

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)	Tasks, multitasking and interruptions among the surgical team in an operating room: a prospective observational study
AUTHORS	Göras, Camilla; Olin, Karolina; Unbeck, Maria; Pukk Harenstam, Karin; Ehrenberg, Anna; Tessma, Mesfin; Nilsson, Ulrica; Ekstedt, Mirjam

VERSION 1 - REVIEW

REVIEWER	Prof Nick Sevdalis KCL, UK
REVIEW RETURNED	15-Oct-2018

GENERAL COMMENTS	<p>The authors report a prospective observational study of multitasking in operating theatres, using a previously validated tool (WOMBAT). They report a range of descriptive findings on multitasking and interruptions, particularly to the surgical team in theatre.</p> <p>I found the paper of interest, and well written. It is within the scope of the journal and it should be attractive to a large readership. The use of a previously developed tool is an asset of the research. I do not have any major reservations – but there are a few points that require clarification in my view for the paper to be acceptable for final publication – as follows:</p> <p>The main issue that requires attention in my view is that the novelty of the research is not well articulated. I am not clear in other words what is really new in the aims of the study and how it advances our understanding of multitasking and interruptions in surgical work (p 6). The authors need to make a clearer statement about how this study advanced the evidence base.</p> <p>Related to the above point: the Discussion of the paper is quite descriptive. It lacks depth and articulation of implications of the findings. What do the findings on multitasking and the interruptions the authors recorded mean for surgical safety or workflow? How should these be managed? They comment on a newly introduced policy that reduced interruptions from external phone calls and bleeps to the surgical team – but they do not reflect on its necessity. So does surgical work involve multitasking by nature – or could it be designed differently? These issues require more thought.</p> <p>P 9: I am unclear whether ethical approval was sought and provided. Informed consent is not enough. This is a major issue as</p>
-------------------------	--

	<p>I would not recommend publication of a study unless it has been ethically reviewed.</p> <p>Finally, inter-rater reliability: can you provide some more detail on p 9 on the reliability across tasks? You offer a ballpark figure, but in my own experience of doing this research some tasks are easier to rate than others.</p>
--	--

REVIEWER	Scott R Walter Macquarie University, Australia
REVIEW RETURNED	27-Oct-2018

GENERAL COMMENTS	<p>This is a descriptive study of work conducted by nurses, registered nurse anaesthetists and surgeons in operating rooms. WOMBAT software was used to collect detailed information on work tasks, including interruptions and multitasking. There is relatively little in the literature on these aspects of work performed in ORs and in that respect this paper is a useful contribution. Also the identification of interruption sources is something rarely done in studies of clinical work and conveys important information about interruptive events. There are, however, several issues with the manuscript which I detail below. These fall into two broad groups: 1) the framing and justification of the study question needs refinement and clarification, part of which involves a more accurate and critical assessment of evidence from the literature including interruption concepts, and 2) the details of the data analysis need to be clarified and, in some instances, possibly reanalysed, which will have implications for the results and hence the discussion. While the comments are extensive, the basic design of the study is sound, and with revision the paper will contribute an interesting and useful set of findings.</p> <p>STRENGTHS AND LIMITATIONS</p> <p>1. Second dot point: I am not sure what is meant by an 'objective' definition of interruption. The term interruption has been used to mean many things in the literature. However, if you meant that the study definition was clearly operationally defined, then that may be better expressed with terms other than 'objective'.</p> <p>BACKGROUND</p> <p>2. In paragraph 3, there are some issues with the way multitasking and interruptions are described.</p> <p>a) In general the first part of the paragraph makes claims about multitasking that are not exactly representative of the cited literature. Reference 19 did not use the term 'interruption' yet it appears in the sentence, while multitasking was a relatively infrequent strategy, rather than one that is 'often used'. The claims related to reference 2 (Skaugset et al.) are not supported by that paper, that is, the relevant sentence appearing in that article provides no citations. The authors should consider revising this section to more accurately reflect the literature, and to avoid the temptation to make strong claims about negative effects without clear supporting evidence.</p> <p>b) "Even though in earlier studies multitasking may have contributed to medical errors,</p>
-------------------------	---

lately it has also been recognized as an integral part of and skill for execution of daily practices, especially in acute care.” Once again the claim of a clear link between multitasking and error is tenuous, and further there are no citations for this sentence.

c) The paragraph seems to be a ‘catch-all’ for various aspects of multitasking in healthcare, but it is rather unfocused and the main point of the paragraph is not clear.

