Eight characteristics of rigorous multilevel implementation research: a step-by-step guide

Seminar for the VA QUERI Implementation Research Group (IRG) March 7, 2024

Dr. Rebecca Lengnick-Hall & Dr. Mark Ehrhart

Our overarching premise:

Given the multilevel nature of healthcare and public health service delivery, <u>we propose that</u> <u>implementation researchers should always start</u> <u>with the default assumption that their research</u> <u>design will need to address multilevel context</u> and related methodological issues, *moving away* from this assumption only after confirming that all the methodological decisions made place the study design completely in "single-level" research territory. aff et al. Implementation Science (2023) 18:52 aoi.org/10.1186/s13012-023-01302-2

Implementation Science

DEBATE

Eight characteristics of rigorous multilevel implementation research: a step-by-step guide



Open Access

Rebecca Lengnick-Hall^{1*†}, Nathaniel J. Williams^{2†}, Mark G. Ehrhart³, Cathleen E. Willging⁴, Alicia C. Bunger⁵, Rinad S. Beidas⁶ and Gregory A. Aarons⁷

Abstract

Background Although healthcare is delivered in inherently multilevel contexts, implementation science has no widely endorsed methodological standards defining the characteristics of rigorous, multilevel implementation research. We identify and describe eight characteristics of high-quality, multilevel implementation research to encourage discussion, spur debate, and guide decision-making around study design and methodological issues.

Recommendations Implementation researchers who conduct rigorous multilevel implementation research demonstrate the following eight characteristics. First, they map and operationalize the specific multilevel context for defined populations and settings. Second, they define and state the level of each construct under study. Third, they describe how constructs relate to each other within and across levels. Fourth, they specify the temporal scope of each phenomenon at each relevant level. Fifth, they align measurement choices and construction of analytic variables with the levels of theories selected (and hypotheses generated, if applicable). Sixth, they use a sampling strategy consistent with the selected theories or research objectives and sufficiently large and variable to examine relationships at requisite levels. Seventh, they align analytic approaches with the chosen theories (and hypotheses, if applicable), ensuring that they account for measurement dependencies and nested data structures. Eighth, they ensure inferences are made at the appropriate level. To guide implementation researchers and encourage debate, we present the rationale for each characteristic, actionable recommendations for operationalizing the characteristics in implementation research, a range of examples, and references to make the characteristics more usable. Our recommendations apply to all types of multilevel implementation study designs and approaches, including randomized trials, quantitative and qualitative observational studies, and mixed methods.

Conclusion These eight characteristics provide benchmarks for evaluating the quality and replicability of multilevel implementation research and promote a common language and reference points. This, in turn, facilitates knowledge generation across diverse multilevel settings and ensures that implementation research is consistent with (and appropriately leverages) what has already been learned in allied multilevel sciences. When a shared and integrated description of what constitutes rigor is defined and broadly communicated, implementation science is better positioned to innovate both methodologically and theoretically.

Keywords Multilevel, Research methods, Research reporting, Guideline, Research best practices

[†]Rebecca Lengnick-Hall and Nathaniel J. Williams contributed equally to this work.

*Correspondence: Rebecca Lengnick-Hall rlengnick-hall@wustl.edu Full list of author information is available at the end of the article



The Author(§ 2023. Open Access This article is licensed under a Creative Commons Attribution 40 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statute regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of the licence, with thttp://creativecommons.org/licenses/by/40./The Creative Commons Fublic Domain Dedication waive (http://creatil.ps/accentil.ps/acce

The predicament of the implementation scientist interested in conducting multilevel research

Goals for our presentation

- 1. Introduce you to this paper and how it came about
- 2. Provide a high-level review of the eight characteristics
- 3. Hurt your brain (just a little)
- 4. Whet your appetite and help you feel supported in pursuing more

Eight characteristics of rigorous multilevel implementation research: a step-by-step guide

WHAT THIS PAPER CAN DO

- Raise your awareness about the complexities (and fun?) of conducting multilevel implementation research
- Give you some core building blocks (the 8 characteristics) that you can pursue more deeply
- Get you thinking about both quantitative and qualitative considerations
- Translate existing literature with imp sci-specific examples and applications

