Author's response to reviews

Title: Measuring factors affecting implementation of health innovations: A systematic review of structural, organizational, provider, patient, and innovation level measures

Authors:

Stephenie R Chaudoir (schaudoir@gmail.com)
Alicia Dugan (alicia.dugan@uconn.edu)
Colin Barr (colin.barr@uconn.edu)

Version: 5 Date: 23 October 2012

Author's response to reviews: see over
October 22, 2012

Dear Dr. O’Connor,

On behalf of my co-authors, I’d like to thank you for providing detailed feedback about our manuscript entitled “Measuring factors affecting implementation of health innovations: A systematic review of structural, organizational, provider, patient, and innovation level measures.” We are pleased that many of the reviewers viewed the manuscript as a valuable contribution to the growing field of implementation science measurement. We are also pleased that the reviewers offered detailed and critical feedback in which they outlined ways in that the manuscript can be strengthened.

In the pages below, we identify how we revised the manuscript in order to address each of the reviewer’s comments and critiques. Our responses appear in boldface font beneath each of the reviewer’s comments below. In the text, we denote new content with the use of red font, and we utilize comments to indicate when information has been moved to a new location in the manuscript.

The revised manuscript includes a number of substantive changes in response to your comments and those of the four reviewers, including:

- Substantial revision of the introduction to better clarify the goals of the current review and provide a stronger literature review.
- Clarification of important terminology, including the use of the term ‘factor’ to refer to the five high-order categories or classes in our working framework, and the term ‘construct’ to refer to the lower-order variables that are subsumed within these factors.
- Substantial expansion of the search strategy to include records in which keywords are located in the title or abstract and additional records published in the past 18 months, updating our search from March 2011 to August 2012.

In addition, we wanted to draw your attention to a comment made by Reviewer #4 that we were unable to resolve. In Points 2 and 4, this reviewer notes that many important details about our search methodology currently reside in our Results section on pages 12-15 rather than in our Method section. As you recall, this information was included in the Method section of our paper at our original submission, but we were asked to move it to the Results section in pre-review correspondence dated April 19, 2012. If you would prefer that we leave methodology details about our record search in the Method section, we are happy to do so; we are also happy to move it back to its original location in the Method section, per the request of Reviewer 4. Which option would you prefer?

We greatly appreciate the opportunity to revise the manuscript, and we appreciate the time and effort you and the reviewers have contributed in order to help strengthen this manuscript for publication.

Please do not hesitate to contact me if any of our responses require additional clarification or if there are any other issues that you would like us to address. Thank you again for the opportunity to revise this manuscript and strengthen it based on the feedback you have provided.

Sincerely,

Stephenie Chaudoir, Ph.D.
EDITOR’S RESPONSE

Dear Dr Chaudoir,

Thank you for submitting your manuscript “Measuring factors affecting implementation of health innovations: A systematic review of structural, organizational, provider, patient, and innovation level measures” for consideration for publication in Implementation Science. Your manuscript has now been peer reviewed by three content experts and an information scientist. The peer review comments are accessible in PDF format from the links below. Do let me know if you have any problems opening the files.

At this time we have been unable to decide on acceptance or rejection of your manuscript pending your consideration of the major essential, minor essential and discretionary revisions recommended by the reviewers. If you wish to submit a revised manuscript for further consideration by the journal addressing the issues raised in peer review, please provide a cover letter giving a point-by-point response to the issues and how you have addressed them in your revised manuscript. Please ensure that you highlight (with ‘tracked changes’/coloured/underlines/highlighted text) all changes made when revising the manuscript to make it easier to give you a prompt decision on your manuscript.

Please also note that I considered the comments raised by Wright to constitute minor essential (rather than discretionary) revisions, and I would ask you to either consider revising the search based on (some or all of) these comments (eg. re. use of MeSH, truncation and adjacency operators in the search strategy etc) or including (some or all of) these issues as shortcomings in the Discussion.

In response to the comments raised by Wright, we have expanded our search to include the search of title fields. We provide a detailed response below outlining other features of our search rationale, and we have added each of these points as limitations in the Discussion section.

Furthermore, the search is over 12 months old which in my view is another potential shortcoming. We have revised our search to include articles and their corresponding measures published through August 12, 2012.

If submitting a revised manuscript for consideration, please ensure that it conforms to the journal style. It is important that your files are correctly formatted.

We look forward to receiving your revised manuscript by 22 October 2012. If you imagine that it will take longer to prepare please give us some estimate of when we can expect it. You should upload your cover letter and revised manuscript through http://www.implementationscience.com/manuscript/login/man.asp?txt_nav=man&txt_man_id=2143304929698873. You will find more detailed instructions at the base of this email. Please don’t hesitate to contact me if you have any problems or questions regarding your manuscript.

With best wishes,
Denise O’Connor, Associate Editor

*******************************************************************************
We would be grateful if you could address the comments in a revised manuscript and provide a cover letter giving a point-by-point response to the concerns.

Please also highlight (with 'tracked changes'/coloured/underlines/highlighted text) all changes made when revising the manuscript to make it easier for the Editors to give you a prompt decision on your manuscript.

