

PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (<http://bmjopen.bmj.com/site/about/resources/checklist.pdf>) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	A review of 99 self-report measures for assessing well-being in adults: Exploring dimensions of well-being and developments over time
AUTHORS	Linton, Myles-Jay; Dieppe, Paul; Medina-Lara, Antonieta

VERSION 1 - REVIEW

REVIEWER	Ian McDowell University of Ottawa, Canada
REVIEW RETURNED	20-Dec-2015

GENERAL COMMENTS	<p>Previous reviews in this field have been selective and have been analytic, comparing the validity of small samples of measures. The authors note that this leaves space for a more comprehensive and descriptive survey of the whole range of well-being measures, covering their historical evolution and comparing their coverage. This is potentially a useful exercise if it can propose ways to consolidate this area of measurement.</p> <p>Scope of the Review. A central challenge for any review concerns how to set boundaries for which scales to include. The authors have been inclusive rather than narrow, which is probably a good thing, but their selection raises some questions. It appears that a scale was included if it was identified via the keyword 'well-being' in their literature search – but who knows how such terms are applied in the databases? The result is that some depression scales are included (CES-D, BDI) while others (Hamilton, Zung, etc) are not; few anxiety scales are included. Many anxiety scales include positive items that could be considered 'well-being' just as the items in the CES-D or BDI. If depression forms part of well-being why does anxiety not? It would be good, therefore, if the authors could more explicitly define their initial concept of 'well-being' – it appears to be anything that was labeled as such by authors, or maybe by the coders who work for PubMed, etc. Should well-being be viewed as a spectrum running from pathology to wellness? As including both positive and negative items? Does it include measures of wealth and material possessions; security; self-esteem?</p> <p>Clarity of the Text. Many parts of the text suffer from imprecise phrasing and simple grammatical errors. I have added numerous "sticky note" comments to my pdf review copy of the MS to suggest clarifications.</p> <p>For example: The authors set out to classify the range of aspects of well-being</p>
-------------------------	--

covered by each instrument; this required first coming up with a single conceptual framework. Their procedure is described on pg 8, but they do not define their distinctions between 'dimensions', 'domains', 'categories' and 'themes'. Phrases like "the themes that their dimensions reflect" (Table 1) make the reader ponder what the distinction may be. The distinction between dimension and themes only eventually becomes clear on page 16; it needs to be explained much earlier. But the 'domains' and 'categories' remain uncertain.

Content of the review article.

Historical review. A section on pg 14 concerns the development of instruments over time. This largely consists of counting of how many scales were developed or modified during particular time-periods, which is not very informative. Why not discuss the way of instruments have evolved over time (e.g. the themes they include, their design, length or scoring method, etc)? Has there been, for example, an evolution from profiles towards indexes? Qualitative to quantitative?

Conceptual structure. The very brief section 'Dimensions as determinants, states and consequences' considers the conceptual structure of well-being. This could with advantage be developed further as a contribution, based on this extensive review, to how we should approach the whole topic: as a hierarchy of themes, or as a linear process over time? With feed-backs? As a complex system? And so forth. They present their 6 themes as separate columns in Table 1, but without reference to ideas such as the ICF from the WHO, which shows the links across the columns, linking health to personal circumstances and to functioning. Perhaps the authors will undertake this in a subsequent publication (I would encourage this!).

The Framework (pp 16-17). Further confusion over themes and dimensions occurs with Figure 3. This figure seems to indicate that a given dimensions (which I am interpreting as the content of a scale as indicated by the original author) may fall under several themes (the broader categories applied by the current authors). For example Future life satisfaction is shown linked to 6 themes. But this seems confusing, as Table 2 seems to indicate that the themes are broader groupings of dimensions, and the dimensions indicated in the table do not seem to recur in more than one theme.

This many-to-many mapping is quite complex and parts of the text on page 17 are unclear. Phrases such as "The clear cluster of 'physical well-being' dimensions highlights the influence of the health related quality of life (HRQoL) approach on the topic of well-being" confused me: physical well-being is a single dimension in the figure so in what way is it a cluster? Perhaps the authors simply mean that many instruments include physical well-being; if so, say this!

Discussion section. The authors open with the comment that previous reviewers underestimated the numbers of scales. Very likely, but they do not address the counter-argument that with 100 scales perhaps the goal should be to encourage studies to be selective and use a smaller number of good-quality instruments. The authors might at least comment on the relative contributions of their extensive review, versus focused reviews that highlight the better quality or more widely used scales: both may play a role.

Hence, much of the Discussion is descriptive rather than analytic and prescriptive. Could they not propose a working definition of well-being, and suggest which dimensions a scale should include? Should studies in this field use broad-scope, generic measures, or

	<p>should they use narrow, more precise scales of each facet of well-being? What are the roles for each type?</p> <p>Conclusion. The conclusion might be more directive: “In conclusion, ambiguity surrounding how well-being is conceptualised and measured prompted us to review the measurement options available...” As anticipated, they found a wide range of dimensions in the measures reviewed. But the real question concerns what to do to improve on this ambiguity: they propose a conceptual approach in the “MeTheD-WB”. But this does not suggest ways to clarify the scope of a ‘well-being’ measure, and is really only a listing of the dimensions covered in current instruments. The article misses an opportunity, based on this wide-ranging review, to propose ways forward in overcoming the ambiguity that they highlight.</p> <p>Miscellaneous comments. If publication space is no object, Figure 1 could be retained, but it adds little to the text. Figure 2 shows a sharply declining line, and this seems visually unfortunate. Why not flip the figure so that the line showing the growth in numbers of measurements rises to the right?</p> <p>The reviewer also provided a marked copy with additional comments. Please contact the publisher for full details.</p>
--	--

REVIEWER	Jutta Lindert University of Emden; Brandeis University
REVIEW RETURNED	22-Dec-2015

GENERAL COMMENTS	<p>Authors recognize that their definition of well-being was broad. I wonder whether a review becomes meaningless in case it is too broad. The conceptual clarity is clearly missing.</p> <p>The topic is important, however in this paper the focus is very broad and precision is highly needed.</p> <p>Lack of precision starts with the title:</p> <ul style="list-style-type: none"> - Are indeed 100 ways of measuring well-being available? - Is this paper a systematic review or a narrative review? - Are these instruments “adult instruments” or “instruments measuring adult’s well-being?” <p>The lack of precision continues with the abstract:</p> <ul style="list-style-type: none"> - Which disciplines are included? - The design includes systematic searches and a narrative review. This should be clarified. Why is “published in English” an inclusion criterion? Guidelines for systematic review propose that no language exclusion criterion should be applied. - Objectives and results do not match. <p>Furthermore, the whole paper raises questions:</p> <ul style="list-style-type: none"> - What is meant by “recent” years? Which years authors refer to? - The definition of well-being used by these authors is very
-------------------------	--

