomedcentral.com/articles/10.1186/s12913-016-1919-z The feasibility study was undertaken at six wards in one acute Trust in the North of England during 2012 – 2013. This Trust was not one of the three trusts where the PRASE intervention was tested in a 33 ward cluster RCT. From this early work, we did not sense that staff mobility was a critical factor. Our published logic model from the O'Hara et al paper specifies what we believed to be the key moderating factors going forward into the full RCT. We modified the intervention based on the findings of the feasibility work. We have added a sentence to the second paragraph of the Discussion to make it clear that our earlier feasibility work led to the development of some of the facilitative processes in the RCT.

Einar Hovlid - 6	Page 5, line 6. Your main research question is "where does the intervention work, how and why". If I have understood your manuscript correctly the intervention did not have a significant effect on outcomes. Given the fact that the intervention did not work, your research question seems strange. What I as a reader and researcher would be interested to know more about is why the intervention did not have the intended effect. Was it because it was not implemented or did the intervention not work as intended? It can be summarized in the following questions: To what extend were the different key elements of the intervention implemented? (If not, why was it not implemented) Did the elements of the intervention have the intended effect when they were implemented? (If not, why did it not work) I suggest that you consider re-phrasing your research question.	We agree with this reviewer that the main research question could be considered confusing to the reader given that the intervention did not have a significant effect on outcomes. This was the research question that we specified a priori in our published research protocol in 2014. Therefore, we are reluctant to change our research question post hoc. We have, however, added a sentence to the first paragraph of the Methods which explains that the purpose of the process evaluation was to understand in what ways the intervention did *not* work and how and why this was the case. We understand the interest this reviewer has in discrete elements of intervention implementation. The answers to the questions which this reviewer raises are already covered in our recently published paper: Lawton R, O'Hara J, Sheard L et al (2017) Can patient involvement improve patient safety? A cluster randomised control trial of the Patient Reporting and Action for a Safe Environment (PRASE) intervention, *BMJ Quality & Safety* http://qualitysafety.bmj.com/content/qhc/early/2017/02/03/bmjqs-2016-005570.full.pdf This paper reports the results of the RCT where the PRASE intervention was tested and details findings about the fidelity of the intervention. We do not wish to replicate what the Lawton et al paper describes, in this current paper. We have added a sentence to the fifth paragraph of the Methods to signpost the reader elsewhere if they are interested in implementation fidelity.

Einar Hovlid - 7	Methods. P5 line 15. Did you transcribe the taped discussions between ward staff? Did you have any predefined guide or structure for what you should observe and document in the field notes? What kind of "context knowledge" do the observers have? Are they familiar with clinical work? Did you transcribe the telephone interviews? Did you develop an interview guide for the telephone interviews, how was it developed, did you use theory?	Taped discussions were not transcribed verbatim. LS and CM listened to all voice files and made detailed notes about the content of the conversation (as already described in the second paragraph of the Methods). We structured our notes under the headings: - Issues seen as important where actions were made (and why) - Issues seen as important where no actions were made (and why not) - Issues dismissed and reasons for this - Comments made on PRASE process/ study/ team - Comments made about ward or hospital context We have now added this detail into the paper concerning how we structured our notes. The guide for the facilitators' field notes is already documented in paragraph three of the Methods. We concentrated on: implicit dynamics, environmental factors, facilitator's overall impressions The facilitators are all experienced applied health services researchers with differing disciplinary backgrounds: sociology, psychology and sustainability. All have worked with teams of ward staff previously and have an understanding of clinical work, particularly in relation to patient safety and patient experience. We do not feel it is necessary to include this detail in the manuscript. Telephone interviews were not transcribed verbatim. Instead, researchers listened to the voice file and made detailed notes. There was a structured interview guide developed, which centred on whether action plans had been successfully implemented or not (and why). Five sentences have been added to the fourth paragraph of the Methods section to address this comment.
Einar Hovlid - 8	P 5 line 54 We created a basic structure for the pen portraits…. How did you develop this basic structure? Did you use existing theory, models, frameworks? Example: (Damschroder, Aron et al. 2009, Kaplan, Brady et al. 2010, Kaplan, Provost et al. 2012) What did you include in the portrait and what did you exclude?	We have added three sentences to the end of the fifth paragraph of the Methods to address these set of comments.

