PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	How capacity building of district health managers has been designed, delivered and evaluated in sub-Saharan Africa: a scoping review and best fit framework analysis
AUTHORS	Bosongo, Samuel; Belrhiti, Zakaria; Ekofo, Joël; Kabanga, Chrispin; Chenge, Faustin; Criel, Bart; Marchal, Bruno

VERSION 1 – REVIEW

REVIEWER	Nkeshimana, Menelas University of Rwanda
REVIEW RETURNED	08-Feb-2023

GENERAL COMMENTS	I read with a great interest this paper which is touching on one of the most under explored topics in the Sub-Saharan Africa. A lot is done, and a lot still needs to be measured in terms of impact. It is unfortunate that few articles explored the health outcomes, but I am confident that the gaps highlighted by this paper will be timely addressed and sealed. The learning method is shifting from a seating classroom to the virtual platforms (online training) which could not be explored in this study. This limitation should be also documented. Just a small correction on the line 38 of the page 7: I believe the articles included in this review were published from 1991 to 2022 instead of 1991 to 2021 (there are 2 articles that were published in 2022). Allow me to congratulate the authors for this great review. Thank you.

REVIEWER	Geerts, Jaason Canadian College of Health Leaders, Research and Leadership Development
REVIEW RETURNED	07-Mar-2023

GENERAL COMMENTS	Peer Review for BMJ Open – 2022 – 071344 By Dr. Jaason Geerts, Mar 6, 2023 My recommendation is to accept it for publication... with some important modifications and a few minor ones. Overall Thank-you for the opportunity to review this paper. Overall, the review is excellent and the writing is very good. I think this definitely should be published, but needs modification to enhance its potential benefit to research and practice. I have established expertise in this area (scholarly and professionally) and my recommendation is accept with minor edits (which are reviewed again before publishing, either by peer reviewers or by the editor) and I have no conflicts to declare.

	High-level comments - Great introduction: well organized and well written - Excellent search strategy and supporting materials - Good to have reported programs with limited effects too - Great lessons learnt section, as well as the success factors section - Overall, the discussion section is excellent Major suggestions - Reported outcomes: there is no mention of program outcomes for the first many pages, including in your objectives or research questions. While you can't change your research questions at this point, outcomes should be made central to the argument, since they are the mark of success in this endeavour - Theory vs. evidence: a key premise of your argument is that few programs are designed based on theory. Without dismissing this consideration, you leave out altogether the notion that programs should be based on established evidence of program designs that have been linked to outcomes. Theory without discussing evidence could be perceived as contributing to the theory/practice divide and theory can fail to describe the specifics of how to translate it into programming - Lack of quality assessment: while this may be common in scoping reviews, it is a significant limitation of the findings, especially in a field where the vast majority of published literature is of anecdotal quality. Not separating reliable evidence methodologically from anecdotal testimonials can exacerbate the problem. This can be supplemented by referencing relevant systematic reviews that did assess for quality (two are suggested here), especially since they reached several of the same conclusions regarding evidence-supported aspects of interventions. The former also includes recommendations for robust evaluation framework, which is presented as having potential to contribute to future research and to providers. - Suppl. File 4: while this is helpful to include, there are key details missing, which should be included. These include the study design for several articles, program activities/methods, number of participants, and reported outcomes (by type). - Rationale section, motivation subsection, sentence starting with "Most programs assumed that strengthening…": This is the inverse order. When individual skills, knowledge, behaviours, practices, etc. are strengthen, then one hopes that capacity increases, not the inverse. - Use of "sound" versus evidence-based/informed? - Levels, modes, and approaches (sentences 1 and 2) – are you suggesting that no programs focused on individual and organizational focuses? That seems unlikely. Minor comments - Sample population details: while you do describe them, it would be helpful for readers to get a sense of the range of the sizes of their organizations (e.g., are they in charge of 10,000 people, 25, etc.?) and a rough estimate of their level of leadership (e.g., comparable to executives of a major academic healthcare centre, or much smaller?). Number of staff, number of beds, number of sites, etc. are some common measures that would give readers a clearer sense of the sample population (including their range). - Search terms: you include your search terms in the supplementary materials, but it might be good to include them for the sake of transparency in the text. - Consistency in terms: in the abstract, you refer to "conceptualized, operationalized, and evaluated", and later you

