



OPEN ACCESS

EDITED BY

Antonio Neme,
National Autonomous University of Mexico,
Mexico

REVIEWED BY

James Lorenz Merle,
University of Utah Hospital, United States
Joy D. Doll,
Creighton University, United States

*CORRESPONDENCE

Dean L. Fixsen
✉ dfixsen1@gmail.com

RECEIVED 24 June 2024

ACCEPTED 04 October 2024

PUBLISHED 16 October 2024

CITATION

Fixsen DL, Van Dyke MK and Blase KA (2024)
Is implementation science a science? Not yet.
Front. Public Health 12:1454268.
doi: 10.3389/fpubh.2024.1454268

COPYRIGHT

© 2024 Fixsen, Van Dyke and Blase. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Is implementation science a science? Not yet

Dean L. Fixsen*, Melissa K. Van Dyke and Karen A. Blase

Active Implementation Research Network, Inc., Chapel Hill, NC, United States

Getting the science right for implementation is critical for making the processes for improving outcomes more predictable and effective in global public health. Unfortunately, “implementation science” has become a catchphrase for ideas, assumptions, and findings concerning the science to service gap and how to close it. The purpose of this paper is to explore the dimensions of a “science of implementation” that meets the definitions of a science and is focused on implementation variables (i.e., purposeful processes to put innovations into effect so that intended benefits can be realized). A science of implementation is important for accomplishing the goals related to improving the health and well-being of populations around the world. Much of public health involves interaction-based interventions. In a typology of science, interaction-based interventions are created by specifying the nature of certain exchanges between and among individual people or groups. The complexity of developing interaction-based independent variables requires meeting benchmarks for fidelity to assure the presence and strength of implementation independent variables. The paper presents information related to the following tenets: (1) A science of implementation is based on if-then predictions. Science is cumulative. As predictions are made, tested, and elaborated, the facts accumulate to form the knowledge base for science and practice. (2) Implementation variables are interaction-based inventions and, therefore, must be created and established so the specific set of activities related to implementation can be studied. (3) A science of implementation is based on theory that organizes facts, leads to testable predictions, and is modified or discarded based on outcomes. (4) A science of interaction-based implementation depends on frequent measures of independent and dependent variables specific to implementation methods and outcomes. Two examples illustrate the implications for theory, research, and practice. The paper advocates a paradigm shift to a new mental model that values fidelity over tailoring, has one size fits all as a goal, and is concerned with the function of evidence rather than the form of evidence based on RCTs. Global health fundamentally requires scaling implementation capacity so that effective innovations can be used as intended and with good effect to achieve population benefits.

KEYWORDS

science, implementation, independent variable, measurement, theory, fidelity

Introduction

Since its beginnings in policy research and behavioral sciences in the 1960s (1–10) “implementation science” has become a catchphrase, a label for loosely related ideas, assumptions, and findings. There is nothing wrong with implementation science as a label, but it should not be confused with a science of implementation.

The National Institutes of Health (NIH) defines implementation science as the study of how to integrate evidence-based practices into routine health care and public health

settings. The goal of implementation science is to improve population health outcomes. In contrast, a “science of implementation” is broader and open to learning about implementation in any domain.

In its current form, “implementation science” lacks organizing themes, agreed-upon language, and research focus. For example, Beidas et al. (11) summarized several “self-critical assessments” and added their own assessment of how recent decades of implementation science have not realized its goals of “achieving population health impact and social justice at scale.” Instead, decades of research has seen “the emergence of a new field with no common body of facts.” To counter “concerns that the field may be stagnating,” they advocated for less emphasis “on becoming a legitimate science” and encouraged the pursuit of simpler implementation strategies while acknowledging “the important role of multilevel context in implementation.” In contrast, a “science of implementation” is based on adhering to the basic tenets of science and acknowledges that implementation methods must account for the complexity of implementation challenges.

Science is based on testing predictions and accumulating facts so that phenomena of interest become more predictable and effective. A “science of implementation” is focused on factors related to changing practitioner, organization, and system behavior so that innovations can be used fully and effectively to reliably produce desired outcomes on a useful scale.

A science of implementation is important for accomplishing the goals for improving the health and well-being of populations around the world. Implementation is one of three factors that interact to produce socially important benefits for populations. The formula for success postulates (12–14):

Effective Innovations X Effective Implementation X Enabling Contexts = Socially Significant Benefits.

In the formula for success, an Effective Innovation is one that is defined and operationalized, and includes a way to assess fidelity of its use in practice; Effective Implementation refers to the interactive factors required to support the full and effective use of innovations in practice; and an Enabling Context is a system where the components are aligned so they support one another in coherent and purposeful ways so that improved system performance and improved outcomes can be achieved year after year. Socially Significant Benefits are outcomes that make a meaningful difference for the population of interest.

The formula is clear: if Effective Implementation is weak, socially significant benefits will be modest. The current “evidence-based movement” (15, 16) has focused on effective innovations and not effective implementation (17, 18) while giving worried attention to enabling contexts (19–21). The current extensive focus on “evidence-based interventions” has not led to substantial (purposeful, replicable, sustainable) improvements in public health. There are now many “evidence-based” innovations and too few socially significant benefits.

The purpose of this paper is to explore the dimensions of a “science of implementation” that make the processes for creating change more predictable and effective. In this paper, “implementation science” and “science of implementation” will be used without the quotes and are intended to mean two distinct things with (sadly) very little in common. The following sections

explore the characteristics of a science of implementation and offer illustrative examples.

Science of implementation

A science of implementation is based on if-then predictions. Science is cumulative. As predictions are made, tested, and elaborated, the facts accumulate to form the knowledge base for science and practice.

What qualifies as “science?” That is, how does science differ from other human activities: what discriminates astronomy from astrology, chemistry from alchemy, facts from beliefs (22, 23)? And, what discriminates a science of implementation from implementation science?

The philosophy of science is concerned with what science is and the logic for developing scientific knowledge. Fundamentally, science requires clearly stated predictions (if-then) and testable hypotheses (24–26). The predictions are falsifiable, and outcomes are replicable. For example, Galileo predicted (1604) that if any two or more objects of any size or weight are dropped from any height, then (without the interference of atmosphere) the objects will reach the ground at the same time. The prediction was supported (not falsified) as it was tested in Earth’s atmosphere. Newton developed a mathematical formula (1687) to calculate more precisely the predicted effects of gravity anywhere in the universe. Newton’s formula was used to determine the thrust to launch the astronauts from Earth and plot the complex four-day course to land them on the moon (1969). In 1971 (367 years after 1604) astronauts were able to meet the “without the interference of atmosphere” condition and tested Galileo’s prediction in the zero atmosphere conditions on the moon.¹ Einstein (1905) expanded the theory of gravity and calculated its predicted effects on spacetime. This prediction was tested during the solar eclipse in 1919, the predicted position of the stars did appear to shift, and the theory of relativity gained major support.

Science is cumulative as exemplified in the discussion of gravity. Predictions generate facts, facts are organized in theories, and knowledge accumulates as facts reveal more of the fundamental truths and operations of general principles. Predictions tell us *if* something works (e.g., Galileo). Then scientists figure out *how* it works so it can be repeated (e.g., Newton). Finally, science may discover *why* something works (e.g., Einstein: all objects with mass bend and curve the fabric of the universe, called spacetime, and that curvature is felt as gravity).

Making predictions, testing predictions in experiments, and revising the knowledge base is not a common part of research under the current banner of implementation science. A search in Google Scholar for “implementation science” in the years 2005–2024 produced nearly 500,000 returns in June 2024. Prediction: no more than one published paper in any randomly selected set of 100 “implementation science” papers will include an explicit if-then statement regarding the purpose of the paper. Readers are encouraged to test each prediction in this paper as a way to hone

1 <https://www.youtube.com/watch?v=ZVfhztmK9zI>

the distinctions between a science of implementation and implementation science.

Discoveries and inventions

A science of implementation is based on interaction-based inventions that describe when, where, with whom, and how specified interactions between people should occur to support the full and effective use of an innovation.

Implementation variables often are described as complex “reciprocal interactions” among multiple actors (27). “While the components of a system, namely the agents and their artifacts, are important, they are often secondary to the relationships between these components. In such systems, agents communicate and learn from each other and from their environment and adjust their behavior accordingly. However, there are many cross-cutting interconnections and influences” [(19), p. 5]. The assorted ideas that characterize implementation science likely result from attempts to deal with the inherent complexity of human behavior noted in this paragraph.

The source of this complexity can be better understood as the types of science are understood. Science is divided into two types as shown in Figure 1: discoveries and inventions (28). Previously unknown things that exist in nature can be discovered and science has generated facts and theories to describe and understand those discoveries. For example, geography and cartography and seismology evolved from the discovery of new lands and their shifting topographies. Chemistry evolved from the discovery of elements in the earth, water, and air. In these and other “natural sciences” such as biology, botany, physics, and astronomy the subject matter already exists and is available for study almost anywhere in the world.

