PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Measuring the Outcomes of Volunteering for Education: Development and pilot of a tool to assess health professionals' personal and professional development from international volunteering
AUTHORS	Tyler, Natasha; Collares, Carlos; Byrne, Ged; Byrne-Davis, Lucie

VERSION 1 - REVIEW

REVIEWER	Nigel Gribble Curtin University, Australia
REVIEW RETURNED	12-Dec-2018

GENERAL COMMENTS	Overall, the manuscript attempts to fill a gap in the literature – i.e. to create a tool to measure the outcomes of healthcare professionals volunteering in international locations. There appears to be a need for such a tool. The final 40-statement questionnaire appears to be a reasonable tool that could be useful to measure the outcomes of volunteering internationally. Unfortunately, there are major problems with the lit review, methods, and discussion sections of the manuscript before this could be considered for publication.
	A summary of the concerns is below. The attached PDF of the manuscript includes detailed comments on where the paper can be improved.
	The lit review needs to be significantly improved and rewritten so the reader understands the scope of international placements, the extent and findings of previous research on the benefits and negatives of participating in international placements, what current tools exist, and the gaps in the literature. The reason for the creation of a new measurement tool needs to be much clearer.
	The research question is perhaps too general. You need to be clearer about the aim of this study in this manuscript. It seems you are conducting an item analysis to eliminate poor items so you can finalise the questionnaire. Are you doing any reliability testing (internal consistently) or validity (e.g. construct, convergent, divergent)?
	The research protocol i.e. the tool development procedures - needs much more detail. You need to show that you have followed a logical, systematic and structured approach to tool development. What or whose procedures did you follow in order to design the tool?

A major concern is that there is no explanation on why participants NOT interested in international placements are included. Participants were also included who had never volunteered internationally but might be interested. Wouldn't these participants potentially skew the responses to the questions?

Overall, the methods that have been used to reduce the original 110 statements to 40 and the 10 categories appears rigorous and appropriate.

In the final 40-item tool, when the participants answered the questions, there is no clarity on the context in which they were to consider the outcomes statements e.g. were they told to think about the 'last month' at work, or their international placement, or at home or all of these contexts or were they given no content? When answering outcome statements one needs a context.

There needs to be a detailed reason for the inclusion of the Satisfaction with Life Scale. This does not appear in the lit review.

Two of the final 40 items related to teamwork, appear to have linkage to teamwork and should be in other categories.

Given you are recommending that the tool could be used to measure the benefits of volunteering overseas, is the tool sensitive enough to measure the changes in the various skillsets? More research on this part of the tool is needed.

The discussion is weak and does not discuss who could use the tool, when it could be used; does not discuss the data from the tool could be used by stakeholders; there is no linkage to previous literature; there are no limitations of the study discussed; and no future research is discussed or the next steps in developing your tool and gathering more info on validity and reliability.

Many of the paragraphs are unclear and difficult to follow.

As discussed previously, there appears to be a need for such a tool. The final 40-statement questionnaire appears to be a reasonable tool that could be useful to measure the outcomes of volunteering internationally. Unfortunately, there are major problems which require major revisions.

The reviewer provided a marked copy with additional comments. Please contact the publisher for full details.

REVIEWER	Benjamin J Lough University of Illinois at Urbana-Champaign
REVIEW RETURNED	06-Mar-2019

| GENERAL COMMENTS | --- Page 5 ---

There is significant critical literature touching on the ethical concerns about medical practice abroad in low-income communities, particularly when asserting that staff can practice skills they could not practice in a high-income context. The manuscript should give some coverage to this important body of literature.

The researchers cite the IVIS but there are already quite a few other multi-dimensional scales that measure outcomes for international volunteers. for e.g.: Developmental Model of Intercultural Sensitivity (DMIS); Global Perspectives Inventory (GPI); Global Engagement Survey (GES); Global Citizenship Scale; etc. The authors need to justify why we need yet another survey to measure learning outcomes for people going abroad. |

They do mention that no previous surveys are specific to healthcare; however, the outcomes of this survey also do not appear to be specific to healthcare, as the outcomes covered in this survey are typical of the other surveys.

In addition, the authors frequently refer to their previous study, which appears to use largely the same items. The authors need to clarify why and how this additional study is different from their previous study.