Einar Hovlid - 9	P6, line 14 Could you please provide more detail on how you did the analysis? What guided the analysis, did you use theory? How did you identify the second and third themes in the findings section?	We used 'adaptive theory' (Layder, 1998) to guide our analysis. This is already detailed in the manuscript. The second and third themes arose from the analysis of the pen portraits where we looked between and across the engagement trajectories of each ward in order to understand how strong or weak engagement with the intervention related to local implementation. Essentially, the development of the engagement typologies which are reported in the first theme of the findings were a springboard to uncover the content of the findings for two and three. On revisiting the paper, we have noticed that it is not clear how the second and third themes arose and how we arrived at these findings. In the last paragraph of the first findings section, our approach is implied but we agree with this reviewer that it needs to be made more explicit. To address this, we have provided more detail to the last paragraph of the Methods with the addition of seven sentences.

Einar Hovlid - 10	Findings P 6 line 21 "We now set out to understand context, circumstance and divergence in the ways in which the 17 intervention wards engaged with the intervention." I like your 5 typologies, but I suggest that you define or clarify what you mean by "engage". Does it refer to engaging people in implanting the intervention or does it also include to what extent the intervention actually was implemented/put into use. A key question is to what extent the intervention was implemented/put into use. In my mind "implementation of the actual intervention" is in not necessarily the same as engaging in the process of implementing the intervention. Example: Damschroder et al. (Damschroder, Aron et al. 2009) define engaging as: Attracting and involving appropriate individuals in the implementation and use of the intervention through a combined strategy of social marketing, education, role modeling, training, and other similar activities. If you use a definition like the one above you should distinguish more clearly between who is involved in the implementation process, and to what extent are key elements of the intervention used, and by whom?	From this set of comments, we have realised that the whole paper needs to be more clearly directed towards its focus on engagement as distinct from implementation, therefore we have taken the following steps: -We have specified our definition of 'engagement' for the purposes of this paper as: "the 'depth' and 'nature' of ward teams' approaches and attitudes towards the intervention" in the fifth paragraph of the Methods. We took time to settle on this definition, considering other definitions including the work of Damschroder (2009) as proposed by the reviewer. We thank the reviewer for this contribution; however we did not find this or other existing definitions readily applicable. Rather than referring to ward team engagement with an intervention, we felt that Damschroder's paper refers to how implementation teams engage staff as leaders and champions of an intervention. Other definitions of engagement (e.g.Leiter & Bakker 2010 – 'Work Engagement – Handbook of Essential Theory and Research) refer to psychological levels of motivation – this too is different from what we are referring to. We acknowledge that the need for us to define our use of the term 'engagement' was important and we hope we have succeeded in removing any confusion. -We have included a distinction between our definition of 'engagement' (as above) and 'implementation fidelity' which is adherence to the intervention components. This is captured in the Lawton et al (2017) paper. -We have changed the title of the paper so that the engagement theme is clear. The new title is: "Exploring how ward staff engage with the implementation of a patient safety intervention: A qualitative process evaluation" -It has been reinforced that the theme of the paper as being about engagement and not strictly implementation in sentences where this was ambivalent -We have taken out the term "context", which we accept could imply analysis of something distinct from 'engagement' and replaced it with "setting" where appropriate. This is not to remove any implication of additional factors at play, but simply to state that 'engagement' differs by (ward) setting. We have not marked up some of these minor changes (which may only relate to one or a few words) as they were present throughout the document.