Inventions are not “natural,” they are created. For example, light bulbs, telephone systems, integrated circuits, pain pills, highways, and office buildings are inventions. Computer science, engineering sciences, and so on evolved from the millions of inventions that have been created (29). In his book *The invention of science*, Wootton (28) documents the creation of science as it is understood today and notes that science itself is an invention, not a product of nature.

This distinction is recognized in law as well as in science. For example, a patent may *not* be issued for the discovery of a previously unknown phenomenon of nature. A patent can be issued only for an invention that applies a “law of nature” to a new and useful end.

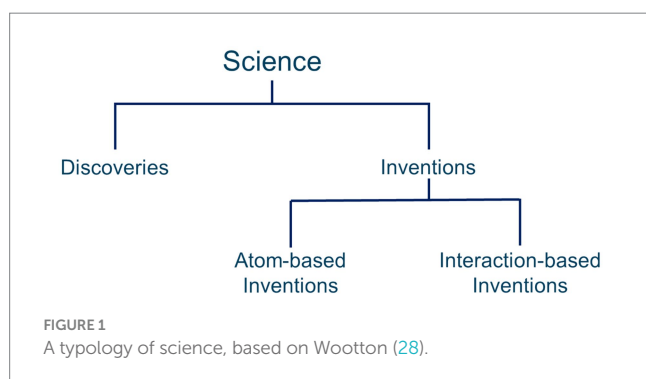
In science, inventions are divided into two major subtypes as shown in Figure 1: atom-based inventions and interaction-based inventions [(30, 31), p. 22].

Atom-based inventions are created by using discoveries in new ways. For example, chemicals may be combined to produce a serum or a plastic object that is not found in nature but is based on natural (already discovered) elements. Or natural metals and crystals may be combined in new ways to form a transistor or silicon-based integrated circuit. After they are created, atom-based inventions are very stable (firmly established, enduring, resistant to change). As independent variables (if this) in tests of predicted outcomes (then that), atom-based inventions do not change from one use to the next or from one user to the next. Once they are created, they are available for study at the convenience of the scientist.

Gertner (29) described the success of Bell Labs, perhaps the most productive scientific enterprise of all time, averaging more than one patent per day for over 50 years. “If an idea begat a discovery, and if a discovery begat an invention, then an innovation defined the lengthy and wholesale transformation of an idea into a technological product (or process) meant for widespread practical use... An innovation could fail for technical reasons (if it proved unreliable) or for manufacturing reasons (if it proved difficult to reproduce consistently or cheaply).” Once they are created, atom-based innovations are relatively stable, predictable, and replicable.

Interaction-based inventions are created by specifying the nature of certain exchanges between and among individual people or groups (e.g., health workers and neighbors, implementation specialists and therapists, directors, and staff). The history of a piece of iron ore makes no difference in the makeup of an atom-based invention. However, in interaction-based inventions the history and ongoing life experiences of each individual or group will influence the quality of each example of an invention. Interaction-based inventions reflect the complexity of human behavior. After they are created, interaction-based inventions may not be stable as “recursive feedback constantly changes both agents, which in turn changes the other, again and over again” (32). As independent variables (if this) in tests of predicted outcomes (then that), the interaction-based invention may change from one use to the next and from one user to the next. Interactions are short-lived, and therefore interaction-based inventions are not available for study without careful preparation and monitoring by the scientist (33–35).

Compaoré et al. (36) studied the use of the WHO recommended intervention for *Plasmodium falciparum* malaria entitled Seasonal Malaria Chemoprevention (SMC). “Paired community distributors (CDs) dispense a full course of SMC drugs each month to targeted children using a door-to-door approach under the supervision of formal health workers. Each monthly drug distribution or SMC cycle lasts over 4 days, with an interval of exactly one month between two cycles. A complete course of SMC comprises a single treatment with SP and three daily doses of AQ; the CDs provide the first dose of AQ with SP to each 3- to 59-month-old child under directly observed treatment (DOT) [fidelity] in the absence of any contraindications.” Interaction-based innovations involve people working with people in planned ways where plans could go awry for many reasons. For example, Compaoré et al. (36) reported “difficulties due to insufficient training of community distributors, inadequate supply of inputs and insufficient financial resources for remuneration, advocacy and supervision, [and] the contextual constraints due to the rainy season.”



None of these factors would influence the integrity of an atom-based innovation.

Understanding that interaction-based inventions are relatively unstable when compared to atom-based inventions is not an invitation to “tailoring” as commonly advocated in implementation science. Instead, it places the onus on implementation scientists to standardize interaction-based implementation methods to improve the stability and reliability of independent variables. In science, altering methods alters outcomes. Fidelity benchmarks set the minimum acceptable standard and exceptions are managed so that performance is sustained within an acceptable range so that predicted outcomes can be achieved more consistently. Meeting benchmarks for fidelity establishes a new normal distribution of competencies and outcomes and scientists can learn from the 15% top performers as they work to find ways to improve the 15% poor performers. In this way, fidelity provides a firm foundation for improvement using PDSAC improvement cycles (37) or Total Quality Management methods (38).

Assuring the integrity of the independent variable is not a common part of research under the banner implementation science. Prediction: no more than one published paper in any randomly selected set of 100 “implementation science” papers in the years 2005–2024 will clearly identify an implementation independent variable and make an if-then prediction of its effects.

Implementation

Implementation variables are interaction-based inventions and, therefore, must be created and established so that the specific set of activities related to implementation can be studied.

Implementation is the purposeful process of putting something into effect. Implementation is:

“a specified set of activities designed to put into practice an activity or program of known dimensions. According to this definition, implementation processes are purposeful and are described in sufficient detail such that independent observers can detect the presence and strength of the ‘specific set of activities’ related to implementation. In addition, the activity or program being implemented is described in sufficient detail so that independent observers can detect its presence and strength. When thinking about implementation, the observer must be aware of two sets of activities (intervention-level activity and implementation-level activity) and two sets of outcomes (intervention outcomes and implementation outcomes)” [(39), p. 5].

As reflected in this definition, implementation is universal and applies to any situation where there is an attempt to “put into practice an activity or program [an innovation] of known dimensions.” For any innovation, implementation methods (independent variables) can be studied to see the effects they may have on “putting the innovation into practice.” For a science of implementation, the focus is not on innovations but rather on factors that support the full and effective use of innovations in practice. Like statistics, implementation is universal and exists independently of any particular area of application. The support methods, not the innovations, are the focus of a science of implementation. Thus, a science of implementation can benefit from studies done in any field.

Interaction-based independent variables in a science of implementation

Implementation factors that support the full and effective use of innovations are interaction-based inventions. They do not exist in nature waiting to be discovered; they are created and specify the interactions between and among people.

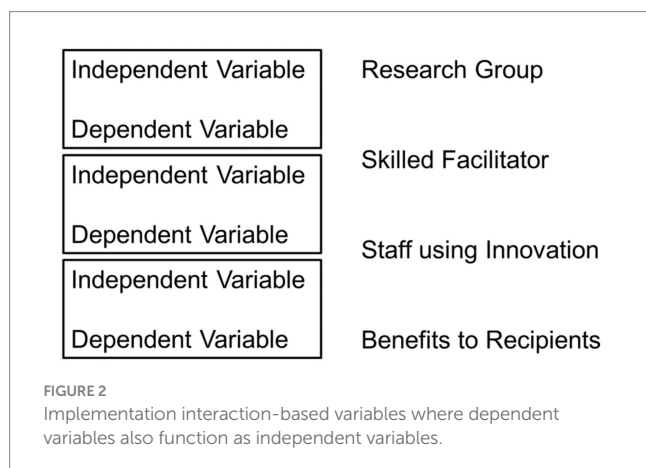
Interaction-based independent variables add considerable difficulty to the development of a science of implementation. The implementation independent variable must be established by a research team at the time and place an experiment is to be conducted, and in sufficient quantity and with sufficient quality to satisfy the requirements of an experimental design. This, in itself, is an implementation challenge. In a science of implementation every dependent variable at one level is an independent variable at the next level [(31), p. 13ff].

“A research group must be sufficiently skilled in using implementation practice (the independent variable) so that practitioners will use an innovation with fidelity (the dependent variable). At the next level, the practitioners’ use of an innovation with fidelity (the independent variable) is assessed in terms of benefits to recipients (the dependent variable). [Thus], the fidelity of practitioners’ use of an innovation is both a dependent variable and an independent variable.”

This means a research team must be knowledgeable and skilled in the use of implementation best practices so that interaction-based independent variables can be established at sufficient strength to be studied. In effect, the research group functions as an implementation team that uses current implementation best practices to establish variables that can be studied to create new implementation knowledge. This is similar to atom-based software development where it is common practice to develop software using the software being developed, such as using and improving barely functional MS Word 0.0 *en route* to developing MS Word 1.0 for the intended end users (40, 41).

