PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	A realist evaluation of UK medical education quality assurance
AUTHORS	Crampton, Paul; Mehdizadeh, Leila; Page, Michael; Knight, Laura; Griffin, Ann

VERSION 1 - REVIEW

REVIEWER	Anna Kalbarczyk Johns Hopkins Bloomberg School of Public Health
REVIEW RETURNED	29-Aug-2019

GENERAL COMMENTS	Thank you for your work on this paper addressing quality assurance in medical education and training. Many components of the paper are strong and I have a series of comments which I hope your team finds helpful: Rationale: It would be helpful to provide information on these different contexts in medical training and what makes some more or less challenging than others. You mention that quality is needed to assure safe training - but there are other reasons as well, right? To ensure students are meeting education/training competencies? To ensure they are prepared to conduct medical practice? Throughout this paper I felt that the trainee component was missing. page 3, approximately line 29, insert "the following" between includes and components From the background it remains unclear what the purpose of the research is. The write-up does not lead the reader clearly to the specific aim. I would have liked to see more description of the QAF and its applications. Methodology: Data collection - The sentence starting with "we posed..." has too many components and is easily read. Please consider revising. For pilot testing - Piloted with whom? More details are needed on the pilot. Were changes made based on the pilot? Why were interviews conducted remotely? Can the authors please provide a rationale? This becomes more clear in the results when you mention people in international settings but this should be mentioned here. Results: I'm not sure that table 1 is needed as is. It may be more useful to present the total counts of QA partners, sectors, location, rather than per interviewee.

	I prefer that quotes all be presented in a consistent manner. Here, some are italicized and shown as separate paragraphs while others are embedded within a paragraph. Personally, I find this approach distracting, wondering why some are presented in one way or another. Page 7, Do you mean that people were skeptical that there was a lack of clarity? Or that they felt there was a lack of clarity? Page 7, line 24: Can you provide specific examples here? The authors mention context and the need for standards to be applicable in different settings. Therefore examples of flexible standards vs. non flexible and how they have actually influenced innovation or created too much variation are needed. Page 8, line 15, revise first sentence for grammar Page 8, institutional visits - Were there any interviewees who felt differently? Did most agree with this sentiment? I would imagine that some might have had something else to say about institutional visits. Page 8, paragraph starting with "the component..": This paragraph was challenging to read. I would recommend the authors think critically bout what quotes should be included - those that say something better than you yourself can say it. Many of these didn't need to be here and muddied takeaways. Overall I felt that the training component was missing - many of the themes and quotes focus on the mechanisms and processes of QA and their various challenges and unintended consequences. However, there's not a clear link between these mechanisms and ensuring safe, consistent training for medical trainees in different settings. Based on my reading of the results and discussion, it would make more sense if the aim were to understand the barriers to quality assurance. The step to what works, how, and why is unclear. However, I think if authors could provide more specific examples throughout the results, particularly for how one approach to QA hinders or supports innovation, learning, training, safety, etc. the connection could be better argued.

REVIEWER	Sivan Spitzer-Shohat Azrieli Faculty of Medicine, Bar Ilan University, Israel
REVIEW RETURNED	18-Sep-2019

GENERAL COMMENTS	The paper describes a realist evaluation carried out to understand the implementation of GMC's medical education QA process has on achieving standarization of quality ultimately protecting patients. The manuscript is clearly written and provides information on the mechanisms driving implementation. However, I have a few concerns: 1. The authors state that prior to participation a 15 min information video was presented to participants (page 4). It is not clear what was presented in this video and whether this may have created bias in the following interview questions. Further clarification is needed. 2. The study aimed to identify what works for WHOM in what situation and why. While the authors differentiate between different context (eg undergraduate versus post graduate education), ultimately their modified program theory (outlined also in figure 4) does not clearly show the different CMO configurations for each context. It was not clear if there are different positive and negative

	mechanisms in the CMO configuration when the context is undergraduate organization versus a trust for example. It may be useful for the reader to have a table showing the different configurations. 3. The researchers state that this study fills a gap on QA efficacy (page 11). However, it is not clear how the authors ascertained the efficacy of this intervention. Rather, they looked at the facilitators and barriers (positive and negative mechanisms) in implementation but not the efficacy of the intervention.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1

Reviewer: Thank you for your work on this paper addressing quality assurance in medical education and training. Many components of the paper are strong and I have a series of comments which I hope your team finds helpful:

Response: Thank you for this positive comment. We have addressed the comments below.