	refer to “designed, delivered, and evaluated” (terms more familiar to me) and a few variations. It might be good to remain consistent - Line 22 of the introduction (first sentence of the third paragraph): you mention the “systemic level”, but I don’t recall you using this term throughout (only individual and organizational). Delete or include/address throughout - Line 54 of the introduction (“Most of these evaluations…”): a good deal of the literature is based on post-test only and many programs are not evaluated at all - Just before Step 3 in the bullet related to “Be”: no need to capitalize management and leadership. For models of theories (MoTh): clarify “of CBPs” - Step 3 line 18 in the sentence “Disagreements between”: you mean “among” not “between”, since there are more than two - Step 4 first sentence – state the number of selected models (2) - Table 2 – “Actors” does not distinguish between the providers and the faculty/instructors, who are often different - Step 5 first sentence – Described where? - Results paragraph one: according to your review, there is basically only one model of evaluation. What does this tell you? - Rationale for conducting… Motivation… sentence 1: “weak” determined by whom? - Line 55 of the same section: what does “profile” mean in this context? - Line 57 of the same section: (2) inadequate “efficacy of” leadership and management courses? - Health outcomes: confused as to why the “quality of care outcomes” seem separate from “health outcomes” - Lessons learnt (sentence 1): Unclear whether these were reported by the included studies’ authors or if they were deduced by the scoping review authors - Accountability: this is a very important concept. It would be helpful to provide more information on how accountability was included in the programmes - Success factors section: “(2) DHMs to actively participate…” this draws on the principles of adult learning (might ref. here). The how part of (4) is not clear - Discussion (para 2): “Therefore, while designing…” you might consider adding “and evidence” to the end of the sentence. Theory alone can be insufficient. - Discussion (para 2, last sentence): how does it provide a framework for evaluation? (again, the Social Science & Medicine paper above includes details of an evaluation framework) - Discussion (para 3): this is where the summary of the best available evidence from the SS&M article aligns nicely with your findings (after having assessed included studies for quality) - Discussion (para 3): action learning also enables participants to benefit from faculty support after having attempted to apply their learning (versus programs where application is expected after program completion) - Discussion para beginning with “this review highlighted the diversity”: yes, it’s possible that Kwamie is right. It’s also possible that they lack basic management skills/knowledge if they have had little training in the area before - Discussion: again, “basing them on explicit theories” and evidence

REVIEWER	Vincent, Charles University of Oxford, Experimental Psychology

REVIEW RETURNED	17-Mar-2023

GENERAL COMMENTS	This paper addresses the important issue of building management and leadership capacity in health systems, in particular sub-Saharan Africa. The authors present a clear rationale for their study, pointing out the critical important of leadership and management in both the routine operation and improvement of health systems. The authors have carried out an impressively thorough and detailed review and clearly a great deal of work has been carried out. However, the huge amount of findings generated and the dense presentation pose challenges for the reader. General comments • The core research questions address the conceptualisation, operationalisation and evaluation of capacity building programmes (CBPs). The authors are particularly concerned with identifying the role of models and theories in CBPs, but it was not quite clear to me how this influenced the inclusion and exclusion criteria. Curiously examining the content of programmes does not appear to be a core research question, although the results contain a short section on content. • The scoping review has been carefully conducted following standard guidelines, is clearly described. The authors use a combination of generating an initial priori framework as a starting point which seems a very useful approach. • The results contain a huge amount of material, densely presented mostly in text form, giving a good overview but not allowing any deeper understanding of the key issues. The authors summarise the main findings effectively in the discussion, but it is nevertheless difficult to identify the key messages. As a reader I was asking 'so, what are the key messages for someone who wants to improve leadership and management? What would I do if I was designing a programme?'. I suspect the authors could give a valuable answer to this critical question, but it is not contained in the paper. Recommendations • I suggest the authors give serious consideration to splitting this manuscript into two papers, the first on theory, content and mode of implementation, the second on evaluation and outcome. There is so much material and I do not believe a single paper can do it justice. I think this would be relatively simple to do. This would also allow a lot more explanation of the results with much more explanation of modes of delivery, content, barriers and so on. At the moment these are mainly lists in the text and it is hard to get a sense of what is truly important. • Splitting into two papers would allow an expansion and illustration of key issues, perhaps with examples of programmes, table of key themes, consideration of what the practical implications are for people building capacity. • Whether or not the paper is split, I strongly suggest a much greater use of boxes and tables of findings, rather than long lists in the text. This would just be much easier to follow and produce a stronger and more influential paper. Examples of good practice would also be helpful. • The authors should offer more interpretation and summary of the findings to help the reader understand the strengths and limitations of the evidence, what the key findings are and what practically might be done to enhance capacity. I suspect the authors have

	much to say on these points and I would welcome a stronger sense of their voice and perspectives. Specific points to address • The authors explain that capacity building is conceptualised in different ways. It would be helpful if they explained the working definition they adopted for the searches and the review. • Should the research questions include one on content of programmes? This has clearly been examined and seems a rather critical issue. • I found it strange that the coding framework did not address the content of the programmes. Surely this should be included? • I did not understand the exclusion of non-theoretical/technical models. This seems odd when the results on p8 specifically say that none of the papers included a theoretical model, which also does not seem to accord with the selection criteria (p5) which has theory as an inclusion criterion. Please explain all this more fully. • Identification of frameworks (p5). There are three criteria. Did papers have to meet all of these criteria or just one or two of them?

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
Dr. Menelas Nkeshimana, University of Rwanda
Comments to the Author:
I read with a great interest this paper which is touching on one of the most under explored topics in the Sub-Saharan Africa. A lot is done, and a lot still needs to be measured in terms of impact. It is unfortunate that few articles explored the health outcomes, but I am confident that the gaps highlighted by this paper will be timely addressed and sealed.