WHAT THIS PAPER CAN'T DO

- Discuss and summarize every important issue you will encounter when you do multilevel implementation research
- Give detailed technical quantitative or qualitative methodological guidance on any specific topic
- Tell you what levels are most important for your project

In sum, we echo Molina-Azorin and colleagues [2], with the intent of addressing the needs of the diverse implementation research community:

Our approach will be to see the 'forest' rather than some particular 'trees.' We examine the big picture, indicating the main elements of multilevel research. An exhaustive analysis of all the elements of multilevel research goes beyond the purpose of this methodological insight, but we provide key references in the literature that could be used…[with the hope that]…multilevel research brings us closer to the reality of [implementation] practice.

To conduct rigorous, high-quality multilevel implementation research…

- 1. Map and operationalize the specific multilevel context for defined populations and settings.
- 2. Define and state the level of each construct under study.
- 3. Describe how constructs relate to each other within and across levels.
- 4. Specify the temporal scope of each phenomenon at each relevant level.
- 5. Align measurement choices and construction of analytic variables with the levels of theories selected (and hypotheses generated, if applicable).
- 6. Use a sampling strategy consistent with the selected theories or research objectives and sufficiently large and variable to examine relationships at requisite levels.
- 7. Align analytic approaches with the chosen theories (and hypotheses, if applicable), ensuring that they account for measurement dependencies and nested data structures.
- 8. Ensure inferences are made at the appropriate level.



Rebecca Lengnick-Hall The Brown School Washington University in St. Louis



Nate Williams School of Social Work Boise State University



Mark Ehrhart Industrial/Organizational Psychology Program University of Central Florida



Catie Willging Pacific Institute for Research and Evaluation



Alicia Bunger Department of Internal Medicine The Ohio State University



Rinad Beidas Department of Medical Social Sciences Northwestern University



Greg Aarons Department of Psychiatry University of California- San Diego

How we organized the material in this paper

In the text, we provide:

- Brief rationale for why we decided this characteristic was important and necessary
- Concrete action steps ("our recommendation for implementation researchers" section)

Each characteristic also has a Supplementary File with:

- Implementation research-specific examples that apply the high-level concepts we introduce in the text
- Practical considerations
- Prompts to use to spur discussion with your research team as you work through this material
- Additional selected references specific to the issues discussed in that characteristic
- Considerations for both quantitative and qualitative methods

We also provide:

- A glossary of terms
- A summary table of our characteristics that can be used for both planning and evaluating multilevel implementation projects
- A real example that illustrates all 8 characteristics (ASPIRE trial in Additional File 9)

Map and operationalize the specific multilevel context for defined populations and settings

Characteristic 1

Characteristic 1: Map and operationalize the specific multilevel context for defined populations and settings

What do we mean in plain language?

Comprehensively think about: (1) what layers of context are important for your implementation research question (2) what they look like in your study.

Why is it important?

Not thinking through and acknowledging relevant levels can lead to blind spots in your analysis and interpretation of results.

Where can you start?

Visually map out the contextual levels you are thinking about (frameworks like CFIR & EPIS can help!). See Table 2 in the paper.

Comprehensively think about: (1) what layers of context are important for your implementation research question (2) what they look like in your study.

Phase: Early implementation				
EPIS Construct	What it looks like in my study			
Ser Level: Statent	State-level Office of Mental Health, statewide implementation processes			
	Managed Care Plans			
Interorganizational	Academic partner + EBP developer + agencies implementing a particular EBP			
environment	Office of Mental Health + payor (managed care plans) + implementing agencies			
FPIS Construct	What it looks like in my study			
OrganiLevel: Agency _{stics}	Infrastructure to support EBP trainings			
Leadership	Agency leadership – communication and diffusion of information re: EBP and implementation activities			
Level: Individual clinicians and leaders with clinician characteristics				
EPIS Construct	What it looks like in my study			
EBP developers	Entities involved in EBP development, training, certification – expectations for and formalization re: data collection, fidelity monitoring			
EBP characteristics	How structured/manualized EBP is; infrastructure requirements for training and implementation			
EPIS Construct	What it looks like in my study			
	Academic partnership (relational type of bridging factor)			
	Contract to pay for EBP training across agencies in the state (formal arrangement type of bridging factor)			
	EPIS Construct Ser Level: Statent Interorganizational environment FPIS Construct Organi Level: Agencystics Leadership el: Individual clinicians an EPIS Construct EBP developers EBP characteristics			

Example of Additional File help:

Prompts to use with your research team to make a map of contextual levels

Prompts to consider when creating your own map of contextual levels:

When identifying and justifying which levels and units (external to the organization) could be relevant to the study:

□ Will environmental influences affect lower-level units (e.g., within the organization)? How will we account for these influences in our analysis and inferences?

