Linking Measures to Mechanisms of Action in Behavior Change: A Qualitative Analysis of Expert Views

Lilly Derby[1,2], Lauren Connell Bohlen[3], Susan Michie[4], Marie Johnston[5], Jeffrey L. Birk[1],

Alexander J. Rothman[6], Talea Cornelius[1]

[1]Center for Behavioral Cardiovascular Health, Columbia University Irving Medical Center, New York, NY, 10032, USA
[2]Department of Psychology, Rutgers University, New Brunswick, NJ, 08901
[3]Department of Behavioral and Social Sciences, Brown University School of Public Health, Providence, RI, 02903, USA
[4]Centre for Behaviour Change, University College London, London, UK
[5]Aberdeen Health Psychology Group, University of Aberdeen, Aberdeen, UK
[6]Department of Psychology, University of Minnesota, Minneapolis, MN, 55455, USA

Word Count: 4,000
Tables: 2
Figures: 0

Corresponding author: Lilly Derby (ld2778@cumc.columbia.edu). Center for Behavioral Cardiovascular Health PH9-312, Columbia University Irving Medical Center, New York, NY, 10032, USA

Abstract

Rationale: Testing mechanisms of action (MoAs) hypothesized to underlie behavior change can enhance intervention effectiveness. Rigorous measurement of putative mechanisms is critical to this effort, but measures are rarely validated with respect to target MoAs. **Objective**: This study aimed to elucidate challenges of linking measures to putative MoAs and to identify priorities for future research. **Method**: This study was a systematic exploration of written comments by experts in behavioral intervention research and theories of behavior change ($N = 20$) capturing their opinions about a task querying whether self-report measures from the Science Of Behavior Change (SOBC) Measures Repository were related to a set of MoAs identified by the Human Behaviour Change Project (HBCP). **Results**: Six themes were identified: 1) Study Value, 2) Measure Properties, 3) Mechanism Properties, 4) Miscellaneous Measure Concerns, 5) Conceptual Challenges, and 6) Approaches to Developing Measure-Mechanism Links. Experts noted challenges such as lack of measure validation, poor measure properties (e.g., double-barreled items), overly broad MoA definitions that limited their utility, lack of clarity around the term "related," and more. Nonetheless, experts expressed the importance of the exercise. Suggestions included developing and refining measures that are validate for assessing MoAs, clarifying and elaborating MoA definitions, and conducting further, more granular research. **Conclusion**: This systematic examination of expert comments highlights issues that need further investigation to advance behavioral science, specifically pertaining to identifying valid measures of MoAs in behavioral and process research. This study highlights the challenges and opportunities for future research on linking measures and MoA in behavioral science and subsequently enhancing the efficacy of behavioral interventions.

Introduction

Understanding mechanisms of action (MoAs) through which behavioral interventions exert their effects can bolster intervention efficacy (Riddle & Science of Behavior Change Working Group, 2015; Sumner et al., 2018), improving health and quality of life. Valid measurement of hypothesized MoAs is critical to this effort.

To advance the measurement of MoAs, researchers at the NIH Science Of Behavior Change (SOBC) initiative and the Human Behaviour Change Project (HBCP) collaborated in conducting an expert opinion study, which identified 167 links between 44 measures in the online SOBC Measures Repository and 26 MoAs identified by the HBCP (links meaning that the measure and MoA were "related"; Cornelius et al., 2023). These links were incorporated into the SOBC repository and the HBCP Theories and Techniques Tool (TaTT), allowing researchers to navigate between these scientific resources as they formulate hypotheses about behavioral mechanisms and consider study design.

The SOBC Measures Repository includes measures of MoAs hypothesized to underlie behavior change across three broad domains: stress resilience and stress reactivity, interpersonal and social processes, and self-regulation (SOBC Measures Repository, 2017). The repository includes information about evidence of the measures' use within an experimental medicine approach, open science documentation, citation information, question items, and scoring. The HBCP TaTT (Theories and Techniques Tool, 2018) summarizes evidence for putative links between 74 behavior change techniques frequently found in intervention research and 26 MoAs. Links were identified by triangulating data from a literature analysis and an expert consensus study (Bohlen et al., 2019; Carey et al., 2019; Johnston et al., 2021).