The learning method is shifting from a seating classroom to the virtual platforms (online training) which could not be explored in this study. This limitation should be also documented.
Thanks. We have amended our previous manuscript version and added the following phrase in the limitations sub-section: *"Third, the fact that none of the included papers have reported on online learning CBP, particularly in the digital and Covid-19 era may be a limitation of this review."* (Revised manuscript, page 19 lines 7-8)

Just a small correction on the line 38 of the page 7: I believe the articles included in this review were published from 1991 to 2022 instead of 1991 to 2021 (there are 2 articles that were published in 2022).
Following a comment from reviewer #2, we have amended our previous manuscript version and replaced the text with Table 5 (untitled characteristics of included papers), where the year 2022 is clearly mentioned (Revised manuscript, page 8 line 34)

Allow me to congratulate the authors for this great review. Thank you.
Thank you for your insightful comment that have increased the quality of our reporting.

Reviewer: 2
Dr. Jaason Geerts, Canadian College of Health Leaders, Cass Business School

Comments to the Author:
Peer Review for BMJ Open – 2022 – 071344
By Dr. Jaason Geerts, Mar 6, 2023
My recommendation is to accept it for publication... with some important modifications and a few minor ones.

Overall
Thank-you for the opportunity to review this paper. Overall, the review is excellent and the writing is very good. I think this definitely should be published, but needs modification to enhance its potential benefit to research and practice. I have established expertise in this area (scholarly and professionally) and my recommendation is accept with minor edits (which are reviewed again before publishing, either by peer reviewers or by the editor) and I have no conflicts to declare.

High-level comments
- Great introduction: well organized and well written
- Excellent search strategy and supporting materials
- Good to have reported programs with limited effects too
- Great lessons learnt section, as well as the success factors section
- Overall, the discussion section is excellent
We thank the reviewer for this appreciation.

Major suggestions
- Reported outcomes: there is no mention of program outcomes for the first many pages, including in your objectives or research questions. While you can't change your research questions at this point, outcomes should be made central to the argument, since they are the mark of success in this endeavour.
Thanks for this comment. To address the reviewer's concern, we first refer to the several mentions of outcomes: (1) performance and organizational performance and effectiveness in page 3 line 42, (2) outcomes of CBPs in table 4 page 7, (3) expected outcomes of CBP in page 9 line 25 to 31 and (4) reported outcomes in table 10, pages 14 and 15. Second, we have added the programme outcomes in the third research question as follows: *"(3) How have such CBPs been evaluated and what were the outcomes?"* (see page 4 lines 19 and 20)

- Theory vs. evidence: a key premise of your argument is that few programs are designed based on theory. Without dismissing this consideration, you leave out altogether the notion that programs should be based on established evidence of program designs that have been linked to outcomes. Theory without discussing evidence could be perceived as contributing to the theory/practice divide and theory can fail to describe the specifics of how to translate it into programming.
Thanks for this comment. We agree that theory and evidence must go together and reinforce each other. Indeed, programmes must be based on evidence-informed theories. Theories, in turn, need to be empirically tested to generate more established evidence and become more advanced theories. To address the reviewer's concern, we have added the word "**evidence'** in the following sentences of the introduction section: *We focused on identifying the underlying assumptions and evidence behind CBPs at the district level* (page 4 line 1).

- Lack of quality assessment: while this may be common in scoping reviews, it is a significant limitation of the findings, especially in a field where the vast majority of published literature is of anecdotal quality. Not separating reliable evidence methodologically from anecdotal testimonials can exacerbate the problem. This can be supplemented by referencing relevant systematic reviews that

did assess for quality (two are suggested here), especially since they reached several of the same conclusions regarding evidence-supported aspects of interventions. The former also includes recommendations for robust evaluation framework, which is presented as having potential to contribute to future research and to providers.
Thanks for this comment. We acknowledge that quality appraisal is yet not required in scoping reviews (H. Arksey & L. O'Malley, 2005). Qualitative appraisal in scoping reviews, at the contrary of systematic reviews, is often motivated by the need to interpret various forms of evidence from heterogenous study designs with a broader coverage of literature than in conventional systematic review. Yet, there is still little consensus on the need for quality appraisal in interpretative and theory driven synthesis (Dixon-Woods, 2004; Pawson, 2002) Some scholars argues that even poor quality papers may bring "nuggets of wisdom". (Dixon-Woods, 2004; Sandelowski et al., 1997).

- Suppl. File 4: while this is helpful to include, there are key details missing, which should be included. These include the study design for several articles, program activities/methods, number of participants, and reported outcomes (by type).
Thanks for this suggestion. We have amended supplementary file 4 of our previous version of the manuscript by adding two new columns related to the *"CBP approaches"* and *"Reported Outcomes"*. We also completed some missing details where data were available. (see Suppl. File #3 of revised manuscript)

- Rationale section, motivation subsection, sentence starting with "Most programs assumed that strengthening…": This is the inverse order. When individual skills, knowledge, behaviours, practices, etc. are strengthen, then one hopes that capacity increases, not the inverse.
Thanks for this suggestion. We have addressed the reviewer's concern and have amended our previous manuscript version as follows: "*Most programmes assumed that strengthening the leadership and/or management knowledge, skills, and practices of health managers would increase their leadership and/or management capacities".* (page 9, line 13-15)