--- Page 6 ---

"lots of the things" -- the use of more professional language would help establish legitimacy in this article

--- Page 8 ---

I would suggest omitting text on the three categories, as this is not essential information.

--- Page 9 ---

The authors could possibly reduce, or clean up, the explanations of PCA and IRT.

--- Page 10 ---

There is redundancy in the sections "creating the tool" and "creating the questionnaire". If these are different, it is not clear.

--- Page 12 ---

Text describing the participants can be greatly reduced, as all this information is contained in the table. The authors should limit this text to any surprising, outstanding or inconsistent findings about the participants.

--- Page 13 ---

Eliminate some of the unnecessary text under the PCA section, such as which marginal items were retained/dropped.

--- Page 14 ---

It is not clear how goodness of fit statistics in the text are associated with the fit statistics in table 4.

--- Page 15 ---

For precision curves and estimates and associated figures of the theta spectrum, readers need to know why this matters. What does this information mean in practical terms for the constructs? I might suggest omitting these figures unless their practical relevance can be explained.

--- Page 16 ---

The summary information in the first paragraph of the conclusion section can be deleted, as it is fully redundant. I might suggest

<table>
<tr><td></td><td>starting this section with what is currently your third paragraph, which describes a few of the limitations. The limitations section should perhaps be a separate sub-section.

There are incomplete sentences, incorrect spellings, etc. Overall, the manuscript should have been edited better before submitting it for review.

--- Page 17 ---

The discussion of moderating variables may be more appropriate in the front portion of the manuscript as it is an odd fit in its current location.

--- Page 24 ---

In table 6, it is not clear what the standard error or p-values are referring to. Also, the table should be clearer about what estimate is being reported on — I assumed these are lambda coefficients but they appear quite high so the table must be reporting on IRT? In any case, please clarify. At minimum, the table title should clarify whether these are PCA or IRT results -- and could be significantly shortened.

--- Page 25 ---

Is there a plan to further reduce the number of items under Confidence, as it seems there are more items than needed in this construct? What other future modifications are planned based on these results?

--- Page 32 ---

As with the precision curves, information functions need to be explained if included. Why is this important information for the readers to know from a practical standpoint?</td></tr>
</table>

VERSION 1 – AUTHOR RESPONSE

Reviewer 1

Dear Reviewer 1, in addition to the addressing the general comments below, we addressed every individual comment in the PDF. In almost all cases be agreed with your comments and adapted the paper accordingly. However, we have included a table with extra discussion of one.

No	Comment	Action
1	Unfortunately, there are major problems with the lit review, methods, and discussion sections of the manuscript before this could be considered for publication	We have significantly changed the background section, we have re-order the methods and made specific changes in line with all of the comments on PDF document. We have completely revised the discussion to reflect the new

		introduction and take into consideration both reviewers comments.
2	The lit review needs to be significantly improved and rewritten so the reader understands the scope of international placements, the extent and findings of previous research on the benefits and negatives of participating in international placements, what current tools exist, and the gaps in the literature. The reason for the creation of a new measurement tool needs to be much clearer.	The literature review has been rewritten to encompass all of the reviewers suggestions. Thank you for the comments, we feel the introduction is now much clearer.
3	The research question is perhaps too general. You need to be clearer about the aim of this study in this manuscript. It seems you are conducting an item analysis to eliminate poor items so you can finalise the questionnaire. Are you doing any reliability testing (internal consistently) or validity (e.g. construct, convergent, divergent)?	We have made the aim more specific and mentioned reliability testing. Thanks for pointing this out.
4	The research protocol i.e. the tool development procedures - needs much more detail. You need to show that you have followed a logical, systematic and structured approach to tool development. What or whose procedures did you follow in order to design the tool?	We have followed a traditional approach to item creation and reduction for the creation of psychometric measurement tools. We have now added a reference and more detail to the design subsection of the methods section
5	A major concern is that there is no explanation on why participants NOT interested in international placements are included. Participants were also included who had never volunteered internationally but might be interested. Wouldn't these participants potentially skew the responses to the questions?	The population who would eventually be answering the questionnaire are health professionals of all cadres – some of these would have had international experience and some who would not (if we were assessing PPD before international experience to compare with after or if we were comparing these PPD outcome in groups of people with and without international experience). We have made this point clear in the methods section.
6	In the final 40-item tool, when the participants answered the questions, there is no clarity on the context in which they were to consider the outcomes statements e.g. were they told to think about the 'last month' at work, or their international placement, or at home or all of these contexts or were they given no content? When answering outcome statements one needs a context.	All participants were told to think about their last month, regardless of whether this was international or not. We have added this detail to the creating the tool sub-section of the results section
7	There needs to be a detailed reason for the inclusion of the Satisfaction with Life Scale. This does not appear in the lit review.	We have a detailed reason for inclusion in the methods that we would like to draw to