□ What sources of theory and evidence support our prioritization of outer context levels and units? How will we report this information?

When identifying and justifying which hierarchically nested intra-organizational levels and units to include in the study:

□ What theoretical or practical basis is there for specifying the intra-organizational levels and units in our study?

□ Who in the organization should we consult to identify the most appropriate and relevant levels and units?

□ Are there any organizational records (e.g., organizational charts) we could draw upon?

□ Do we need to address the presence of informal structures (e.g., informal social groups) in the organization that may affect our implementation research questions? If so, how will we identify those structures and address their influence?

Define and state the level of each construct under study

Characteristic 2

Characteristic 2: Define and state the level of each construct under study

What do we mean in plain language?

Figure out: (1) what levels you are going to deal with in your study, (2) what constructs you are going to consider for each level, (3) how you are going to define each construct for your study.

Why is it important?

It provides the basis for the accurate: construction of measures (Characteristic 5), treatment of analytic variables (Characteristic 7, appropriate interpretation of results (Characteristic 8).

Where can you start?

For each construct you are considering: (1) Define it (2) Identify the level (3)Provide an explanation or "mini theory" for the level.



Construct: Organizational culture

• Define its substantive meaning

"A pattern of shared basic assumptions learned by a group as it solved its problems of external adaptation and internal integration, which has worked well enough to be considered valid and, therefore, to be taught to new members as the correct way to perceive, think, and feel in relation to those problems" (Schein pg. 18).

• Identify the level at which it resides in my study

Hospital level (organizational culture is a characteristics of the hospital)

• Provide an explanation or "mini theory" that clarifies why organizational culture is assigned to the hospital level

[Drawing upon Schein's theorizing…] Medical providers work together, observe each other, and learn from each other. They see how people react to their own behavior and the behavior of their colleagues. They notice what policies, goals, and organizational processes are formalized and enforced.

Through this, medical providers develop a shared understanding of what the norms and values of working at this hospital are.

Culture at the hospital level 'emerges' from these individual provider level experiences and behaviors.

Example of Additional File help: ASPIRE trial example

Construct	Substantive Definition	Level/ Population Unit	Theoretical rationale for leve
Implementation strategy condition	This is the study's primary antecedent variable. It is defined and operationalized as covariate constrained random assignment to <i>Nudge</i>	Primary care practice	This variable occurs at the leve of primary care practices because (a) randomization occurred at the level of primary care practices, and (b) the
	(electronic health record intervention) vs. <i>Nudge</i> + (electronic health record intervention + external facilitation).		implementation strategies (modification of electronic health record and facilitation) target entire primary care practices for change.
Practice adaptive reserve	This is the study's primary (only) quantitative mediator for the analysis of mechanisms. It is defined as the collective capacity within the practice to make and sustain change in support of implementing newly introduced clinical interventions. Measured using the Practice Adaptive Reserve Scale [2], a staff-reported measure that assesses relationship infrastructure, facilitative leadership, sense- making, teamwork, work environment, and culture of learning.	Primary care practice	This variable occurs at the leve of primary care practices because theory on adaptive reserve indicates it is a collective (i.e., shared) characteristic of practices, not of individual clinicians within the practices.
Reach of S.A.F.E. Firearm	This is the study's primary implementation outcome. It is defined and operationalized as youth-caregiver dyad's receipt or nonreceipt of <i>S.A.F.E.</i> <i>Firearm</i> as documented in the electronic health record.	Youth- caregiver dyad	This variable occurs at the youth-caregiver dyad because S.A.F.E. Firearm is designed to be delivered to individual yout and their caregivers during primary care clinical encounters.

Table A2. Construct definitions and levels in the ASPIRE trial.