Please also ensure that your revised manuscript conforms to the journal style (http://www.implementationscience.com/info/instructions/). It is important that your files are correctly formatted.

We look forward to receiving your revised manuscript by 22 October 2012. If you imagine that it will take longer to prepare please give us some estimate of when we can expect it.

You should upload your cover letter and revised manuscript through http://www.implementationscience.com/manuscript/login/man.asp?txt_nav=man&txt_man_id=2143304929698873. You will find more detailed instructions at the base of this email.

Please don't hesitate to contact me if you have any problems or questions regarding your manuscript.

With best wishes,
The Implementation Science Editorial Team

REVIEWER #1: DAMSCHRODER

The authors used a “five-factor” framework to conduct a theory-guided systematic review in order to identify measures available to assess constructs hypothesized to predict implementation outcomes. This topic has very high importance but, unfortunately, the paper suffers from unclear writing (e.g., redundant sentences and paragraphs), lack of conceptual clarity, and poorly developed ideas and recommendations.

Thank you for your detailed and constructive feedback.

MAJOR COMPULSORY EDITS
1. Page 5, first full paragraph: you introduce a “5 factor” framework. It is unclear whether you developed this as a result of your literature synthesis or whether you developed this a priori.

As we now explain in the “Multi-level Framework” section starting on pg. 6, we utilize two existing frameworks—Damschroder et al., 2009 and Durlak & DuPre, 2008—as the basis of our framework. As we now explain on pg. 7-8, we expand these converging frameworks to include a 5th factor representing patient-level factors.

Why is this multi-level and what are “varying levels of analyses” – are these one and the same? We use the term “multi-level” to refer to the fact that factors can be conceptualized as being hierarchically nested within other factors. In the revised analysis, we have omitted the phrase ‘varying levels of analysis’ in order to avoid confusion.
a. Related to this, throughout the paper you refer to 5 factor5, 5 types of factors, and constructs – seemingly interchangeably. In places, they seem to refer to different terms but in other places you clearly are using them synonymously. This is an example of the language/terminology that needs to be cleared up – especially for a paper that has an implied goal of cleaning up some of the terminology mess that is rife in the literature.

We agree that the previous terminology was confusing. We have now clarified our terminology throughout the text such that we use the term ‘factor’ to refer to a high-order category or class and the term ‘construct’ to refer to the lower-order variables that are subsumed within factors. We delineate between factors and constructs on pg. 4-5 and then further describe the distinction when we review factors (vs. constructs within each factor) on pg. 6-8.

b. It is not necessary to point out “impact” versus “process” theories. These terms are obscure for people not familiar with Grol’s terminology and do not add anything. If you do want to clarify your point here, consider using synonyms to help clarify these terms (e.g., impact or explanatory).

Thank you for this suggestion. We see that introduction of the impact vs. process theories terminology detracts from the flow of our ideas. For purposes of clarity, we now only mention parenthetically that many existing frameworks are impact frameworks on pg. 4.

c. Two papers on implementation outcomes is not sufficient as your sole reason for focusing on predictors of implementation outcomes. It is more compelling to simply state this is your focus and perhaps mention that other papers have attempted to define/measure implementation outcomes – though this is also a topic in need of development.

We have now revised the introduction and reorganized our thoughts in order to provide rationale for our approach. This revised text appears on pg. 4 of the introduction.

d. Page 12: Definitions for your 5 factors can be moved up earlier to help those earlier sections. It is unclear whether you derived these factors from your literature search or whether you developed them a priori. The latter is stated (that you used this 5-factor framework to guide your synthesis) but the former is implied (you are presenting definitions under results).

We agree that providing conceptual definitions of the 5 factors should occur much earlier in our introduction. We have moved these definitions to the introduction (see pg. 6-7) in the revised manuscript.

2. Page 6, 1st para: The rationale for the mix of theories surveyed in this paragraph is unclear. You refer to Roger’s work as a “framework” but then call it a “theory” (random use of these terms is another example of muddled terms in this paper) ... RE-AIM is not appropriate here and is not an implementation framework. You could cite PRISM (Feldstein, et al 2008) which incorporates RE-AIM – except PRISM is an example of an explicitly multi-level model/framework and you are trying to say there are none.

We have clarified our use of the terms ‘framework’ and ‘theory’ throughout the text. We now utilize the term ‘framework’ to describe the existing models we review in the introduction. Further, we now end our conclusion section of the discussion by noting the fact that the development of additional theories—which take categories of constructs (as outlined in existing frameworks or otherwise) and provide greater specificity about their hypothesized interrelations—is necessary for the advancement of the field of implementation science.

3. Page 7 – you list your five “factors” here. You want to be very clear in your rationale for these factors – define them explicitly, and position them clearly micro-, meso-, macro-levels and rationale for doing
We have reorganized our discussion of the five factor model such that we now first introduce the concept of multi-level frameworks broadly, and we then discuss how the Durlak & DuPre, 2008 and Damschroder et al., 2009 frameworks offer two such frameworks which converge in identifying structural, organizational, provider, and innovation-level factors.

a. For “organizational” why do you say “i.e., prevention delivery system” – do you mean “e.g.,” here; an example? Or do you literally mean to say that you are focused on prevention delivery systems? Even as an example, it doesn’t serve well in this context. Organizational should include things like leadership, culture, etc.