- Use of "sound" versus evidence-based/informed?
Thanks for this suggestion. The adjective *"sound"* in our previous manuscript version means *"good".* After carefully considering this question, we realized that "*sound*" may be confusing. We have replaced it with "*good*" to avoid misunderstanding and address the reviewer's concern. (page 9, lines 18, 19, and 22)

- Levels, modes, and approaches (sentences 1 and 2) – are you suggesting that no programs focused on individual and organizational focuses? That seems unlikely.
Thanks for this question. After carefully rereading our manuscript, we realized that the individual and organizational levels were the two main entry points of CBPs reported in our review. To avoid any misunderstandings and to address the reviewer's concern, we have modified the statement as follows: *"We found that CBPs reported in the included papers of this review had two entry points: the individual and organisational levels. Nine CBPs focused on strengthening individual health managers' knowledge and skills.*[17 67 68 72 74 77 86 93 100] *The remaining CBPs took an organisational entry point to strengthen the capacity of the health management teams to perform their managerial functions and achieve health outcomes."* (page 11, line 5-9)

Minor comments
- Sample population details: while you do describe them, it would be helpful for readers to get a sense of the range of the sizes of their organizations (e.g., are they in charge of 10,000 people, 25, etc.?) and a rough estimate of their level of leadership (e.g., comparable to executives of a major academic healthcare centre, or much smaller?). Number of staff, number of beds, number of sites, etc. are some common measures that would give readers a clearer sense of the sample population (including their range).

Thanks for this suggestion. In supplementary file 4 of our previous manuscript, related to the description of included studies, we have added some details of the number of participants where available (see Suppl. File #3 of revised manuscript). However, this helpful information is only available in some included papers.

- Search terms: you include your search terms in the supplementary materials, but it might be good to include them for the sake of transparency in the text.
Thanks for this suggestion. We have moved the both search strategies (for primary studies and for models, theories and framework) from the supplementary file 2 into the main text (see page 5, table 1 line 1 and table 2 line 11).

- Consistency in terms: in the abstract, you refer to "conceptualized, operationalized, and evaluated", and later you refer to "designed, delivered, and evaluated" (terms more familiar to me) and a few variations. It might be good to remain consistent
Thanks for this suggestion. We have amended our previous manuscript version and kept consistency by replacing the word "conceptualized" by *"designed"*, and "operationalized" by "*delivered*" in the title, abstract and elsewhere throughout the revised version of our manuscript. (page 3 line 47, page 4 lines 18 and 19)

- Line 22 of the introduction (first sentence of the third paragraph): you mention the "systemic level", but I don't recall you using this term throughout (only individual and organizational). Delete or include/address throughout
Thanks for this suggestion. We did not use the term "systemic level" throughout our manuscript because we did not come across papers dealing with capacity building at the systemic level during our scoping review.

- Line 54 of the introduction ("Most of these evaluations…"): a good deal of the literature is based on post-test only and many programs are not evaluated at all
Thanks for this suggestion. We modified it as follows: *"A good deal of the literature of CBP evaluation is based on pre- and post-test only and many programs are not evaluated at all."* (see page 3, line 42 to 44)

- Just before Step 3 in the bullet related to "Be": no need to capitalize management and leadership. For models of theories (MoTh): clarify "of CBPs"
Thanks for this suggestion. We have modified the text accordingly. (page 5, line 6-9)

- Step 3 line 18 in the sentence "Disagreements between": you mean "among" not "between", since there are more than two
Thanks for this suggestion. We have replaced "*between*" by "*among*". (page 6, line 8)

- Step 4 first sentence – state the number of selected models (2)
Thanks for this suggestion. We have added the number two in the first sentence as follow: "*Based on the **two** selected models,[63] [64] we generated a list of a priori themes and codes related to the rationale, process (strategies, implementation, and evaluation), and outcomes of CBPs"* (page 7, line 3).

- Table 2 – "Actors" does not distinguish between the providers and the faculty/instructors, who are often different
Thanks for this remark. By actors, we mean people engaged in the CBP processes. They are the participants (who receive the CBP) and the providers or facilitators (who provide the CBP). In the sub-section "Implementation of capacity building programmes", we describe different categories of institutions from which were the facilitators as follows: "*The programmes were provided by facilitators from the Ministry of Health at the national, regional or district level,[4] [49] [67] [75] [76] [81] [91] [97] [99] [102] [104] academic*

and research institutions,[9 68 74 77 80 83 86 88 89 96] international non-governmental organisations, [79 93] or a mix of these institutions.[17 70 82 87 90 94 95]" (page 13, line 2 to 5)

- Step 5 first sentence – Described where?
Thanks for this question. We described the main characteristics of the included studies using descriptive statistics (number and percentage) by texts in our previous manuscript). However, we have replaced the text with the table 5 for more clarity according to the recommendation of the reviewer #3 (see page 8, line 34).

- Results paragraph one: according to your review, there is basically only one model of evaluation. What does this tell you?
Thanks for this question. Using the BeHEMoTh approach for searching theories, models or frameworks related to CBPs, we came across only two similar models. This scarcity of theories, models or frameworks of CBP in health reinforces our recommendation for explicit theories underpinning CBPs and theory-driven evaluations.