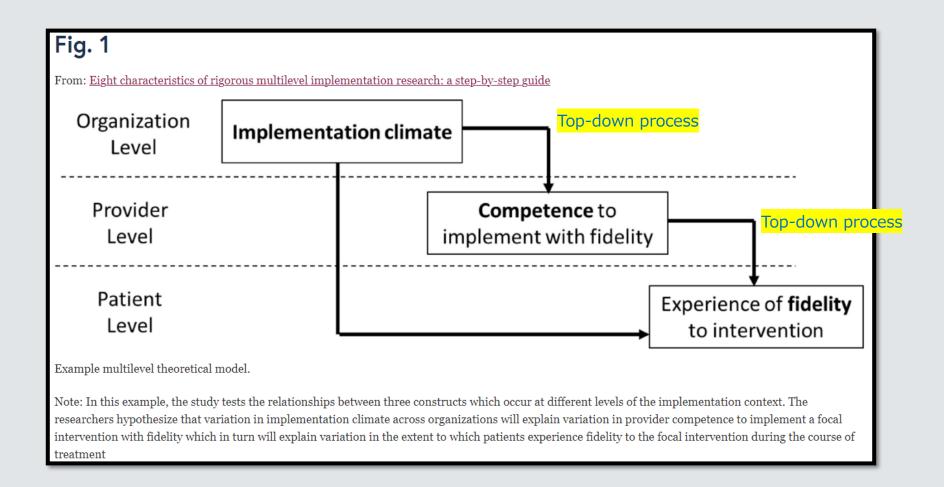
Describe how constructs relate to each other within and across levels

Characteristic 3

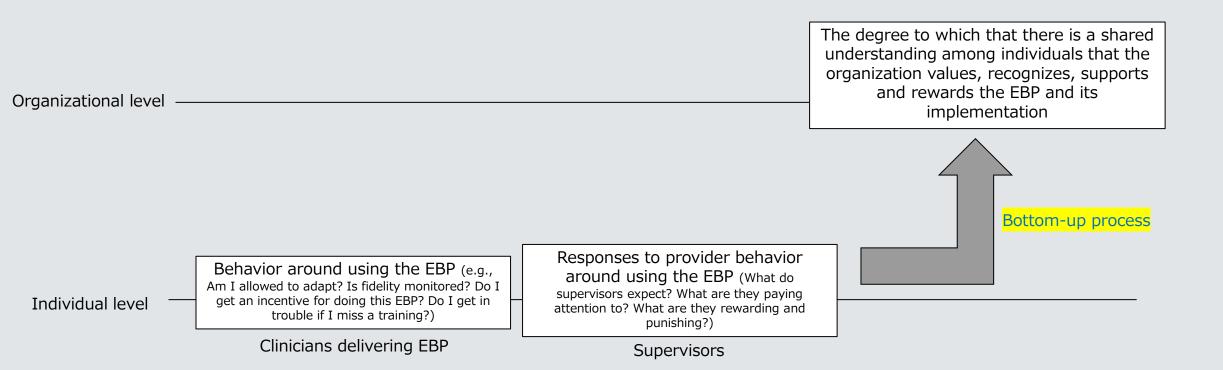
Characteristic 3: Describe how constructs relate to each other within and across levels

What do we mean in plain language?

Identify and describe top-down and bottom-up processes that explain how the levels in your study are connected to and influence each other.



Another implementation climate-focused example multilevel theoretical model



Ehrhart MG, Aarons GA, Farahnak LR. Assessing the organizational context for EBP implementation: the development and validity testing of the Implementation Climate Scale (ICS). Implement Sci. 2014 Oct 23;9:157. doi: 10.1186/s13012-014-0157-1.

Weiner BJ, Belden CM, Bergmire DM, Johnston M. The meaning and measurement of implementation climate. Implement Sci. 2011 Jul 22;6:78. doi: 10.1186/1748-5908-6-78.

Characteristic 3: Describe how constructs relate to each other within and across levels

What do we mean in plain language?

Identify and describe top-down and bottom-up processes that explain how the levels in your study are connected to and influence each other.

Why is it important?