3. Paragraph 4

a) As with the comment above, this paragraph seems to attempt a comprehensive literature review, but could benefit from being more focused and concise.

b) For the sentence: “Relationships have been found between increased interruptions and error rates”, the cited paper by Westbrook et al. does not support this claim as they did not analyse errors. They did look at the effect of interruptions on task completion time which is a different question, and a reanalysis of this data generated a different finding in that regard (see Walter et al., 2016). A more recent paper, and one of the very few to find such an association, is by Westbrook et al. (2018), who observed increased prescribing error rates associated with interruptions during prescribing.

c) In the Wiegmann paper (ref. 29) the analysis ignored length bias. The longer you observe the higher the number of both errors and interruptions, which means the variation in counts of both errors and interruptions could partly be due to the duration of the observed procedure. The result is that it is possible to see a correlation in counts of the two measures even if there is no real association. That is, there is a reasonable chance that the results are false positives.

d) The discussion of interruption definitions is unclear. In this regard the literature is a mess and a summary of literature needs to be more critical as well as drawing out the aspects of the literature that are relevant to this study. Arguably applying the term itself in healthcare is problematic for a number of reasons (see Walter, 2018), but with a clearer definition of the term for this study, it is then easier to draw out the studies that have studied similar concepts (even if labelled with a different term). This would also clarify aspects of the results and discussion related to interruptions.

METHODS

4. In the Tools and Definitions section, sentence 2: it would be helpful to specify that the dimension and categories were user-defined, that is, customised by the researchers.

5. Is there any overlap between the categories Medication-Document and Documentation?

6. The definition of an interruption is typical of other studies, but similarly ambiguous. The term interruption is used as part of the definition of an interruption which seems like circular logic. The term has been used to mean the external event prompting a change in workflow (e.g. the surgeon’s request to change table position), the decision to cease one task and start another in response to an external event (e.g. stopping the infusion task and assisting with the repositioning), the task performed in response to the external prompt (e.g. repositioning the patient) or the whole

interaction which combines all of the above elements. Whatever the definitions used, they need to be clear and used in a consistent way throughout the paper.

7. The section on inter-rater reliability could be considerably improved. Agreement on number of tasks is very basic and does not capture agreement on aspects of the data that are relevant to the analysis, e.g. ignores timing and temporal order as well as agreement on dimensions of interest (What, Who, etc.). The latest emerging approach to quantifying agreement is to reformat the data from the parallel observation sessions into 1 second time windows, align the time-windows from each observer, then apply either univariate measures (e.g. Cohen's kappa, Fleiss' kappa) to individual variables to generate multiple kappa scores, or apply a multivariate measure (see Janson & Olsson, 2001).

8. The Data Analysis section seems to be rather brief. Even simple summary measures with CIs are not straightforward with WOMBAT data, but the current description is insufficient for me to say whether the analyses were appropriate or not.

RESULTS

9. Task counts and proportions of tasks are not particularly useful measures since they tell us as much about the way tasks were defined as it does about the work. In general, time-based measures (e.g. proportions of time) provide a more informative summary of observed work.

10. Table 3

a) The column title "Frequency of tasks per total observed time", is not entirely clear. Is it the rate of tasks per hour?

b) There may be no need to say "per profession" in every column title, as it is covered in the table title.

c) The proportions of time at the bottom of the table (33.6, 44.4, 22.0) use a different denominator to the rest of the column and may not be particularly useful, since they essentially summarise observation time which tells us nothing about the relative work profiles of the three professions.

d) The proportions of multitasking seem overstated. E.g. if total task time is 261 hours and total observation time is 169 then the overall proportion of time spent multitasking can be no more than 27% [$0.5 \times (261 - 169) / 169$]. This is quite different to the reported 48%, suggesting there may be some multiple counting of multitasking intervals that was not taken into account.