Einar Hovlid - 11	P 7 line 47 As I understand it, your findings are drawn from the pen portraits which are written on a ward level. Where do the findings about the Trust come from?	The findings that we describe in our second theme - *Trust-level support for an intervention does not predict the strength of ward-level engagement*' arise predominantly from the pen portraits. However, we have utilised certain tacit knowledge gained throughout the set-up, roll out and implementation of the intervention pertaining to the support shown at the level of the Trust. This tacit knowledge is used briefly to describe Trust A in order to contrast the ethos of this Trust with the engagement styles of the four individual wards involved in the trial. We experienced differing levels of senior management support and enthusiasm at the organisational level and anticipated that corresponding levels of support and enthusiasm in wards in each Trust may be exhibited. The fact that no such Trust patterns of engagement existed in our findings was interesting to us. It also speaks to wider debates that we know others in this area are interested in - about the link between ward teams and senior management. We therefore felt it was a finding worthy of documenting even though we had not carried out specific analysis of Trust support mechanisms. We acknowledge that this could be more directly explored in other studies and have now added a sentence (last line of the third paragraph of the Discussion) to reflect this.
Einar Hovlid - 12	There are parts of your findings section where it seems like you start the discussion of your findings, example p 8 line 16-20.	We do not think this is a problem as the sentence in question is used to bridge the content between a lack of multi-disciplinary involvement and the reasons we identified from the data as to why this may have occurred
Einar Hovlid - 13	I would like to see a more thorough discussion that more specifically addresses your main findings. Was the intervention implemented? Did the intervention work? You should also consider relating your findings more specifically to previous research and theory/frameworks.	The main findings arising from this paper are concerned with staff engagement with the intervention. As previously discussed in our response to comment 6, implementation fidelity is covered in the Lawton et al (2017) paper. The Lawton paper focusses explicitly on whether the intervention worked (there was no significant effect of the intervention on outcomes). We do not want to replicate this work in the current manuscript which is about staff engagement. We now relate our findings to Damscroder's (2009) 'inner setting' construct in the first paragraph of the Discussion.

Einar Hovlid - 14	P 10 line 20 Our process evaluation found that context was so varied within the intervention group that this led to a general 'dilution' of intervention implementation. Can you be more specific, how did context vary and what do you mean by "dilution". Was the intervention not implemented, or were only some elements of the intervention implemented, or was it the interventions itself that did not work? I think that you can develop the link between the implementation trajectories and theme 2 and 3 when explaining the dilution effect.	We have re-written most of the first paragraph of the Discussion and rather than describing variability of context, we now use Damscroder's four domains of the 'inner setting' in order to situate our findings in the published implementation science literature. We have included a definition of 'dilution' in the first paragraph of the Findings section to clarify that by this we mean "'non-standardisation' of the intervention group thereby reducing the potential for this to be meaningfully compared with a control group". This definition, along with our clarification of our definition of 'engagement', and removal of the term 'context' (see response to comment 10 above) should address this comment. Again, we do not wish to replicate the Lawton paper by detailing intervention implementation and fidelity results.
Einar Hovlid - 15	P11 line 16 We believe this a simplistic view which does not take into account the wealth of positive benefits which patients and staff gained. Your intervention did not have a significant outcome. I think you should consider developing the argument for why the examples you mention are positive benefits.	We have slightly enlarged the sixth paragraph of the Discussion to add further detail to the arguments put forward about positive benefits of staff and patients taking part in the study despite a non-significant outcome.

Einar Hovlid - 16	Conclusion P 12 line 6: Again, what do you mean by a dilution effect of the intervention. Line: 6 and 7. …ward interacting with the intervention in highly divergent manner.. What do you mean by interacting with the intervention. Was it implemented, was part of it implemented, did it not work as intended, why did it not work as intended? Line 8 and 9: How was the facilitative process inadequate to fully "embed" the intervention? What to you mean by embed, is it different from implement? Line 10 and 11: … how ward staff on the ground engaged.. again what do you mean by engage?	The term "dilution" is now defined in the paper (see response to comment 14) We have changed the phrase "interacting" to "engaging" or "engaging with" throughout the paper to focus the reader's attention on our explicit emphasis on "engagement". We have removed the phrase "embed" from the paper to avoid confusion. We have now defined what we mean by "engagement" throughout the paper (see response to comment 10).
Einar Hovlid - 17	What I miss in your manuscript is a better and more precise description of the following: To what extend were the different key elements of the intervention implemented/put into use? What factors contribute to explain implementation/lack of implementation? Did the intervention work as intended where it was implemented? Why did it work/not work? Given the answers to the above questions; how do you explain that the intervention had no significant effects on the outcomes? It seems to me like a major problem was that the action plans that the wards developed based on feedback from the patients not necessarily were implemented. Maybe you should consider developing this finding more in your manuscript.	As previously explained, this paper is about staff engagement with the intervention rather than strictly about implementation. Again, we would prefer that readers who are interested in the questions which this reviewer raises consult the Lawton et al (2017) paper detailed in comment 6.