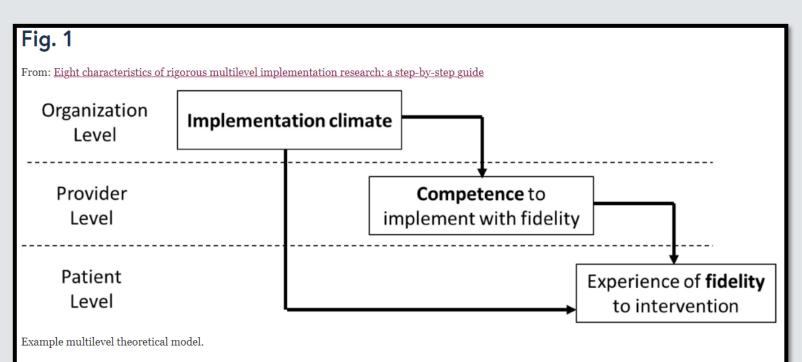
This is an essential step for planning your data collection and analysis.

Where can you start?

Draw out your theorizing like I just showed you!

As you visualize:

- Depict each level and what's going on at that level (constructs).
- Draw those top-down, bottom-up relationships.
- Write out what the crosslevel relationships mean (articulate hypotheses if relevant).
- Use theory to explain why you chose these specific levels and cross-level relationships.



Note: In this example, the study tests the relationships between three constructs which occur at different levels of the implementation context. The researchers hypothesize that variation in implementation climate across organizations will explain variation in provider competence to implement a focal intervention with fidelity which in turn will explain variation in the extent to which patients experience fidelity to the focal intervention during the course of treatment

Example of Additional File help: Practical considerations for describing how constructs relate to each other within and across levels

Practical considerations: Klein et. al.[3] and Rousseau [4] present typologies of relationships that may occur between variables at different levels within multilevel settings; these are useful for spurring investigators' thinking in this area. Lewis et al. [5] offer guidelines for developing causal pathway diagrams which describe the relationships between constructs in implementation in this important to note that when providing theoretical justification for each cross-level effect to be examined in the study, existing implementation process and determinant frameworks will likely not be sufficient as they typically do not specify causal and interactive relationships between levels. We suggest supplementing these types of implementation frameworks with 'big T' and 'little t' theorizing from your own field or discipline.

Don't forget, we made a Glossary too!

Characteristic #3: Describe how constructs relate to each other within and across levels. **Bottom-up processes** – A sequence or series of events, or actions taken, in a specific order toward a specific outcome, which begin at a lower level and terminate at a higher level. An example is increased motivation among individual clinicians within a team to use a screening tool may lead to increased leader advocacy for funding for use of the tool (in response to the groundswell of support from clinicians), which may lead to increased funding available for the tool and greater reach of the tool to more patients within the organization. **Top-down processes** – A sequence or series of events, or actions taken, in a specific order toward a specific outcome, which begin at a higher level and terminate at a lower level. An example is focused organizational implementation climate increasing individual clinicians' selfefficacy to deliver an intervention with fidelity resulting in patients experiencing high-fidelity to the intervention during service interactions.

Specify the temporal scope of each phenomenon at each relevant level

Characteristic 4

Characteristic 4: Specify the temporal scope of each phenomenon at each relevant level

What do we mean in plain language?

For each level in your study, ask yourself: How <u>quickly</u> can I expect to see change?

Characteristic 4: Specify the temporal scope of each phenomenon at each relevant level

- At each level… how long do I need to wait before I first measure change?
 How often do I need to measure things to pick up on these changes?
- How is change at this level going to match up with change at the other levels I care about? (e.g., seeing changes sooner in one level and accounting for that in measurement plan)
- Is there something expected (e.g., planned major leadership change) or unexpected (e.g., global pandemic) going on that could affect the timing and pace of change at the different levels?

Characteristic 4: Specify the temporal scope of each phenomenon at each relevant level

What do we mean in plain language?

For each level in your study, ask yourself: How <u>quickly</u> can I expect to see change?

Why is it important?

People, teams, organizations and systems change at different rates. Rule of thumb: lower level (e.g., individual) likely to change more quickly than higher level (e.g., organization).

Where can you start?

Look at Additional File 4 for prompts to consider when deciding and explaining the frequency and timing of measurements at different levels.