Our conceptualization of what we term the ‘organizational-level’ factor encompasses what Durlak & DuPre, 2008 call the “prevention delivery system organizational capacity” and what Damschroder et al., 2009 call attributes of the “inner setting” (pg. 6). As we go on to explain later on that same page, we view organizational-level variables as representing aspects of the organization in which the innovation is being implemented.

4. I’ve never seen the term “structural” used to mean “outer setting”. Using this term is quite confusing. You need to provide the basis for doing this or use another term more descriptive to the outer/external context/setting you are trying to convey.

Though the term ‘structural’ was used in Damschroder et al., 2009 to represent aspects of the “inner setting” or the organization in which the innovation is being implemented, here we use the term ‘structural’ to refer to the structure of the larger sociocultural context in which the organization exists (e.g., funding climate, innovation-relevant policies). The term ‘structural’ is used quite commonly in research examining behavioral innovations designed to prevent and treat HIV/AIDS, and we now include additional citations from this literature to further clarify our use of the term.

5. The argument for including “patient” in the way you propose has problems. What is “implementation” at the patient level? It usually means use of the “clinical intervention” – or being engaged or any of the other myriad measures used in intervention trials. These measures are covered in randomized controlled trials typically as summative outcomes. Implementation research focuses on what it takes to implement those proven effective interventions. Thus, implementation is typically focused on process and the measures in your other four factors. You entire rationale applies wonderfully to developing clinical interventions (e.g., behavior change interventions). For example, health literacy may predict “efficacy” of the intervention itself. Your use of the term “implementation” in this context is inappropriate – what you do mean here? It may mean degree of use of the intervention by the patient, or degree of engagement or participation or other summative outcome measure that is unique to the intervention. Implementation in the context of “implementation research” is defined something akin to, “Efforts designed to get interventions and related products into use via appropriate activities” –thus, the focus is on targeted stakeholders, typically in organizations – ideally, informed by patients and with involvement by patients. But actually patient measures as you describe are tested as *summative* outcomes in earlier trials and perhaps confirmed in later effectiveness-implementation hybrid trials (see Curran, et al 2012).

Thank you for this suggestion. We have provided an expanded rationale to justify why patient-level factors are appropriate constructs to assess as predictors of implementation outcomes. Our discussion, which appears on pg. 7-8 of the revised manuscript, focuses on clarifying how the very conceptualization of fidelity—a common implementation outcome—directly implicates involvement
(and effects of) patients and how some implementation approaches, such as Community Based Participatory Research, also identify patient-level factors as important in this research.

a. Your last sentence in the 1st paragraph on page 8 is covered by other models under other constructs (ORCA: Evidence/Patient experience; CFIR: Patient Needs & Resources)
In our revised manuscript, this sentence no longer exists.

6. Why do you focus only on providers in your 5-factor framework? What about other clinicians (e.g., health psychologists) or staff (e.g., RN care manager)?
We use the term ‘providers’ as an omnibus term to refer to anyone who implements an innovation with a patient or client. We now include additional examples (e.g., counselors, nurses, physicians) to clarify that we are using provider as a general term on pg. 7.

7. Literature Search Results: you say you exclude education but then include it later on (page 15) and it’s listed in your Table 1
Thank you for identifying this discrepancy. The goal of the search was to identify measures utilized to assess factors predicting implementation of health-related innovations. Therefore, we retained measures that were utilized in studies examining health-related innovations even if they were in non-health contexts (e.g., education). However, we excluded measures that were utilized in studies examining non-health-related innovations (e.g., end-user computing systems) in any context. We have updated the text (pg. 14) and Additional File 2 in order to clarify this point.

a. You included nonpublished work in your count?
On pg. 14 we indicate that we excluded measures that were not published. This means that the search article that came up in our search used a measure that is not yet published (e.g., a measure from an unpublished dissertation). Thus, while the search article itself was published, the measure it utilized was not. In these cases, we excluded the measure from further consideration.

8. Your ORCA example under Patient doesn’t map to “patient feasibility” – and it isn’t clear what “patient feasibility” means
This sentence has been moved to the Results section in the revised manuscript (pg. 13). We have edited the sentence to refer to “patient needs” so that it better matches the example item we provide in parentheses.

9. Page 16: why are “institutional resources” part of “structural”? The last sentence seems to imply that “patient preferences” are also part of structural?
Thank you for noting this discrepancy. We re-reviewed our coding for that measure and agree that the institutional resources construct should be coded as representing the organizational-level factor. These changes have been made in text and in Table 1. We have also clarified the sentence regarding patient preferences to better reflect that we are discussing measures available to assess two separate factors.