11. Interruptions section

a) The section title does not seem to make sense. If there are "interrupted tasks", "sources of interruption" and "responses after interruption", then what does the term "interruption" refer to? This relates to the earlier comment about the definitions of interruptions. It would be helpful to have a clearer definition of terms earlier in the paper so that the use of terms in the results is unambiguous.

b) With the percentages reported in the interruptions section, it may be easier to follow if the percentages were also shown in Table 4. That is, rather than showing the count only, the column percentages could be included in the format of n (%), for example.

c) An interesting aspect of the data would be the combination of the type of interrupted task and the source of interruption. For

example, interruptions sources from other professionals may be more likely to occur during certain tasks. This would not necessarily require a cross tabulation of the two variables (which might be a bit unwieldy), but the authors may consider reporting the important combinations in the text, or a table of the common combinations only.

d) Table 4: In the title, are 'sources' and 'causes' the same thing? What does that mean for the overall interruption definition? Again, a clear system of terminology is needed.

DISCUSSION

12. Paragraph 2: Following on from comment 9, I am not sure task counts and proportions need to be emphasised in the discussion, I would suggest minimising or even omitting this.

13. The discussion of multitasking may change after reanalysing the proportions of time on multitasking. However, surgery is a setting that is naturally more suited to multitasking as there is a 'hands on' component while there is also regular communication, and tasks of different modality (e.g. manual vs, verbal) are more easily combined. It may be that for OR staff much of the multitasking is experienced more like a single task and may thus be less cognitively demanding (and error prone) than multitasking between tasks with differing content or with less compatible modalities. This should be considered when comparing results from other studies, particularly non-OR settings.

14. Paragraph 6: "Out of all interruptions identified in this study, patient-related and procedure-related interruptions often arose in situations where safe and smooth intra-operative care processes needed to be secured – e.g. when patient positioning was altered for better visibility or changed operative plans required new equipment. This is an example of how team adaptation can counteract the negative impact of increased complexity introduced by interruptions or new medical challenges"

To me these things sound like a normal part of team flexibility, and this highlights the difficulty with lumping a diverse set of behaviours under one term, such as 'interruption'. Teams need to be able to adapt to changes in circumstance, and that is an essential aspect of resilience in healthcare. Looking through a Safety II lens you could say that teams coping well with unpredicted changes is the key finding here. In any case, it is important to consider the way categories of work are defined when interpreting the results, so that connotations encoded into task definitions (e.g. "interruptions=negative effects") don't produce results that reinforce preconceived ideas.

15. The discussion of the OR as a CAS is rather general and does not seem to add value to the discussion. The paper would hold together without this, but if the authors want to retain this thread then the discussion may need some development to emphasise the aspects of OR revealed in this study that support the idea of a CAS.

16. Paragraph 8: The idea of "the dual nature of interruptions" seems to set up a binary concept, presumably beneficial/detrimental. However, the events typically lumped under the umbrella of interruption are diverse and their effects may not simply be good or bad.

	<p>REFERENCES</p> <p>Walter SR, Dunsmuir WTM, Westbrook JI. 2016. Assessing the effect of interruptive events on task completion time: a multi-site study. In: Proceedings of the Human Factors, Ergonomics and Patient Safety conference (HEPS 2016): Toulouse, France, 4-7 October.</p> <p>Westbrook JI, Raban MZ, Walter SR, Douglas HE. Task errors by emergency physicians are associated with interruptions, multitasking, fatigue and working memory capacity: a prospective, direct observation study <i>BMJ Qual Saf</i> 2018;27:655-663.</p> <p>Walter SR. 2018. Interruptions in Emergency Medicine: Things Are Not Always What They Seem. <i>Academic Emergency Medicine</i>, in press, doi: 10.1111/acem.13505</p> <p>Janson H, Olsson U. 2001. A measure of agreement for interval or nominal multivariate observations. <i>Educational and Psychological Measurement</i>, 61(2): 277-289.</p>
--	--

VERSION 1 – AUTHOR RESPONSE

	Comments reviewer I	Revisions made by authors	Page
1)	<p>Background:</p> <p>The novelty of the research is not well articulated. I am not clear in other words what is really new in the aims of the study and how it advances our understanding of multitasking and interruptions in surgical work (p 6). The authors need to make a clearer statement about how this study advanced the evidence base</p>	Revised according to comments.	P.5, line 35
2)	<p>Discussion: The discussion of the paper is quite descriptive. It lacks depth and articulation of implications of the findings. What do the findings on multitasking and the interruptions the authors recorded mean for surgical safety or workflow? How should these be managed? They comment on a newly introduced policy that reduced interruptions from external phone calls and bleeps to the surgical team – but they do not reflect on its necessity. So does surgical work involve multitasking by nature – or could it be designed differently? These issues require more thought.</p>	The discussion section is revised and clarified according to comments, especially in paragraph 3 on page 16.	P.15, lines 14 -
3)	<p>Ethical approval: I am unclear whether ethical approval was sought and provided. Informed consent is not enough. This is a major issue as I would not recommend</p>	A separate section that clarifies ethical approval is added to manuscript.	P.10, line 1