Tara Lamont – 18	This is a thorough piece of qualitative research which adds to our understanding of how and why patient safety and service improvement initiatives take hold in service settings. It is clear, well-written and is appropriately grounded in relevant literature. I just have a few observations/comments about how this could be strengthened, but no substantive criticisms.	Thank you. We are delighted with these comments.
Tara Lamont – 19	The description of approach to analysing the qualitative data using adaptive theory is rather vague and generalised. This could be strengthened, with some examples of emerging themes and how team members worked to challenge or test emerging framework with the data.	We agree that the description of our analysis is too vague. Comment 9 and 31 also attest to this. We have added seven sentences to the last paragraph of the Methods to address this. See our response to comment 9 for a more thorough reply.
Tara Lamont – 20	The study is rather atheoretical and could explain how our knowledge of how change and improvements spread (or don't) by use of theoretical frameworks like NPT and how this study is positioned in relation to other work in the field of implementation science.	We have re-written parts of the Discussion in relation Dixon –Wood's theory and Damschroder's framework. Whilst we agree that NPT is a useful framework to understand how improvement is spread, we feel that to properly incorporate NPT into this paper we would need to re-analyse the findings using NPT as our coding framework and reframe the focus of the paper to be about implementation spread. Our focus is about how staff engaged with the intervention rather than the above so we have chosen not to bring NPT into the paper.
Tara Lamont – 21	The engagement taxonomy is interesting - again, as above (2) would help to place this in relation to other frameworks and categories for judging levels of implementation or embedding. I think the temporal dimension is interesting in terms of engagement or disengagement over time - which was not immediately obvious from the labels of `upward' and `downward' engagement.	Dixon-Woods et al (2011) and Damschroder et al (2009) have now been brought into the Discussion in order to situate the findings of our study. As per previous comments relating to the Lawton et al (2017) paper, implementation fidelity is discussed there and here in this paper we want to focus demonstrably on the concept of engagement. We have changed the labels of upward and downward to 'increasing' and 'decreasing'.

Tara Lamont – 22	The discussion section is helpful but I would have liked to see slightly more high-level analysis of what this adds to existing knowledge. I thought the findings were very interesting around the lack of fit between trust-level support and `good' ward engagement. It would be good to ground this in related literature on eg high-performing wards and clinical microsystems. Similarly, although some of Mary Dixon-Wood's later work is cited, I think directly relevant is her original work in Michigan looking at how units adopted improvements http://onlinelibrary.wiley.com/doi/10.1111/j.1468-0009.2011.00625.x/abstract. It would be helpful to relate the findings here to this seminal work on patient safety improvement.	We situated our Discussion in relation to the patient safety literature in order to understand how the findings of the process evaluation apply more broadly to the current patient safety agenda rather than adopting an overt focus on the improvement science literature. To address this comment, we have added a paragraph to the Discussion in order to discuss our findings in relation to the recommended Dixon-Woods et al (2011) paper. We are grateful for this suggestion as it helps situate more clearly our second key finding regarding the relationship between Trust management and frontline staff with respect to ward engagement.
Birgit Heckemann – 23	While this paper is interesting to read, particularly due to the novel methods the authors used for their data analysis and presentation, I have major concerns about the study's lack of theoretical grounding and scientific rigour. The discussion of results and the conclusions appear to be somewhat superficial. Furthermore I consider the paper poorly balanced. The description of the process evaluation, data collection and analysis is lacking and the results are not discussed in depth, while too much emphasis is given to the description of the actual intervention (which has been reported elsewhere). The rigour in reporting could be enhanced (by adhering to COREQ reporting criteria).	We anticipate that in addressing this reviewer's more detailed comments in relation to specific parts of the manuscript that we are able to alleviate most of the concerns raised.