Reviewer: Rationale: It would be helpful to provide information on these different contexts in medical training and what makes some more or less challenging than others. You mention that quality is needed to assure safe training - but there are other reasons as well, right? To ensure students are meeting education/training competencies? To ensure they are prepared to conduct medical practice?

Response: We have broadened the rationale to incorporate the focus on medical students and doctors in training: 'Therefore, the challenge for regulators is to mediate the quality of education and training across these spaces in order to assure the public that education and training is safe, that medical students are prepared for practice and that doctors are fit to practise.'

Reviewer: Throughout this paper I felt that the trainee component was missing.

Response: We have added more about the trainee in the introduction, results and discussion. The postgraduate context is largely about the trainee aspects which we have already discussed throughout the article.

Reviewer: page 3, approximately line 29, insert "the following" between includes and components

Response: Change made.

Reviewer: From the background it remains unclear what the purpose of the research is. The write-up does not lead the reader clearly to the specific aim. I would have liked to see more description of the QAF and its applications.

Response: The specific aim of this research was to explore what components of the GMC's quality assurance framework work, for whom, in what circumstances and how? This is explicitly stated at the end of the background section. Given the word count limit we have detailed the QAF in the introduction but readers are able to follow the references for further information.

Reviewer: Methodology: Data collection - The sentence starting with "we posed..." has too many components and is easily read. Please consider revising.

Response: We have shortened the sentence for clarity. The following sentences have also been clarified to explain what we mean by candidate theories.

Reviewer: For pilot testing - Piloted with whom? More details are needed on the pilot. Were changes made based on the pilot?

Response: We have added information on the piloting. This was completed with a quality assurance manager at our own institution (UCL). Colleagues from our own institution were not included in the formal study to avoid any potential conflict of interest.

Reviewer: Why were interviews conducted remotely? Can the authors please provide a rationale? This becomes more clear in the results when you mention people in international settings but this should be mentioned here.

Response: We have added a sentence in the methods to explain the need for telephone interviews. Participants were geographically dispersed across the world so face-to-face interviews were not feasible.

Reviewer: Results: 'm not sure that table 1 is needed as is. It may be more useful to present the total counts of QA partners, sectors, location, rather than per interviewee.

Response: After consideration we have retained table 1. Given the detailed level of analysis we feel this table provides a valuable insight into the participants demographics. If total counts are provided this would reduce the ability of a keen reader to look closely at the data. As highlighted by the reviewer we have a large number of characteristics including QA partners, sectors, location, etc therefore this richness adds to the trustworthiness of the analysis and findings.

Reviewer: I prefer that quotes all be presented in a consistent manner. Here, some are italicized and shown as separate paragraphs while others are embedded within a paragraph. Personally, I find this approach distracting, wondering why some are presented in one way or another.

Response: Thank you for this comment. We have standardised the approach for ease of reading. We have placed full sentence quotes in separate paragraphs and kept shortened phrases and terms embedded. All are in italics.

Reviewer: Page 7, Do you mean that people were skeptical that there was a lack of clarity? Or that they felt there was a lack of clarity?

Response: We are trying to make the latter point so have clarified this in the text.

Reviewer: Page 7, line 24: Can you provide specific examples here? The authors mention context and the need for standards to be applicable in different settings. Therefore examples of flexible

standards vs. non flexible and how they have actually influenced innovation or created too much variation are needed.

Response: We have added brief exemplars to help clarify the point made in the paragraph. Standards that are overly prescriptive, rigid and inflexible prevent providers from being adaptable to need and innovation. For example standards which focus on particular aspects (e.g. student diversity) may detract attention from other areas of need (e.g. widening participation). Conversely, less binding standards (e.g. not detailing specific teaching methods) triggered mechanisms of ambiguity, openness and flexibility creating too much variation in education across contexts and producing new risks to quality.

Reviewer: Page 8, line 15, revise first sentence for grammar

Response: We have made the sentence structure active.

Reviewer: Page 8, institutional visits - Were there any interviewees who felt differently? Did most agree with this sentiment? I would imagine that some might have had something else to say about institutional visits.

Response: Interviewees were on the whole were in favour of institutional visits and the way in which they functioned. In the section we highlight the characteristics associated with this effectiveness and its limits.