	publication of a study unless it has been ethically reviewed.		
4)	Finally, inter-rater reliability: can you provide some more detail on p 9 on the reliability across tasks? You offer a ballpark figure, but in my own experience of doing this research some tasks are easier to rate than others.	More information is added according to comments.	P.9, lines 1-
	Comments reviewer II	Revisions made by authors	
1)	Strengths and limitations: Second dot point: I am not sure what is meant by an 'objective' definition of interruption. The term interruption has been used to mean many things in the literature. However, if you meant that the study definition was clearly operationally defined, then that may be better expressed with terms other than 'objective'.	According to comments, the term objective is exchanged to 'operationalized' in manuscript.	P.3, line 6
2)	<p>Background: In paragraph 3</p> <p>a) In general the first part of the paragraph makes claims about multitasking that are not exactly representative of the cited literature. Reference 19 did not use the term 'interruption' yet it appears in the sentence, while multitasking was a relatively infrequent strategy, rather than one that is 'often used'. The claims related to reference 2 (Skaugset et al.) are not supported by that paper, that is, the relevant sentence appearing in that article provides no citations. The authors should consider revising this section to more accurately reflect the literature, and to avoid the temptation to make strong claims about negative effects without clear supporting evidence.</p> <p>b) "Even though in earlier studies multitasking may have contributed to medical errors, lately it has also been recognized as an integral part of and skill for execution of daily practices, especially in acute care." Once again the claim of a clear link between multitasking and error is tenuous, and further there are no citations for this sentence.</p> <p>c) The paragraph seems to be a 'catch-all' for various aspects of multitasking in healthcare, but it is rather unfocused and the main point of the paragraph is not clear.</p>	a-c) The literature has been reviewed once again and according to this the authors agree that there are no supporting evidence and that strong claims cannot be made. Major revisions have been done according to this.	P.4, lines 20-

3)	<p>Background: Paragraph 4</p> <p>a) As with the comment above, this paragraph seems to attempt a comprehensive literature review, but could benefit from being more focused and concise.</p> <p>b) For the sentence: “Relationships have been found between increased interruptions and error rates”, the cited paper by Westbrook et al. does not support this claim as they did not analyse errors. They did look at the effect of interruptions on task completion time which is a different question, and a reanalysis of this data generated a different finding in that regard (see Walter et al., 2016). A more recent paper, and one of the very few to find such an association, is by Westbrook et al. (2018), who observed increased prescribing error rates associated with interruptions during prescribing.</p> <p>c) In the Wiegmann paper (ref. 29) the analysis ignored length bias. The longer you observe the higher the number of both errors and interruptions, which means the variation in counts of both errors and interruptions could partly be due to the duration of the observed procedure. The result is that it is possible to see a correlation in counts of the two measures even if there is no real association. That is, there is a reasonable chance that the results are false positives.</p> <p>d) The discussion of interruption definitions is unclear. In this regard the literature is a mess and a summary of literature needs to be more critical as well as drawing out the aspects of the literature that are relevant to this study. Arguably applying the term itself in healthcare is problematic for a number of reasons (see Walter, 2018), but with a clearer definition of the term for this study, it is then easier to draw out the studies that have studied similar concepts (even if labelled with a different term). This would also clarify aspects of the results and discussion related to interruptions.</p>	<p>a – d) The literature has been reviewed once again and according to this the authors agree that there are no supporting evidence and that strong claims cannot be made. Major revisions have been done according to this.</p> <p>To clarify the mess of interruption definitions we have early in the paper, added a table (1) with relevant concepts and operational definitions.</p>	<p>P.4, lines 20-</p> <p>P.7, line 6</p>
4)	<p>Methods: In the Tools and Definitions section, sentence 2: it would be helpful to specify that the dimension and categories were user-defined, that is, customised by the researchers.</p>	<p>Revised according to comments.</p>	<p>P.6, line 33</p>