	<p>broad (“quality and state of a person`s life) and might be meaningless.</p> <ul style="list-style-type: none"> - I wonder about the sentence: “The assessment of well-being in clinical trials, other empirical studies, and to comple4ment indices of social progress....” Grammar and content are unclear. This might be due to the unclear concept of well-being. Well-being and quality of life are different concepts (at least in Public Health) and cannot and should not be used as synonyms. - Furthermore, I agree that there are a large number of measures available. However, I disagree with the authors that there are no criteria for what makes a good instrument. The criteria are clear and psychometric properties can be investigated. - Maybe it makes sense to differentiate between economics and Public Health? - Authors state that the most influential theoretical influences remain unclear. I would disagree and refer to the amount of theoretical literature. - The search terms should be listed in the additional material; as 6 databases were used these search terms which have been used across databases are interesting. - Exclusion criteria are not according to standards (exclusion of paper not published in English). - Was a priori-defined study protocol used? Can authors add this to the additional material? - The procedure remains unclear. - As regards data extraction I wonder who extracted the data. - I wonder who did the thematic analysis and what the exact steps were. The description of methods remains vague. - Authors include a sense of coherence scale, however, sense of coherence is a different concept and inclusion of sense of coherence does not match the introduction to this paper. - It should read “Likert” not “likert”. - Theoretical influence: I wonder about this part of the paper. Can “theoretical influence” be assessed? Investigating “theoretical influence” might need a paper of its own. - The part on <i>dimensions as determinants, states and consequences</i> is unclear. Conceptual clarity about causality might be necessary to improve clarity. - I wonder how authors can say that the “number of available instruments has been underestimated”. The inclusion and exclusion criteria make the process of including and excluding very clear. I agree that in this review the process of including and excluding is vague. - Authors recognize that their definition of well-being was broad. I wonder whether a review becomes meaningless in case it is too broad. The conceptual clarity is clearly missing. <p>Overall, language and grammar should be checked in the paper.</p>
--	--

	There are a range of grammatical mistakes.
--	--

VERSION 1 – AUTHOR RESPONSE

Comment number #	Reviewer comment	Response
Reviewer 1: Prof. Ian McDowell		
Comment #1	Previous reviews in this field have been selective and have been analytic, comparing the validity of small samples of measures. The authors note that this leaves space for a more comprehensive and descriptive survey of the whole range of well-being measures, covering their historical evolution and comparing their coverage. This is potentially a useful exercise if it can propose ways to consolidate this area of measurement.	Thank you for highlighting this point. Our objective in this MS has not been to consolidate the options, but instead to organise and to recognise the vast amount of measures available. Our primary reason for not consolidating was the absence of a widely accepted definition of well-being. We believe this would be needed to aid consolidation. Further, it may only be possible to reach an accepted definition within disciplines, yet the objective of the current MS was to look at well-being without one single discipline in mind. We have attempted to provide researchers in the area of well-being a more informed choice of the different available dimensions and their definitions (in the enclosed glossary).
Comment #2	A central challenge for any review concerns how to set boundaries for which scales to include. The authors have been inclusive rather than narrow, which is probably a good thing, but their selection raises some questions. It appears that a scale was included if it was identified via the keyword 'well-being' in their literature search – but who knows how such terms are applied in the databases? The result is that some depression scales are included (CES-D, BDI) while others (Hamilton, Zung, etc) are not; few anxiety scales are included. Many anxiety scales include positive items that could be considered 'well-being' just as the items in the CES-D or BDI. If depression forms part of well-being why does anxiety not? It would be good, therefore, if the authors could more explicitly define their initial concept of 'well-being' – it appears to be anything that was	We appreciate the reviewer's observation that we have attempted to be inclusive. Although the application of terms in databases can be an issue, this is a challenge faced by all searches in this area. In order to accommodate this we attempted to keep our search strategy as simple as possible. Further, we undertook hand searching to ensure we were not solely reliant on results from the databases. We have included the Zung scale that the reviewer stated has not been reviewed. We took the approach to exclude depression and/or anxiety tools that explicitly state in their instructions that they were to be used in individuals already diagnosed mental health conditions (for example, the Hamilton scale for depression). Nevertheless, Table 1 includes the Hamilton-Snaith Pleasure Scale, which is a more generic measure developed by Hamilton. We

	labeled as such by authors, or maybe by the coders who work for PubMed, etc. Should well-being be viewed as a spectrum running from pathology to wellness? As including both positive and negative items? Does it include measures of wealth and material possessions; security; self-esteem?	have made this clearer in the MS.
Comment #3	Clarity of the Text. Many parts of the text suffer from imprecise phrasing and simple grammatical errors. I have added numerous "sticky note" comments to my pdf review copy of the MS to suggest clarifications.	We appreciate the time it took to go through the manuscript highlighting these points. All comments (those that appeared within your feedback and those that appeared as sticky notes on the manuscript) have been addressed in the amended MS.
Comment #4	The authors set out to classify the range of aspects of well-being covered by each instrument; this required first coming up with a single conceptual framework. Their procedure is described on pg 8, but they do not define their distinctions between 'dimensions', 'domains', 'categories' and 'themes'. Phrases like "the themes that their dimensions reflect" (Table 1) make the reader ponder what the distinction may be. The distinction between dimension and themes only eventually becomes clear on page 16; it needs to be explained much earlier. But the 'domains' and 'categories' remain uncertain.	We agree that in our MS the use of some terms was not clear or consistent. In line with this feedback we provide an earlier explanation of 'dimensions' and 'themes' (page 6). 'Domains' and 'categories' were instances of loose phrasing; therefore they have been removed from the paper.
Comment #5	Content of the review article. Historical review. A section on pg 14 concerns the development of instruments over time. This largely consists of counting of how many scales were developed or modified during particular time-periods, which is not very informative. Why not discuss the way of instruments have evolved over time (e.g. the themes they include, their design, length or scoring method, etc)? Has there been, for example, an evolution from profiles towards indexes? Qualitative to quantitative?	Thank you for your comment. In a previous version of the MS we had included a paragraph on how the tools have developed over time. In accordance with your suggestion we have added this back into the results section (page 14).
Comment #6	Conceptual structure. The very brief section 'Dimensions as determinants, states and consequences' considers the conceptual structure of well-being. This could with advantage be	Our focus in this MS has been on the measurement aspect of the topic alone. The intended contribution of this work is to provide researchers with a broader spectrum of dimensions and tools. We

	developed further as a contribution, based on this extensive review, to how we should approach the whole topic: as a hierarchy of themes, or as a linear process over time? With feed-backs? As a complex system? And so forth.	hope the presented work will encourage others to select measures of well-being that are conceptually best suited to their study question.
Comment #7	They present their 6 themes as separate columns in Table 1, but without reference to ideas such as the ICF from the WHO, which shows the links across the columns, linking health to personal circumstances and to functioning. Perhaps the authors will undertake this in a subsequent publication (I would encourage this!).	Thank you. We too are very interested in the WHO ICF framework. We are currently discussing how a subsequent manuscript could take a more in depth look at the interactions across themes in order to explore the dynamic nature of well-being.
Comment #8	Further confusion over themes and dimensions occurs with Figure 3. This figure seems to indicate that a given dimensions (which I am interpreting as the content of a scale as indicated by the original author) may fall under several themes (the broader categories applied by the current authors). For example Future life satisfaction is shown linked to 6 themes. But this seems confusing, as Table 2 seems to indicate that the themes are broader groupings of dimensions, and the dimensions indicated in the table do not seem to recur in more than one theme. This many-to-many mapping is quite complex and parts of the text on page 17 are unclear.	We have made revisions throughout the document in order to add clarity to the text. In table 2 we have removed the numbering so that the table only includes the definitions of the themes.
Comment #9	Phrases such as “The clear cluster of ‘physical well-being’ dimensions highlights the influence of the health related quality of life (HRQoL) approach on the topic of well-being” confused me: physical well-being is a single dimension in the figure so in what way is it a cluster? Perhaps the authors simply mean that many instruments include physical well-being; if so, say this!	Thank you for pointing out that this was unclear. We have revisited the MS and have removed this section.
Comment #10	Discussion section. The authors open with the comment that previous reviewers underestimated the numbers of scales. Very likely, but they do not address the counter-argument that with 100 scales perhaps the goal should be to	We agree with the reviewer that our discussion section was missing a comment on the possible benefits of both extensive and focused reviews. We have included this on page 17.

