Reviewer comments

The authors propose a manuscript reviewing several characteristics of DiGA listed in the DV directory maintained by BfArM, in particular, with respect to evidence.

The manuscript could be an interesting source of information given that it also summarizes mainly publications in German; so it could be a good source of information for a German non-speaking audience interested in the DiGA framework. Additionally, I found the rational process applied for the literature review (screening, use of dedicated tool, etc...) a good aspect of the work.

> → Reply: Thank you for your kind words.

1. My major comment is about the readability and clarity of the text which I found challenging and should be significantly improved for the authors to reach their objective. In its present form, the manuscript is very hard to read and follow. Communication level must be increased significantly throughout the manuscript. This is important to avoid misunderstanding. Example of a misleading sentence in page 13: "It is critical that in the current process, there is reimbursement and clinical use of products for which there is no evidence".

> → Reply 1: Thank you for this reasonable comment. Assisted by a native speaker, we have tried to improve the readability and clarity of the text.

> → Changes in the text: There were numerous language-related changes as well as some re-structuring throughout the text. Please see the tracked changes within the manuscript. A point-by-point listing of all changes would exceed a reasonable size for this answer.

2. My following point might resolve naturally while tackling the previous one, but I found quite hard to understand what the "take home messages" are. At first, the manuscript reads very negative with respect to the evaluated sources while I am not sure this was the intention of the authors. It is also somehow paradoxical to read about the so many important limitations around the topic of evidences while reviewing products which had to go through a dedicated framework which include "evidence" as one of the key pillar. The authors indicate some positive aspects (e.g., page 8 and 10) but overall the tone is very critical. I would recommend the authors to clarify their position with respect to that point choosing to re-balance statements or - if indeed they wish to highlight a true problem with the current framework - to express it more clearly. At the current state, the reader is left alone with what the main conclusions could be. The paragraph "Limited external generalizability" is symptomatic of that problem. It is hard to understand what the authors want to communicate as they state first that studies "met internal validity criteria" and that "external generalizability was also met", and in addition to that, they mention a "moderate to large effect", but they concluded this paragraph by stating that "the review showed that the manufacturers provided only limited high-quality evidence for PIC". This reads as a contradiction.

Another confusing sentence: page 12: "Reasons for a notable dropout rates may include potential systematic errors". Reason or a consequence?

> → Reply 2: Thank you for your remarks. We have tried to clarify our manuscript and hope that the language-related updates will have taken care of some of the points you mention (e.g., related the take-home message, to the "confusing sentence" on page 12 and similar problems).

> Related to your concern about the manuscript being largely critical or even negative: This is due to the concerns raised by the authors of the papers included in our review. We were also

able to corroborate these (unfortunately) negative points by comparing the results to an analysis of apps listed in the directory that we recently conducted (which we had previously included in the manuscript in somewhat adapted form as a citation (11), but which is now shown as an external reference (29) where appropriate, as the corresponding paper has now been published).

While we certainly applaud the intent of the DVG and its provisions for making apps prescribable, including the call for evidence as specified in the law as one of the pillars for reimbursable apps, we were unable to identify more positive points related to assessing the available evidence for available DiGA from either the literature or our own evaluation of the studies for apps permanently included in the directory (which we now cite as an external reference only). This is not to say that we are critical of the process of evidence generation per se, but are nevertheless bound by the concerns voiced in the studies in our review, and by the findings from the DiGA directory described in our previous publication.

We have, however, tried to use more neutral wording, where possible, specifically related to the four thematic clusters that were identified from the review (see below), and hope that our intention of simply evaluating the literature and compiling and presenting the corresponding findings is now clearer.

→ Changes in the text:

- Numerous language-related corrections.

- Extensive restructuring

- More elaborate subsection 1.2 Rationale and knowledge gap.

- Rewrite of subsection 1.3 Objective to more clearly state the focus of the presented work

- Removal of content related to previous work, i.e. the paragraph describing our own analysis of studies related to DiGA listed in the directory from the Methods section, removal of subsection "3.2 Review of studies provided for apps listed in the DiGA directory" from the Results section

- More positive wording, e.g.: a) Results section, second paragraph, b) renaming of the thematic clusters (Adequacy of scientific study designs, Adequacy of evidence validation tools, Potential for bias, External generalizability); this is also reflected in Table 1.

- Adaptations to the conclusion in the "External generalizability" part, as well as numerous changes to the paragraphs in the other thematic clusters. The four headings for the thematic clusters have now been promoted to second level headings.

- The previous subsection 4.3 Explanation of findings has been replaced by a more elaborate subsection 4.3 Implications and actions needed (previously section 4.4).

- More concise Conclusions.

- Please see the tracked version of the paper for a more detailed look into the changes.

3. I couldn't access reference 11. I understood it is in preprint. Could you nonetheless discuss what are the main differences between the content of this reference and your contribution?

→ Reply 3: Thank you for this remark. While meanwhile, the referenced paper has finally been published and is freely available (see https://ebooks.iospress.nl/doi/10.3233/SHTI23, CC-by-NC license, which would also have allowed us to reuse its content in the originally planned manner), we decided to remove the table and related content (mainly in the Methods and Results sections) from the current manuscript. Interested readers can now easily access the paper (previously referenced by (11), now (29)). The relevant aspects have now sim-

ply been integrated into the discussion part, with added references where appropriate. The literature review was not part of this previous work.

→ Changes in the text: Table 2 and related content (specifically, within the Methods and sub-section 3.2, "Review of studies provided for apps listed in the DiGA directory") were deleted, and appropriate content was added to section "4.2 Assessment of the findings in context with a recent DV evaluation".

Minor comments:

- please avoid when possible the use of "( )" (parenthesis) in the abstract.

→ Reply: Thank you for this suggestion. We have tried to avoid the use of "()" as far as possible.

→ Changes in the text: Multiple, please see the tracked version of the paper.

- Some of key sentences are repeated in the result and in the discussion. Example: "typically complex interventions whose success....". Please avoid repetition. Generally, I find it hard to understand the added value of having your "result" section AND the "discussion". They seem to say very similar things.

→ Reply: We have tried to adapt the text accordingly, and are now doing a more extensive discussion of the literature review in context with the findings from our previous publication (which is no longer presented in the Results section).

→ Changes in the text: Removal of the content related to the results of our previous evaluation from the results section (also see our answers to the reviewer's remarks above), integration of the points worth discussing as a corroboration of the result from the literature review.

- Should the last section of the results "Review of studies provided for apps listed in the DiGA directory" come as first section (of the result)?

→ Reply: Thank you for this suggestion.

→ Changes in the text: We deleted this last section of the results.

- Page 12: you wrote "RTCs" instead of "RCT".

→ Reply: Thank you for identifying this typo.

→ Changes in the text: We corrected the typo.

- Table 1: there is a "*" associated with #13 but no corresponding footnote.

→ Reply: This "*" should have mirrored the footnote for #14.

→ Changes in the text: Entries #13 and #14 are now using the same marker to reference the footnote, which is identical for both.