			your attention- An additional existing scale was used within the tool, the satisfaction with life scale (SWLS) [25]. This is a five-item scale that has been used frequently to measure satisfaction with life. This replaced a number of statements from the core outcome set about satisfaction with life, since the questions had already been refined and tested for validity and reliability[25]. We have also included a reference to guidance.
8	Two of the final 40 items related to teamwork, appear to have linkage to teamwork and should be in other categories.		We agree that the title 'team work' is not the best description of these items. We have changed this to 'attitude to work'
9	Given you are recommending that the tool could be used to measure the benefits of volunteering overseas, is the tool sensitive enough to measure the changes in the various skillsets? More research on this part of the tool is needed.		We have not tested the sensitivity to change of the measure and have added this as an important next step to the end of the conclusion.
10	The discussion is weak and does not discuss who could use the tool, when it could be used; does not discuss the data from the tool could be used by stakeholders; there is no linkage to previous literature; there are no limitations of the study discussed; and no future research is discussed or the next steps in developing your tool and gathering more info on validity and reliability.		We agree that these are really important points. We have altered the discussion and now discuss: who could use the tool, linkages back to previous literature. In particular, we are grateful for the reviewers pointing out that our tool is similar to IVIS, a really important point which we have now brought out in the discussion.
11	Many of the paragraphs are unclear and difficult to follow.		We are grateful to the reviewer for improving readability of the paper. We have worked on all of the comments in the PDF.

Additional comments from PDF document, not changed directly in the text

A1	Is this an outcome statement? What if someone has been	This came from the paper on which we are basing this paper (Tyler et al 2018) – which did ask a large group of experts if this was something that was an outcome of 'international' work, so we

	exposed to multiple ethical dilemmas and had no idea what to do? As healthcare professionals, we are exposed to ethical dilemmas almost every minute of every day.	were not in a position to judge whether its inclusion was appropriate. However, if the item showed no variability when tested in the pilot, it would not be taken forward into the tool – as in order to be included an item would have to show some variation.

Reviewer 2

12	There is significant critical literature touching on the ethical concerns about medical practice abroad in low-income communities, particularly when asserting that staff can practice skills they could not practice in a high-income context. The manuscript should give some coverage to this important body of literature.	This has now been discussed in the limitations section. We have also included reference to a more in depth discussion of such issues in a book published by members of our research team. Thank you for bringing it to our attention that we haven't included this crucial debate in this publication.
13	The researchers cite the IVIS but there are already quite a few other multi-dimensional scales that measure outcomes for international volunteers. for e.g.: Developmental Model of Intercultural Sensitivity (DMIS); Global Perspectives Inventory (GPI); Global Engagement Survey (GES); Global Citizenship Scale; etc. The authors need to justify why we need yet another survey to measure learning outcomes for people going abroad. They do mention that no previous surveys are specific to healthcare; however, the outcomes of this survey also do not appear to be specific to healthcare, as the outcomes covered in this survey are typical of the other surveys.	This is a very important point and we have added discussion of the need to consider health workers differently in the introduction and added the interesting point that the findings could be applicable outside health workers and, in fact, are very similar to the items in the other surveys.
14	In addition, the authors frequently refer to their previous study, which appears to use largely the same items. The authors need to clarify why and how this additional study is different from their previous study.	The previous study was a way of synthesising potential outcomes from the literature, it was not a tool. It has been published since this was submitted; which would clarify this. However, the purpose of this study is now stated in the introduction- 'In a recent meta-synthesis and Delphi study, we reported a list of 116 outcomes from a review of literature on international placements for healthcare professionals'