	encourage studies to be selective and use a smaller number of good-quality instruments. The authors might at least comment on the relative contributions of their extensive review, versus focused reviews that highlight the better quality or more widely used scales: both may play a role.	
Comment #11	Hence, much of the Discussion is descriptive rather than analytic and prescriptive. Could they not propose a working definition of well-being, and suggest which dimensions a scale should include? Should studies in this field use broad-scope, generic measures, or should they use narrow, more precise scales of each facet of well-being? What are the roles for each type?	We acknowledge that the discussion section was not directive enough. We have amended it by including an addition paragraph concerned with the characteristics and benefits of: short overall well-being instruments, broad-scoped/cross themed instruments, and more narrowly focused instruments (page 18).
Comment #12	Conclusion. The conclusion might be more directive: "In conclusion, ambiguity surrounding how well-being is conceptualised and measured prompted us to review the measurement options available..." As anticipated, they found a wide range of dimensions in the measures reviewed. But the real question concerns what to do to improve on this ambiguity: they propose a conceptual approach in the "MeTheD-WB". But this does not suggest ways to clarify the scope of a 'well-being' measure, and is really only a listing of the dimensions covered in current instruments. The article misses an opportunity, based on this wide-ranging review, to propose ways forward in overcoming the ambiguity that they highlight.	Our work revealed an unexpectedly high level of ambiguity, and we agree with the reviewer's point that reducing the ambiguity in the field is an important research goal. But the main aim of this review was to inform other researchers about the vast range of measurement options; we did not set out to find ways through this ambiguity. However in the discussion section (page 19) we have now indicated that tackling this ambiguity in further research should be a priority.
Comment #13	Miscellaneous comments. If publication space is no object, Figure 1 could be retained, but it adds little to the text.	We have moved the sentence introducing the content of the PRISMA diagram to the start of the paragraph with the expectation that it is better placed and more clearly explained (page 8).
Comment #14	Figure 2 shows a sharply declining line, and this seems visually unfortunate. Why not flip the figure so that the line showing the growth in numbers of measurements rises to the right?	Thank you, we agree. Figure 2 has now been restructured, in order to show the growth of tools in an ascending line, instead of a declining line.

Comment #15	... dimensions they cover. (Dimensions aren't really 'available' within a measurement: they may underlie its structure.)	This has been rephrased in line with the reviewer's suggestion (page 2)
Comment #16	A rather awkward sentence. Perhaps recast as "This is demonstrated by ..."?	This sentence has now been removed from the MS.
Comment #17	is is a 'growing agenda' or an increase in prominence of this topic? I find some of the phrasing a bit loose!	We have revisited the manuscript and amended this sentence according to the suggestions of the reviewer (page 4).
Comment #18	but where would the guidance come from? There are few explicit theories, and few consensus statements have been prepared. 'Guidance' seems a vague and overall term: you might indicate potential mechanisms using only a few extra words.	In line with the reviewer's recommendations, we have clarified in the MS (page 4) that we are specifically talking about a lack of guidance in existing literature.
Comment #19	This first sentence seems naive. Consider saying "A fundamental issue is that disagreement has surrounded both the definition..."	This sentence has been rephrased in line with the suggestion made (page 4).
Comment #20	Are they alternative schools, or does this reflect an evolution in thinking, perhaps prompted by shifting goals with rising aspirations as the overall level of health improves?	We agree with the reviewer that the evolution of ideas over time has resulted in a succession of perspectives on how 'the good life' is best understood. However we also note, that in any given time, there have been multiple, sometimes competing schools of thought on the topic. The attention paid to both hedonic and eudaimonic theories of well-being is just one recent example of this. This may be due to the presence of discipline specific theories, or the slow and imperfect process of theory testing. In order to reflect both of these possibilities, the sentence has been rephrased (page 4).
Comment #21	... is how to choose between the large number...'	This sentence has been rephrased in line with the suggestion made (page 4).
Comment #22	There is a confusion here: a gold standard could be (and often is) quite separate from the measurement instruments. And with purely subjective themes, such as well-being, the notion of gold standard would be seen by many as not applicable. Most link gold standard to criterion validation, and there is no accepted criterion for measuring well-being. This, in fact, likely encouraged the proliferation of measurements.	The reviewer is right to point out that in the context of well-being, the notion of a gold standard is largely not applicable. We agree that a 'perfect' instrument is unrealistic, regardless of how many people in the field seek out such a tool. We have now rewritten this sentence accordingly (page 4). We have also attempted to recast the MS in general to be more analytical and informative.

	By slightly recasting your text you could indicate some of the mechanisms that underlie the reality you are describing. That would make the text more interpretive and therefore helpful.	
Comment #23	or there are widely differing disciplinary perspectives on the theme of well-being. As with health, mental, physical, social, but also environmental, ecological, etc.	We agree with the reviewer; however we feel that this is a separate point (made by the reviewer in comment #27). Therefore we have addressed this issue in the response #27.
Comment #24	this is interesting: if there are many disciplinary perspectives, each tool could be quite different in focus. But now you imply that there are also comprehensive measures (covering several perspectives?). It would seem valuable to separate these ideas.	Thank you, we agree. We have clarified in the MS that we indeed mean that some newer tools attempt to cover multiple perspectives, as with the Flourishing Scale which draws on Ryff's psychological well-being, but also older ideas from the humanistic tradition.
Comment #25	Should it have been capped? Is this a cancerous growth? Or is it a positive sign?	We have rephrased this sentence to clarify that that growth has continued (page 4). With the current manuscript we are highlighting to other researchers the many options available, without suggesting whether the growth is inherently good or bad, therefore we have removed the term 'uncapped' from the sentence.
Comment #26	But it may not be "in addition to" - it may be because of the differing perspectives on well-being.	We agree with the reviewer that the ambiguity we refer to in this instance may be the result of differing perspectives on the topic. This point has now been added (page 5)
Comment #27	or unstated?	This sentence has been rewritten in line with the suggestion made (page 5).
Comment #28	I find this sentence unclear. Do you mean that dimensions such as life satisfaction are mentioned but not translated into explicit scores by the measurement instruments? Clarification needed!	Thank you. This sentence has been restructured in order to make the paragraph clearer (page 5).
Comment #29	"As with most domains of health measurement, there is no agency or group that oversees and coordinates the development of the field of well-being measurement."	This has been rephrased in line with the reviewer's suggestion (page 5).
Comment #30	this is heartening indeed. But the contrast between your previous sentence (which suggested anarchy in the field) and this sentence seems too strong. Why not say that in the past researchers were	We have rephrased this sentence in order to improve the flow of the paragraph (page 5)