10. The discussion needs to be considerably tightened up. There are many redundant and unneeded sentences:
   a. Example, Page 19: “...these measures might also provide important predictive information if they are used as predictors of implementation outcomes.”
We have edited that sentence to reduce redundancy.
b. Bottom paragraph, page 19: If measures of culture are readily available then why not include them in your synthesis?
As we have clarified in the text, we omitted measures utilized in research examining non-health-related innovations because they were beyond the scope of the paper.
c. You have three paragraphs (the last one is on page 20) that say your review is limited in seemingly the same general ways.
We have now restructured the Discussion to include each of our identified limitations in a new section called ‘Limitations and Future Directions’ (pg. 22-24)

11. Why did you exclude information technologies in a healthcare setting?
As noted above, we have revised the text to clarify that measures utilized in research examining non-health-related innovations were excluded from the current review.

MINOR ESSENTIAL REVISIONS
1. First paragraph, last sentence: are you saying that the US needs “new and innovative approaches…[the] most” out of other countries in the world? Is that true?
We have now eliminated the last clause of that sentence “where it is needed most” in order to avoid making a comparison between implementation needs in the U.S. versus other parts of the world.

2. The first paragraph starts off with the usual statement of need for implementation research but then leaps to the assumption that conducting theory-based implementation research will solve the problem. This needs better development. (I agree with this, by the way – you just need to make your case more clear)
We have revised the introduction substantially. In this revision, we have removed our assertion that theory-based implementation research will solve the problem. Instead, in paragraph 2 (pg. 4) we move to discuss current measurement limitations and how our article is positioned to help contribute to these limitations.

3. Paragraph 3, 1st sentence. This is true but also add that many studies are single/small sample case studies that produce new theories, adding to the mélange.
This sentence has been edited.

4. The role of qualitative data needs to be acknowledged somewhere in the paper. Simply developing psychometrically sound quantitative measures will not completely solve the problem.
We agree. We have added a sentence to this effect on pg. 24 in our ‘Limitations and Future Directions’ subsection.

5. Page 6, 1st paragraph: why “quality improvement: QUERI” in your parenthetical? QUERI is not synonymous with quality improvement.
We have edited this sentence accordingly.

6. You switch to using “broad types of factors” to indicate your five “factors” and then sometimes you refer to individual factors.
As noted above, we have clarified our use of the terms ‘factors’ and ‘constructs’ throughout the manuscript, including this specific instance.

7. Examples of unclear or repetitive writing:
a. 1st full paragraph on page 9: is this necessary? What point are you making?
We have now deleted this paragraph and incorporated its information in more relevant places in the text.

b. The following paragraph also has unclear meaning. It meanders and the last sentence skips to implementation outcomes which you said you weren’t including.

The purpose of the paragraph is to provide an overview of previous systematic reviews that have identified implementation-related measures so that the reader has an idea of how the current review fits into, and advances, the current literature.

c. Last paragraph under Methods – unclear paragraph. Unclear how you got from 330 to 146 if you used the same criteria for both.

As we describe in the paragraph, we first screened titles and abstracts for the inclusion criteria. Then, we took this remaining set of records, obtained full-text copies of the articles, read the articles, and applied the same inclusion criteria in order to further hone the sample of records. Figure 1 also provides a visual representation for this process.

d. Last sentence, main paragraph, page 14: very unclear

8. Methods: Your list of 13 is actually 12 because “research utilization” is listed twice

Research Utilization is, in fact, listed twice, once with the U.S. spelling (utilization) and once with the U.K. spelling (utilisation). We included both as keywords because some databases do not automatically search for both types of spelling.

9. Organization definition: your parenthetical seems to present a new idea, not an example for either leadership or culture

We have edited this example accordingly.

10. Last paragraph on page 13, Sentence beginning “Though these two perspectives...” – I thought I knew what you were saying but you lost me here

We have edited this paragraph (and the one that precedes it) in order to be clearer that we are coding items and that we assessed the content of each item, not the perspective (provider vs. organizational administrator) being assessed.

REVIEWER #2: KATH WRIGHT

Discretionary Revisions

Please note that I do not have access to the EBSCO database interface so cannot comment fully on how the search strategies as described could be put into operation. Nevertheless, can still comment on the overall approach. The choice of databases was appropriate and was supplemented by a handsearch of the Implementation Science journal.

Thank you for this feedback about our search strategies.

MEDLINE search

All of the search statements have been restricted to the abstract field. Not clear why the title field hasn’t also been searched; I would have expected this to be an equally useful field.

We appreciate your specific feedback about our search strategies. We agree that searching for our primary keywords in the title field would expand our search in a useful way. In the revised
manuscript, we now search for these keywords in the title of our MEDLINE and Implementation Science search. Please see pg. 11 of our Method section and Additional File 1 for these changes.

Although MeSH does contain relevant terms (e.g. questionnaires) none of the search statements have utilised this facility.

Neither has any of the search statements made use of either truncation or adjacency (e.g. using the phrase “diffus$ adj2 innovations“ would have identified many more records and increased the overall sensitivity of the search).