The purpose of this qualitative study was to systematically identify experts' opinions of study strengths, challenges, and limitations to illuminate issues in the curation and development of measures of MoAs and to guide future research efforts to improve measurement, rigor, and transparency in behavioral intervention research.

Methods

Study Design

A three-round expert opinion exercise was conducted to establish the extent to which 44 self-report measures from the SOBC Repository were "related" to 26 MoAs in the HBCP TaTT. The second round of the consensus exercise was an anonymous, asynchronous discussion of 10 measure-MOA pairs for which experts disagreed on, the comments from which are the focus of the present qualitative analysis.

Participants

Thirty international experts (15 US-based) were recruited through professional listservs and scientific networks/societies that focus on health behavior research. Participants had self-reported high levels of expertise in theories of behavior change research and represented diverse disciplines and geographic locations. Participants could earn up to $200 USD. Of the 30 experts, 23 participated in Round 2, and 20 provided comments (Table 1).

Procedure

This study was approved by Columbia University Institutional Review Board. In Round 1, experts rated whether 84 measure-MoA pairs pre-identified by the study team were "related" and suggested additional pairs. Pairs reaching 50% agreement, or new pairs suggested by ≥50% experts, were considered links. Round 2 was a two-week, asynchronous, online discussion about 10 selected measure-MoA pairs that had not reached 50% agreement in Round 1 (see

). This study analyzed

comments from these 10 measure-MoA pair discussion boards and one general board (11

discussion boards in total; Cornelius et al., 2023).

Code Development and Analytic Plan

Codes were developed using an inductive approach (Thomas, 2006). [LD] and [TC] read

all comments to immerse themselves in the data and identify overarching themes. Preliminary

definitions for each theme were developed iteratively via discussion to ensure saturation and

distinctness. A final dictionary was created and approved by [LCB]. All data was independently

coded by [LD and TC], resolving discrepancies through discussion with [LCB]. Interrater

reliability was calculated using NVivo (QSR International Pty Ltd., 2020). Agreement ranged

from $\kappa = 0.46$ to 0.86.

Personal details at the time of this study that may influence data interpretation are

provided. [LD] is a White cis-gender female with a bachelor's degree in neuroscience, a clinical

psychology doctoral trainee, and a senior-level project coordinator for SOBC. [TC] is a White,

cis-gender female, with a doctoral degree in social psychology, master's degrees in clinical

social work and biostatistics, and was Assistant Professor of Medical Sciences and co-

investigator at SOBC. [LCB] is a White cis-gender female with a doctoral degree in

biobehavioral health, and was Assistant Professor of Behavioral and Social Sciences, contributed

to identifying and coding MoAs with the HBCP, and developing the TaTT.

<div align="center">

Results

</div>

The 10 measure-MoA forums had a mean of 17 comments (*Range* = 12, 19). Six themes

were identified: 1) Study Value, 2) Measure Properties, 3) Mechanism Properties, 4)

Miscellaneous Measure Concerns, 5) Conceptual Challenges, and 6) Approaches to Developing

Measure-Mechanism Links. Elaborated definitions are in Table 2.

1. Study Value. Despite concerns (elaborated below), experts agreed that validating measures

with respect to MoAs is a crucial but often neglected task, presenting an impediment to scientific

rigor and progress. Relevant expert comments included: "*the linking of measures to mechanisms*

is an important and valuable undertaking" (P12), and "*insights will come out of this exercise,*

even given the [study] limitations" (P18)*.*

2. Measure Properties. Poor measurement properties were cited as a serious barrier to

conducting rigorous empirical tests of mechanistic hypotheses in behavioral intervention

research. Experts stressed the foundational importance of valid and reliable measurement and

appropriately cautioned scientists against using flawed measures, stating, *"the items have a lot of*

issues and it is not a measure I would recommend for assessing the construct" (P18), and

"Addressing the construct validity of the measures is an issue that should be front and center in

having robust literature, repository, etc. of measures of MoAs" (P12). Another expert stated that,

"many of the measures I viewed were fundamentally flawed in their items and/or development.