- Rationale for conducting… Motivation… sentence 1: "weak" determined by whom?
Thanks for this question. We acknowledge that this phrase was a bit less understandable. We have replaced it by the following: "*A good deal of the literature included in this review have reported weak leadership and/or management capacities of DHMs as the most frequent reason for conducting CBPs."* (page 9 lines 4 and 5)

- Line 55 of the same section: what does "profile" mean in this context?
Thanks for this question. "Profile" in this context means the description or (a set of) attributes or features of DHMs. For more clarity, we have added *"professional profile"* (see page 4 line 31; page 7 table 4 line 7; page 9 line 9)

- Line 57 of the same section: (2) inadequate "efficacy of" leadership and management courses?
Thanks for this suggestion that makes the phrase more understandable. We have added *"efficacy of"* between inadequate and leadership. (see page 9 line 11, page 18 line 4)

- Health outcomes: confused as to why the "quality of care outcomes" seem separate from "health outcomes"
Thanks for this remark. After carefully considering this remark and rereading our manuscript, we realized that we have not expressed ourselves clearly enough in our text. To avoid any misunderstandings and following the recommendation of the reviewer #3, we have replaced text with Table 10, where quality of care is included in the health outcomes (page 14 line 17 table 10).

- Lessons learnt (sentence 1): Unclear whether these were reported by the included studies' authors or if they were deduced by the scoping review authors
Thanks for this remark. The lessons learnt are those reported by the included studies' authors. To make the sentence clearer, we have reformulated it as follow: *"Lessons learnt from CBPs reported in the included papers of this review are …"* (Page 16, line 21).

- Accountability: this is a very important concept. It would be helpful to provide more information on how accountability was included in the programmes
Thanks for this remark. Indeed, we have struggled to keep within the word limit required by the journal (and this is a challenging task). For the reviewers' information, the CODES project, which aimed at strengthening district-based health systems management in Uganda, involved the community through dialogue based on Citizen Report Cards and U-reports as a feedback mechanism. This community engagement improved care-seeking and social accountability from the service providers (See Katahoire et al, 2015).

- Success factors section: “(2) DHMs to actively participate…” this draws on the principles of adult learning (might ref. here). The how part of (4) is not clear
Thanks for these remarks. We have added a sentence related to the adult learning *“Such interactions require facilitators to have good relational skills, which are central in the adult learning process (Mezirow J., 1993)”* (page 17, line 4 and 5). We have also clarified the how part of (4) as follow: *“Adaptability and flexibility of CBP processes make them more responsive as they consider the needs of DHMs and their context, which contributes to increased perceived relevance and sense of ownership by DHMs”* (page 17, line 6-8).

- Discussion (para 2): “Therefore, while designing…” you might consider adding “and evidence” to the end of the sentence. Theory alone can be insufficient.
Thanks for this suggestion. We have added the term “evidence” in the sentence: *“Therefore, while designing a CBP, it is good to make explicit the theoretical assumptions and* ***evidence*** *explaining the pathway to the expected outcomes.”* (Page 17, line 29)

- Discussion (para 2, last sentence): how does it provide a framework for evaluation? (again, the Social Science & Medicine paper above includes details of an evaluation framework)
Thanks for this question. According to Pawson and Tilley, programmes are theory incarnate. The programme theory is a set of assumptions explaining how an intervention is supposed to bring about the expected outcomes within a given context. For more clarity, we have replaced the previous sentence by the following: “*Making this programme theory explicit allows for a better understanding of the programme functioning by different stakeholders and will facilitate its evaluation”.* (Page 17, lines 30 and 31*)*

- Discussion (para 3): this is where the summary of the best available evidence from the SS&M article aligns nicely with your findings (after having assessed included studies for quality)
Thanks for this appreciation.

- Discussion (para 3): action learning also enables participants to benefit from faculty support after having attempted to apply their learning (versus programs where application is expected after program completion)
Thanks for this suggestion. We have added the suggested sentence in the paragraph 3: “*It (action learning) also enables participants to benefit from faculty or supervisor support after having attempted to apply their learning.”* (page 17, lines 38 and 39)

- Discussion para beginning with “this review highlighted the diversity”: yes, it’s possible that Kwamie is right. It’s also possible that they lack basic management skills/knowledge if they have had little training in the area before
Thanks for this suggestion. We have added the suggested sentence in the paragraph 5 of discussion section: “*It is also possible that the focus on management is because most DHMs are clinicians who need more basic management knowledge and skills since they have had little training in the area before. In any case, the content of CBPs for DHMs must consider the balance...”* (page 18, line 24-26)

- Discussion: again, “basing them on explicit theories” and evidence
Thanks for this suggestion. We have added the term “evidence” in the sentence: *“…who call for strengthening CBP evaluations by basing them on explicit theories and* ***evidence*** *that describe how a CBP is supposed to lead to expected outcomes.”* (Page 18, line 35)

Reviewer: 3
Dr. Charles Vincent, University of Oxford

Comments to the Author:
This paper addresses the important issue of building management and leadership capacity in health systems, in particular sub-Saharan Africa. The authors present a clear rationale for their study, pointing out the critical important of leadership and management in both the routine operation and improvement of health systems. The authors have carried out an impressively thorough and detailed review and clearly a great deal of work has been carried out. However, the huge amount of findings generated and the dense presentation pose challenges for the reader.
Thanks for this appreciation.