	left to select instruments largely through personal preference, but recent initiatives, such as that of OECD, are offering guidance in the selection of instruments...?	
Comment #31	proposes' ?	This sentence has now been removed from the MS.
Comment #32	(grammar glitch). Also, consider the tenses you use: you use the present, but their work is now complete; why not past tense?	This grammar glitch has been resolved.
Comment #33	Hmmm: you have earlier implied that there are needlessly large numbers of instruments, so isn't a goal to discard some? So surely a goal is to reduce the numbers of scales that are mentioned!	Thank you for highlighting this issue. We have responded to the concern over consolidating the measures mentioned in our response to comment #1.
Comment #34	or "without clear analysis of how their content differs"? Not sure you can defend claim it was not thought about...	This has been rephrased in line with the reviewer's suggestion (page 5).
Comment #35	This justifies a review... (drop "requirement")	'Requirement' has been dropped from this sentence (page 5).
Comment #36	or "has not been clearly described."	This has been rephrased in line with the reviewer's suggestion (page 5).
Comment #37	why 'methodological'? Does this mean you are linking the dimensions to the ways in which they were derived statistically? I would think a conceptual framework might come first. Consider distinguishing inductive vs. deductive approaches at this point?	We agree with the reviewer and have relabelled this as a conceptual framework (page 6).
Comment #38	One wonders why this particular year? You review various of the older instruments, so surely you must have searched older literature? Consider explaining this: was the intention to discard older instruments that have not been referenced since 1993?	The current review was largely concerned with instruments referenced in the last 20~ years, as the reviewer has pointed out. We have clarified this point in the MS (page 6).
Comment #39	(word missing)	Thank you for pointing this out. This sentence has now been removed from the MS.
Comment #40	(or available in an English translation?) [published in is not clear: do you mean originally published in?]	Yes, available in an English translation was what we meant, we have rephrased the point accordingly (pages 2 & 6).
Comment #41	in brief, you excluded age- and disease-specific instruments.	We have simplified the phrasing in the inclusion/exclusion criteria paragraph. We did not exclude instruments developed with older adults in mind; therefore we have retained the age

		related phrasing.
Comment #42	as specified by the original author, or according to your analysis?	We have rephrased this for clarity (page 7). Prior to our analysis, we noted dimensions as stated by their authors.
Comment #43	(a) as with last query, might this approach conflict with claims of the author? (b) Consider distinguishing more clearly between dimensions & domains. (c) In table 1 the domains seem to have become 'themes' - are these the same?	(a) Whilst we agree with the reviewer that it is theoretically possible that this approach may conflict with the claims of the original authors of an instrument, we undertook the current work to re-analyse dimensions within instruments. Our objective was to provide a practical structure that would be informative to other researchers. If readers are interested in claims made by the original authors, we have provided references to each of the instruments analysed (in table 1/reference section) Table 1 includes references back to each of the instruments reviewed. (b) We have revisited the manuscript in order to make this distinction clearer. (c) This has been addressed by sticking to consistent phrasing 'themes' throughout.
Comment #44	identified by whom? (original author, or via your analysis?)	As with comment #42, we began with dimensions as their original authors had defined them. While familiarising ourselves with the dimensions and the definitions, the review team decided to simplify the wording of the dimensions with the goal of making the work clearer and more useful for other researchers. This however only happened in a minority of cases. The process was largely iterative, and our understanding and interpretation of the dimensions developing during the analysis.
Comment #45	conceptually, or because of the items included in the scale? [This is a very important distinction - see literature on life satisfaction vs. morale]	We agree with the reviewer that this is a very important distinction. We looked at both the author's definition and the items that had been used. Where multiple members of the review team agreed that considerable overlap existed between dimensions, they were combined, in order to limit (where pragmatically possible) the existence of multiple dimensions covering the same aspect of well-being. We are aware of the literature on life satisfaction vs. morale, particularly the work of James Horley. In review team discussions we were not combining any dimensions as distinct as those discussed in Horley's work (life

		satisfaction, happiness morale etc.). The combining of dimensions took place with examples such as the dimension 'social relationships' (AQOL) with the dimension 'relationships' (PWI-A, BBC-SWB etc). Another example is where we placed 'habitual levels of physical activity' (HWB) under the same dimension heading as 'physical activity' (QWB-SA).
Comment #46	Now we have themes, categories, domains & dimensions. Increasing need to explain these subtleties!	We agree that clarity in the use of the terms through this MS needed tightening. As stated earlier (#4), we have thus revised the whole document and now only themes and dimensions remain. They are defined on page 6.
Comment #47	Capital "L". But some use 5-point answer scales that are not Likert scales, so be careful! [Likert = "strongly agree", "agree", etc]. Likert never used 11-point answers.	Thank you for pointing this out, this has been rephrased accordingly (page 13). This comment corresponds to comment #75.
Comment #48	a bit overstated... Yes, developed by some economists, but used very widely and not only for health economic analyses.	We have removed this sentence from the MS.
Comment #49	strange grammar! Why not simple "4 of the instruments were the Warwick..."	We have removed this sentence for clarity, and in order to fit within the word limit.
Comment #50	your use of "found" seems to imply that 'dimensions' are your classification, not that of the original author. Am I right?	The dimensions as they are named in this MS are a product of both the original authors and the analysis undertaken in the current MS. In the vast majority of cases, dimensions are phrased exactly as authors had named them within their own papers.
Comment #51	when	This has been rephrased in line with the reviewer's suggestion (page 14).
Comment #52	reported' now seems to conflict with 'found' earlier in the paragraph! (see previous note)	We hope we have clarified (as in our response to your previous note, comment #50) that the dimensions were largely left unchanged, and in the vast majority of cases reflect the phrasing of authors. As such, we felt it would be beneficial here to note which dimensions were most frequently found across instruments.
Comment #53	why 'specific'? Surely the outcomes could be non-specific? Again, I keep tripping up over the apparently loose phrasing!	Thank you for pointing this out, we agree and have rephrased this sentence accordingly (page 13).
Comment #54	oops. There goes all my Latin training!	Oops! Correct. This sentence has now been removed.

Comment #55	provided. Need to run a grammar check on whole MS. Likewise, should be 'resources' in previous sentence. (MS seems to have been prepared in great haste!)	Thank you for pointing this out, these errors have been corrected.
Reviewer 2: Prof. Jutta Lindert		
Comment #56	Are indeed 100 ways of measuring well-being available?	We agree that this makes the title unclear. We have rephrased it accordingly.
Comment #57	Is this paper a systematic review or a narrative review?	This paper used mixed methods. In identifying the instruments we applied systematic review methodology in respect of the search strategy, screening, and data extraction. We then used both thematic analysis and narrative synthesis to summarise our findings.
Comment #58	Are these instruments “adult instruments” or “instruments measuring adult’s well-being?”	The instruments are 'instruments measuring adult's well-being'. We have adjusted the title to reflect this and we hope this is clearer now.
Comment #59	Which disciplines are included?	We apologise but we are unsure of the context of this question. We have not attempted to include or exclude any specific disciplines. The databases searched represent literature sources across the social and clinical sciences (including psychology, economics and medicine and sociology).
Comment #60	The design includes systematic searches and a narrative review. This should be clarified. Why is “published in English” an inclusion criterion? Guidelines for systematic review propose that no language exclusion criterion should be applied.	Thank you for pointing this out. We have now explicitly stated that in addition to systematic searches, a narrative review was undertaken (page 7). We have also clarified the point concerning papers 'published in English' (pages 2 & 6). This corresponds to comments #40 and #69.
Comment #61	Objectives and results do not match	In accordance with the reviewer's comments we have refined the review objectives/results in order to reflect the content presented (page 6).
Comment #62	What is meant by “recent” years? Which years authors refer to?	By recent years, we are referring to the spike in interest in the last 20 years. We have clarified this in the text (page 4).
Comment #63	The definition of well-being used by these authors is very broad (“quality and state of a person’s life) and might be meaningless.	It is true that the definition used in our review is broad. A more restrictive definition of well-being might have resulted in fewer instruments; we chose an inclusive approach, in order to ensure we were able to identify the diverse range of instruments across all