Example of Additional File help: Checklist of what to report in your research plan to help you specify the temporal scope of phenomena at different levels

Checklist of what to report in your research plan:

□ When do you expect to observe change in each relevant outcome at each relevant level (e.g., of system- or organization-level implementation strategies)?

□ How frequently and when will you measure constructs to capture these changes?

□ How will changes in outcomes at different levels align with each other in the research design? For example, how long might it take for a leadership strategy to change supervisor's behavior and what implications would that have for when we should see change in providers' fidelity and ultimately in patient outcomes?

 \Box What are the theoretical rationales for these choices (formal 'big T' theories or informal 'little t' theories)?

Example of Additional File help:

Illustrating the issue of temporal scope with an implementation research-specific example

Example of how the sequence or pace of change at one level may differ from that at another level:

Consider a study in which a clinic-level implementation strategy is employed to improve clinic implementation leadership as a means of enhancing clinician fidelity to a targeted psychotherapy practice. The primary implementation outcome in this study is clinician fidelity and the primary clinical outcome is patient improvement in symptoms. Designing this study requires careful consideration of how the change process is expected to unfold across levels over time. It may take two to four months to change clinic leaders' behaviors. Once clinic leadership is modified, it may take clinicians one to two months to change their behavior and increase fidelity in response to leadership changes. For individual patients receiving therapy, it may take three to six months of high-fidelity treatment before they begin experiencing symptom relief. Given this sequence, the implementation strategy might not be expected to improve fidelity until four months after completion of the leadership training and fidelity may not be expected to improve until 6-8 months after completion of the leadership training. All of these temporal dynamics must be taken into account in designing multilevel measurement and analysis plans.

Align measurement choices and construction of analytic variables with the levels of theories selected (and hypotheses generated, if applicable)

Characteristic 5

Characteristic 5: Align measurement choices and construction of analytic variables with the levels of theories selected (and hypotheses generated, if applicable)

What do we mean in plain language?

Be sure that the measures and how they will be collected are consistent with the levels and theories of interest. Measurement must align with the level of theory!

Why is it important?

How questions are asked impacts how participants respond and the validity of those responses. It is hard to disentangle measurement issues from substantive findings after the fact.

Characteristic 5: Align measurement choices and construction of analytic variables with the levels of theories selected (and hypotheses generated, if applicable)

Where can you start?

For each construct in the model, review items or questions and ensure that:

- The referent is consistent with the level of theory
- Participants can report on the construct
- Aggregate individual-level data to the unit level if appropriate (and with evidence)

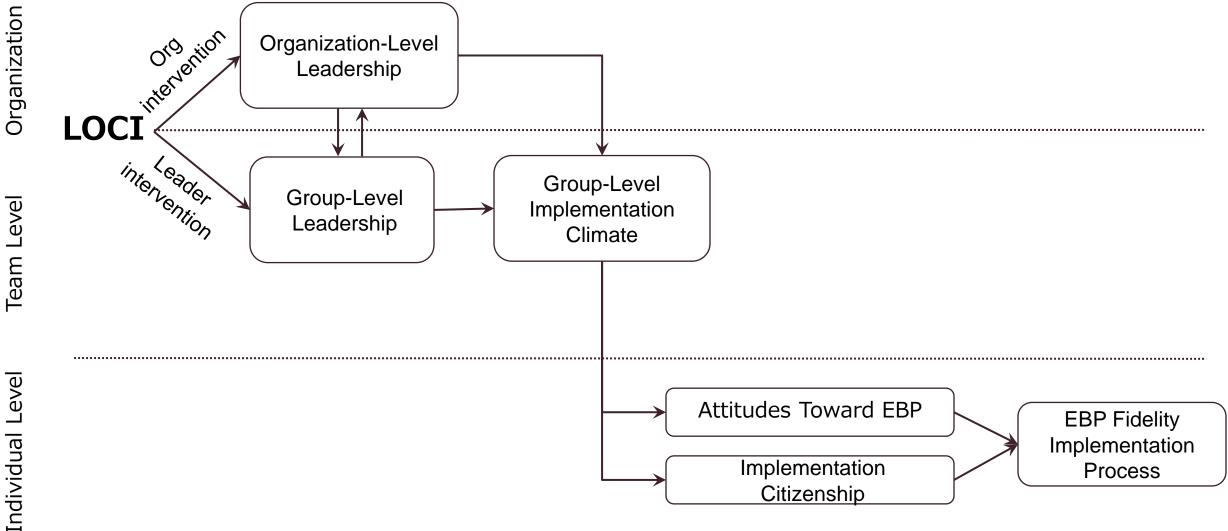
Three categories of multilevel constructs