Implementation applies to itself as the first users of new knowledge are the scientists themselves. Armed with new knowledge, scientists can establish and test more robust and complete interaction-based independent variables to test riskier predictions. A science of implementation depends on implementation practice.

The logic for research on implementation variables is shown in Figure 2. In this example “facilitation” is the implementation independent variable. Skilled facilitators must be established by the research group so that facilitation is available for study. The intended outcome of facilitation (the implementation independent variable) is increased use of a specific innovation (the implementation dependent variable). Use of the innovation as intended (the innovation independent variable) is expected to result in benefits to recipients (the innovation dependent variable). Given that implementation is universal, the same implementation methods (e.g., training, coaching, fidelity to develop competencies) can be used to establish the independent variable at each level of use (research group creating skilled facilitators; skilled facilitators creating staff skills to use the innovation). Thus, skilled facilitators in this example (simultaneously) are the outcome of the research group and the input for staff using the innovation.



With respect to fidelity of the independent variable at each level, An et al. (42) recommend establishing pre-set benchmarks that must be met before a study is conducted. For example, a large-scale study (over 4,000 households) employed specially trained community health workers (CHWs) who delivered individual and combined water, sanitation, handwashing (WSH) and child nutrition interventions to the homes and families of pregnant women in rural villages in Bangladesh. In this study, the fidelity of delivery of the innovation was assessed proactively from the beginning. After three months, fidelity (in the 30–60% range) was found to be well below the pre-set benchmark of 80% for each independent variable. The research group made adjustments to improve the delivery of CHW services, and the fidelity scores improved (in the 86 to 93% range). At the end of the study the multiple benefits of the complex WSH intervention were substantial (43, 44). The research group produced skilled CHWs who produced high fidelity examples of the combined intervention and, therefore, the independent variable was available for study and was found to be effective for producing improvements in water quality, handwashing, sanitation, and nutrition.

Assessing fidelity and using 80% as a benchmark for fidelity helps to assure the intervention “is there.” Given that implementation variables are interaction-based inventions, implementation scientist teams must create sufficient implementation practitioner expertise so that implementation variables can be established at sufficient strength for study.

Implementation science does not insist on fidelity assessments to assure the presence and strength of interaction-based independent variables. Methods for establishing independent variables are glossed over if they are included at all (e.g., “the practitioners were trained to use the innovation”). Thus, learning “what works” for putting something into effect has not been cumulative for practice or science. Prediction: no more than one published paper in any randomly selected set of 100 “implementation science” papers in the years 2005–2024 will clearly identify the implementation independent variable and report the fidelity of its use during the study.

Theory

A science of implementation is based on theory that organizes facts, leads to testable predictions, and is modified or discarded based on outcomes.

In science, theory is a source of predictions (if-then) that lead to observations to confirm or disconfirm theory-based predictions. Theory is improved or discarded based on the outcomes of theory-based predictions. In this way, science is cumulative with today’s knowledge built on the foundations of past predictions that were tested and led to improved theory (45).

Five criteria that define a theory were outlined by Carpiano and Daley (46).

- 1 Logic: The major concepts and relations should be logically coherent. Terms must be clearly defined so that they can be understood by those examining the theory.
- 2 Causality: The goal of theoretically based research is to identify the systematic components of a set of factors that produce change in the phenomena being studied. Causal drivers and a sense of causal process should be clearly identified.
- 3 Falsification: The major propositions should be empirically falsifiable. All useful theories suggest ways in which they may be subjected to empirical assessment.
- 4 Scope: Although it can change over time, the scope of the theory should be clear and relatively broad. It must be focused on generic processes and not unique characteristics of any specific situation or case.
- 5 Productivity: The theory should promote non-obvious implications (“risky predictions”) and produce a relatively large number of predictions per assumption. Popper (25) (p. 39) adds another criterion:
- 6 Prohibition: Every good scientific theory forbids certain things to happen. The more a theory forbids, the better it is.

The explicit predictions made in this paper are examples that meet Popper’s prohibition criterion.

With respect to Carpiano and Daley’s criterion regarding scope, the Improved Clinical Effectiveness through Behavioral Research Group (47) distinguishes among grand theories, mid-range theories, and micro-theories. These are summarized in Table 1.

Implementation frameworks can qualify as theory. Carpiano and Daley (46) define a conceptual framework as a specific set of variables and the relations among them that are presumed to account for a set of phenomena. There are multiple implementation frameworks (48–50) that could be assessed as fitting the definition of mid-range theories and could serve as a source of predictions (if-then) in a science of implementation.

To use theory to advance science, Gelman and Shalizi (45) suggest “we build a ... model out of available parts and drive it as far as it can take us, and then a little farther. When the model breaks down, we dissect it and figure out what went wrong.” “The goal is to learn about general laws, as expressed in the probability that one model or another is correct.”

Given that implementation practice and science are focused on outcomes (*if we do this, then that occurs so that people benefit*), consequential validity is an essential test of any theory-based prediction. Consequential validity favors external validity over internal validity (51, 52). Galea (53), working in a health context, described consequential validity:

A consequentialist approach is centrally concerned with maximizing desired outcomes A consequentialist epidemiology inducts new trainees not around canonical learning

TABLE 1 Levels of theory as defined by the improved clinical effectiveness through behavioral research group (47).

Scope	Definition	Purpose
Grand or Macro Theory	A grand or macro theory is a very broad theory that encompasses a wide range of phenomena	A grand theory is a general construction about the nature and goals of a discipline
Mid-Range Theory	A mid-range theory is more limited in scope, less abstract, addresses specific phenomena, and reflects practice. It encompasses a limited number of concepts and a limited aspect of the real world.	Mid-range theory is designed to guide empirical inquiry. Mid-range theories are made up of relatively concrete concepts that are operationally defined and relatively concrete propositions that can be empirically tested.
Micro Theory	A micro-practice, or situation-specific theory (sometimes referred to as prescriptive theory) has the narrowest range of interest.	Micro theory focuses on specific phenomena that reflect clinical practice and are limited to specific populations or to a particular field of practice.

but rather around our goals. Our purpose would be defined around health optimization and disease reduction, with our methods as tools, convenient only insofar as they help us get there.

By thinking of “our methods as tools, convenient only insofar as they help us get there,” the consequential validity question is “so what?” and the prediction/conclusion statement is “so that” (a phrase used throughout this paper). Once an implementation variable is postulated, it is incumbent on the researcher (the if-then predictor) to provide data that demonstrate how knowing that information “helps us get there” (i.e., behavior change so that outcomes are achieved).

A theory often is mentioned and rarely tested, improved, or discarded in implementation science. Prediction: no more than one published paper in any randomly selected set of 100 “implementation science” papers in the years 2005–2024 will clearly state how the study was designed to test a prediction based on an explicit theory of implementation.

Measurement

A science of interaction-based implementation depends on frequent measures of independent and dependent variables specific to implementation methods and outcomes.

Given the relative instability of interaction-based innovations, accurate measurement is paramount. A scientist (and the audience) must be assured that the implementation factor (if this: the independent variable) is present and at sufficient strength so that the results (then that: the dependent variable) reasonably can be attributed to the implementation factor (54–58). For interaction-based innovations, the independent variable must be measured repeatedly throughout an experiment with the same accuracy and care as the dependent variable.

For example, McIntosh et al. (59) examined fidelity of the use of Positive Behavior Interventions and Supports (PBIS). PBIS is an evidence-based multifaceted whole school intervention to reduce student discipline problems and improve academic outcomes. PBIS is conducted in the context of complex local and state education systems. The PBIS innovation involves all teachers and all students in each school along with parents, community members, and education system staff. The study included 5,331 schools located in 1420 school districts in 37 states. Each school that met fidelity criteria for the first time between 2005 and 2009 was included. The

fidelity data for each school was tracked for the following Years 1–5. The researchers found four distinct patterns of the use of PBIS over five years. There were two patterns of sustained implementation (sustainers and slow starters) and two patterns of abandonment (late abandoners and rapid abandoners). In Year 1, 3,904 of the schools continued to use PBIS as intended (i.e., met fidelity criteria). In Year 3, 2,735 schools met fidelity criteria and in Year 5, 2,239 schools met fidelity criteria. Given the patterns of use shown in the data, different schools did or did not meet fidelity criteria each year, the exception being the subgroup of sustainers (29% of the 5,331 schools). Thus, for individual schools, PBIS “was there” in some years and “not there” in others. With multiple interaction-based variables in play, knowledge of the fidelity of the use of PBIS (the independent variable) is essential for interpreting any outcome data, and assessing fidelity at one point in time was insufficient.