5)	Methods: Is there any overlap between the categories Medication-Document and Documentation?	In case of deeper analysis concerning medication, the authors deliberately separated documentation concerning medication and other documenting activities.	
6)	Methods: The definition of an interruption is typical of other studies, but similarly ambiguous. The term interruption is used as part of the definition of an interruption which seems like circular logic. The term has been used to mean the external event prompting a change in workflow (e.g. the surgeon's request to change table position), the decision to cease one task and start another in response to an external event (e.g. stopping the infusion task and assisting with the repositioning), the task performed in response to the external prompt (e.g. repositioning the patient) or the whole interaction which combines all of the above elements. Whatever the definitions used, they need to be clear and used in a consistent way throughout the paper	To clarify the mess of interruption definitions we have early in the paper, added a table (1) with relevant concepts and operational definitions.	P.7, line 6
7)	Methods: The section on inter-rater reliability could be considerably improved. Agreement on number of tasks is very basic and does not capture agreement on aspects of the data that are relevant to the analysis, e.g. ignores timing and temporal order as well as agreement on dimensions of interest (What, Who, etc.). The latest emerging approach to quantifying agreement is to reformat the data from the parallel observation sessions into 1 second time windows, align the time-windows from each observer, then apply either univariate measures (e.g. Cohen's kappa, Fleiss' kappa) to individual variables to generate multiple kappa scores, or apply a multivariate measure (see Janson & Olsson, 2001).	More information has been added on interrater reliability. The suggestions of the multivariate kappa scores will be taken into consideration in further publication and possible future study design.	P.9, line 1
8)	Methods: The Data Analysis section seems to be rather brief. Even simple summary measures with CIs are not straightforward with WOMBAT data, but the current description is insufficient for me to say whether the analyses were appropriate or not.	Revised according to comments.	P.10, line 8 -

9)	<p>Results: Task counts and proportions of tasks are not particularly useful measures since they tell us as much about the way tasks were defined as it does about the work. In general, time-based measures (e.g. proportions of time) provide a more informative summary of observed work</p>	Revised according to comments.	P.11, lines 19 - 20
10)	<p>Results: a) The column title “Frequency of tasks per total observed time”, is not entirely clear. Is it the rate of tasks per hour?</p> <p>b) There may be no need to say “per profession” in every column title, as it is covered in the table title.</p> <p>c) The proportions of time at the bottom of the table (33.6, 44.4, 22.0) use a different denominator to the rest of the column and may not be particularly useful, since they essentially summarise observation time which tells us nothing about the relative work profiles of the three professions.</p> <p>d) The proportions of multitasking seem overstated. E.g. if total task time is 261 hours and total observation time is 169 then the overall proportion of time spent multitasking can be no more than 27% $[0.5 \cdot (261 - 169) / 169]$. This is quite different to the reported 48%, suggesting there may be some multiple counting of multitasking intervals that was not taken into account.</p>	<p>a – b) Revised according to comments.</p> <p>c) Removed according to comments.</p> <p>d) In this study, the total category specific multitasking time was 173 hours 46 minutes. Using the formula suggested by the reviewer ($0.5 \cdot 173 / 169$), the result is 51%. However, as there were instances when more than two tasks were performed simultaneously – the total multitasking time was 82 hours 6 minutes, instead of 86 hours 5 minutes, resulting to the 48% reported in the manuscript.</p> <p>During several of the observed procedures, students were present. When this was the case, the observed profession had “supervision” as an ongoing task during the time when observed professional was in the OR. This resembles what was observed (active, continuous performance), and may be one – but not the main - explanation, why the proportion of multitasking is extensive. The seemingly large amount of multitasking seems indeed to be a very descriptive feature of operation room teamwork, especially regarding communication.</p> <p>Based on reviewer’s comment on multiple counting of multitasking intervals, the data was checked accordingly, and as a result some changes were made in category specific task time and category specific multitasking time. The changes were not significant and thus did not affect the discussion or conclusions. However, authors are extremely grateful for reviewer’s insight on the method.</p>	P.12, table 4 P.13, line 3