General comments
• The core research questions address the conceptualisation, operationalisation and evaluation of capacity building programmes (CBPs). The authors are particularly concerned with identifying the role of models and theories in CBPs, but it was not quite clear to me how this influenced the inclusion and exclusion criteria. Curiously examining the content of programmes does not appear to be a core research question, although the results contain a short section on content.
Thanks for this comment. We have done two parallel literature searches: (1) the search of primary studies using PCC approach, and (2) the search of theories, models or frameworks using BeHEMoth approach (see figure 1, page 4 line 16). The search of theories, models and frameworks is an element of the best fit framework strategy, which allows a systematic data extraction (Carroll et al.,2013; Carroll et al.,2011), but does not influence the inclusion/exclusion of primary research papers that is driven by our research question. Content is one of the components of evaluation capacity building model used as the coding framework in our review (see table 4, page 7 line 7).

• The scoping review has been carefully conducted following standard guidelines, is clearly described. The authors use a combination of generating an initial priori framework as a starting point which seems a very useful approach.
Thanks for this appreciation.

• The results contain a huge amount of material, densely presented mostly in text form, giving a good overview but not allowing any deeper understanding of the key issues. The authors summarise the main findings effectively in the discussion, but it is nevertheless difficult to identify the key messages. As a reader I was asking ‘so, what are the key messages for someone who wants to improve leadership and management? What would I do if I was designing a programme?’. I suspect the authors could give a valuable answer to this critical question, but it is not contained in the paper.
Thanks for these questions. We agree with the reviewer about the density of our findings, which may prevent a deeper understanding of our key messages. Indeed, we have struggled to keep within the word limit required by the journal (and this is a challenging task). To address this important reviewer concern, we have amended our previous manuscript version and added the box #3 which summarizes the key implications of our review for practice and research (see page 18, line 44).

Recommendations

• I suggest the authors give serious consideration to splitting this manuscript into two papers, the first on theory, content and mode of implementation, the second on evaluation and outcome. There is so much material and I do not believe a single paper can do it justice. I think this would be relatively simple to do. This would also allow a lot more explanation of the results with much more explanation of modes of delivery, content, barriers and so on. At the moment these are mainly lists in the text and it is hard to get a sense of what is truly important.

• Splitting into two papers would allow an expansion and illustration of key issues, perhaps with examples of programmes, table of key themes, consideration of what the practical implications are for people building capacity.
• Whether or not the paper is split, I strongly suggest a much greater use of boxes and tables of findings, rather than long lists in the text. This would just be much easier to follow and produce a stronger and more influential paper. Examples of good practice would also be helpful.
• The authors should offer more interpretation and summary of the findings to help the reader understand the strengths and limitations of the evidence, what the key findings are and what practically might be done to enhance capacity. I suspect the authors have much to say on these points and I would welcome a stronger sense of their voice and perspectives.

We are particularly grateful for these recommendations from reviewer #3. They are really relevant. After discussion with the research team, we have concluded that splitting the paper may be challenging in due time required by the Editor. However, we have addressed the reviewer's recommendations by using boxes and tables of findings. Once more, we are thankful for receiving this valuable idea from the reviewer.

Specific points to address

• The authors explain that capacity building is conceptualised in different ways. It would be helpful if they explained the working definition they adopted for the searches and the review.
Thanks for this observation. We agree with the reviewer the necessity of making explicit the concept "capacity building", the main concept of our review. To address the reviewer's concern, we have added our working definition of capacity building in the methods section, step 2 as follows: *The main concept is "capacity building", i.e., any programme or intervention whose aim is to enable an individual or organisation to achieve its stated objectives.*[75] *CBP comprises both hard or measurable (e.g., knowledge and skills, organisational structure, procedures and resources, etc.) and soft or intangible (e.g., leadership, motivation and organisational culture) component".* (see page 4, line 35 to 39)

• Should the research questions include one on content of programmes? This has clearly been examined and seems a rather critical issue.
Thanks for this observation. We agree with the reviewer that the content of CBPs is one of the critical issues in this review. To address the reviewer's concern, we have rephrased our research question as follows: *(1) "How has capacity building of DHMs in sub-Saharan Africa been designed in terms of theory, mode, level, approach and contents."* (see page 4 line 17-18)
• I found it strange that the coding framework did not address the content of the programmes. Surely this should be included?
Thanks for this remark. After carefully rereading our coding framework, we found that the code *"content"* is included under the theme *"strategies of CBPs"*. (see page 7, table 4 and line 7)

• I did not understand the exclusion of non-theoretical/technical models. This seems odd when the results on p8 specifically say that none of the papers included a theoretical model, which also does not seem to accord with the selection criteria (p5) which has theory as an inclusion criterion. Please explain all this more fully.
Thanks for this observation. According to Carroll et al. (2013), the term "non theoretical/ technical" models refer to terms often used in biomedical research such as "epidemiological model", "disease model", "care model" or "statistical model" that do not fit the theoretical focus of the best fit framework strategy. For more details on the methodology please refer to Carroll et al. (2013). As stated in the first general comment of the reviewer #3, the best fit framework methodology allowed us to identify a coding framework for systematic data extraction, but does not influence the inclusion/exclusion of primary research papers that is driven by our research question. Furthermore, we have added the box

#1 where we have defined theory, model and framework from Bergeron et al. (2017) inspired by Nilsen (2015). (see page 6, line 18)

• Identification of frameworks (p5). There are three criteria. Did papers have to meet all of these criteria or just one or two of them?
Thanks for this question. The two first criteria were critical. To be included, papers had to meet one of the two first criteria and the third criteria related to the language.