The relative instability of interaction-based implementation methods makes fidelity an essential implementation variable in every study. Fidelity answers the question: Is “it” there? When “it” is there a scientist can ask: Does “it” matter? Assessing fidelity is important for any independent variable (54), and especially important for interaction-based implementation variables. For example, are the “methods to assure executive leadership understanding and support” being used as intended (with fidelity)? Are the “methods to support practitioner competency development” being used as intended (with fidelity)? Are the “methods to create effective implementation teams” being used as intended (with fidelity)? If “it” is there, then the outcomes of “it” can be studied. Without an assessment of fidelity, it is difficult to know the extent to which all, some, or none of any intended method has been delivered. “Without fidelity measures, treatment becomes a mysterious black box: We do not know precisely what the intervention is, how to implement it, and what quality of it has been delivered. The black-box approach represents pre-scientific clinical care” [(60), p. 881].

Fortunately, fidelity is being recognized as a critical implementation measure across a number of fields using interaction-based inventions (61–71). Nevertheless, fidelity is not a requirement and is not commonly assessed in implementation science. Prediction: no more than one published paper in any randomly selected set of 100 “implementation science” papers in the years 2005–2024 will clearly state how the independent variable in the study was created/established and include an assessment of the presence and strength of the independent variable during the study.

Two examples

As outlined in this paper, the science of implementation will advance as evidence is generated to test predictions based on theory, interaction-based implementation variables are measured to assure their presence and strength, and theory is adjusted or discarded based on new evidence.

Evidence is “facts, information, etc. that give reasons for believing that something is true or present”.² Currently, implementation science judges evidence by the type of research design (72). This focus on form (evidence is the product of a research design) is seen to “contribute to the ... implementation gap” (73). In a science of implementation evidence is judged by its function (evidence that a variable depends on and varies with another). External validity and consequential validity are primary considerations and “our methods [are] tools, convenient only insofar as they help us get there.”

To establish a science of implementation, we propose the application of the following nine criteria to guide the development of research studies, as well as the analysis of the quality of the research design. In a science of implementation, evidence (the function of the evidence) can be judged using the following nine criteria.

- 1 Is the implementation independent variable clearly identified?
- 2 Is the implementation independent variable explicitly related to a theory?
- 3 Is the implementation independent variable explicitly related to implementation (i.e., the purposeful process of putting something into effect)?
- 4 Are the methods clearly stated for purposefully establishing the implementation independent variable (e.g., creating staff and management readiness, developing staff competencies, changing organization routines)?
- 5 Are the methods for establishing the implementation independent variable assessed (e.g., pre-post training tests of knowledge and skill, coaching done by a skilled coach)?
- 6 Is the implementation independent variable measured to detect its presence and strength (e.g., facilitation was done as intended)?
- 7 Is the implementation independent variable assessed frequently during the study (e.g., a minimum of three times, no more than three months apart)?
- 8 Does the implementation dependent variable focus on changes in behavior (e.g., practitioners, organizations, or systems)?
- 9 Are the results used to confirm, modify, or discard a theory?

These questions are based on the forgoing discussion of science, interaction-based variables, theory, and measurement. To illustrate the application of these questions, two reviews are provided. Research conducted by Acosta et al. (74) is an excellent example that meets all nine criteria. Research conducted by Seers et al. (75) was augmented by post-hoc analyses (76, 77) and is presented as the second example. In the comments below, the answers to the nine questions primarily are excerpts (without quote marks) or paraphrased from the cited documents. Thus, any first-person pronouns (e.g., I, we) in the

answers below refer to Acosta et al. or Seers et al. The studies cited below were not conducted with the questions in mind. A science of implementation will benefit from using the criteria as a guide for planning future implementation research.

Prediction: no more than one published paper in any randomly selected set of 100 “implementation science” papers in the years 2005–2024 will meet five or more of the nine “(function of) evidence” criteria outlined for a science of implementation.

- 1 Is the implementation independent variable clearly identified?
 - a Yes. Assets Getting To Outcomes (AGTO) is an implementation support intervention that consists of a manual of text and tools, face-to-face training, and onsite technical assistance, focused on activities shown to be associated with obtaining positive results across any prevention program.
- 2 Is the implementation independent variable explicitly related to a theory?
 - a Yes. The current study is the first to evaluate the Getting to Outcomes Framework (GTO) on both individual capacity and program performance in a way flexible enough to account for the ‘complex, interacting, multi-level, and transient states of constructs in the real world’. The Getting to Outcomes Framework is a mid-level theory of implementation.
 - b The AGTO model operationalizes the Consolidated Framework for Implementation Research (CFIR) to ensure that all the major domains that influence implementation are considered. Consistent with CFIR, the study attempts to evaluate AGTO in terms of its impact on three of the CFIR domains: inner setting, individual characteristics, and implementation process, while using measures of the outer setting as covariates. CFIR is a mid-level theory of implementation.
- 3 Is the implementation independent variable explicitly related to implementation (i.e., the purposeful process of putting something into effect)?
 - a Yes. As assessed in the survey, prevention capacity was defined as efficacy and behaviors of practitioners and relates to CFIR’s individual characteristics domain. To assess AGTO’s potential impact on each program’s implementation process, we first documented the degree to which each program engaged in the AGTO intervention through the AGTO participation index, which is the sum of six true/false items added to the mid and post survey.
- 4 Are the methods clearly stated for purposefully establishing the implementation independent variable (e.g., creating staff and management readiness, developing staff competencies, changing organization routines)?
 - a Yes. The AGTO intervention includes three types of assistance which are adapted to fit the needs and priorities of the individuals involved, as well as the inner and outer setting: a manual of text and tools, face-to-face training, and onsite technical assistance (TA). These three types of assistance aim to improve the implementation process for each program. Two full-time, Maine-based staff, one with a master’s and one with a bachelor’s degree, provided AGTO tools, training, and TA to the intervention coalitions and

² <https://dictionary.cambridge.org>

- programs during the two-year intervention period. The tools are in the Search Institute-published manual, *Getting To Outcomes with Developmental Assets: Ten steps to measuring success in youth programs and communities*, which all intervention participants received. The training was delivered separately to each coalition over a full day after baseline, and covered the AGTO model, tools in the manual, and an introduction to the TA process. Based on TA literature, the AGTO-based TA involves three structured steps, including an initial diagnosis of program functioning, development of a logic model, and development of a plan for how the TA and program staff were to make improvements, carried out during and in between biweekly TA visits. TA staff provided consultation and feedback to practitioners on conducting AGTO tasks as applied to their program.
- 5 Are the methods for establishing the implementation independent variable assessed (e.g., pre-post training tests of knowledge and skill, coaching done by a skilled coach)?
 - a Yes. To assess AGTO's potential impact on each program's implementation process, we first documented the degree to which each program engaged in the AGTO intervention through the AGTO participation index, which is the sum of six true/false items added to the mid and post coalition survey. Based on the Hall et al. model of categorizing the degree to which individuals 'use' an innovation, these items assess key markers of use including participation in training, reading the materials, planning, discussing the model with colleagues, securing resources, and receiving TA. Exposure to AGTO was also documented by TA providers recording hours of TA they delivered to each program, by AGTO step. The Participation Index and hours of TA have been shown to be related to prevention capacity and performance in a previous study of GTO.
 - 6 Is the implementation independent variable measured to detect its presence and strength (e.g., facilitation was done as intended)?
 - a Yes, see above. A structured interview was used to assess the impact of AGTO on the implementation process, administered on the same timeline as the survey.
 - 7 Is the implementation independent variable assessed frequently during the study (e.g., a minimum of three times, no more than three months apart)?
 - a Yes, see above, twice for AGTO and incrementally as GTO segments were taught and used.
 - b The survey assessed prevention capacity defined as efficacy and behaviors of practitioners and relates to CFIR's individual characteristics domain. The Asset Behaviors scale includes five items assessing whether individuals are motivating both adults and youth to become asset-builders; incorporating asset building into existing youth programs; influencing community leaders to implement asset-aligned policies.
 - 8 Does the implementation dependent variable focus on changes in behavior (e.g., practitioners, organizations, or systems)?
 - a Yes. Capacity of coalition members and performance of their programs were compared between the randomly selected groups (n = 6 in each) across the baseline, one-, and two-year timepoints.
 - b A structured interview was used to assess the impact of AGTO on the implementation process, administered on the same timeline as the coalition survey. Prevention practitioners performance of tasks associated with high-quality prevention targeted by AGTO were captured through the interview. Whole programs are rated, not individuals, because programs operate as a unit.
 - c The amount of change between the intervention and control groups did not significantly differ across the three time points for assets efficacy, GTO behaviors, or for any of the other prevention capacity scales in tests of condition by time interaction.
 - 9 Are the results used to confirm, modify, or discard a theory?
 - a Yes. Using the implementation research typology outlined in CFIR, this study evaluated the AGTO intervention's impact on the capacity of individual prevention practitioners (i.e., CFIR's individual characteristics) and the performance of whole programs (CFIR's implementation process), while accounting for several factors in CFIR's inner setting domain.
 - b The methods used in the current study also have implications for broader implementation research using the CFIR. The study used measures to assess the quality of the implementation process as well as the relevant capacity of practitioners (i.e., individuals involved) that could be adapted to many other interventions and exist already for interventions in the areas of substance abuse prevention, positive youth development, prevention of teen pregnancy and sexually transmitted infections, and homelessness prevention. These measures were used in a randomized controlled trial and our analytic approach addressed certain challenges of conducting research on real-world implementation (e.g., the turnover of membership at participating coalitions, differential response rates between intervention and control groups). More innovative methods and more acceptance of nontraditional research methods will be needed to capture the complex, multi-level and transient status of constructs in the real-world as the study of implementation evolves.
- The following nine questions and answers pertain to the Seers et al. studies.
- 1 Is the implementation independent variable clearly identified?
 - a Yes. In the context of the PARIHS framework, facilitation refers to the process of enabling (making easier) the implementation of evidence into practice. Thus, facilitation is achieved by an individual carrying out a specific role (a facilitator), which aims to help others. This indicates that facilitators are individuals with the appropriate roles, skills, and knowledge to help individuals, teams, and organizations apply evidence into practice.... To fulfill the potential demands of the role, facilitators likely will need a wide repertoire of skills and attributes.
 - 2 Is the implementation independent variable explicitly related to a theory?
 - a Yes. The PARIHS framework (78) is a mid-level theory of implementation that predicts a key role for facilitation in producing implementation outcomes. A test of this theory-based prediction was conducted by a PARIHS group of researchers (75–77).