Limiting to “health” in the abstract field is unlikely to identify all potentially relevant records as there will be many instances where reference is made to the specific clinical topic or area of medicine rather than “health” e.g. dentistry, paediatrics, oncology, renal medicine, surgery etc etc

For each of the above three comments, the reviewer points out important considerations regarding additional techniques that could broaden our search. We gave these items careful consideration as we revised our manuscript. Like other researchers conducting systematic reviews, our goal was to identify a scope for our review that would allow us to investigate our core questions while balancing available resources (Higgins & Green, 2011). In essence, our search strategy was designed to offer a thorough and efficient search of the literature, maximizing the “signal to noise” ratio in our search and minimizing false positives.

With regards to MeSH terms, specifically, we did originally consider their use in our search strategy. We pilot-tested use of MeSH terms with our three sets of keywords (implementation science, measure, health) but decided against using them because they did not contribute to the goal of efficiently identifying relevant measures. For example, changing our Medline search (detailed in Additional File 1) to use MeSH terms (with explode) for our implementation science keywords instead of Topic yielded no new search records. Alternatively, using MeSH terms (with explode) for all three types of keywords (implementation science, measure, health) without the validation study restriction yielded 31,000 results. In essence, using MeSH terms for all three types of keywords would have exponentially increased the false positives in the search, thereby diminishing the effectiveness of the search (i.e., 31,000 records cannot be hand screened).

For these reasons, we opted to utilize an a priori list of terms from a review publication by Rabin et al., 2008 rather than to utilize truncation or adjacency search techniques. As our other reviewers have noted, the progress of the field of Implementation Science is hampered by an overabundance of diverse terms that essentially refer to the same phenomena. Thus, we didn’t want our search to contribute to these terminology inconsistencies. Therefore, we used the thirteen “glossary” terms identified by several experts in the field to guide our search.

Our use of the word “health” rather than specific health disciplines was also intentional and we gave this decision much consideration. Our use of this broad term that cross-cuts many individual health subfields (e.g., cancer, HIV/AIDS, diabetes) is designed to allow our search to represent a reasonable cross-section of the field of implementation science. One of the most significant barriers to progress in our field is that dissemination and implementation research has historically been siloed within individual subfields in which findings are not easily disseminated across disciplines. The consequence is that each subfield ends up developing its own discipline-specific frameworks and measures rather than contribute to the advancement of a larger field of study. Because our goal was to conduct a
review that would advance the study of implementation science in health-related domains as a whole, we opted to use the keyword “health” rather than to focus on individual subfields.

We realize, however, that these choices shape the nature of the articles we reviewed and the way in which our results should be interpreted by our readers. In our revised manuscript, we discuss these as points for future research to consider in the ‘Limitations and Future Directions’ subsection starting on pg. 22.

Reference:

PsycINFO search
The search conducted in PsycINFO was restricted to the Tests and Measures field (labelled TM) where tests named in the paper are listed. This approach may have unexpected outcomes however. For example, when a search for “diffusion of innovation” in the TM field is executed the search engine identifies records that include the single word “Diffusion” and not the phrase as specified. So records that describe or refer to the “Career Decision Scale-Identity Diffusion Subscale” are identified. Similarly, searching for “knowledge to action”, “knowledge transfer”, or “knowledge translation” in the TM field identifies records where the one word “knowledge” is part of the test/measure’s name e.g. Number Knowledge Task, Affect Knowledge Test, Infant Feeding Knowledge Test-Form A, Multicultural Awareness-Knowledge- Skills Survey, Metapragmatic Knowledge Test. None of these seem to be of relevance to the project being undertaken.

The Tests and Measures (TM) field code was utilized to limit false positive records that would occur by assessing for these keywords anywhere within the article. As this reviewer correctly notes, the TM field code does have its limitations, however, whereby some false positives would be retained in our search. We did not find these false positives to be prohibitive in our search. Further, these false positives were vetted during subsequent coding phases.

It would have perhaps been preferable to have widened the search to include other fields within the record, especially as “knowledge translation” is used as a key concept and “knowledge transfer” is used as a subject heading.

We agree that expanding the search to obtain records that included the implementation science keywords in any field would have greatly increased the number of identified records. However, this search would have also dramatically increased the number of false positive records retained for further coding, thereby diminishing the overall efficiency of the search. Because we were only interested in assessing measures of implementation science constructs, rather than reviewing the entire literature of implementation science (including qualitative papers, theory or framework papers, etc.), we believe that using the Tests and Measures field to increase the accuracy of our search records is warranted and necessary. See below for a further discussion of how our search results would have changed without the use of this feature.

In addition, not all records in the database include the TM field so perhaps it’s an overly restrictive approach. For example, a paper with title “Knowledge translation in cancer services: Implementing the Research and Evidence in Practice model” sounds potentially relevant but there is no TM field in this record although the key concepts include “knowledge translation” and the subject heading “knowledge transfer”.
We did consider using other search strategies. As mentioned previously, we gave careful consideration in identifying the search strategies that would allow us to most efficiently investigate our core research questions while balancing available resources. For example, searching our dissemination keywords within the abstract or title in PsycINFO, and then limiting results to when the terms questionnaire, scale, measure, tool, or survey was available in the abstract or title, along with the word health (in abstract or title), would have resulted in over 4000 results. Hand reviewing 4000 records would not have been feasible given a reasonable amount of human resources available (Higgins & Green, 2011).