Reviewer: 1
Competing interests of Reviewer: None.

Reviewer: 2
Competing interests of Reviewer: None

Reviewer: 3
Competing interests of Reviewer: None

VERSION 2 – REVIEW

REVIEWER	Geerts, Jaason Canadian College of Health Leaders, Research and Leadership Development
REVIEW RETURNED	11-Jun-2023

GENERAL COMMENTS	Dear authors, Thank-you for your resubmission. Given that this is a resubmission, following extensive first-round peer-review suggestions, I have focused on your point-for-point response to reviewers document. I repeat my original comment that the quality of your original review and writing of the manuscript was high. Overall, you have done an excellent job incorporating our suggestions. I am recommending that the most recent version be accepted. Especially since the two articles I mentioned in the first review were not provided to you, I have proposed considerations for you, which you are welcome to take or leave (and I am not proposing that any modifications are required for publication). 1) Returning to the issue of quality of study assessment in scoping reviews. You argue that it isn't required in scoping reviews and that "nuggets of wisdom" may be included in "poor quality papers". While that may be true, I would argue that anecdotal quality papers do not necessarily mean they have no valuable information, but that their value in terms of quality and usefulness for further scholarship and practice is uncertain. This is a major problem in a field where clarity regarding reliable, not uncertain, evidence is desired. I suggest you consider the following two publications (particularly the first) for two reasons:

	Geerts, J. M., Goodall, A. H., & Agius, S. (2020). Evidence-based leadership development for physicians: A systematic literature review. Social Science & Medicine, 246, 1–17. https://doi.org/10.1016/j.socscimed.2019.112709 Lyons, O., George, R., Galante, J. R., Mafi, A., Fordwoh, T., Frich, J., & Geerts, J. M. (2020). Evidence-based medical leadership development: A systematic review. BMJ Leader, leader-2020-000360. https://doi.org/10.1136/leader-2020-000360 1) These two papers clarify high-quality in our field, and 2) The conclusions of the "best available evidence" support your conclusions and thereby, would strengthen them with support from high-quality evidence in academic literature. If, after reviewing these papers, you choose not to include them, I would still support publishing your manuscript as is. Second, I disagree that programmes are theory incarnate. Many programs are designed without robust theoretical considerations, not without accurate familiarity with evidence-based approaches (which is further exacerbated by unclarity regarding quality publications versus anecdotal ones. I think your point is that if programmes thoroughly articulate their designs and reported outcomes, readers can make their own assessments of them. The Social Science & Medicine paper above provides one approach to evaluating programs that you might consider (without conditions). Otherwise, thank-you for your excellent work.

REVIEWER	Vincent, Charles University of Oxford, Experimental Psychology
REVIEW RETURNED	16-Jun-2023

GENERAL COMMENTS	I thank the authors for their comprehensive and thoughtful response to all the reviews. The paper is certainly much clearer and easier to follow, and the wealth of data is presented in more accessible ways. The approach of using more tables and boxes, accompanied by short commentaries in each sub-section of the results, has greatly improved the paper. I also very much appreciate the addition of Box 3 summarising the main implications and practical lessons for people trying to build capacity in practice. I have one further request which is that the authors do a little more work to clarify their approach of using two parallel reviews and also the issue of non-theoretical models, both discussed in my previous review. The authors have responded and explained their approach to both issues in their response and it sounds reasonable though still a little difficult to follow. However, I am not so much concerned that the authors defend their approach to me as with them explaining to the reader. For instance, the authors explain in their response that they did two parallel literature reviews, but this is still not explained clearly in the resubmission. There is reference to a scoping review which is ‘combined’ with a framework synthesis but its not clear either that there are two reviews or how they are combined. I suggest that the authors go through the whole Methods sections again and see if they can explain their approach clearly and simply to a reader who may not be familiar with Carroll, Behemoth and

	the like. Of course, the reader would need to go to the source papers for more detail, but it will really help to have a clear, basic over-arching explanation of what is going on. This in turn will add to the power of the paper and the value of the authors' findings on this important topic.

VERSION 2 – AUTHOR RESPONSE

Reviewer: 2
Dr. Jaason Geerts, Canadian College of Health Leaders, Cass Business School

Comments to the Author:
Dear authors,
Thank-you for your resubmission. Given that this is a resubmission, following extensive first-round peer-review suggestions, I have focused on your point-for-point response to reviewers document. I repeat my original comment that the quality of your original review and writing of the manuscript was high.