E.g., double-barreled, or conditional items, inappropriate or missing psychometric evaluation,

etc." (P2). Many of the measures included in this study are widely used (e.g., the Emotion

Regulation Questionnaire scale development paper cited 13,643 times; Gross & John, 2003),

which underscores the breadth of the measurement issue, challenges of rectifying flawed

measures, detrimental impacts on mechanistic hypothesis testing, and the urgency for concerted

efforts to improve and uphold measurement standards.

3. Mechanism Properties. Clear definitions are critical when considering the ability of experts

to come to consensus regarding measurement of MoAs or to draw conclusions regarding the

consistency of evidence for (or against) a given MoA underlying behavior change in the published literature. Although experts agreed that these definitions were clear and helpful for the study tasks, they did not necessarily agree with MoA definitions and concerns were raised that some properties of the MoAs were an impediment to creating measure-MoA links that were valid, specific, and useful. Comments such as, "*self-efficacy is related to pretty much every construct in health behavior change process*" (P18) illustrate the limited utility of broad MoAs in formulating specific mechanistic hypotheses. There was also some debate regarding whether certain MoAs were modifiable.

4. Miscellaneous Measure Concerns. Concerns regarding measures that could have impeded the creation of measure-MoA links extended beyond flawed measurement properties alone. Some commented that they identified a link only after *"looking at the details of the specific statements"* (P10) rather than simply relying on the measure title or description, which could be vague or inaccurate. Others pointed that MoAs may differ across subscales (e.g., *"Additionally, I thought each of the sub-scales of the TFEQ better aligned with other mechanisms (e.g., emotional eating = behavioral cueing)"* (P15), and, *"[I] saw them as 'related' based on the subscale 'Future', including 13 items such as, 'I complete projects on time by making steady progress'"* (P13)), which presented a problem insofar as the present study asked experts to link MoAs to measures as a whole. It also became clear that measures cannot be stagnant—rather, validity and rigor in measurement changes over time and across contexts. One comment, in particular, highlighted this issue: *"Some of the language in some of the measures can be outdated and culturally specific, preventing them from universal promotion. Some caveats or boundaries on recommended use of measures might be appropriate for some measures"* (P20).

5. Conceptual Challenges. Several overarching conceptual challenges limited experts' ability to create measure-MoA links. Foremost, there was confusion about the definition of "related" when creating links, specifically whether this meant that a measure had something to do with a given MoA v. whether it actually assessed the MoA. One expert called explicit attention to this, stating that they "*found that it was not clear what was meant by 'related.' Some experts interpreted the task as judging to which degree the measure assesses the MOA, others did not*" (P18). In reply, another wrote, "*I have noticed in responding to the discussions […] that for some people they are responding as to whether the measure in question is measuring the mechanism of action in question where as others (including myself) are responding in relation to whether the measure would be related (but not necessarily measuring) the mechanism. This may be important for explaining some of the variance in responses*" (P9).

Properties of the measures and MoAs also contributed to conceptual challenges. As stated by P2, "*[i]t's dangerous to recommend measures for a construct if those measures are not construct-valid. Although a 'link' between a measure and MOA exists (the links we are drawing, in this consensus exercise), this does not mean that that measure assesses the entire construct of interest and ONLY that construct of interest, for the intended population, context, behavior of interest, etc. Therefore, regardless of a conceptual 'link,' the measures are not valid for use as measures of the MOAs.*" One cannot create a valid link under such circumstances. Overly broad definitions of MoAs compound this conceptual issue.

6. Approaches to Developing Measure-Mechanism Links. Some experts offered suggestions for future research to help develop mechanism-specific measures or improve evaluation of measure-MoA links, such as assessing the relative strength of measure-MoA links (P11) and conducting *"follow-ups of such expert rounds on more pre-selected measures (pre-selected for*

quality and construct validity and assessing the constructs that can be modified through

interventions)" (P18). An acknowledgment to the deep-seated issue of valid measurement of

hypothesized MoAs using existing measures, another stated, "*A more fruitful effort would be to*

develop measures from scratch, for the purpose of assessing each MOA, and then using the

existing measures, that we have been rating, for validation (i.e., for convergent and discriminant

validity in particular)" (P2).