- 3 Is the implementation independent variable explicitly related to implementation (i.e., the purposeful process of putting something into effect)?
- a Yes. In theory, if staff in an organization use facilitation as a process, then the organization will be more likely to translate research evidence into practice.
- 4 Are the methods clearly stated for purposefully establishing the implementation independent variable (e.g., creating staff and management readiness, developing staff competencies, changing organization routines)?
- a Yes. To teach the wide repertoire of skills and attributes needed by facilitators, the research team prepared two different facilitator development programs, each of which involved an initial residential program, followed by virtual support (monthly telephone group supervision and email communication) for the internal facilitators (IFs) in implementing the urinary incontinence (UI) recommendations.
- 5 Are the methods for establishing the implementation independent variable assessed (e.g., pre-post training tests of knowledge and skill, coaching done by a skilled coach)?
- a No. The methods section identifies two implementation independent variables, (1) training (initial residential program) and (2) coaching (followed by virtual support), and one implementation dependent variable (prepare the internal facilitators). The effects of training and coaching were not measured during the experiment or in post-hoc analyses. Staff behavior also was not measured, a critical link between the potential impact of facilitators on staff behavior and patient outcomes.
- 6 Is the implementation independent variable measured to detect its presence and strength (e.g., facilitation was done as intended)?
- a No.
- 7 Is the implementation independent variable assessed frequently during the study (e.g., a minimum of three times, no more than three months apart)?
- a No. Fidelity (facilitation as provided by the internal facilitators) was not measured during the study but was assessed retrospectively with record reviews and interviews (76).
- 8 Does the implementation dependent variable focus on changes in behavior (e.g., practitioners, organizations, or systems)?
- a No. The primary outcome was the documented percentage compliance with continence recommendations produced by the fourth International Consultation on Incontinence. Percentage compliance is calculated for each resident, so outcomes are measured at the resident level.
- 9 Are the results used to confirm, modify, or discard a theory?
- a No. Facilitation as defined by PARIHS was not present and, therefore, the predicted influences of facilitation and the PARIHS theory were not tested in this study.

provides two examples of how those dimensions relate to the evidence on which a science of implementation can be based.

- A science of implementation is based on if-then predictions. Science is cumulative. As predictions are made, tested, and elaborated, the facts accumulate to form the knowledge base for science and practice.
- Implementation variables are interaction-based inventions and, therefore, must be created and established so the specific set of activities related to implementation can be studied.
- A science of implementation is based on theory that organizes facts, leads to testable predictions, and is modified or discarded based on outcomes.
- A science of interaction-based implementation depends on frequent measures of independent and dependent variables specific to implementation methods and outcomes.

The loosely related ideas, assumptions, and findings currently under the label implementation science fall far short of meeting the qualifications to be considered science, and little of the literature relates to purposeful implementation processes created to put something into effect. Despite the massive investment in the “evidence-based movement” in the past three decades, the science to service gap remains and the contributions of implementation science to society have been modest at best. Six explicit predictions are made in this paper (“no more than one published paper in any randomly selected set of 100 “implementation science” papers in the years 2005–2024...”). The predictions likely will hold up if they are tested, a clear indication that implementation science is not a science of implementation. Following the tenets of science, if this-then *not* that, it may be time to give up that paradigm and try something new.

The lack of socially significant outcomes was noted early in the current evidence-based movement. Perl (18) was concerned about the “addiction” to creating new evidence-based programs and Kessler and Glasgow (52) called for a 10-year moratorium on efficacy-focused randomized controlled trials (RCTs) in health and health services research. These authors encouraged an investment in a science of implementation. Calling for a “revolution,” Kruk et al. (79) (p. 1196) note that “Changing health needs, growing public expectations, and ambitious new health goals are raising the bar for health systems to produce better health outcomes and greater social value. But staying on current trajectory will not suffice to meet these demands.”

The revolution begins with changes in how a science of implementation is understood. Specifically, the characteristics of a science as outlined in this paper challenge deeply held beliefs about implementation: what it is and how to develop new knowledge about it.

First, a science of implementation argues against “tailoring” as advocated in implementation science (80, 81). A science of implementation is based on understanding that it is necessary for scientists to assure the fidelity of the independent variable to counter the inherent instability of interaction-based inventions. An assessment of fidelity (is it there?) is a requirement. Encouraging users to alter methods may increase their acceptability to users but not their benefits to recipients. In science, changing methods changes outcomes.

Second, the goal of science is to establish general truths about how things work. In this sense, it is a quest to develop implementation methods where “one size fits all.” Any plan for

Discussion

Is implementation science a science? Not yet. This paper summarizes the dimensions of a science of implementation and

scaling for population impact will consist of many one-size-fits-all processes so that “it” can be done with fidelity many thousands of times with improvement methods built into the process. “Build a ... model out of available parts and drive it as far as it can take us, and then a little farther. When the model breaks down, we dissect it and figure out what went wrong” [(45), p. 32]. As the science and the processes of one size fits all are put into practice, evidence is generated and double loop learning can occur. In this way the pool of effective methods expands to incorporate effective responses to what previously was unanticipated. The expanded methods then can benefit a greater proportion of the variations encountered in communities, service settings, and organizations. The new, more robust “one” then can bene-“fit” a greater proportion of “all.”

Third, if-then predictions can be tested in many ways to demonstrate functional relationships between implementation independent and dependent variables (if this is done, then that happens). The Acosta et al. (74) example demonstrates how each implementation variable (participation in training, reading the materials, planning, discussing the model with colleagues, securing resources) can be assessed for fidelity (is “it” there?) in the context of doing multifaceted work in complex environments over longer periods of time. The first question is, can it be done at all, even once? Given the counterfactuals for interaction-based variables that must be invented (it has never been done, not even once), pre-post experiments may be sufficient to assess strong variables and learn how to create them on demand. Thinking of randomized control trials (“RCTs”) as the gold standard emphasizes form over function. While RCTs may be appropriate later on as the science matures, initially they get in the way of creating powerful independent variables and developing a science of implementation.

These implications mean the “mental model” (82, 83) for implementation research and practice needs to change in pursuit of a science of implementation. A mental model is an intuitive perception about acts and their consequences, tacit assumptions of how the world works. How do we think about problems and potential solutions? As Ashby (84) said, “The fault cannot be in the part responsible for the repair.” Exhortations to “think outside the box” are invitations to change mental models. “Deep change in mental models, or double loop learning, arises when evidence not only alters our decisions within the context of existing frames, but also feeds back to alter our mental models. As our mental models change, we change the structure of our systems, creating different decision rules and new strategies” [(83), p. 509]. Tailoring, one size cannot fit all, and the value of RCTs are accepted truths in implementation science and are impediments to establishing a science of implementation.

Getting the science right for implementation is critical for global public health. Global health fundamentally requires scaling implementation capacity (i.e., purposeful processes to put something in place) so that effective innovations can be used as intended and with good effect for whole populations. As predicted in the formula for success, scaling effective implementation is critical for scaling effective innovations to achieve population benefits.