		disciplines. We did this to ensure that other researchers were able to find measures of well-being within our review that best matched their own definition of well-being and their own research question.
Comment #64	I wonder about the sentence: "The assessment of well-being in clinical trials, other empirical studies, and to complement indices of social progress...." Grammar and content are unclear. This might be due to the unclear concept of well-being. Well-being and quality of life are different concepts (at least in Public Health) and cannot and should not be used as synonyms.	We completely agree that health related quality of life and well-being are different concepts; however we fear that in the wider research community they are frequently used interchangeable. In retrospect we agree that the sentence is unclear therefore we have omitted it from the manuscript.
Comment #65	Furthermore, I agree that there are a large number of measures available. However, I disagree with the authors that there are no criteria for what makes a good instrument. The criteria are clear and psychometric properties can be investigated.	The reviewer is right to point out here that psychometrics are able to provide criteria for selecting instruments. We have rephrased this sentence (page 4) in order to clarify that there is little conceptual agreement concerning what a measure of well-being should contain.
Comment #66	Maybe it makes sense to differentiate between economics and Public Health?	Indeed, there are many disciplinary perspectives on well-being. Initially, we analysed the instruments identified in order to categorise them by discipline, however it did not prove methodologically possible. This is in part due to the issue raised in Professor Ian McDowell's comment #48. It became increasingly difficult in a topic as interdisciplinary as well-being to determine where the boundaries between disciplines lay. Any comprehensive exploration of this would require a lot of new, different work (such as contacting the developers of the instruments), and a separate manuscript.
Comment #67	Authors state that the most influential theoretical influences remain unclear. I would disagree and refer to the amount of theoretical literature.	We have rephrased this sentence hoping that it is now clearer (page 4). However, we wish to clarify that we are not commenting on whether the theory of well-being itself is unclear, we are instead stating that the way in which these many theories have influenced the development of instruments is unclear. For example, we were interested in which theories authors were frequently referencing as influential on their own work.

Comment #68	The search terms should be listed in the additional material; as 6 databases were used these search terms which have been used across databases are interesting	Thank you for pointing this out. We have now moved the search terms to additional material, following the reviewer's suggestion (appendix 1).
Comment #69	Exclusion criteria are not according to standards (exclusion of paper not published in English)	This point corresponds to comment #40 and #60. We agree with the reviewer that papers should not be excluded for not being published in English. We clarify in text that it was not 'papers' that were excluded, but 'instruments not available in an English translation' (page 2 & 6).
Comment #70	Was a priori-defined study protocol used? Can authors add this to the additional material?	This work forms part of a PhD project. A protocol for the PhD itself was developed and approved by a panel at the University of Exeter Medical School, with the help of external reviewers; however the project has changed and developed over time. As such, we feel that it would be inappropriate to publish the original protocol alongside this MS.
Comment #71	The procedure remains unclear.	We hope the amendments and revisions we have made to the methods section, based on the reviewers' specific comments have clarified the procedure undertaken (page 6-7).
Comment #72	As regards data extraction I wonder who extracted the data.	Data were extracted by two members of the review team (ML and AML). This has now been stated in the main body of the paper (page 6); this is also detailed in the Contributorship statement.
Comment #73	I wonder who did the thematic analysis and what the exact steps were. The description of methods remains vague.	We agree that clarity in our explanation of the steps undertaken during the thematic analysis is important. We have edited this section to indicate (with initials) the members of the review team who undertook the various steps of the analysis (page 7).
Comment #74	Authors include a sense of coherence scale, however, sense of coherence is a different concept and inclusion of sense of coherence does not match the introduction to this paper.	We appreciate that the Sense of Coherence Scale is not as explicitly linked to well-being as some of the other measures included. However Aaron Antonovsky based this scale on the concept of salutogenesis, which has at its core a focus on fostering both health and well-being. As such, we believed it would be a conceptually relevant instrument to include.
Comment #75	It should read "Likert" not "likert".	Thank you for pointing this out, this has been rephrased accordingly (page 13). This comment corresponds to comment #47.

Comment #76	Theoretical influence: I wonder about this part of the paper. Can “theoretical influence” be assessed? Investigating “theoretical influence” might need a paper of its own.	Our objective was not to assess the theories as such, but to highlight which theories have influenced the development of instruments. We agree that this could be a paper in itself; however, we consider that a brief reference to the influence of Diener's work and its influence is important.
Comment #77	The part on dimensions as determinants, states and consequences is unclear. Conceptual clarity about causality might be necessary to improve clarity.	We agree with the reviewer that causality is central to the topic of well-being. In the results section of the current MS we are stating that we identified many dimensions that could be considered as determinants of well-being, feelings of well-being, or consequences of well-being. We are currently working on a separate MS using secondary data to investigate causality and determinants of well-being.
Comment #78	I wonder how authors can say that the “number of available instruments has been underestimated”. The inclusion and exclusion criteria make the process of including and excluding very clear. I agree that in this review the process of including and excluding is vague.	This sentence regarding the number of tools being underestimated has now been removed from the MS. In line with multiple comments related to the clarity of the methods section, we hope it is now clearer.
Comment #79	Overall, language and grammar should be checked in the paper. There are a range of grammatical mistakes.	Thank you for highlighting this. We have revised the manuscript with grammar in mind and hope the language is now clearer.

VERSION 2 – REVIEW

REVIEWER	Ian McDowell University of Ottawa, Canada
REVIEW RETURNED	08-Mar-2016

GENERAL COMMENTS	<p>This is potentially a useful article, but please see my concerns in the attached review. I really feel that your review work has put you in a position to help clarify the direction we should take in developing a clearer conceptual formulation of "well-being", and how it relates to topics such as depression, health, happiness, etc.</p> <p>[All my references refer to the pdf copy, by page & line number in the format pg x; line number y. E.g. “4; 6-9” = page 4, lines 6-9]</p> <p>This article has some potential usefulness, but I kept wishing that the authors would make more explicit proposals for advancing the rudimentary state of conceptualising ‘well-being’.</p> <p>Unfortunately the article remains weakened by imprecise language; I have cited examples below. Because this field is in great need of clear thinking it is all the more important that language is used</p>
-------------------------	--

	<p>precisely!</p> <p>The search strategy, described as systematic, relied on mention of 'well-being' in the title or keywords of an article. This has led to some seemingly arbitrary selections. For example, Cella's Functional Assessment of Cancer Therapy is included (presumably because it mentions well-being as a sub-heading in the scale and someone added it as a keyword). But Schipper's very similar Functional Living Index – Cancer, which covers very similar ground, is not included; it does not happen to include the phrase 'well-being'. Similarly, there are several depression scales that might equally have been included (Carroll, Montgomery-Asberg, etc) but were not, presumably because of phrasing of an article title or someone's choice of keywords. I get the feeling that "mechanical" might be a better descriptor for the selection than "systematic". I kept wondering whether a more reasoned selection procedure might have been a valuable contribution? The authors note (page17, line 19) that their review identified the largest number of scales, but sheer numbers may not be the core issue.</p> <p>I understand that including only scales that explicitly mention 'well-being' is a practical and defensible approach. But if the purpose is to help refine the conceptualisation of well-being (objective 4), it would at least be worth comment in the Discussion section that many of these scales could equally well be described as 'general health status measures'. Table 2 includes several themes that are central to general health status measures. Well-being is such a loose concept that a valuable contribution of an article such as this would be to present a clear argument as to how it should be used and how it relates to concepts such as health, suffering, well-being, satisfaction, spirituality, contentment and happiness. This type of discussion begins on page 15, lines 38-54, but it would be good to see more.</p> <p>I do not find the tracking of the years of publication or of the discipline of the authors very helpful. More interesting might be to trace evolving definitions of well-being over time, but I am far from sure that such an effort would be worthwhile.</p> <p>I was underwhelmed by the "Measures, Themes and Dimensions" framework. First, the elements appear to be in the wrong order: Themes form the broadest level, and each can include several dimensions. For each dimension there can be several measures. So why not "TDM"? Second, some phrases (e.g. Overall well-being) are both a theme and a dimension, which seems unhelpful. Third, some themes seem redundant (life satisfaction, future life satisfaction, past life satisfaction). Fourth, where a dimension can map onto many themes (e.g. life satisfaction), should this not imply its promotion to theme status?</p> <p>Throughout this article is a sub-text that the field of measuring well-being is in a parlous state. I fully agree. But the Discussion section misses the opportunity to offer some really useful direction to this field; it is not prescriptive at all. I do not mean prescriptive in the sense of page 19, lines 3-19, in terms of suggesting which instrument may be most suitable. I mean including a reasoned discussion over whether 'well-being' remains a useful concept and if so, how it should be defined and differentiated from other, related constructs.</p>
--	---