Reviewer: Page 8, paragraph starting with "the component..": This paragraph was challenging to read. I would recommend the authors think critically bout what quotes should be included - those that say something better than you yourself can say it. Many of these didn't need to be here and muddied takeaways.

Response: We have reduced some of the text for clarity.

Reviewer: Overall I felt that the training component was missing - many of the themes and quotes focus on the mechanisms and processes of QA and their various challenges and unintended consequences. However, there's not a clear link between these mechanisms and ensuring safe, consistent training for medical trainees in different settings. Based on my reading of the results and discussion, it would make more sense if the aim were to understand the barriers to quality assurance. The step to what works, how, and why is unclear. However, I think if authors could provide more specific examples throughout the results, particularly for how one approach to QA hinders or supports innovation, learning, training, safety, etc. the connection could be better argued.

Response: We have addressed the reviewers point about the training component as mentioned in earlier comments. Specifically, we have we have added more about the trainee in the introduction, results and discussion. The postgraduate context is largely about the trainee aspects which we have already discussed throughout the article. The reviewer is absolutely right the aim of this paper is to explore the elements of the quality assurance framework used by the General medical Council looking at what works in what context. In amending the paper we have tried to make it clearer about what works and for whom. The aim of our paper is to produce an overarching programme theory for how quality assurance could facilitate or undermine educational environments. The reviewer makes an extremely important point about linking approaches to quality insurance and effects on learning,

training, safety et cetera. We should point out that our stakeholder group contained exceptionally senior medical professionals and did not contain either trainees or their supervisors. Therefore, this study didn't provide the data to examine specifically the link between approaches and the impact on learners, their training or indeed patient safety. Making these connections would be an important aspect to follow up with further research now that the programme theory has been developed.

Reviewer: 2

Please leave your comments for the authors below

The paper describes a realist evaluation carried out to understand the implementation of GMC's medical education QA process has on achieving standarization of quality ultimately protecting patients. The manuscript is clearly written and provides information on the mechanisms driving implementation. However, I have a few concerns:

1. The authors state that prior to participation a 15 min information video was presented to participants (page 4). It is not clear what was presented in this video and whether this may have created bias in the following interview questions. Further clarification is needed.

Response: We agree with the reviewer that information about the video would be helpful however due to the word count limit we had to reduce the description. This video was produced from material publically available on the GMC website. In accordance with realist approaches, the video helps to explain the intended consequences of the intervention design. Bias therefore is not applicable in the sense that it influences the interview questions, as the interview guide is in effect testing whether the assumptions from the intervention provider hold true. Hence in our methods we explain that we asked questions to provoke agreement and disagreement in relation to the guiding principles (in accordance with realist interviews).

Reviewer: 2. The study aimed to identify what works for WHOM in what situation and why. While the authors differentiate between different context (eg undergraduate versus post graduate education), ultimately their modified program theory (outlined also in figure 4) does not clearly show the different CMO configurations for each context. It was not clear if there are different positive and negative mechanisms in the CMO configuration when the context is undergraduate organization versus a trust for example. It may be useful for the reader to have a table showing the different configurations.

Response: The programme theory is designed to show the overarching flow of processes which subtly include the various contexts, mechanisms and outcomes. To incorporate the various CMO configurations would further muddy the waters in terms of a reader being able to draw out the key messages from the study. Whilst it may be helpful to provide those interested in specific areas it will negate the overall context in which the GMC's QAF operates. Essentially the various undergraduate, postgraduate contexts are operating simultaneously therefore to delineate any one in particular would lack the understanding of the overall operationalisation.

Reviewer:

3. The researchers state that this study fills a gap on QA efficacy (page 11). However, it is not clear how the authors ascertained the efficacy of this intervention. Rather, they looked at the facilitators and barriers (positive and negative mechanisms) in implementation but not the efficacy of the intervention.

Response: Thank you for this comment. We agree and have re-phrased the term to functionality for clarity. Given the number of factors involved with how the QAF operates it is more important to understand the way it functions rather than judging efficacy.

VERSION 2 – REVIEW

REVIEWER	Anna Kalbarczyk Johns Hopkins Bloomberg School of Public Health, USA
REVIEW RETURNED	26-Oct-2019

GENERAL COMMENTS	My original comments were all adequately addressed in the revision.