Atom-based innovations are scaled by developing manufacturing capacity. For example, Apple corporation has 615 production facilities (implementation capacity) that produce about 240 million iPhones (the innovation) per year. Apple designs the products, specifies the manufacturing processes, and requires 1,200 quality assurance staff in

each contracted production facility to achieve 80% yield rates for products that pass final quality control standards (fidelity).³

Interaction-based innovations are scaled by developing implementation teams (implementation capacity). Scaling to achieve public health outcomes begins by developing a scalable unit (85). A scalable unit is an administrative unit that includes the practitioners who deliver the effective innovation; the effective implementation team that supports the selection, competency development, and fidelity assessment of practitioners; and administrators to assure an enabling context of funding, licensing, and so on. As noted in the description of the formula for success in the introduction, any one part of a scalable unit without the others will not produce and sustain local or (eventually) population benefits. Initial testing of the scalable unit can be done at a single site if that site represents key system components and relationships among components likely to be encountered in the system at full scale (85–87).

The methods to establish a scalable unit are detailed by Titler et al. (87) in their description of the Iowa Model of Research-Based Practice. “The practice is first implemented with a small group of patients, and an evaluation is carried out. The EBP is then refined based on evaluation data, and the change is implemented with additional patient populations for which it is appropriate. Patient/family, staff, and fiscal outcomes are monitored. Organizational and administrative supports are important factors for success in using evidence in care delivery.” Thus, the elements essential to effective innovation, effective implementation, and enabling context factors that reliably produce socially significant outcomes are worked out before attempting scaling to achieve population benefits. Of course, yesterday’s solutions provide the platform for tomorrow’s problems, so the improvement processes never end until population goals are reached and sustained.

Scaling is possible only when scalable units are sustained. If already established units are dropping out or losing their effectiveness as fast as new units can be created, then scaling and socially significant benefits plateau at that point. Metaphorically, the public health container is draining as fast as it is being filled. Thus, testing of the scalable unit is essential. For example, Fixsen, Blase, Timbers, and Wolf (88) found when the scalable unit only focused on developing practitioners to deliver the effective innovation, 17% of the residential treatment units sustained for six or more years. When the focus shifted to developing whole organizations as the scalable unit so that the implementation team and administration also were included, 84% of the residential treatment programs sustained for six years or more. Testing the scalable unit and getting it right made a big difference in sustainability of effective implementation and innovation practices.

With a scalable unit tested and ready, the next task is to develop implementation teams at the district level whose purpose is to develop and support scalable units at the local level. Then, regional implementation teams must be developed to establish more district implementation teams and continually support them. And, finally, a national implementation team must be developed to establish and support a growing number of regional implementation teams. Establishing teams can be done simultaneously (or nearly so) in a top-down and bottom-up sequence so that the linked teams can be created expeditiously. The good news is that a science of implementation is universal, so the work of the various

³ <https://9to5mac.com/2023/01/17/iphone-quality-standards>

implementation teams is similar although with different goals (12, 89). The same concepts for scaling can be found in descriptions of holarchies (90) and fractals (32, 91) where units with similar functions are repeated at various points of scale. Common concepts, common language, and common measures promote clear communication and coordinated effort among the linked implementation teams [(31); Chapter 15].

Developing implementation capacity locally, regionally, and nationally was essential in the successful efforts to eradicate smallpox (accomplished in 1979). As described by Foege (92), thousands of local surveillance teams and containment teams (the effective innovation) were supported by hundreds of regional implementation teams (effective implementation) that were developed with the support of national and global governments and global health organizations (enabling contexts). In India there were 29 states and 386 districts subdivided into “blocks” (roughly 100,000 population in a block; a scalable unit for surveillance). Competency development for many thousand health workers had to be done well and be repeatable in several hundred training sessions provided at each level in each state. In India, training was provided for staff at each level: state, district, and block. In one state, preparations for the first search required over 60 training sessions at the regional level simply to get down to the district level, and an additional 930 training sessions at the district and block levels. This process for developing implementation capacity was repeated in the other 28 states so that effective surveillance teams and containment teams could reach the entire population of India (600 million). Implementation capacity was established, and smallpox was eradicated.

Developing implementation capacity locally, regionally, and nationally also is essential in the successful efforts to implement the National Health Policy in Ethiopia (beginning in 1998). The National Health Policy gives strong emphasis to fulfilling the needs of less privileged rural communities, which constitute about 83% of the total population of 90 million (93). Like the smallpox surveillance teams and containment teams for a “block” of 100,000 population, Ethiopia recruited and developed two HEWs (health education workers) for each “health post” serving 3,000–5,000 population. HEWs are recruited based on nationally agreed criteria that include residence in the village, capacity to speak local language, graduation from 10th grade, and willingness to remain in the village and serve rural and distressed communities. Selection is done by a committee comprising members nominated by the local community and representatives from the woreda (district) health office, the woreda capacity-building office, and the woreda education office. All selected HEWs go through a year-long training, which includes both theoretical training in training institutions and practical training in health centers. By 2014, over 30,000 HEWs had been deployed in health posts nationally.

Wang et al. (93) described linked implementation (“supervisory”) teams consisting of members from different disciplines at the federal, regional, and woreda levels in Ethiopia. The implementation teams are involved in all aspects of program management, including planning, implementation, and monitoring and evaluation. Members of each team are trained in skills needed for supportive supervision (facilitation, interpersonal communication, problem solving, and analytical skills); oriented to various tools and methods (such as peer review and

performance assessment tools); and provided with opportunities to frequently upgrade their technical skills. The implementation team members are trained in a specially designed curriculum. At each level (federal, regional, and woreda), the implementation team prepares its own annual plan, checklists, and detailed schedule for each supervisory visit. The work of implementation teams cascades from regional level down to woreda, health center, and health posts. Implementation teams also actively engage regional and woreda councils for issues that go beyond the health sector itself (93).

As illustrated in these examples, realizing socially significant benefits requires effective innovations supported by effective implementation in enabling contexts. Each factor needs to be purposeful, functional, and improvable so that benefits to whole populations are good to start with and improve with experience. Without a science of implementation, global health and well-being will remain an aspiration and not an achievement.

A science of implementation is within our reach. If experiments focus on implementation variables and if independent and dependent implementation variables are measured accurately and repeatedly to determine their function, then a science of implementation can be developed, and populations can benefit.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

DF: Formal analysis, Writing – original draft, Conceptualization. MD: Conceptualization, Writing – review & editing. KB: Conceptualization, Writing – review & editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Conflict of interest

DF, MD, and KB were employed by Active Implementation Research Network, Inc.