Overall, you have done an excellent job incorporating our suggestions.

I am recommending that the most recent version be accepted. Especially since the two articles I mentioned in the first review were not provided to you, I have proposed considerations for you, which you are welcome to take or leave (and I am not proposing that any modifications are required for publication).

1) Returning to the issue of quality of study assessment in scoping reviews. You argue that it isn't required in scoping reviews and that "nuggets of wisdom" may be included in "poor quality papers". While that may be true, I would argue that anecdotal quality papers do not necessarily mean they have no valuable information, but that their value in terms of quality and usefulness for further scholarship and practice is uncertain. This is a major problem in a field where clarity regarding reliable, not uncertain, evidence is desired.

I suggest you consider the following two publications (particularly the first) for two reasons:

Geerts, J. M., Goodall, A. H., & Agius, S. (2020). Evidence-based leadership development for physicians: A systematic literature review. Social Science & Medicine, 246, 1–17. https://doi.org/10.1016/j.socscimed.2019.112709

Lyons, O., George, R., Galante, J. R., Mafi, A., Fordwoh, T., Frich, J., & Geerts, J. M. (2020). Evidence-based medical leadership development: A systematic review. BMJ Leader, leader-2020-000360. https://doi.org/10.1136/leader-2020-000360

1) These two papers clarify high-quality in our field, and

2) The conclusions of the "best available evidence" support your conclusions and thereby, would strengthen them with support from high-quality evidence in academic literature.

If, after reviewing these papers, you choose not to include them, I would still support publishing your manuscript as is.

We thank the reviewers for the insightful papers suggested. We appreciated their relevance and

referred to them in the discussion section to support our findings (see page 18, lines 11-12, lines 33-35, lines 37-39, and 43-44).

Second, I disagree that programmes are theory incarnate. Many programs are designed without robust theoretical considerations, not without accurate familiarity with evidence-based approaches (which is further exacerbated by unclarity regarding quality publications versus anecdotal ones. I think your point is that if programmes thoroughly articulate their designs and reported outcomes, readers can make their own assessments of them. The Social Science & Medicine paper above provides one approach to evaluating programs that you might consider (without conditions).

We respect the reviewer's opinion. From a realism standpoint, programmes are considered "theory incarnate" (Pawson and Tilley, 1997), i.e., theories (including causal assumptions) held by programme designers, funders, and implementers or the ‘target’s of programmes are always underpinning a programme or intervention, whether they are explicit or implicit. We agree with the reviewer that "many programs are designed without robust theoretical considerations, not without accurate familiarity with evidence-based approaches". That is why we recommended that programme designers make explicit their (evidence-informed) theoretical assumptions that explain how a programme is supposed to lead to the expected outcomes (see page 18, line 1-8).

Otherwise, thank-you for your excellent work.

Reviewer: 3
Dr. Charles Vincent, University of Oxford

Comments to the Author:
I thank the authors for their comprehensive and thoughtful response to all the reviews. The paper is certainly much clearer and easier to follow, and the wealth of data is presented in more accessible ways. The approach of using more tables and boxes, accompanied by short commentaries in each sub-section of the results, has greatly improved the paper. I also very much appreciate the addition of Box 3 summarising the main implications and practical lessons for people trying to build capacity in practice.

I have one further request which is that the authors do a little more work to clarify their approach of using two parallel reviews and also the issue of non-theoretical models, both discussed in my previous review. The authors have responded and explained their approach to both issues in their response and it sounds reasonable though still a little difficult to follow. However, I am not so much concerned that the authors defend their approach to me as with them explaining to the reader. For instance, the authors explain in their response that they did two parallel literature reviews, but this is still not explained clearly in the resubmission. There is reference to a scoping review which is ‘combined’ with a framework synthesis but it's not clear either that there are two reviews or how they are combined.

I suggest that the authors go through the whole Methods sections again and see if they can explain their approach clearly and simply to a reader who may not be familiar with Carroll, Behemoth and the like. Of course, the reader would need to go to the source papers for more detail, but it will really help to have a clear, basic over-arching explanation of what is going on. This in turn will add to the power of the paper and the value of the authors’ findings on this important topic.

We thank the reviewer for this comment. We have explained more clearly the overall process of the scoping review and best-fit framework synthesis as follows: "The process of the scoping review and best-fit framework synthesis is shown in Figure 1. Based on the research questions (step 1), we searched for and selected primary studies (step 2a). Concurrently, we searched for and selected

frameworks, models or theories (step 2b). Next, we summarized the characteristics of primary studies included (step 3a) and generated an a priori coding framework from the selected frameworks, models or theories (step 3b). We then coded data from primary studies against the a priori coding framework (step 4). We performed a thematic analysis for data that could not be coded against the a priori framework (step 5). This resulted in a new framework comprising a priori and new themes supported by the data (step 6)" (see page 4, lines 14-21).
We have also explained what we mean by non-theoretical models as follows: "i.e., terms often used in biomedical research such as "epidemiological model", "disease model", "care model" or "statistical model" that do not fit the theoretical focus of the best-fit framework strategy" (see page 6, lines 1-3)