

Web Appendix 1: Scaling-up frameworks literature review methodology

As background to this research, in order to establish the range and breadth of scaling-up frameworks available, we conducted a review of the published and grey literature through Medline, PubMed (title or abstract), Web of Science (topic or title), Embase, PsycINFO, The Cochrane Library, Cumulative Index to Nursing and Allied Health Literature (CINAHL), and WorldCat. Additional literature was found through the World Health Organization (WHO) affiliated ExpandNet database, and by searching the reference lists of identified publications. Relevant key words used in various combinations as search terms were: 'scaling-up', 'scaling-up', 'going to scale'; 'framework', 'model', 'health systems', 'implementation', 'implementation science'. We limited the search to 2000 to 2016, and to the age-group 0 to 6 years.

We retained frameworks that described strategies, characteristics or domains for scale-up, or a detailed sequential theoretical approach (steps and considerations) to scaling-up innovations across health related disciplines (referred to as 'general frameworks' throughout the report), while the ones focusing on specific disciplines other than ECD were excluded. Only articles/reports that described a conceptual model for scaling-up interventions in LMICs were included, as well as only the most updated framework from each author (that is original frameworks were excluded if the main author published a new framework or an updated version of the original one).

Main themes relevant to successful scale up were identified and listed by VC, and were included only if they were found in six or more of the sources. They were then organized into domains and sub-domains through discussions among VC, TD and MT to reach consensus. These were then used to construct the stakeholders' questionnaire and framework for analysis. Additionally, all identified frameworks were reviewed to integrate their principles and the lessons learned from their development and implementation in the final set of recommendations.

Web Appendix 2: Interview schedule

1. Please tell me about a project/intervention that you have either scaled up, intend to scale up, or planned to scale up. How did you go about scaling-up/ what are the steps that you took?
2. What do you think were the major determinants of the successes and failures of scaling-up? Based on your own personal experience and expertise, what are the key elements for successful scale-up in general? And why does scale up often fail?
3. In meeting the challenges, what worked, what didn't, and why? In hindsight, what would you have done differently?
4. Did you ever consider not to scale up your innovation, or thought about whether the program is worthy of scaling-up (how effective is it at achieving the targeted behavioral change)?
5. There are various frameworks that assist with managing the scaling-up of innovations. Is there any relevant/specific framework that you know of/ have used/ was helpful to you as a guideline to manage your scale up plan?
6. What are the key papers that you think helped to inform the science of scale up, including those that are in the grey literature?
7. What are the big gaps in the field of implementation science—what are some of the research questions that need to be addressed to help fill the gaps in the knowledge base on scale up?
8. How might the quality of pilot programs be preserved and costs kept within national resource limitations while taking them to scale?

Web Appendix 3: Framework analysis

The steps of the framework analysis comprised:

- i. Familiarization – the research team listened to the audio recordings and thoroughly read through each transcript to become familiar with the whole data set. During the familiarization phase, we highlighted contrasting views on barriers to and facilitators of scaling-up projects.
- ii. Identification of a framework – we organized the data using both an a-priori coding framework from the literature review and an inductive approach for emergent categories. After VC coded all transcripts, MT reviewed all codes for consistency and meaning. All coders went through an additional themes refining phase until team consensus was reached.
- iii. Indexing data – VC examined all transcripts and manually coded transcripts line-by-line to organise conceptually-related codes into prominent categories, which were then reviewed by the team and agreed upon.
- iv. Charting – data were summarized into tables where responses of each participant were listed in a single row, with codes for those responses organised into columns within each category. Potentially interesting quotations were highlighted.
- v. Mapping and interpretation – the table was reviewed by making connections within and between participants and categories. Underlying ideas, assumptions, ideologies, and conceptualizations that informed the semantic content of the data were examined to identify themes on the basis of their prevalence, richness and the importance placed upon them by participants.