We believe, therefore, that despite some limitations, using the TM feature combined with our searches in Medline, CINHAL, and the Journal Implementation Science provided the best balance between completing our research goals with our available resources. As noted previously, we realize that these choices shape the way in which our results should be interpreted by our readers. In our revised manuscript, we discuss these limitations on pg. 22-24.

Furthermore, it is important to note that expansion of our search strategies from our original submission to our revised manuscript yielded only 9 new measures. In our first submission, our search criteria yielded 330 records for further hand screening, a pool of records that yielded 52 measures (16% “hit” rate). In our revision, our search criteria yielded an additional 268 records but this pool of records only yielded 9 new measures (62 total measures; 10% “hit” rate). Thus, though we expanded the search, the overall efficiency of the search diminished by 38%. These comparative data suggest that further expansion of the search would yield diminishing returns.

Finally, in order to address these collective concerns in the manuscript, we have added a new paragraph of information on pg. 11 in which we discuss research which supports the utility of using field codes such as those we have adopted in order to efficiently identify high-quality search records related to the topic. To our knowledge, none of these empirical reports provide any caution regarding the possibility that use of these or other field codes could be too restrictive.

Also, please note in Additional File 1 that we have changed our search details to reflect the fact that searching Medline using EBSCO HOST was no longer available when we updated our search in August 2012. Thus, we switched from using EBSCO HOST to using Web of Science. However, this did not affect the overall number of search records located. The updated search identified all of the previously obtained results from the original Medline search in EBSCO HOST, and it obtained 86 new results.

REVIEWER #3: HUTCHINSON

Thank you for the opportunity to review this manuscript. This article provides a valuable overview of the instruments available to measure constructs related to implementation of innovations in health care. The findings are of practical value to researchers and implementers who wish to examine and understand constructs relevant to implementation efforts.

Thank you for your feedback
• Major Compulsory Revisions
  Nil

• Minor Essential Revisions
  1. Background, paragraph 5, should read “the goal of the current paper is to
We have now clearly outlined the three goals of the paper in the abstract, introduction (pg. 5), and discussion.

• Discretionary Revisions
1. Given that this article is directed towards an international audience, I strongly recommend that the opening sentence of the abstract and the opening paragraph of the manuscript be re-worded so that the focus is not on the U.S.

Thank you for this feedback. We now reference the need for broader implementation in a wider international context, and we have added two additional references in support of these claims.

2. I suggest that you clarify that your review aimed to identify high-level or broad constructs (which you later refer to as factors) to predict implementation outcomes. Within each broad construct/factor (for example, organizational) are multiple constructs that potentially predict implementation.

We agree that the previous terminology was confusing. We have now clarified our terminology throughout the text such that we use the term ‘factor’ to refer to a high-order category or class and the term ‘construct’ to refer to the lower-order variables that are subsumed within factors. We delineate between factors and constructs on pg. 4-5 and then further describe the distinction when we review factors (vs. constructs within each factor) on pg. 6-8.

3. The section labelled Conclusions seems to be discussion and conclusions and may be better labelled as such.

We have labeled these sections of the manuscript accordingly.

4. In additional File 2 you list 11 instruments that did not assess any of the 5 factors included in your theory-informed framework. Could you comment on the types of factors they did assess – presumably there are dimensions, outside your framework, that can be measured and are also important to implementation efforts? This would be of interest to implementers and implementation scientists.

We now provide a parenthetical example of this exclusion criterion on pg. 14.

Minor issues not for publication
1. Typographical error noted in Method, paragraph 1 - “…sections of recordds via PsycINFO…”.

Thank you for identifying this typo. We have fixed this error.

REVIEWER #4: HEATHER COLQUHOUN
Reviewer’s report:
I appreciated the opportunity to review this interesting and valuable work. I think the paper has excellent merit and if the authors are able to address the major compulsory revisions outlined below, the paper has good potential for publication.

Thank you very much for this feedback.

Major Compulsory Revisions
1. Background: The background section had several places where the purpose of the SR is stated but the objective of investigating criterion validity was not indicated. The criterion validity analysis seemed to be introduced in the results section. There are several places in the background in which the issue of causal
pathways was discussed so perhaps this could be inserted in those areas.

We have revised our manuscript to clarify the primary goals of our paper: 1) identify a multi-level framework that captures the predominant factors that impact implementation outcomes, 2) conduct a systematic review of available measures assessing constructs subsumed within these primary factors, and 3) determine the criterion validity of these measures in the search articles. These goals are now articulated more clearly within the abstract, the introduction (pg. 5), and again in the discussion section (pg. 18).