Discussion

Building on previous research advocating for valid measurement of MoAs to uphold

rigorous and transparent science (e.g., Nielsen et al., 2018), this study is the first to elucidate the

complexities of utilizing existing measures to assess MoAs hypothesized to underlie behavior

change by examining expert feedback. Themes identified in this study confirm enthusiasm for

efforts to link measures to MoAs and offer novel insight into challenges that hinder progress in

this domain. Challenges spanned measure- and MoA-specific concerns, which compounded

conceptual difficulties in identifying measure-MoA links. These measures are frequently used in

behavioral science, which underscores the urgency of developing, refining, and curating

measures that are appropriate for assessing MoAs (Flake & Fried, 2020).

The task of identifying measure-MoA links is not straightforward, and experts' comments

highlighted the nuance of these challenges. Factors such as poor measure quality (e.g., lack of

construct validity, double-barreled items) extend beyond criticism of the measure alone—these

issues limit accurate assessment of intended MoAs and there may be substantial heterogeneity in

individuals' understanding of the underlying construct. Furthermore, measures must undergo

continuous validation across time and contexts to ensure that the underlying MoA assessed

remains the same. One cannot create valid measure-MoA links if the measure is interpreted

differently by different readers. Similarly, the broadness of some MoA definitions (e.g., self-efficacy) leads to a lack of specificity, such that multiple MoAs may be captured within a single definition, and measures that assess these broad MoAs would be—by design—non-specific, non-comprehensive, and heterogeneous, leaving us with the same problem we began with. Experts also had varied interpretations of the term "related" when judging measure-MoA pairs. This is understandable (it is a broad term) and was an intentional decision when designing the study to err on the side of including links, given the preliminary nature of the work. Indeed, the quantitative results of this work were intended to stimulate critical thinking about measures of MoAs and improve the interoperability of SOBC and HBCP resources for behavioral scientists, rather than to recommend the use of specific measures. Beyond cautioning against overreliance on the quantitative results of this study (i.e., confirming measure-MoA links using the term "related"), discussion of the term "related" highlights the larger issue of the subjectivity inherent in an expert consensus study, particularly when grappling with broad and potentially ambiguous constructs. Future efforts should carefully consider these limitations and explore strategies to enhance clarity, such as pre-selecting measures with demonstrated quality, using comprehensive, clear, and specific MoA definitions, and implementing more stringent criteria in creating measure-MoA links.

We are not the first to draw attention to poor measurement practices (Flake & Fried, 2020). This remains an urgent issue that undermines behavioral intervention research at its core: if we are not measuring the constructs we intend to measure, and if these definitions are not shared across studies, then conclusions of hypothesis tests and research syntheses are not valid. Results of our qualitative study illustrate the inherent challenges of linking measures to MoAs, but this knowledge will enable us to address these challenges with greater efficiency and

precision. Extending these efforts to include a larger set key MoAs and to identify a core set of reliable and valid measures—whether these measures already exist, must be validated or refined, or must be created—is a huge but necessary undertaking, requiring the collective effort of numerous behavioral scientists. Furthermore, as MoAs become more narrowly defined within a specific study context, so, too, must the measure of that MoA. Rigorous testing of content validity and discriminant content validity is paramount. We must ensure that measures not only assess the target construct and the hypothesized MoA (completeness of content validity; Bell et al., 2017; Dixon & Johnston, 2019; Johnston et al., 2014) but also comprehensively capture it, avoiding contaminations with other constructs. For example, Williams (2010) noted that measures of "self-efficacy" may be contaminated with the construct "outcome expectancy," confounding subsequent attributions to self-efficacy alone. Increasing journal requirements for describing measure selection and measure development studies, and/or increased funding for measure development and validation research may help to incentivize these efforts, ultimately enhancing the rigor, reproducibility, and conceptual transparency of our research.

Limitations & Future Considerations

Expert comments were limited to 44 self-report measures from the SOBC Measures Repository, which were not required to meet any minimum quality standard. Although the 26 MoAs identified by the HBCP represent those frequently found in the literature, these are only a small fraction of those represented in the MoA Ontology (Schenck et al., 2023), which is part of the Behaviour Change Intervention Ontology (Michie et al., 2020). Additionally, only 10 measure-MoA pairs were discussed, 20 of 30 experts participated in the qualitative component of the study, and experts' psychometrics expertise was not assessed.