Birgit Heckemann – 24	The title does not clearly reflect the topic under study: while the title implies that this study is about organizational context, the results presented are about ward/staff engagement and how this is influenced (or not) by facilitating processes and providing support at the level of healthcare trusts. Organizational context is a much more complex construct. It may be helpful for the authors to draw on the CFIR resource (Consolidated Framework for Implementation Research, http://www.cfirguide.org/constructs.html) to better define what they are actually looking at in their study.	Having taken this comment into account with several comments from Einar Hovlid about the definition of engagement, we have decided to change the title of the paper to make explicit the focus on engagement. The new title is: "Exploring how ward staff engage with the implementation of a patient safety intervention: A qualitative process evaluation" We also refrained from using the term "context" in the paper in the sense of 'organisational/ward context' and have replaced this with "setting"
Birgit Heckemann – 25	The objectives of this study are not clearly stated, there is too much focus on the PRASE intervention rather than on the objectives of this process evaluation. Design: Lacks clarity Setting and participants: ok Data: the authors merely mention their data sources, not how these were analysed (pen portraits, adaptive theory). Findings: The authors present results about staff engagement and some influencing factors, but leave out further dimensions that the organizational context includes. Conclusion: These appear to be findings rather than conclusions. I was expecting some recommendations for process evaluation in general / to inform similar projects. Strengths and limitation are not compelling. I am not sure whether the novel approach taken (point 3) is a strength or a limitation. The same is true for point 4 which is speculation and point 5, engagement as a key factor: Engagement has been reported as a key factor in literature on implementation (see above, CFIR).	We have amended the abstract in light of this reviewer's comments, specifically the Objectives and Data section. We have used our Discussion section to situate our findings in the patient safety agenda/ literature to try and ensure that this paper has broad applicability. Therefore, we have steered away from making implementation science recommendations for other similar projects as we prefer to discuss what our findings mean for patient safety more generally. The fourth point of the Strengths and Limitations section has been toned down so as not to introduce speculation.

Birgit Heckemann – 26	The background section lacks focus: line 4 to 16 are concerned with the PRASE trial and the importance of patient feedback. Instead, the authors could have focussed on introducing and discussing process evaluation in more depth.	We felt that it was necessary for the reader to have an explanation of the PRASE trial in order to put the process evaluation methods and findings in context. Without this detail upfront, it will be difficult for the reader to reconcile the findings of this paper with the context of the wider study. We do already provide material about other process evaluations of complex interventions in the Background.
Birgit Heckemann – 27	The manuscript could be enhanced by a brief mention of an overall guiding framework of the PRASE project, such as the MRC framework or a similar theoretical framework, to demonstrate how the process evaluation was planned from the outset and embedded within the overall trial.	We drew upon elements of Grant et al's (2013) framework for designing process evaluations of cluster randomised trials. We have added a sentence into the first paragraph of the Methods section to make this clear.
Birgit Heckemann – 28	Furthermore, the researchers' underlying assumptions with regard to assumed specific barriers and facilitators to successful implementation are not discussed (these could have been explored systematically in the preparation of the trial by drawing on literature on implementation research). The background section therefore does not lead to specific, focussed research questions, but to the rather general question about how and why the intervention works with a focus on staff engagement.	As detailed in the response to comment 5, barriers and facilitators to implementation were previously explored in the feasibility work undertaken prior to the full RCT and have been published elsewhere (O'Hara et al, 2016). We feel that our general research question about how and why the intervention works, with a subsequent focus on engagement, is not problematic. The process evaluation methods were devised a priori to the start of the trial in order to capture a wide range of qualitative data on the implementation of this complex intervention. Research questions which were too tightly focussed may have led to the wrong questions being asked.
Birgit Heckemann – 29	The authors state that they conducted 'short structured phone interviews' (p. 5, l. 32). I would have appreciated a short description of how these phone calls were structured. Was there a guide, maybe? And if so, how was this developed?	Five sentences have been added to the fourth paragraph of the Methods section to address this comment.