- Global constructs: originate at the unit level and represent objective, easily observable characteristics of the unit (e.g., hospital ward, clinic)
- Shared constructs: originate at the individual level but are shared across unit members (e.g., clinic implementation climate)
- Configural constructs: originate at the individual level and represent a pattern of individual characteristics within the unit (e.g., clinician experience)

Avoid misalignment!

For instance, avoid using individually varying scores to measure a theoretically shared organizational characteristic such as organizational climate)

Leadership and Organizational Change for Implementation (LOCI) Proposed Effects and Mechanisms (R01DA03846)



Aarons, G. A., Ehrhart, M. G., Moullin, J. C., Torres, E. M., & Green, A. E. (2017). Testing the leadership and organizational change for implementation (LOCI) intervention in substance abuse treatment: a cluster randomized trial study protocol. Implementation Science, 12:29.

Use a sampling strategy consistent with the selected theories or research objectives and sufficiently large and variable to examine relationships at requisite levels

Characteristic 6

Characteristic 6: Use a sampling strategy consistent with the selected theories or research objectives and sufficiently large and variable to examine relationships at requisite levels

What do we mean in plain language?

Need to make sure that the sample for the study is adequate at all levels of interest. Need to also pay attention to issues of variability and representativeness at each level.

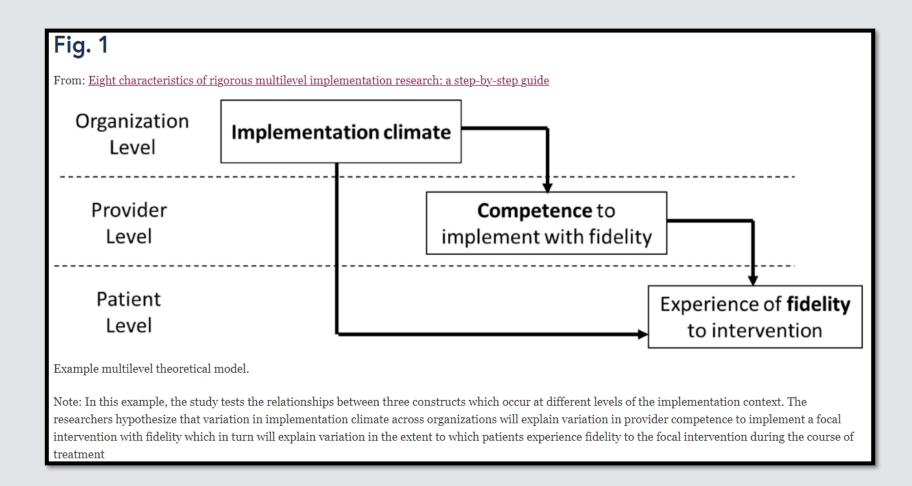
Why is it important?

Implementation research often considers interventions and outcomes that cut across levels. Poorly planned sampling strategies can have disastrous results for research findings. Characteristic 6: Use a sampling strategy consistent with the selected theories or research objectives and sufficiently large and variable to examine relationships at requisite levels

Where can you start?

Build on other characteristics and ensure that sampling choices align with the constructs and relationships being studied, as well as the measurement choices.

See Addition File 6 for prompts to consider when designing your multilevel study and checklist for reporting sampling plan.



Align analytic approaches with the chosen theories (and hypotheses, if applicable), ensuring that they account for measurement dependencies and nested data structures

Characteristic 7

Characteristic 7: Align analytic approaches with the chosen theories (and hypotheses, if applicable), ensuring that they account for measurement dependencies and nested data structures

What do we mean in plain language?

There is no single best way to analyze data from multilevel implementation studies. Decisions regarding how best to analyze data can be aligned or misaligned with theories/hypotheses.

Why is it important?

A lack of alignment between proposed theories or hypotheses and how the data are analyzed can lead to erroneous conclusions.