	<p>Trivia & Editorial Comments</p> <p>Overall, there is a lot of imprecise language and “woolly wording” (“I guess I kinda understand what I think they mean...”). The article would greatly benefit from careful editing, from the excision of unnecessary words, and from casting sentences into the active voice.</p> <p>The Abstract contains some good examples: Page 2, line 7: “different” is unnecessary; Line 9: what does “they” refer to – the measures or the researchers or even the disciplines? Line 26: “English translation”. Do you mean an English version? Translation would seem to imply excluding scales that were originally developed in English...! Line 27: “197 different dimensions”. But how different are they in reality? You count related things like past, present and future life satisfaction separately, or anhedonic depression, depression and depression-happiness separately (see Fig 3). Line 34: “individual circumstances”. Surely ‘personal’ – individual could imply excluding family or shared circumstances.</p> <p>Other scattered editorial comments: Page 4; 6-9 “are clear evidence of this”, but you don’t really give clear evidence. I did not, for example, understand the reference to “growing policy implications” – have democratic governments not always been (somewhat) concerned over the well-being of their population? Has this suddenly changed? Macmillan claimed “You’ve never had it so good” in the 1950s...</p> <p>4; 15 “guidance” → consensus?</p> <p>4;21-23 Consider re-wording? “Fundamental to the challenge of measuring well-being is the extent of disagreement over its definition and theoretical basis”...?</p> <p>4; 46: no universally accepted measure. Perhaps that is too high an expectation – there will most likely be a small set of measures, each tailored to different purposes. (The authors imply this on page 6; lines 17-19)</p> <p>4; 54: “multiple perspectives” or is it actually differing perspectives? Many of the measures do not actually reflect multiple perspectives, but those of a particular discipline. And therein lies the dilemma of the profusion of measures. Perhaps we need some multiple perspectives, but if so, why not argue for this explicitly?</p> <p>5; 3: “within” or between (or among) instruments? I doubt that the dimensions vary within an instrument.</p> <p>5; 5-7: dimensions defined in terms of items on sub-scales. This seems circular (dimension = subscale) and misses an opportunity to provide an external definition of a dimension (e.g. in terms of separate aspects of health and function, distinguished by different body systems or adaptive processes), which you do offer in the Glossary .</p> <p>5; 11: “nature” or dimensions?</p>
--	--

	<p>5; 17: “how and where dimensions ... can be found...” this appears close to meaningless. Surely we aspire to more than the author saying that “items 5-10 measure spiritual well-being”? This whole paragraph seems to miss the point that there are 2 requirements: defense of the dimensions considered relevant in a measure of well-being, and a working definition of each dimension to drive the content validation of the sub-scales measuring each.</p> <p>5; 31: advice on navigating the literature?? Please! Not sure your librarian needs assistance from the OECD. Woolly phrasing!</p> <p>5; 44-46: herein lies a challenge. You have just implied that it is not desirable to invent yet another well-being measure; from that I infer that you would encourage researchers to focus on a narrower set of high-quality scales. But now you argue that previous reviews have been selective (or, at least, ‘unsystematic’ which could imply something different). But those reviews may have deliberately been selective, suggesting that we ignore some scales that are for various reasons unsuitable. It would be good if you could argue the need for a comprehensive review in this context of an excess of scales. The notion of ‘systematic’ could be explained: do you mean comprehensive, or perhaps selective, based on clear selection criteria?</p> <p>12; 14: I was surprised that the Zung depression scale is indicated as measuring social well-being rather than mental. I have just re-checked the Zung items and see only 1 or 2 that have anything to do with social relations or well-being.</p> <p>Table 1: I then notice that the CES-D is indicated as covering physical, social and activities themes. OK, but these are only included as potential symptoms of depression. Back to Zung: why did this scale not also get a tick under physical? Zung items cover constipation, tachycardia, weight loss, sleep difficulties, etc. I began to worry about consistency of the check-marks in Table 1.</p> <p>I then pondered situations in which a scale is shown as measuring all of the themes except for Overall well-being (e.g. Cantril and others). As the Cantril (presumably his ladder scale) is only 1 item, evidently these variants must be alternative forms, so I would have thought that this method could equally well be applied to overall well-being?</p> <p>Regrettably, I did not check the themes indicated for all 100 scales but suggest that the authors should probably confirm that Table 1 is accurate.</p> <p>10; 5: Should be “Health Utilities Index” (plural).</p> <p>The paragraph on page 18, lines 11-28 offers little of value – readers are likely to understand that scales have various uses and that longer ones offer more comprehensive assessments. Consider deleting!</p>
--	--

REVIEWER	Jutta Lindert University of Emden, Germany
REVIEW RETURNED	02-Mar-2016

GENERAL COMMENTS	Thank you for submitting a revised Version. Unfortunately, I do not see that questions of reviewers were addressed in the revised version, e.g. sense of coherence is still included in the measures. In my understanding a major limitation of this review is that it is both, too broad and herewith unspecific and too narrow (e.g. leaving out Spanish databases). The concept of this paper might be an excellent contribution to a book on well-being. In this book the variety of concepts of well-being might be specified. The psychometric properties of instruments need further detailed investigation which could be done if not too many databases would have been included. Maybe you consider making two papers?
-------------------------	--

VERSION 2 – AUTHOR RESPONSE

Review of “10 ways to measure...”

[All my references refer to the pdf copy, by page & line number in the format pg x; line number y. E.g. “4; 6-9” = page 4, lines 6-9]

Comment 1: This article has some potential usefulness, but I kept wishing that the authors would make more explicit proposals for advancing the rudimentary state of conceptualising ‘well-being’.

Response: *We thank the reviewer for prompting us to take a more directive approach. We have addressed his point in the discussion.*

Comment 2: Unfortunately the article remains weakened by imprecise language; I have cited examples below. Because this field is in great need of clear thinking it is all the more important that language is used precisely!

The search strategy, described as systematic, relied on mention of ‘well-being’ in the title or keywords of an article. This has led to some seemingly arbitrary selections.

For example, Cella’s Functional Assessment of Cancer Therapy is included (presumably because it mentions well-being as a sub-heading in the scale and someone added it as a keyword). But Schipper’s very similar Functional Living Index – Cancer, which covers very similar ground, is not included; it does not happen to include the phrase ‘well-being’. Similarly, there are several depression scales that might equally have been included (Carroll, Montgomery-Asberg, etc) but were not, presumably because of phrasing of an article title or someone’s choice of keywords. I get the feeling that “mechanical” might be a better descriptor for the selection than “systematic”. I kept wondering whether a more reasoned selection procedure might have been a valuable contribution? The authors note (page 17, line 19) that their review identified the largest number of scales, but sheer numbers may not be the core issue.