Publisher’s note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Azrin NH. A strategy for applied research: learning based but outcome oriented. *Am Psychol.* (1977) 32:140–9. doi: 10.1037/0003-066X.32.2.140
- Fairweather GW. *Methods for experimental social innovation.* Oxford: John Wiley + Sons (1967).
- Flanagan SG, Cray ME, Van Meter D. A facility-wide consultation and training team as a catalyst in promoting institutional change. *Anal Interv Dev Disabil.* (1983) 3:151–69. doi: 10.1016/0270-4684(83)90012-5
- Hall GE, Loucks SF. A developmental model for determining whether the treatment is actually implemented. *Am Educ Res J.* (1977) 14:263–76. doi: 10.3102/00028312014003263
- Havelock RG. *Planning for innovation through dissemination and utilisation of knowledge.* Ann Arbor, MI: Center for Research on Utilisation of Scientific Knowledge (1969).
- Paine SC, Bellamy GT, Wilcox B. *Human services that work: From innovation to standard practice.* Baltimore, MD: Paul H. Brookes Publishing (1984).
- Pressman JL, Wildavsky A. *Implementation: How great expectations in Washington are dashed in Oakland.* Berkeley, CA: University of California Press (1973).
- Rossi PH, Wright JD. Evaluation research: An assessment. *Annu Rev Sociol.* (1984) 10:331–52. doi: 10.1146/annurev.so.10.080184.001555
- Stolz SB. Adoption of innovations from applied behavioral research: "does anybody care?". *J Appl Behav Anal.* (1981) 14:491–505. doi: 10.1901/jaba.1981.14.491
- Van Meter DS, Van Horn CE. The policy implementation process: a conceptual framework. *Adm Society.* (1975) 6:445–88. doi: 10.1177/009539977500600404
- Beidas RS, Dorsey S, Lewis CC, Lyon AR, Powell BJ, Purtile J, et al. Promises and pitfalls in implementation science from the perspective of US-based researchers: learning from a pre-mortem. *Implement Sci.* (2022) 17:55. doi: 10.1186/s13012-022-01226-3
- Fixsen DL, Blase KA, Metz A, Van Dyke MK. Statewide implementation of evidence-based programs. *Except Children.* (2013) 79:213–30. doi: 10.1177/0014402913079002071
- Fixsen DL, Van Dyke MK, Blase KA. Implementation science for evidence-based policy In: BC Welsh, SN Zane and DP Mears, editors. *The Oxford handbook of evidence-based crime and justice policy.* Oxford: Oxford University Press (2024). 58–75.
- Søreide E, Morrison L, Hillman K, Monsieurs K, Sunde K, Zideman D, et al. The formula for survival in resuscitation. *Resuscitation.* (2013) 84:1487–93. doi: 10.1016/j.resuscitation.2013.07.020
- Roberts MC, Hinton-Nelson M. Models for service delivery in child and family mental health In: MC Roberts, editor. *Model programs in child and family mental health.* Mahwah, NJ: Lawrence Erlbaum Associates (1996). 1–21.
- Sackett DL, Rosenberg WMC, Gray JAM, Haynes RB, Richardson WS. Evidence-based medicine: what it is and what it isn't. *Br Med J.* (1996) 312:71–2. doi: 10.1136/bmj.312.7023.71
- Clancy CM, Slutsky JR, Patton LT. AHRQ update: evidence-based health care 2004: AHRQ moves research to translation and implementation. *Health Serv Res.* (2004) 39:15–23. doi: 10.1111/j.1475-6773.2004.00286.x
- Perl HI. Addicted to discovery: does the quest for new knowledge hinder practice improvement? *Addict Behav.* (2011) 36:590–6. doi: 10.1016/j.addbeh.2011.01.027
- Braithwaite J, Churrua K, Long JC, Ellis LA, Herkes J. When complexity science meets implementation science: a theoretical and empirical analysis of systems change. *BMC Med.* (2018) 16:63. doi: 10.1186/s12916-018-1057-z
- Kruk ME, Gage AD, Joseph NT, Danaei G, García-Saisó S, Salomon PJA. Mortality due to low-quality health systems in the universal health coverage era: a systematic analysis of amenable deaths in 137 countries. *Lancet.* (2018) 392:2203–12. doi: 10.1016/S0140-6736(18)31668-4
- Pfadenhauer LM. Conceptualizing context and intervention as a system in implementation science: learning from complexity theory. *Int J Health Policy Manag.* (2021) 11:1570–3. doi: 10.34172/ijhpm.2021.147
- Bortolotti L. *An introduction to the philosophy of science.* Cambridge: Polity Press (2008).
- Kincaid H. *The Oxford handbook of philosophy of social science.* Oxford, UK: Oxford University Press (2012).
- Beard MT. Theory construction and testing: An introduction and overview In: MT Beard, editor. *Theory construction and testing.* Lisle, IL: Tucker Publications, Inc (1995). 1–18.
- Popper K. *Conjectures and refutations: The growth of scientific knowledge.* New York: Harper Torchbooks (1963).
- Popper K. *The logic of scientific discovery.* New York: Routledge (2002).
- Greenhalgh T, Robert G, MacFarlane F, Bate P, Kyriakidou O. Diffusion of innovations in service organizations: systematic review and recommendations. *Milbank Q.* (2004) 82:581–629. doi: 10.1111/j.0887-378X.2004.00325.x
- Wootton D. *The invention of science: A new history of the scientific revolution.* New York: Harper Collins (2015).
- Gertner J. *The idea factory: Bell labs and the great age of American innovation.* New York: The Penguin Press (2012).
- Fixsen DL, Blase KA, Fixsen AAM. *Scaling effective innovations.* *Criminol Public Policy.* (2017) 16:487–99. doi: 10.1111/1745-9133.12288
- Fixsen DL, Blase KA, Van Dyke MK. *Implementation practice and science.* 1st ed. Chapel Hill, NC: Active Implementation Research Network (2019).
- Morin E. *On complexity.* Creskill, NJ: Hampton Press (2007).
- Forgatch MS, DeGarmo DS. Sustaining Fidelity following the Nationwide PMTO implementation in Norway. *Prev Sci.* (2011) 12:235–46. doi: 10.1007/s11121-011-0225-6
- Schoenwald SK, Brown TL, Henggeler SW. Inside multisystemic therapy: therapist, supervisory, and program practices. *J Emot Behav Disord.* (2000) 8:113–27. doi: 10.1177/10634266000800207
- Tiruneh GT, Karim AM, Avan BI, Zemichael NF, Wereta TG, Wickremasinghe D, et al. The effect of implementation strength of basic emergency obstetric and newborn care (BEmONC) on facility deliveries and the met need for BEmONC at the primary health care level in Ethiopia. *BMC Pregnancy Childbirth.* (2018) 18:123. doi: 10.1186/s12884-018-1751-z
- Compaoré R, Yameogo MWE, Millogo T, Tougri H, Kouanda S. Evaluation of the implementation fidelity of the seasonal malaria chemoprevention intervention in Kaya health district, Burkina Faso. *PLoS ONE.* (2017) 12:187460. doi: 10.1371/journal.pone.0187460
- McNicholas C, Lennox L, Woodcock T, Bell D, Reed JE. Evolving quality improvement support strategies to improve plan–do–study–act cycle fidelity: a retrospective mixed-methods study. *Br Med J Quality Safety.* (2019) 28:356–65. doi: 10.1136/bmjqs-2017-007605
- D'Andreamatteo A, Ianni L, Lega F, Sargiacomo M. Lean in healthcare: a comprehensive review. *Health Policy.* (2015) 119:1197–209. doi: 10.1016/j.healthpol.2015.02.002
- Fixsen DL, Naoom SF, Blase KA, Friedman RM, Wallace F. *Implementation research: A synthesis of the literature.* National Implementation Research Network. Tampa, FL: University of South Florida (2005).
- Appleman D. *How computer programming works.* Berkeley, CA: Apress (2000).
- Humble J, Farley D. *Continuous delivery.* Boston: Pearson Education (2011).
- An M, Dusing SC, Harbourne RT, Sheridan SM. What really works in intervention? Using Fidelity measures to support optimal outcomes. *Phys Ther.* (2020) 100:757–65. doi: 10.1093/ptj/pzaa006
- Rahman M, Ashraf S, Unicomb L, Mainuddin AKM, Parvez SM, Begum F, et al. WASH benefits Bangladesh trial: system for monitoring coverage and quality in an efficacy trial. *Trials.* (2018) 19:360. doi: 10.1186/s13063-018-2708-2
- Tofail F, Fernald LC, Das KK, Rahman M, Ahmed T, Jannat KK, et al. Effect of water quality, sanitation, hand washing, and nutritional interventions on child development in rural Bangladesh (WASH benefits Bangladesh): a cluster-randomised controlled trial. *Lancet Child Adolescent Health.* (2018) 2:255–68. doi: 10.1016/S2352-4642(18)30031-2
- Gelman A, Shalizi CR. Philosophy and the practice of Bayesian statistics. *Br J Math Stat Psychol.* (2013) 66:8–38. doi: 10.1111/j.2044-8317.2011.02037.x
- Carpiano RM, Daley DM. A guide and glossary on postpositivist theory building for population health. *J Epidemiol Community Health.* (2006) 60:564–70. doi: 10.1136/jech.2004.031534
- ICEBeRG Group. Designing theoretically-informed implementation interventions. *Implement Sci.* (2006) 1:4. doi: 10.1186/1748-5908-1-4
- Meyers DC, Durlak JA, Wandersman A. The quality implementation framework: a synthesis of critical steps in the implementation process. *Am J Community Psychol.* (2012) 50:462–80. doi: 10.1007/s10464-012-9522-x
- Nilsen P, Birken SA. *Handbook on implementation science.* Cheltenham, UK: Edward Elgar Publishing (2020).
- Tabak RG, Khoong EC, Chambers DA, Brownson RC. Bridging research and practice: models for dissemination and implementation research. *Am J Prev Med.* (2012) 43:337–50. doi: 10.1016/j.amepre.2012.05.024
- Green LW. Making research relevant: if it is an evidence-based practice, where's the practice-based evidence? *Fam Pract.* (2008) 25:i20–4. doi: 10.1093/fampra/cmn055
- Kessler RC, Glasgow RE. A proposal to speed translation of healthcare research into practice: dramatic change is needed. *Am J Prev Med.* (2011) 40:637–44. doi: 10.1016/j.amepre.2011.02.023
- Galea S. An argument for a consequentialist epidemiology. *Am J Epidemiol.* (2013) 178:1185–91. doi: 10.1093/aje/kwt172
- Dobson L, Cook T. Avoiding type III error in program evaluation: results from a field experiment. *Eval Program Plann.* (1980) 3:269–76. doi: 10.1016/0149-7189(80)90042-7