I wondered if there are actually three objectives: 1) Combine existing theory-based frameworks to develop a comprehensive framework in which to organize important groupings of constructs (factors), search for measures that will measure constructs within these factors, and establish whether or not the constructs have been shown to be predictive (within the studies obtained). Consider outlining these activities in the section just before “A multi-level framework guiding IS research”. As it is now, there is too much activity related to how you developed your five-factor framework in the background section and the exact nature of how this proposed framework fits into your study objectives is uncertain. You do seem to be proposing a new framework here, not just utilizing an existing one. I really liked that this work was two-fold. It provides a framework on which to understand important factors and it provides information on measuring constructs within those factors. It makes more sense to highlight this framework as part of the study objectives.

We agree that clearly articulating our goals would strengthen the manuscript. As we note above, we have revised our manuscript to clarify the primary goals of our paper: 1) identify a multi-level framework that captures the predominant factors that impact implementation outcomes, 2) conduct a systematic review of available measures assessing constructs subsumed within these primary factors, and 3) determine the criterion validity of these measures in the search articles. These goals are now articulated more clearly within the abstract, the introduction (pg. 5), and again in the discussion section (pg. 18).

2. Methods: In general, there were many concepts in your results that should have been stated in the methods (how measures were excluded, how coding discrepancies were resolved, identifying the criteria and definitions of each of the five factors, the process of coding for criterion validity). You seemed to have a process of excluding studies, followed by a process of excluding measures. I think outlining how you excluded for both of these activities are part of the methods. In general, the results section really seemed to begin at “Our search found 52 measures” - 3.5 pages into the results. We agree that this detail might be best positioned in the Method section. However, layout requirements differ between journals. The information provided in the methods and results section complies with the requested format from Implementation Science.

3. More information was needed as to what variables were extracted from the articles and how certain coding was achieved. Did you use the construct name reported in the articles or did you have a prior listing of constructs that you slotted the measures (or items) into? Did a measure only need to have one item related to a construct to be included? We have now clarified our coding procedure in the Results section (pg. 14) and have elaborated on our item-focused coding procedure.

4. Results: As indicated above under the methods, there seemed to be a lot of method oriented information in the results section. As we note in the response letter to the editor, we were asked to move this information from the Method to the Results section.
5. Literature Search Results (first paragraph): The paper indicated that you excluded based on education but then also used education as one of your descriptions (paragraph before “Factors Assessed in Measures” and in paragraph preceding “Conclusions”. This requires some clarification. Did you include education as related to health?

Thank you for identifying this discrepancy. The goal of the search was to identify measures utilized to assess factors predicting implementation of health-related innovations. Therefore, we retained measures that were utilized in studies examining health-related innovations even if they were in non-health contexts (e.g., education). However, we excluded measures that were utilized in studies examining non-health-related innovations (e.g., end-user computing systems) in any context. We have updated the text (pg. 14) and Additional File 2 in order to clarify this point.

6. You indicated that most of the measures were in a health care setting, which was true, but this was only 73%. So 27% were not in a health care setting. Can you address why so many were found outside of health care given your search strategy?

Please see response to Item #5 above. In addition, we have added language to the paragraph in which we discuss results based on implementation context (pg. 17) to clarify that there was a minority of studies that evaluated health innovations in non-health contexts.

7. Limitations: There does not seem to be a paragraph on limitations. You touch on a few limitations in the conclusion but having them all presented in one place as limitations would be useful.

We have now created a ‘Limitations and Future Directions’ subheading within the Discussion (pg. 22-24) which outlines a series of limitations of the current review and how future research could address these limitations.

Minor Essential Revisions

1. Abstract: The issue of a need for ‘standardized constructs’ was raised in the abstract but this issue was not necessarily fully addressed by the study (proposed framework and systematic review [SR]). The framework does provide a standardized list of factors in which multiple constructs will fall but it remains a high level list and doesn’t necessarily provide a comprehensive list of constructs or even a set of sub-categories that might be used to standardize constructs. I think it is worth indicating (perhaps the conclusion or even when the five-factor framework is proposed) that this work progresses our goal towards more standardized constructs by suggesting a structure on which we can start to standardize constructs.

Thank you for this point. We now address the fact that the current review advances conceptualization and measurement of factors and constructs subsumed within factors on in the Discussion on pg. 18-20.

We agree that the previous terminology was confusing. We have now clarified our terminology throughout the text such that we use the term ‘factor’ to refer to a high-order category or class and the term ‘construct’ to refer to the lower-order variables that are subsumed within factors. We delineate between factors and constructs on pg. 4-5 and then further describe the distinction when we review factors (vs. constructs within each factor) on pg. 6-8.

2. Abstract and Background (first paragraph): Can you add a reference from another country, in addition to the US figures? Obviously, issues of implementation extend beyond the US, so revising these sentences to reflect that would be useful and ensure an international context and appeal. The same
applies to the last sentence of the first paragraph in the background – you could potentially just remove the “in the US where it is most needed”.

Thank you for this feedback. We now reference the need for broader implementation in a wider international context, and we have added two additional references in support of these claims.

3. Background (last paragraph): I was not sold on the term ‘impact theory’ but I cannot come up with a viable alternative. Do you have a reference for the term ‘impact theory’? That might help as opposed to introducing what seems like a new term into our theory lexicon.