Birgit Heckemann – 30	Furthermore the description of how the data were analysed is scant. It is interesting and novel that the authors used a relatively new method, pen portraits, to synthesise their data. However, the authors' assertion that there is a lack of methodological literature pertaining to the construction of said pen portraits (compare p. 5 l. 50-52) does not justify the lack of rigour in the description of how the pen portraits were created in this particular study. The authors speak of a 'basic structure' that was created to write a linear account of how each ward engaged in the intervention. I do not get a clear idea on how data were synthesised from this description. General rules ensuring rigour in qualitative research can also be applied to novel methods (e.g. creating a guide for the creation of the pen portraits, creating an audit trail, member checking, etc.).	We have added three sentences to the end of the fifth paragraph of the Methods to address these set of comments.
Birgit Heckemann – 31	The authors explain 'that they used techniques derived from adaptive theory' (p.6 l. 14/15) citing Layder 1998 [18] to further analyse the pen portraits. However, Layder (1998) states in his book: 'The generation of adaptive theory operates at each and every moment of the research from the preparation and planning of data collection (including choice of methods and techniques, problems of access and so on), through every phase of the actual collection and analysis of the data' (Layder, 1998, p. 174.) It is not clear how the authors ensured this kind of continuity throughout the process evaluation. The manuscript would be enhanced if the authors established a link between data sources, pen portraits and adaptive theory.	Our published protocol (Sheard et al, 2014) was written before the process evaluation data collection took place and in this we cite “techniques derived from adaptive theory (Layder, 1998)” as our analytical strategy. This was specified a priori. However, we were working within the constraints of a process evaluation of a cluster RCT. Therefore, we use the phrase “techniques derived from” as we were particularly interested in applying Layder’s approach to the data analysis. Specifically, his notion of moving between empirical data and theoretical interpretations working in a continuous cycle. Comment 9 and 31 also express uncertainty over our analytic approach. We have added seven sentences to the last paragraph of the Methods to address this. See our response to comment 9 for a more thorough reply on this issue.

Birgit Heckemann – 32	The authors aim to describe 'context, circumstance and divergence' in the findings section. However, they actually present findings about staff engagement along with some influencing factors, such as trust level support and facilitative processes. While the findings regarding staff engagement are interesting, the number of influencing factors examined is limited and based on authors' assumptions, which are mentioned only in this section (e.g. p. 7, l. 37- 39, l. 42-43). These assumptions are essential and should be presented much earlier in the paper, for example in the background section. It would also be interesting to learn why these particular assumptions were made and considered most relevant with a view to existing implementation research (see CFIR, as noted above).	We have altered the first paragraph of the Findings to focus it on engagement. Our assumptions come directly from the feasibility study we undertook prior to commencement of the full RCT. A sentence has been added to the Background section to refer the reader to the O'Hara et al (2016) paper. Whilst we feel it is necessary to signpost the reader, we do not want to take up space in this paper detailing material that is published elsewhere.
Birgit Heckemann – 33	The authors discuss their findings in relation to patient safety. They state that 'the relationships between different parts and levels of the organization from senior management to ward teams to individuals were vital in achieving success.'(p. 10, l 43-45). This statement appears trivial, because it would be true for any complex intervention. Furthermore, the discussion lacks in-depth exploration about how interventions regarding patient safety differ from other complex interventions in clinical practice	We understand that this reviewer may find the sentence they highlighted as trivial but making this broad statement (based on our findings) leads us into and then allows us to make a discussion point about organisational alignment in relation to what we found. We do not feel that one of the main points we wish to make in our Discussion section is about the difference between patient safety interventions and other clinical interventions. We are concerned here with situating our findings in the broad patient safety literature.
Birgit Heckemann - 34	The first paragraph on the limitations section seems to be speculation. I do not really understand what the authors are implying here. The second limitation statement concerning the pen portrait methodology is plausible, but I think there is room for improvement in the rigour of reporting.	We have re-written this sentence to tone down any elements of speculation. Further description has been added to the paper concerning how the pen portraits were created (as per comment 30)