Response: *The reviewer highlights important clarification points.*

Firstly, we acknowledged that the nature of a systematic search in a topic like well-being will run into some issues. Key terms were carefully chosen after scoping review provided over 300,000 hits (articles). Given the amount of hits, a pragmatic strategy needed to be devised. The search terms used in this review were agreed after careful consideration and consultation with a senior information specialist (this is not reported in the paper). However, we do not claim to have devised the perfect search strategy, in fact, we believe that in a topic is intricate as well-being; all searches will understandably contain imperfections (page 20).

It appears that there has been a misinterpretation by the reviewer with our use of the term systematic and his suggestion that we should instead use the term mechanical. We followed the principles of the CRD for systematic reviews. This review is systematic as it has a defined set of research questions or aims, specific search terms being used in each of the databases, and the searches are replicable. Although title and keywords were mainly searched in the first scanning of the hits, the majority of abstracts were also read. It is also systematic as the results are reported in a PRISMA figure. For these reasons, the term mechanical would not be the correct phrasing to use.

Secondly, the reviewer is right to suggest that there are some tools which may have been inadvertently missed. However, the Functional Living Index-Cancer that was mentioned by the reviewer does not reflect this review's selection criteria as this is a cancer specific tool. The Functional Assessment of Cancer Therapy-General Population on the other hand was included, as this version of the tool was developed with a non-clinical population in mind. We do however appreciate that this review may be missing a number of scales for the measurement of depression.

Sheer numbers is not necessarily the core issue. But highlighting a large number of tools allows the current work to compliment the more focused reviews already in existence. It is surely important that reviews on the topic reflect both the depth and the breadth of measurement instruments available.

Comment 3: I understand that including only scales that explicitly mention 'well-being' is a practical and defensible approach. But if the purpose is to help refine the conceptualisation of well-being (objective 4), it would at least be worth comment in the Discussion section that many of these scales could equally well be described as 'general health status measures'. Table 2 includes several themes that are central to general health status measures. Well-being is such a loose concept that a valuable contribution of an article such as this would be to present a clear argument as to how it should be used and how it relates to concepts such as health, suffering, wellbeing, satisfaction, spirituality, contentment and happiness. This type of discussion begins on page 15, lines 38-54, but it would be good to see more.

Response: The reviewer pointed out that our fourth aim was to refine the conceptualisation of well-being. However our aim was to: "(4) organise the dimensions identified into a conceptual framework, highlighting key themes of well-being, and the instruments that measure them".

We are currently preparing a separate manuscript that deals exclusively with the definitions, theories and concepts within the topic.

We agree with the reviewer that measures of well-being could easily be described as measures of health. We have amended the discussion section in order to include a reflection on the similarities found between measures of health and measures of well-being (page 20). We also include a section on how the term well-being should be used (page 20). Namely, we recommend that 'well-being' may be best understood as an umbrella term that encapsulates its many constituent parts, including but not limited to, happiness and life satisfaction. We hope that our expanded discussion section provides a more focused dialogue on how well-being and its related concepts should be understood.

Comment 4: I do not find the tracking of the years of publication or of the discipline of the authors very helpful. More interesting might be to trace evolving definitions of wellbeing over time, but I am far from sure that such an effort would be worthwhile.

Response: We thank the review for his suggestions. While we appreciate that this figure may not be new information for experts in the field who are already familiar with the trends, there are many other researchers (possibly the majority) who may not be aware of how and when the increased development of tools took place. After editing the figure in line with the previous set of reviewer's comments, we feel that we are providing a worthwhile and undemonstrated depiction of the field.

Comment 5: I was underwhelmed by the “Measures, Themes and Dimensions” framework. First, the elements appear to be in the wrong order: Themes form the broadest level, and each can include several dimensions. For each dimension there can be several measures. So why not “TDM”?

Response: *The framework represents one approach to providing a degree of order to an unmistakably complex web of available options. We feel that the framework is of practical usage, and may be most beneficial to researchers who are unfamiliar with the literature on well-being. We do however appreciate the reviewer pointing out that the name of the framework may be unhelpful. For simplicity, we have simply referred to it in text as ‘a thematic framework of well-being’.*

Comment 6: Second, some phrases (e.g. Overall well-being) are both a theme and a dimension, which seems unhelpful.

Response: *This theme has now been rephrased as ‘Global well-being’, to reflect that it covers well-being across themes, in a more general sense, opposed to the more domain specific themes, for example social well-being.*

Comment 7: Third, some themes seem redundant (life satisfaction, future life satisfaction, past life satisfaction).

Response: *Although we agree with this comment; the developer of the tool included separate subscales for life satisfaction with different time references in mind. We certainly feel that it may be the objective of further work to question whether these and other distinctions are indeed appropriate. We have highlighted in the discussion the importance of further research in order to determine the conceptual differences between dimensions discussed in this paper (Discussion, page 22).*

Comment 8: Fourth, where a dimension can map onto many themes (e.g. life satisfaction), should this not imply its promotion to theme status?

Response: *To base the themes on the dimensions that reflected many themes would be a different methodology altogether, and diverge from the methodological guidance (Braun & Clark, 2006) that we followed. In line with this comment we have amended this dimension, so it reflects ‘global well-being’, instead of separate themes.*

Comment 9: Throughout this article is a sub-text that the field of measuring well-being is in a parlous state. I fully agree. But the Discussion section misses the opportunity to offer some really useful direction to this field; it is not prescriptive at all. I do not mean prescriptive in the sense of page 19, lines 3-19, in terms of suggesting which instrument may be most suitable. I mean including a reasoned discussion over whether ‘well-being’ remains a useful concept and if so, how it should be defined and differentiated from other, related constructs.

Response: *We clearly share many of the reviewer’s frustrations with the field. In order to provide a more prescriptive case and to highlight why well-being remains a useful concept a paragraph has been included in the introduction that summarises why the concept of well-being remains a worthwhile outcome to consider (page 5-6) and we have expanded the discussion (page 20).*

Trivia & Editorial Comments

Overall, there is a lot of imprecise language and “woolly wording” (“I guess I kinda understand what I think they mean...”). The article would greatly benefit from careful editing, from the excision of unnecessary words, and from casting sentences into the active voice.

Response: *We thank the reviewer for these comments. The manuscript has now been reviewed by a medical writer.*

The Abstract contains some good examples:

Comment 10: Page 2, line 7: “different” is unnecessary;

Response: *This word has now been removed from the sentence.*

Comment 11: Line 9: what does “they” refer to – the measures or the researchers or even the disciplines?

Response: *This has now been clarified (page 2).*

Comment 12: Line 26: “English translation”. Do you mean an English version? Translation would seem to imply excluding scales that were originally developed in English...!

Response: *English translation has been replaced with ‘English version’ (page 2).*

Comment 13: Line 27: “197 different dimensions”. But how different are they in reality? You count related things like past, present and future life satisfaction separately, or anhedonic depression, depression and depression-happiness separately (see Fig 3).

Response: *‘different’ has now been removed from this sentence. We hope it is now clearer, and we agree, that there may be dimensions that are not explicitly different from others, however we were attempting to preserve as much as we could from the original instruments. We have highlighted that many terms in well-being are used interchangeably (Results, page 15-16), and discussed the need to develop a better understanding of the distinctions between dimensions (Discussion, page 20). A subsequent separate manuscript is currently being prepared that deals more specifically with this topic in a more focussed way.*

Comment 14: Line 34: “individual circumstances”. Surely ‘personal’ – individual could imply excluding family or shared circumstances.