Future studies should take a more granular approach by examining the MoA Ontology (Schenck et al., 2023), separating measures by subscale, and even testing the correspondence of single items to MoAs (e.g., to test whether a given measure captures one v. multiple mechanisms). Future work should also consider the use of a larger and more diverse pool of experts with established psychometric expertise to enhance the reliability and generalizability of findings. Developing and implementing objective criteria or checklists to standardize the creation of measure-MoA links would limit subjectivity and build greater consensus.

Conclusion

We report an in-depth analysis of behavior-change experts' reflections on one of the first efforts to link measures to MoAs in behavior change research. Our analysis of expert comments revealed significant conceptual hurdles due to poor measure quality, broad MoA definitions, and ambiguities in terminology. Careful attention to quality, clarity, and consistency in measurement practices and even the definition and selection of MoAs for hypothesis testing is needed. By following the steps outlined in our proposed future directions, we hope to pave the way for more robust and reliable measurement of MoAs, ultimately advancing the field of behavior change research towards greater rigor, transparency, and cumulative progress.

<div align="center">**References**</div>

Bell, C., Johnston, D., Allan, J., Pollard, B., & Johnston, M. (2017). What do Demand-Control
and Effort-Reward work stress questionnaires really measure? A discriminant content
validity study of relevance and representativeness of measures. *British Journal of Health
Psychology*, *22*(2), 295-329.

Bohlen, L. C., Carey, R., de Bruin, M., Rothman, A., Johnston, M., Kelly, M. P., & Michie,
S. (2019). Links between behaviour change techniques and mechanisms of action: an
expert consensus study. *Annals of Behavioral Medicine*, 53(8), 708– 720.

Carey, R. N., Connell, L. E., Johnston, M., Rothman, A. J., De Bruin, M., Kelly, M. P.,
& Michie, S. (2019). Behavior change techniques and their mechanisms of action: a
synthesis of links described in published intervention literature. *Annals of Behavioral
Medicine*, 53(8), 693– 707.

Cornelius, T., Derby, L., Connell Bohlen, L., Birk, J. L., Rothman, A. J., Johnston, M., &
Michie, S. (2023). Linking measures to mechanisms of action: An expert opinion study.
British journal of health psychology, 28(1), 98-115.

Dixon, D., & Johnston, M. (2019). Content validity of measures of theoretical constructs in
health psychology: Discriminant content validity is needed. *British Journal of Health
Psychology*, *24*(3), 477-484.

Flake, J. K., & Fried, E. I. (2020). Measurement schmeasurement: Questionable measurement
practices and how to avoid them. *Advances in Methods and Practices in Psychological
Science*, *3*(4), 456-465.

Human Behaviour Change Project (HBCP) Theory and Techniques Tool. (2018). [Available

from: https://theoryandtechniquetool.humanbehaviourchange.org/

Johnston, M., Carey, R. N., Connell Bohlen, L. E., Johnston, D. W., Rothman, A. J., de Bruin, M., Kelly, M. P., Groarke, H., & Michie, S. (2021). Development of an online tool for linking behavior change techniques and mechanisms of action based on triangulation of findings from literature synthesis and expert consensus. *Translational Behavioral Medicine*, *11*(5), 1049– 1065.

Johnston, M., Dixon, D., Hart, J., Glidewell, L., Schröder, C., & Pollard, B. (2014). Discriminant content validity: A quantitative methodology for assessing content of theory-based measures, with illustrative applications. *British Journal of Health Psychology*, *19*(2), 240-257.

Michie, S., West, R., Finnerty, A. N., Norris, E., Wright, A. J., Marques, M. M., ... & Hastings, J. (2020). Representation of behaviour change interventions and their evaluation: Development of the upper level of the Behaviour Change Intervention Ontology. *Wellcome Open Research*, 5.

Nielsen, L., Riddle, M., King, J. W., Aklin, W. M., Chen, W., Clark, D., ... & Weber, W. (2018). The NIH Science of Behavior Change Program: Transforming the science through a focus on mechanisms of change. *Behaviour Research and Therapy*, *101*, 3-11.