Birgit Heckemann – 35	While I understand what the findings mean in terms of the PRASE trial, the conclusion could be enhanced by an explanation of how the authors' findings could inform similar projects.	The Conclusion section offers a brief concluding summary of the paper as a whole. We would not want to introduce new material in this brief section where the material has not been detailed elsewhere in the paper.
Birgit Heckemann – 36	References - Could be enhanced by including literature on implementation science.	We have now included Dixon-Woods et al (2011) and Damschroder et al (2009)
Birgit Heckemann – 37	Appendices 1,2 . These are difficult to read and should be, in my opinion, more concise. Appendix 3: This is a very nice illustration of the findings.	We have now streamlined the appendices. Appendix 1 and 2 have now merged content and appears as Appendix 1.

VERSION 2 – REVIEW

REVIEWER	Birgit Heckemann Maastricht University
REVIEW RETURNED	10-Mar-2017

GENERAL COMMENTS	Thank you for the opportunity to review this much improved paper (incl. appendices) again. The new title is well chosen and I appreciate the consideration and effort with which each of the four reviewers' comments have been addressed.

REVIEWER	Einar Hovlid University of Bergen, Norway
REVIEW RETURNED	07-Mar-2017

GENERAL COMMENTS	Comment 11 Einar Hovlid P 7 line 47 My comment: As I understand it, your findings are drawn from the pen portraits which are written on a ward level. Where do the findings about the Trust come from? Your answere: The findings that we describe in our second theme - Trust-level support for an intervention does not predict the strength of ward-level engagement' arise predominantly from the pen portraits. However, we have utilised certain tacit knowledge gained throughout the setup, roll out and implementation of the intervention pertaining to the support shown at the level of the Trust. I agree that this is a relevant and important finding. My concern is that your finding in part is based on what you describe as “tacit knowledge”. I would recommend that you include a few sentences in the method section that describes you data source and analysis.

VERSION 2 – AUTHOR RESPONSE

Thanks to Einar Hovlid and Birgit Heckemann for reviewing the paper again.

In response to the final outstanding comment, we have written two new sentences in the third paragraph of the Methods section. This demonstrates how the ‘tacit knowledge’ we utilised in theme

two was captured.

VERSION 3 – REVIEW

REVIEWER	Birgit Heckemann Maastricht University
REVIEW RETURNED	03-Apr-2017

GENERAL COMMENTS	No further comments

REVIEWER	Jamie Murdoch University of East Anglia, United Kingdom
REVIEW RETURNED	05-Apr-2017

GENERAL COMMENTS	The authors have addressed all reviewer's previous comments. The article provides an interesting contribution to the process evaluation literature, particularly the challenges of trial implementation within secondary care.

REVIEWER	Einar Hovlid University of Bergen, Norway
REVIEW RETURNED	03-Apr-2017

GENERAL COMMENTS	I have no further omments

REVIEWER	Tara Lamont University of Southampton Chilworth Park SO16 7NS
REVIEW RETURNED	07-Apr-2017

GENERAL COMMENTS	This is a second review of a revised manuscript, so I have checked against my original concerns. Overall, I think the paper is now more clearly framed as a process evaluation, with clearer implications for the service and the authors have grounded this work more closely in other relevant patient safety and quality improvement literature. I think it is acceptable as a qualitative study offering insights into service development and staff engagement with improvement. A small note - I do not like the term `disinterested' to mean lack of interest (as opposed to neutral or without a stake in something_. It might be better for the final part of their scale of engagement to be labelled `disengaged or lack of interest' or similar. There are a few other typographical/grammatical errors, which perhaps will be picked up by copyediting - for instance, on p12 the sentence starting `Thirdly...most staff do want action plan'.

VERSION 3 – AUTHOR RESPONSE

Thanks to all four reviewers for looking over the paper again.

In response to Tara Lamont's comment, we have removed the phrase "disinterested" in relation to the engagement typology of two wards. We have replaced this with the phrase "disengaged" where appropriate throughout the document.