15	"lots of the things" -- the use of more professional language would help establish legitimacy in this article	Thank you for bringing this to our attention, now changed to 'elements'
16	I would suggest omitting text on the three categories, as this is not essential information.	Thank you for this suggestion, we agree it's not essential and have removed it
17	The authors could possibly reduce, or clean up, the explanations of PCA and IRT.	We agree with this comment and have removed half of each paragraph.
18	There is redundancy in the sections "creating the tool" and "creating the questionnaire". If these are different, it is not clear.	We separated this section across methods and results, in an attempt to follow standardised reporting. But have chosen to report the results of the pre-pilot and creating the tool in the methods to allow the reader to better understand the process.
19	Text describing the participants can be greatly reduced, as all this information is contained in the table. The authors should limit this text to any surprising, outstanding or inconsistent findings about the participants.	Thanks for this advice, we have removed a lot of this section.
20	Eliminate some of the unnecessary text under the PCA section, such as which marginal items were retained/dropped.	We have removed this section, thank you
21	It is not clear how goodness of fit statistics in the text are associated with the fit statistics in table 4.	These are for the MIRT, we have now made this clear by adding an additional heading.
22	For precision curves and estimates and associated figures of the theta spectrum, readers need to know why this matters. What does this information mean in practical terms for the constructs? I might suggest omitting these figures unless their practical relevance can be explained.	We have included some practical information about the figures in the results section, thanks for this comment.
23	The summary information in the first paragraph of the conclusion section can be deleted, as it is fully redundant. I might suggest starting this section with what is currently your third paragraph, which describes a few of the limitations. The limitations section should perhaps be a separate sub-section.	Limitations are now a separate sub section. We have completely revised the discussion, adding more detail and reduced the conclusion section, avoiding redundancy.
24	There are incomplete sentences, incorrect spellings, etc. Overall, the manuscript should have been edited better before submitting it for review.	We apologise for this and have proof read again.
25	The discussion of moderating variables may be more appropriate in the front portion of the manuscript as it is an odd fit in its current location.	We have moved this into the introduction and revisited in the discussion.

26	In table 6, it is not clear what the standard error or p-values are referring to. Also, the table should be clearer about what estimate is being reported on — I assumed these are lambda coefficients but they appear quite high so the table must be reporting on IRT? In any case, please clarify. At minimum, the table title should clarify whether these are PCA or IRT results - and could be significantly shortened.	We have changed the title to address your queries.
27	Is there a plan to further reduce the number of items under Confidence, as it seems there are more items than needed in this construct? What other future modifications are planned based on these results?	No, we made a judgement to keep any items within the domain that the statistics suggest. At this stage there are no plans to refine the tool, however we may do based on later results. This has been added to the discussion
28	As with the precision curves, information functions need to be explained if included. Why is this important information for the readers to know from a practical standpoint?	We have added a sentence to the results about the practical implications of this information, thank you for this idea.

VERSION 2 – REVIEW

GENERAL COMMENTS	The revamped paper is a vast improvement on the original submission. My congratulations to the authors for taking the time and effort to improve the paper as the outcomes are worthwhile for readers and practitioners in this area. Accepted with only a few grammar/spelling errors found: page 6, line 19 – 'we' should 'be ' page 9, line 20 – 'sixteen' should be written as '16' page 10, line 25 – '9' should 'nine' page 16, line 25 – didn't should be 'did not'

VERSION 2 – AUTHOR RESPONSE

2. Reviewer 1:

The revamped paper is a vast improvement on the original submission. My congratulations to the authors for taking the time and effort to improve the paper as the outcomes are worthwhile for readers and practitioners in this area.

We thank you for the positive feedback and agree it's much improved. Thanks for your help.

3. Reviewer 1:

Accepted with only a few grammar/spelling errors found:

page 6, line 19 – 'we' should 'be '

page 9, line 20 – 'sixteen' should be written as '16'

page 10, line 25 – '9' should 'nine'

page 16, line 25 – didn't should be 'did not'

We have changed the final 3 typos. The first typo is an incomplete sentence and the 'we' that's on page 6, line 19 is not a typo, therefore has not been changed.