REVIEWER	Sivan Spitzer-Shohat Bar-Ilan University, Israel
REVIEW RETURNED	18-Nov-2019

GENERAL COMMENTS	The authors claim that context is a critical factor: In the results section the authors write 'We found that depending on context, the same interventions triggered a range of mechanisms leading ti positive or negative outcomes. (page 36 of 44). Similarly, in the opening of the discussion the authors state: "We found that intervention components support or undermine QA for different organizations, and at different times in undergraduate and postgraduate contexts" (page 39 of 44). Yet in their modified program theory (figure 4) they fail to show the effect of contexts on the CMO's. In their response to the review, the explanation to the lack of acknowledgment to different CMO's is that it would 'muddy the waters' for the reader. I remain unconvinced. If the aim is to show the effect of contexts on the program theory, one cannot assume a unified program theory for different contexts without accounting for their differences and how these trigger a unique set of positive and negative mechanisms, and in turn, positive and negative outcomes. I find figure 4 to be a 'catch all' and not a conceptual model that helps the reader understand the linkages between C-M-O.

VERSION 2 – AUTHOR RESPONSE

Editorial requests/reviewer comments	Response
Reviewer(s)' Comments to Author: Reviewer: 1 Reviewer Name Anna Kalbarczyk Institution and Country Johns Hopkins Bloomberg School of Public Health, USA	Thank you to the reviewer for providing this review. We are pleased to have addressed your comments with the revisions.

Please state any competing interests or state 'None declared': None declared Please leave your comments for the authors below My original comments were all adequately addressed in the revision.	
Reviewer: 2 Reviewer Name Sivan Spitzer-Shohat Institution and Country Bar-Ilan University, Israel Please state any competing interests or state 'None declared': None Please leave your comments for the authors below The authors claim that context is a critical factor: In the results section the authors write 'We found that depending on context, the same interventions triggered a range of mechanisms leading to positive or negative outcomes. (page 36 of 44). Similarly, in the opening of the discussion the authors state: "We found that intervention components support or undermine QA for different organizations, and at different times in undergraduate and postgraduate contexts" (page 39 of 44). Yet in their modified program theory (figure 4) they fail to show the effect of contexts on the CMO's. In their response to the review, the explanation to the lack of acknowledgment to different CMO's is that it would 'muddy the waters' for the reader. I remain unconvinced. If the aim is to show the effect of contexts on the program theory, one cannot assume a unified program theory for different contexts without accounting for their differences and how these trigger a unique set of positive and negative mechanisms, and in turn, positive and negative outcomes. I find figure 4 to be a 'catch all' and not a conceptual model that helps the reader understand the linkages between C-M-O.	The reviewer is absolutely right. The research demonstrated that context was a critical factor determining how QA components worked. The modified programme theory contains a rich amount of detail to understand the numerous ways in which contexts can lead to a range of outcomes. Through revealing such insights policymakers will be in an enlighted position to think critically about the ways in which to enhance quality assurance processes. In our study we had a broad range of contexts as well as a broad range of components making "generalisable findings" highly problematic. However, we are grateful for the reviewers urge to create a more conceptual model. We have sought to address this comment in numerous ways. Firstly, we have removed the sentence ('We found that depending on context, the same interventions triggered a range of mechanisms leading to positive or negative outcomes) in the results section and explained at a later stage so that readers can grasp the detail first before learning about the key finding. Secondly, we have amended figure 4 so that it incorporates a greater visualisation of the contexts in which positive and negative outcomes and unintended consequences were present. We are cautious not to overgeneralise the contextual features of the undergraduate and postgraduate contexts. Therefore, the linkages between CMO's were understood through the concept of contexts with associative and dissociative features. Those contexts that broadly aligned to the General Medical Council's QAF, or its individual components

	were conceptualised as associative contexts - in which case they demonstrated adherence to the General Medical Council's framework resulting in triggering positive mechanisms and outcomes. Contexts with dissociative features were contexts that were unable or unwilling to integrate the General Medical Council's QAF into their approach to QA – and therefore triggered differing mechanisms and outcomes. Thirdly, we have added further description on the modified programme theory to help readers further understand the links.

VERSION 3 - REVIEW

REVIEWER	Sivan Spitzer Shohat Department of Population Health, Azrieli Faculty of Medicine, Bar-Ilan University
REVIEW RETURNED	04-Dec-2019

GENERAL COMMENTS	The authors have adequately addressed my comments in this revision.