11)	<p>Results: Interruption section</p> <p>a) The section title does not seem to make sense. If there are “interrupted tasks”, “sources of interruption” and “responses after interruption”, then what does the term “interruption” refer to? This relates to the earlier comment about the definitions of interruptions. It would be helpful to have a clearer definition of terms earlier in the paper so that the use of terms in the results is unambiguous.</p> <p>b) With the percentages reported in the interruptions section, it may be easier to follow if the percentages were also shown in Table 4. That is, rather than showing the count only, the column percentages could be included in the format of n (%), for example.</p> <p>c) An interesting aspect of the data would be the combination of the type of interrupted task and the source of interruption. For example, interruptions sources from other professionals may be more likely to occur during certain tasks. This would not necessarily require a cross tabulation of the two variables (which might be a bit unwieldy), but the authors may consider reporting the important combinations in the text, or a table of the common combinations only.</p> <p>d) Table 4: In the title, are ‘sources’ and ‘causes’ the same thing? What does that mean for the overall interruption definition? Again, a clear system of terminology is needed.</p>	<p>a) The previous definitions of interruption, interrupting task, interrupted task and task after interruption have, in table 1, been updated and clarified in manuscript.</p> <p>b) Revised according to comment.</p> <p>c) Interesting suggestions which we further can analyze and report in an upcoming publication.</p> <p>d) Revised according to comment.</p>	<p>P.7, line 6, table 1</p> <p>P.14, table 5</p> <p>P.14, table 5</p>
12)	<p>Discussion: Paragraph 2: Following on from comment 9, I am not sure task counts and proportions need to be emphasised in the discussion, I would suggest minimising or even omitting this.</p>	<p>Removed according to comments.</p>	

13)	<p>Discussion: The discussion of multitasking may change after reanalysing the proportions of time on multitasking. However, surgery is a setting that is naturally more suited to multitasking as there is a 'hands on' component while there is also regular communication, and tasks of different modality (e.g. manual vs, verbal) are more easily combined. It may be that for OR staff much of the multitasking is experienced more like a single task and may thus be less cognitively demanding (and error prone) than multitasking between tasks with differing content or with less compatible modalities. This should be considered when comparing results from other studies, particularly non-OR settings.</p>	<p>The reanalysis of multitasking did not provide results that requires revisions in the discussion. However, the comment is highly relevant and has been addressed carefully.</p> <p>The discussion section is revised and clarified according to comments.</p>	P.15, line 1-
14)	<p>Discussion: Paragraph 6: "Out of all interruptions identified in this study, patient-related and procedure-related interruptions often arose in situations where safe and smooth intra-operative care processes needed to be secured – e.g. when patient positioning was altered for better visibility or changed operative plans required new equipment. This is an example of how team adaptation can counteract the negative impact of increased complexity introduced by interruptions or new medical challenges" To me these things sound like a normal part of team flexibility, and this highlights the difficulty with lumping a diverse set of behaviours under one term, such as 'interruption'. Teams need to be able to adapt to changes in circumstance, and that is an essential aspect of resilience in healthcare. Looking through a Safety II lens you could say that teams coping well with unpredicted changes is the key finding here. In any case, it is important to consider the way categories of work are defined when interpreting the results, so that connotations encoded into task definitions (e.g. "interruptions=negative effects") don't produce results that reinforce preconceived ideas</p>	<p>Revised according to comments – negative impact has been removed.</p>	P. 15 line 1 -

15)	Discussion: The discussion of the OR as a CAS is rather general and does not seem to add value to the discussion. The paper would hold together without this, but if the authors want to retain this thread then the discussion may need some development to emphasise the aspects of OR revealed in this study that support the idea of a CAS.	The concept of the OR as a CAS is considered important from a greater perspective as complexity theory and CAS are the main theoretical framework in a thesis that this paper is part of. To clarify the importance of the concept for the OR revisions were made according to comments.	P.15, line 1-
16)	Discussion: Paragraph 8: The idea of “the dual nature of interruptions” seems to set up a binary concept, presumably beneficial/detrimental. However, the events typically lumped under the umbrella of interruption are diverse and their effects may not simply be good or bad	Revised according to comments, dual nature was instead replaced by diverse nature which means that it is not only binary but could be positive, negative or for example neutral.	P.17, line 4

VERSION 2 – REVIEW

REVIEWER	Professor Nick Sevdalis KCL, UK
REVIEW RETURNED	22-Jan-2019

GENERAL COMMENTS	The revisions have addressed the points I raised, no further comments
-------------------------	---