QSR International Pty Ltd. (2020) NVivo (released in March 2020), https://www.qsrinternational.com/nvivo-qualitative-data-analysis-software/home

Riddle, M., & Science of Behavior Change (SOBC) Working Group. (2015). *News from the NIH: using an experimental medicine approach to facilitate translational research*. Oxford University Press.

Schenk, P. M., Wright, A. J., West, R., Hastings, J., Lorencatto, F., Moore, C., ... & Michie, S.

    (2023). An ontology of mechanisms of action in behaviour change

    interventions. *Wellcome Open Research*, 8.

Science Of Behavior Change (SOBC) Measures Repository. (2017). Retrieved

    from https://measures.scienceofbehaviorchange.org/

Sumner, J. A., Carey, R. N., Michie, S., Johnston, M., Edmondson, D., & Davidson, K.

    W. (2018). Using rigorous methods to advance behaviour change science. *Nature Human*

    Behaviour, **2**(11), 797– 799.

Williams, D. M. (2010). Outcome expectancy and self-efficacy: Theoretical implications of an

    unresolved contradiction. *Personality and Social Psychology Review*, *14*(4), 417-425.

Table 1. Round 2 (*N* = 20) Expert Demographics.

ID	Country	Work Sector	Discipline
P1	USA	Voluntary/charity sector	Public Health
P2	USA	University	Psychology
P3	Netherlands	University	Nursing and epidemiology
P4	USA	University	Psychology
P5	USA	University	Public Health
P6	Ireland/UK	Public Sector	Public Health
P7	Canada	University	Behavioral Medicine
P8	USA	Academic Medical Center	Psychology
P9	Canada	University	Kinesiology
P10	Canada	Public Sector	Public Health
P11	UK	University	Psychology
P12	USA	University	Psychology
P13	USA	University	Intervention Research
P14	USA	University	Psychology
P15	USA	University	Psychology
P16	Australia	University	Public Health
P17	Israel	University	Public Health
P18	Canada	University	Psychology
P19	USA	University	Psychology
P20	Canada	University	Psychology

Table 2. Theme definitions and illustrative examples.

Theme (Number of Comments; Number of Commenters)	Definition	Examples
Study Value (3; 3)	Importance of linking measures to mechanisms of action (MoAs)	"…linking of measures and mechanisms is an important and valuable undertaking."
Measure Properties (14; 6)	Issues with measure design, including validity and other psychometrics (e.g., double-barreled items)	"Lastly, this measure seems problematic in its lumping of quite distinct constructs... If the construct validity of a measure depends on levels of a moderator, then it is not a good measure for all (and only for appropriate populations, contexts, etc). "…many of the measures I viewed were fundamentally flawed in their items and/or development. E.g., double-barreled or conditional items, inappropriate or missing psychometric evaluation, etc."
Mechanism Properties (20; 10)	Concerns regarding MoAs, including definitions (e.g., too broad) and malleability	"... Regardless, it raises a related question about how much change should be observed in the mechanisms (e.g., Are some more transient than others?)." "Motivation as its defined here is so broad that it seems like it may be tied to almost anything including this measure."
Miscellaneous Measure Concerns (27; 15)	Other measure concerns, including discrepancies between measure content and title, differences across subscales, language usage (e.g., cultural appropriateness), and measure specificity	"If one sticks close to the item wording… then this scale does not assess intentions but rather indeed beliefs." "…these measures are not appropriate for MOA, because they are theoretically not malleable with intervention."

Conceptual Challenges (25; 14)	Conceptual issues with creating measure-MoA links, including use of the term "related," corresponding measures to MoAs that were not created with the MoA definition in mind, and impediments due to measure quality	"I'm concerned this process won't be helpful overall and poses many risks (in that the outcome will mislead researchers, towards inappropriate measurement of theoretical constructs)." "Therefore, regardless of a conceptual 'link', the measures are not valid for use as measures of the MOAs."
Approaches to Developing Measure-Mechanism Links (5; 4)	Suggested approaches to developing, and improving evaluation of, measure-MoA links	"A more fruitful effort would be to develop measures from scratch, for the purpose of assessing each MOA, and then using the existing measures, that we have been rating, for validation (i.e., for convergent and discriminant validity in particular)."