We see that introduction of the impact vs. process theories terminology detracts from the flow of our ideas. For purposes of clarity, we now only mention parenthetically that many existing frameworks we draw upon in this manuscript are impact frameworks on pg. 4.

4. You indicated including only self-report measures. A rationale for this choice is warranted. Could there be useful organization measures that are not self-report?

Thank you for identifying this discrepancy. We have removed the phrase ‘self-report’ from this sentence as it is not an accurate description of our search strategy. Though many of the measures identified in the review were self-report, this was not used as an inclusion criterion.

5. You searched for both the use of a measure and the development of a measure. I can see the importance of searching for both but I was not certain as to how newly developed measures would be integrated into your analysis, particularly for the criterion validity objective. Newly developed instruments might still be in the process of initial development and we would not expect to have predictive work done on them just yet. Is this a limitation of your analysis?

We now include a sentence in the results section to orient the reader to frame their interpretation of criterion validity based on the age of the measure (pg. 19): “In addition, it is important to keep in mind that the criterion validity of recently developed scales may be weak solely because these measures have been evaluated less frequently than more established measures.” Further, we have also added comments that address limitations of our criterion validity assessment (pg. 25) in the discussion.

6. Methods (first few paragraphs): The spelling of CINAHL needs to be corrected in two places. (CINHAL to CINAHL)

Thank you for identifying these typos. They have been corrected in the text.

7. I am not sure that your chosen databases would have allowed you to find all organization or structural related measures. Consider introducing this as a limitation in the conclusion.

In the ‘Limitations and Future Directions’ section of the Discussion, we now address the limitations of our database search (pg. 23-24): “Our use of multiple databases (i.e., MEDLINE, CINAHL, PsycINFO) which span multiple disciplines (e.g., medicine, public health, nursing, psychology) and search of the Implementation Science journal provides a broad cross-section of the empirical literature examining implementation of health-related innovations. Despite this broad search, it is possible that additional relevant literature and measures could be identified in future reviews through use of other databases such as ERIC or Business Source Complete, which could catalogue additional health- and non-health implementation research from other disciplines such as education and business, respectively.

8. Results (first paragraph): Your reliability was high for what seems like some un-straightforward variables. Can you clarify in the paper that the reliability was calculated based on everything you extracted?
We have now elaborated on our assessment of interrater reliability by describing that reliability was assessed by independently coding a subset (25%) of items at each of the five stages of coding. This expanded description now appears on pg. 12 of the revised manuscript.

9. Results (first paragraph): Some clarification was needed for “provided insufficient information for review”. What types of things were lacking such that the measure was excluded for this? We have now provided parenthetical examples for each of the measures exclusion criteria noted on pg. 14-15, including “provided insufficient information.”

10. The last paragraph before conclusions seems a bit out of place. I wonder if it should have been placed just before the section on criterion validity.
Thank you for the suggestion. We have made this change such the implementation context information now precedes the criterion validity information (pg. 17).

11. Conclusion (first paragraph): You switched to using ‘characteristics’ instead of ‘constructs’. I suggest keeping the language at constructs. You used the term characteristics earlier in the paper (abstract and background) but once you switch to constructs, I would keep with that term. You also indicated in this paragraph that you examined five constructs but I’m not sure you did. Rather, you investigated constructs found within five factors. It would be useful to stay with five factors, constructs within each factor, constructs representing these 5 factors, etc.
Thank you for pointing out these inconsistencies in our terminology. As mentioned above, we now use ‘factor’ and ‘construct’ consistently throughout the text, including text within the Conclusion section of the manuscript.

12. The background made the case for more theory-based measurement approaches and you used a theory-guided framework to structure your SR but this was not reconsidered in light of the findings in the conclusion. Commenting on the value of using the framework would make a strong case for why theory-based approaches are needed.
Thank you for the suggestion. We now conclude our paper by discussing the need for theory in order to advance implementation science.

13. It would be useful in the conclusion to return to the French et al review to comment on how many of your organization measures were included in that review.
We now include a sentence in the ‘Limitations and Future Directions’ section on pg. 23 regarding the overlap of measures identified in the two reviews: “Indeed, a recent review of a wide array of organizational literatures yielded 30 measures of organizational-level constructs [38], only 13% of which overlapped with the current review.”

14. Table 1: This is a great table. My only comment is that the search article section was not clear. It took me some time to see that it was the manuscript reference number. Consider adding this clarity to the column. Also, I was not certain why some of the source articles from the first column were not also in the list of search articles. I was under the impression that the source articles would have still been in the review.
As noted, there were 2 measures for which the source article was not also listed as a search article. We now include this information in Table 1. Further, confusion might also arise from the fact that for some measures, the source article was not always listed as the first article in the Search Article and Criterion Validity columns, making this presentation a bit confusing. In the revised manuscript, Table 1 now lists all source articles as the first search article. It’s also important to note that there are a few
measures for which we note the source article but also list additional articles from which the measure was adapted (e.g., Competing Values Framework). Because previous versions of the measure were substantively different than the current version, we do not assess criterion validity for them.

Discretionary Revisions
None