REVIEWER	Scott R Walter Macquaire University
REVIEW RETURNED	11-Feb-2019

GENERAL COMMENTS	<p>My previous comments have been largely addressed. There are only a few points that remain.</p> <p>1. Background</p> <p>a) Paragraph 5: I am still not sure what the key message of this paragraph is. Currently it covers many topics - definitions, negative associations, interventions, positive effects and communication-based interruptions. I assume you are trying to build the case for studying interruptions in the OR, but I don't think the paragraph achieves that in its current form.</p> <p>b) Paragraph 6: “These findings reveal interruptions to be a predominantly negative phenomenon in the OR”. As always with interruptions research, results must be interpreted with caution. There is a tendency for researchers to design studies around negative effects, or to emphasise any negative aspects as key findings since they support status quo assumptions about interruptions. The findings referred to in this sentence may not</p>
-------------------------	--

necessarily mean that interruptions are 'predominantly negative in OR', but rather that negative aspects are more likely to be studied and negative results more likely to be published. It may make more sense to say "interruptions are not well understood" as the justification for studying them in the OR, rather than try to construct a narrative about them being negative.

2. Table 1 is a useful addition, but some of the specific definitions are still unclear.

-My understanding of the definition of an 'interruption' in the table is that it refers to the act of suspending a task in response to an external stimulus. That is, it describes the surgeon's response to the stimulus only, as distinct from describing the sequence of: primary task->external event->suspend primary task->commence secondary task.

-The 'cause of an interruption' is a clear enough definition, but other terms have been used in the past to describe the same concept. In Trafton et al.'s 'anatomy of an interruption' this is called the 'alert for the secondary task', while in my papers I use the term 'external prompt' as part of a broader systems of terms that encompasses all types of change in activity during clinical work. The authors may consider whether any of these existing terms adequately describe this concept.

-The 'interrupting task' definition is somewhat less clear. As I understand it, the 'cause of an interruption' is what does the interrupting. If 'interrupting task' is meant to describe the task that follows the cause/alert/prompt, then it might be better described in a different way. A typical term for this concept is 'secondary task' (e.g. Trafton et al.).

- The 'interrupted task' is clear enough, but the authors may consider using existing terms: e.g. primary task.

- 'Task after interruption' is the least clear definition since it seems to describe the same concept as 'interrupting task', i.e. the task following the cause/alert/prompt. In the results section 'tasks after interruption' and 'tasks following interruption' are mentioned, but then 'interrupting task' is used elsewhere seemingly to describe the same concept. If these are distinct concepts then more explanation is needed. If not, then fewer terms are needed.

Trafton JG, Altmann EM, Brock DP, Mintz FE. 2003. Preparing to resume an interrupted task: effects of prospective goal encoding and retrospective rehearsal. *Int. J. Human-Computer Studies*, 58: 583–603.

Walter SR, Raban MZ, Dunsmuir WTM, Douglas HE, Westbrook JI. 2017. Emergency doctors' strategies to manage competing workload demands in an interruptive environment: An observational workflow time study. *Applied Ergonomics*, 58: 454-460.

3. The IRR section is much better. The main thing that remains unclear is how tasks were aligned for the kappa calculation. To my knowledge there is no way to determine with certainty which pairs of tasks from two observers refer to the same observed activity, and previous studies of this type never report those details. The ICC approach ignores the time ordering of the tasks but avoids this problem of task alignment since it is applied at an aggregated level.

	4. Data Analysis: It would be useful to cite the WOMBAT Analysis Guide so that readers know what you are referring to: Walter SR, Li L, Westbrook JI (2018) A Guide to the Analysis of Data from the Work Observation Method by Activity Timing (WOMBAT) System. Sydney: Australian Institute of Health Innovation, Macquarie University.
--	--