Characteristic 7: Align analytic approaches with the chosen theories (and hypotheses, if applicable), ensuring that they account for measurement dependencies and nested data structures

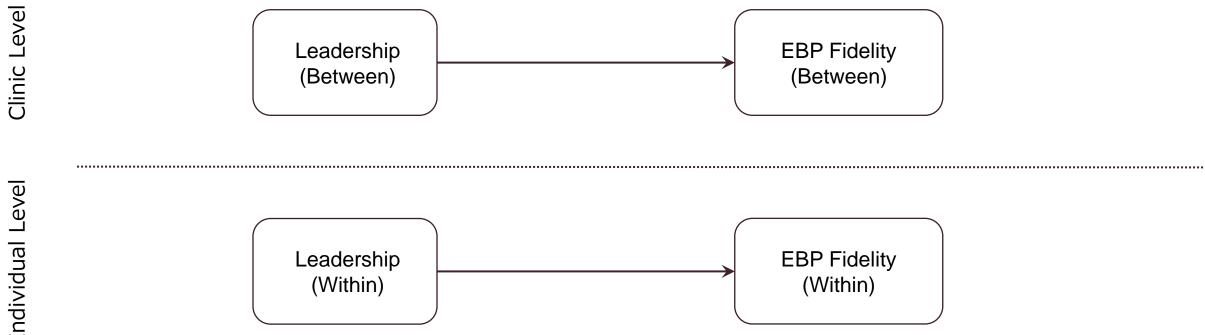
Where can you start?

Clarity on levels of theory, constructs, and measurement should make analyses more straightforward.

Select analyses that account for the dependencies in hierarchically sampled observations (or make a strong case they aren't necessary).

Be clear in write-up on what decisions were made and why.

A multilevel SEM example: The effects of leadership fidelity between and within units



Ensure inferences are made at the appropriate level

Characteristic 8

Characteristic 8: Ensure inferences are made at the appropriate level

What do we mean in plain language?

Ultimately, we want to reach conclusions about our findings and address their implications for practice – we need to make sure those implications reflect what we actually found.

Why is it important?

If we infer effects at the wrong level, we can inadvertently limit or misdirect the advancement of implementation science, as well as create bad policies and waste resources. Characteristic 8: Ensure inferences are made at the appropriate level

Where can you start?

Follow the suggestions from the first seven characteristics! Doing so should make the level of the inferences we make clear.

Take care in writing up the results to ensure that readers understand what our findings mean and how to properly interpret them.

Characteristic 8: Ensure inferences are made at the appropriate level

Ensure inferences are made at the appropriate level(s)

- Atomistic fallacy: Analyze the association between variables at the individual level and then inappropriately make inferences about a higher level of analysis (e.g., groups, organizations)
- Ecological fallacy: Studies conducted at a higher level of analysis (e.g., group, organization) and inappropriately make inferences about lower-level units (e.g., individuals)

- Be precise in language and descriptions
 - > Imprecise: "Higher readiness for change was associated higher fidelity"
 - Precise: "Higher clinic-level readiness for change was associated with higher providerlevel fidelity"



Mark's Take Away Points

- Be warned once you start looking through a multilevel lens it is hard to stop!
- In implementation research, multilevel issues are the norm. Acknowledging and operationalizing the complexity can advance the field.
- It helps to be as familiar as possible with the setting or settings where you will be working to understand the dynamics and levels most important for sustained change.
- Mixed methods can be very useful in understanding context, how participants perceive levels, and how documented levels are or are not consistent with perceptions.

Rebecca's Take Away points

- Multilevel implementation research is deceptively hard.
- You can't measure and deal with everything in a single study, so you have to have good reasons for the levels and relationships you choose to focus on.
- Specifying and reporting this reasoning (what levels and why) is critical to the rigor of implementation research (and a major goal of us writing this paper).
- Conducting rigorous multilevel research is an important area for growth in the field in terms of how we are trained, who is on our research teams, and what existing literature and thinking we draw upon when conceptualizing and executing our studies.

Thank you!

Dr. Rebecca Lengnick-Hall rlengnick-hall@wustl.edu

Dr. Mark Ehrhart Mark Ehrhart@ucf.edu