Response: *We agree. This theme has now been rephrased.*

Other scattered editorial comments:

Comment 15: Page 4; 6-9 “are clear evidence of this”, but you don’t really give clear evidence. I did not, for example, understand the reference to “growing policy implications” – have democratic governments not always been (somewhat) concerned over the wellbeing of their population? Has this suddenly changed? Macmillan claimed “You’ve never had it so good” in the 1950s... .

Response: *We have made this point clearer in text. While well-being has been a concern for some time, the explicit measurement of subjective forms of well-being on a regular basis by governments is a more recent trend (page 4 & 5/6).*

Comment 16: 4; 15 “guidance” □□ consensus? .

Response: *Consensus has replaced guidance as suggested by the reviewer (page 4).*

Comment 17: 4;21-23 Consider re-wording? “Fundamental to the challenge of measuring wellbeing is the extent of disagreement over its definition and theoretical basis”...?

Response: *This has now been rephrased accordingly (Intro, page 4).*

Comment 18: 4; 46: no universally accepted measure. Perhaps that is too high an expectation – there will most likely be a small set of measures, each tailored to different purposes. (The authors imply this on page 6; lines 17-19)

Response: *We agree with the reviewer, this may indeed be too high of an expectation. We have stated that this expectation may be unrealistic (page 4).*

Comment 19: 4; 54: “multiple perspectives” or is it actually differing perspectives? Many of the measures do not actually reflect multiple perspectives, but those of a particular discipline. And therein lies the dilemma of the profusion of measures. Perhaps we need some multiple perspectives, but if so, why not argue for this explicitly?

Response: *This has been rephrased as ‘differing perspectives’ in line with the reviewer’s suggestion (page 5).*

Comment 20: 5; 3: “within” or between (or among) instruments? I doubt that the dimensions vary within an instrument.

Response: *Within, between has been replaced (page 5).*

Comment 21: 5; 5-7: dimensions defined in terms of items on sub-scales. This seems circular (dimension = subscale) and misses an opportunity to provide an external definition of a dimension (e.g. in terms of separate aspects of health and function, distinguished by different body systems or adaptive processes), which you do offer in the Glossary.

Response: *This has been modified in the text (page 5).*

Comment 22: 5; 11: “nature” or dimensions?

Response: *Dimension, this has been corrected (page 5)*

Comment 23: 5; 17: “how and where dimensions ... can be found...” this appears close to meaningless. Surely we aspire to more than the author saying that “items 5-10 measure spiritual well-being”? This whole paragraph seems to miss the point that there are 2 requirements: defense of the dimensions considered relevant in a measure of well-being, and a working definition of each dimension to drive the content validation of the sub-scales measuring each.

Response: *Thank you for highlighting this. We also recognise the need to examine the defence and definitions of dimensions of well-being. This would largely be the product of a psychometric review as has been suggested by the reviewer. We have highlighted in the discussion section the need for further psychometric research (page 20), and we have made a clearer statement in favour of the need for researchers to defend the conceptual underpinnings of their instruments (page 21).*

Comment 24: 5; 31: advice on navigating the literature?? Please! Not sure your librarian needs assistance from the OECD. Woolly phrasing!

Response: *This has now been rephrased (Intro, page 5).*

Comment 25: 5; 44-46: herein lies a challenge. You have just implied that it is not desirable to invent yet another well-being measure; from that I infer that you would encourage researchers to focus on a narrower set of high-quality scales. But now you argue that previous reviews have been selective (or, at least, ‘unsystematic’ which could imply something different). But those reviews may have deliberately been selective, suggesting that we ignore some scales that are for various reasons unsuitable. It would be good if you could argue the need for a comprehensive review in this context of an excess of scales. The notion of ‘systematic’ could be explained: do you mean comprehensive, or perhaps selective, based on clear selection criteria?

Response: *Thank you for highlighting that this is unclear. We do indeed state that authors are at times prone to developing new tools without seeing whether existing tools are appropriate. However we are not stating that researchers should necessarily focus on a narrow set of tools, but should instead explore the many tools in existence in order to identify instruments with the most relevant dimensions for their specific research question.*

The ‘systematic’ element refers to the fact that the review had a structured and specific search strategy. This fits in with the best practice guidance and understanding provided by the Centre for Reviews and Dissemination (CRD’s) https://www.york.ac.uk/media/crd/Systematic_Reviews.pdf. We have clarified this in text (Methods, page 6).

As stated in the discussion section, and in our last round of responses to reviewer’s comments, we believe there is a place for both focused and comprehensive measures. The former are able to provide more targeted in depth insight, while the latter are able to survey the breadth of the field. We do not feel that either trumps the other. Both provide valid insight. both may be informative.

Comment 26: 12; 14: I was surprised that the Zung depression scale is indicated as measuring social well-being rather than mental. I have just re-checked the Zung items and see only 1 or 2 that have anything to do with social relations or well-being.

Table 1: I then notice that the CES-D is indicated as covering physical, social and activities themes. OK, but these are only included as potential symptoms of depression. Back to Zung: why did this scale not also get a tick under physical? Zung items cover constipation, tachycardia, weight loss, sleep difficulties, etc. I began to worry about consistency of the check-marks in Table 1.

Response: We thank the reviewer for pointing out this error in table 1, this has now been corrected.

Comment 27: I then pondered situations in which a scale is shown as measuring all of the themes except for Overall well-being (e.g. Cantril and others). As the Cantril (presumably his ladder scale) is only 1 item, evidently these variants must be alternative forms, so I would have thought that this method could equally well be applied to overall wellbeing?

Response: Thank you for pointing this out. We have amended the framework so that instruments that measure life satisfaction (like the Cantril Ladder) now fall under the ‘Global well-being’ theme.

Comment 28: Regrettably, I did not check the themes indicated for all 100 scales but suggest that the authors should probably confirm that Table 1 is accurate.

Response: All of the tools have now been cross checked and the table is accurate.

Comment 29: 10; 5: Should be “Health Utilities Index” (plural).

Response: Thank you, this has been amended.

Comment 30: The paragraph on page 18, lines 11-28 offers little of value – readers are likely to understand that scales have various uses and that longer ones offer more comprehensive assessments. Consider deleting!

Response: Thank you for your feedback. Following consultation with our colleagues and as there was sufficient space left in the manuscript we have retained this paragraph. However if the reviewers feel it would not be of value to some readers (possibly those new to the field) we could delete it.

Comment 31: Thank you for submitting a revised Version. Unfortunately, I do not see that questions of reviewers were addressed in the revised version, e.g. sense of coherence is still included in the measures. In my understanding a major limitation of this review is that it is both, too broad and herewith unspecific and too narrow (e.g. leaving out Spanish databases). The concept of this paper might be an excellent contribution to a book on well-being. In this book the variety of concepts of well-being might be specified. The psychometric properties of instruments need further detailed investigation which could be done if not too many databases would have been included. Maybe you consider making two papers?

Response: Thank you for your feedback. We hope that in our second revision we have more fully addressed the points you made. Firstly, we have removed the sense of coherence scale from the included instruments. We designed this review to be broad in order to compliment the more specific reviews in existence. In reference to your point about not consulting Spanish databases, we had consulted a information specialist and they did not highlight this as being particularly detrimental to our search. We acknowledge that no search strategy will be perfect and agree with the reviewers that systematic searches in this field are a challenge. We have now mentioned this in the discussion section in order to highlight the issue to readers (page 20). Currently a second manuscript is being prepared with the sole objective of exploring the multiple conceptions of well-being.

VERSION 3 – REVIEW

REVIEWER	Jutta Lindert University of Emden, Emden, Germany
REVIEW RETURNED	06-May-2016

GENERAL COMMENTS	The manuscript has significantly improved. Congratulations and good luck!
-------------------------	---