55. Glasgow RE, Klesges LM, Dzawaltowski DA, Bull SS, Estabrook P. The future of health behavior change research: what is needed to improve translation of research into health promotion practice? *Ann Behav Med.* (2004) 27:3–12. doi: 10.1207/s15324796abm2701_2
56. Jensen PS, Weersing R, Hoagwood KE, Goldman E. What is the evidence for evidence-based treatments? A hard look at our soft underbelly. *Ment Health Serv Res.* (2005) 7:53–74. doi: 10.1007/s11020-005-1965-3
57. Korfmacher J, Kitzman HJ, Olds DL. Intervention processes as predictors of outcomes in a preventive home-visitation program. *J Community Psychol.* (1998) 26:49–64. doi: 10.1002/(SICI)1520-6629(199801)26:1<49::AID-JCOP5>3.0.CO;2-X
58. Michie S, Fixsen D, Grimshaw J, Eccles M. Specifying and reporting complex behaviour change interventions: the need for a scientific method. *Implement Sci.* (2009) 4:40. doi: 10.1186/1748-5908-4-40
59. McIntosh K, Mercer SH, Nese RNT, Ghemraoui A. Identifying and predicting distinct patterns of implementation in a school-wide behavior support framework. *Prev Sci.* (2016) 17:992–1001. doi: 10.1007/s11121-016-0700-1
60. Bond GR, Drake RE. Assessing the Fidelity of evidence-based practices: history and current status of a standardized measurement methodology. *Adm Policy Ment Health Ment Health Serv Res.* (2020) 47:874–84. doi: 10.1007/s10488-019-00991-6
61. Aarons GA, Sommerfeld DH, Hecht DB, Silovsky JF, Chaffin MJ. The impact of evidence-based practice implementation and fidelity monitoring on staff turnover: evidence for a protective effect. *J Consult Clin Psychol.* (2009) 77:270–80. doi: 10.1037/a0013223
62. Collier-Meek MA, Sanetti LM, Fallon L, Chafouleas S. Exploring the influences of assessment method, intervention steps, intervention sessions, and observation timing on treatment Fidelity estimates. *Assess Eff Interv.* (2020) 46:3–13. doi: 10.1177/1534508419857228
63. Hamid S, Mureed S, Kayani A, Javed K, Khan A, Awais S, et al. Learning active implementation frameworks: the role of implementation teams in a case study from Pakistan. *Glob Health Action.* (2020) 13:1805164. doi: 10.1080/16549716.2020.1805164
64. Ledford JR, Wolery M. Procedural Fidelity: An analysis of measurement and reporting practices. *J Early Interv.* (2013) 35:173–93. doi: 10.1177/1053815113515908
65. Palacio A, Garay D, Langer B, Taylor J, Wood BA, Tamariz L. Motivational interviewing improves medication adherence: a systematic review and Meta-analysis. *J Gen Intern Med.* (2016) 31:929–40. doi: 10.1007/s11606-016-3685-3
66. Rookes TA, Schrag A, Walters K, Armstrong M. Measures of fidelity of delivery and engagement in self-management interventions: a systematic review of measures. *Clin Trials.* (2022) 19:665–72. doi: 10.1177/17407745221118555
67. Sanetti LMH, Cook BG, Cook L. Treatment Fidelity: what it is and why it matters. *Learn Disabil Res Pract.* (2021) 36:5–11. doi: 10.1111/ldrp.12238
68. Strain P, Fox L, Barton EE. On expanding the definition and use of procedural Fidelity. *Res Practice Persons Severe Disabili.* (2021) 46:173–83. doi: 10.1177/15407969211036911
69. Taylor MJ, McNicholas C, Nicolay C, Darzi A, Bel D, Reed JE. Systematic review of the application of the plan–do–study–act method to improve quality in healthcare. *Br Med J Qual Safety.* (2014) 23:290–8. doi: 10.1136/bmjqs-2013-001862
70. Thurman T, Chowdhury S, Taxman FS. Fidelity measures for risk-need assessment (RNA) tools usage in case plans. *Corrections.* (2021) 6:383–99. doi: 10.1080/23774657.2019.1696252
71. Tong F, Tang S, Irby BJ, Lara-Alecio R, Guerrero C. Inter-rater reliability data of classroom observation: Fidelity in large-scale randomized research in education. *Data Brief.* (2020) 29:105303. doi: 10.1016/j.dib.2020.105303
72. Haskins R., Baron J. (2011). Building the connection between policy and evidence: The Obama evidence-based initiatives. Available online at: www.nesta.org.uk
73. McGinty EE, Alegria M, Beidas RS, Braithwaite J, Kola L, Leslie DL, et al. The lancet psychiatry commission: Transforming mental health implementation research. *The Lancet.* (2024) 11:368–96. doi: 10.1016/S2215-0366(24)00040-3
74. Acosta J, Chinman M, Ebener P, Malone PS, Paddock S, Phillips A, et al. An intervention to improve program implementation: findings from a two-year cluster trial of assets-getting to outcomes. *Implement Sci.* (2013) 8:87. doi: 10.1186/1748-5908-8-87
75. Seers K, Rycroft-Malone J, Cox K, Crichton N, Edwards RT, Eldh AC, et al. Facilitating implementation of research evidence (FIRE): an international cluster randomised controlled trial to evaluate two models of facilitation informed by the promoting action on research implementation in health services (PARIHS) framework. *Implement Sci.* (2018) 13:137. doi: 10.1186/s13012-018-0831-9
76. Harvey G, McCormack B, Kitson A, Lynch E, Titchen A. Designing and implementing two facilitation interventions within the ‘facilitating implementation of research evidence (FIRE)’ study: a qualitative analysis from an external facilitators’ perspective. *Implement Sci.* (2018) 13:141. doi: 10.1186/s13012-018-0812-z
77. Rycroft-Malone J, Seers K, Eldh AC, Cox K, Crichton N, Harvey G, et al. A realist process evaluation within the facilitating implementation of research evidence (FIRE) cluster randomised controlled international trial: an exemplar. *Implement Sci.* (2018) 13:138. doi: 10.1186/s13012-018-0811-0
78. Rycroft-Malone J. The PARIHS framework: a framework for guiding the implementation of evidence-based practice. *J Nurs Care Qual.* (2004) 19:297–304. doi: 10.1097/00001786-200410000-00002
79. Kruk ME, Gage AD, Arsenault C, Jordan K, Leslie HH, Roder-DeWan S, et al. High-quality health systems in the sustainable development goals era: time for a revolution. *Lancet Global Health.* (2018) 6:e1196–252. doi: 10.1016/S2214-109X(18)30386-3
80. Albers B, Rapley T, Nilsen P, Clack L. Editorial: tailoring in implementation science. *Front Health Serv.* (2023) 3:1233597. doi: 10.3389/frhs.2023.1233597
81. Baker R, Camosso-Stepinovic J, Gillies C, Shaw EJ, Cheater F, Flottorp S, et al. Tailored interventions to address determinants of practice. *Cochrane Database Syst Rev.* (2015) 2015:CD005470. doi: 10.1002/14651858.CD005470.pub3
82. Holtrop JS, Scherer LD, Matlock DD, Glasgow RE, Green LA. The importance of mental models in implementation science. *Front Public Health.* (2021) 9:680316. doi: 10.3389/fpubh.2021.680316
83. Stermann JD. Learning from evidence in a complex world. *Am J Public Health.* (2006) 96:505–14. doi: 10.2105/ajph.2005.066043
84. Ashby WR. An introduction to cybernetics. London: Chapman & Hall (1956).
85. Barker PM, Reid A, Schall MW. A framework for scaling up health interventions: lessons from large-scale improvement initiatives in Africa. *Implement Sci.* (2016) 11:12. doi: 10.1186/s13012-016-0374-x
86. Fixsen DL, Blase KA, Van Dyke MK. From ghost systems to host systems via transformation zones. Washington, DC: U.S. Department of Education Office of Vocational and Adult Education (2012).
87. Titler MG, Kleiber C, Steelman VJ, Rakel BA, Budreau G, Everett LQ, et al. The Iowa model of evidence-based practice to promote quality care In: J Rycroft-Malone and T Bucknell, editors. Models and frameworks for implementing evidence-based practice: Linking evidence to action. New York: John Wiley & Sons (2001). 137–46.
88. Fixsen DL, Blase KA, Timbers GD, Wolf MM. In search of program implementation: 792 replications of the teaching-family model. *The Behavior Analyst Today.* (2001/2007) 8:96–110. doi: 10.1037/h0100104
89. Ward CS, Ihlo T, Ryan Jackson K, Farmer S. Effective implementation capacity to impact change within state education systems to support students with disabilities. *J Disabil Policy Stud.* (2022) 34:104–14. doi: 10.1177/10442073221096392
90. Koestler A. The ghost in the machine. 1990 reprint ed. New York: Penguin Group (1967).
91. Gleick J. Chaos: Making a new science. Harmondsworth, Middlesex: Penguin Books, Ltd (1987).
92. Foege WH. House on fire: The fight to eradicate smallpox. Los Angeles, CA: University of California Press, Ltd (2011).
93. Wang H, Tesfaye R, Ramana GNV, Chekagn CT. Ethiopia health extension program: An institutionalized community approach for universal health coverage. Washington, DC: World Bank Group (2016).