VERSION 2 – AUTHOR RESPONSE

	Reviewers comments	Response from authors
1.	a) Paragraph 5: I am still not sure what the key message of this paragraph is. Currently it covers many topics - definitions, negative associations, interventions, positive effects and communication-based interruptions. I assume you are trying to build the case for studying interruptions in the OR, but I don't think the paragraph achieves that in its current form.	Revised according to reviewer's suggestion
	b) Paragraph 6: "These findings reveal interruptions to be a predominantly negative phenomenon in the OR". As always with interruptions research, results must be interpreted with caution. There is a tendency for researchers to design studies around negative effects, or to emphasise any negative aspects as key findings since they support status quo assumptions about interruptions. The findings referred to in this sentence may not necessarily mean that interruptions are 'predominantly negative in OR', but rather that negative aspects are more likely to be studied and negative results more likely to be published. It may make more sense to say "interruptions are not well understood" as the justification for studying them in the OR, rather than try to construct a narrative about them being negative.	Revised according to reviewer's suggestion
2.	Table 1 is a useful addition, but some of the specific definitions are still unclear.	Thank you for your valuable comments and guidance. It has helped us further in avoiding possible misinterpretations in relation to other studies. We have revised the manuscript and concepts according to your suggestions
	-My understanding of the definition of an 'interruption' in the table is that it refers to the act of suspending a task in response to an external stimulus. That is, it describes the surgeon's response to the stimulus only, as distinct from describing the sequence of: primary task->external event->suspend primary task->commence secondary task.	Revised according to reviewer's suggestion

	<p>-The 'cause of an interruption' is a clear enough definition, but other terms have been used in the past to describe the same concept. In Trafton et al.'s 'anatomy of an interruption' this is called the 'alert for the secondary task', while in my papers I use the term 'external prompt' as part of a broader systems of terms that encompasses all types of change in activity during clinical work. The authors may consider whether any of these existing terms adequately describe this concept.</p>	<p>Respectfully the authors would like to keep the original concept (cause). However additional information has been added"the observable cause has in other studies been named as 'alert for the secondary task'(51) or 'external prompt'(40)"</p>
	<p>-The 'interrupting task' definition is somewhat less clear. As I understand it, the 'cause of an interruption' is what does the interrupting. If 'interrupting task' is meant to describe the task that follows the cause/alert/prompt, then it might be better described in a different way. A typical term for this concept is 'secondary task' (e.g. Trafton et al.).</p>	<p>Revised according to reviewer's suggestion – to secondary task.</p>
	<p>- The 'interrupted task' is clear enough, but the authors may consider using existing terms: e.g. primary task.</p>	<p>Revised according to reviewer's suggestion – to primary task</p>
	<p>- 'Task after interruption' is the least clear definition since it seems to describe the same concept as 'interrupting task', i.e. the task following the cause/alert/prompt. In the results section 'tasks after interruption' and 'tasks following interruption' are mentioned, but then 'interrupting task' is used elsewhere seemingly to describe the same concept. If these are distinct concepts then more explanation is needed. If not, then fewer terms are needed.</p>	<p>This is very true. Revised according to reviewer's suggestion – to task after secondary task</p>
	<p>Trafton JG, Altmann EM, Brock DP, Mintz FE. 2003. Preparing to resume an interrupted task: effects of prospective goal encoding and retrospective rehearsal. <i>Int. J. Human-Computer Studies</i>, 58: 583–603.</p> <p>Walter SR, Raban MZ, Dunsmuir WTM, Douglas HE, Westbrook JI. 2017. Emergency doctors' strategies to manage competing workload demands in an interruptive environment: An observational workflow time study. <i>Applied Ergonomics</i>, 58: 454-460.</p>	

	<p>3. The IRR section is much better. The main thing that remains unclear is how tasks were aligned for the kappa calculation. To my knowledge there is no way to determine with certainty which pairs of tasks from two observers refer to the same observed activity, and previous studies of this type never report those details. The ICC approach ignores the time ordering of the tasks but avoids this problem of task alignment since it is applied at an aggregated level.</p>	<p>We agree, that it is difficult to be certain. It is also true that we cannot be certain since Cohen's kappa is for estimating agreement. However, agreement can be determined for categories (eg., observation categories as in many clinical studies). Accordingly, we have now clarified this issue in the manuscript.</p>
<p>4</p>	<p>4. Data Analysis: It would be useful to cite the WOMBAT Analysis Guide so that readers know what you are referring to: Walter SR, Li L, Westbrook JI (2018) A Guide to the Analysis of Data from the Work Observation Method by Activity Timing (WOMBAT) System. Sydney: Australian Institute of Health Innovation, Macquarie University.</p>	<p>Revised according to reviewer's suggestion</p>