



## OPEN ACCESS

## EDITED BY

Noel Kalanga,  
University of Malawi, Malawi

## REVIEWED BY

Suzanne Kerns,  
University of Colorado, United States  
Thomas J. Waltz,  
Eastern Michigan University, United States

## \*CORRESPONDENCE

Dean L. Fixsen  
✉ [dfixsen1@gmail.com](mailto:dfixsen1@gmail.com)

RECEIVED 11 February 2025

ACCEPTED 14 April 2025

PUBLISHED 25 April 2025

## CITATION

Fixsen DL (2025) Fidelity, not adaptation, is essential for implementation.  
*Front. Health Serv.* 5:1575179.  
doi: 10.3389/frhs.2025.1575179

## COPYRIGHT

© 2025 Fixsen. 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.

# Fidelity, not adaptation, is essential for implementation

Dean L. Fixsen\*

Active Implementation Research Network, Inc., Albuquerque, NM, United States

Fidelity is not yet a requirement when developing an evidence-based innovation or when attempting to use an innovation in typical settings. Currently, users are encouraged to adapt innovations to fit existing practitioner skills and organization situations. Instead of adapting innovations, the essential components of an innovation need to be established in the original research and the essential components need to be used in practice with the support of implementation processes so that promised outcomes can be realized. Fidelity is an assessment of the presence and strength of the essential components that define the independent variable (the innovation) and is directly linked to outcomes. A test of any fidelity assessment is a high correlation (0.70+) with outcomes. The functional relationship between fidelity and outcomes ensures that the essential components are effective and ensures that a reliable fidelity assessment is available. Implementation is the planned process of putting something into effect. Evidence that an innovation has been put into effect is provided by the fidelity assessment. High fidelity scores indicate that the essential components of the innovation are in place and good outcomes are expected. A test of any planned process is fidelity of the use of the innovation. At present fidelity assessments are missing or inadequate and, therefore, there is a notable lack of evidence that an independent variable is present.

## KEYWORDS

fidelity, essential components, science, implementation, scaling

## Introduction

Fidelity often is viewed as optional when developing an evidence-based innovation or when attempting to use evidence-based innovations to benefit individuals and society (1). Instead of using an innovation as intended (with fidelity), a “science of adaptation” has been proposed (2) for the process of adapting innovations. An alternative view is that fidelity is essential. Fidelity is integral to the definition of an innovation, is essential when developing an evidence-based innovation, and is the standard to meet when using an innovation (3–5). Instead of adapting the innovation, the goal of implementation is to change practitioner, organization, and system behavior so that innovations can be used with fidelity and good outcomes. What is needed is a science of implementation, not a science of adaptation.

The essential role of fidelity assessment in science, in implementation, and in practice is summarized in this paper.

## Fidelity in science

In science, fidelity is an assessment of the presence and strength of the independent variable in experiments to establish if-then relationships—if “this” is done, then “that”

happens consistently. Experiments that test if-then predictions provide the evidence that is the foundation for any science (6, 7).

An experiment provides evidence of a functional relationship, that is, one variable systematically affects another. “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. For interaction-based innovations, the independent variable must be measured repeatedly throughout an experiment with the same accuracy and care as the dependent variable” (8). Thus, the function of fidelity in science is to assure the scientific community that the independent variable “is there”—we did “this” with a known level of strength, and “that” outcome was produced (or not).

In science, any credible experiment to test if-then relationships provides (a) a clear description of the essential components of the independent variable, (b) indicators that the essential components of the independent variable are present and at sufficient strength to be tested, and (c) evidence that outcomes are directly attributable to the essential components of the independent variable. The direct tie between essential components and indicators of the presence and strength of those components means that fidelity always is specific to an innovation [referred to as program differentiation (9, 10)]. Fidelity provides evidence that the essential components of *this* innovation are present and at sufficient strength to have an impact. An experiment provides evidence that the essential components are highly related to the outcomes (if this, then that). In science, outcomes cannot be attributed to something that is not there, although in practice specious attributions are not uncommon (11).

Thus, a fidelity assessment is a requirement for any research to develop an evidence-based innovation. The scientific community needs to know what “it” is, and “it” needs to be assessed to assure the scientific community that “it” was present and used as described (12). While there are several ways to develop an assessment (9, 13), the test of any fidelity assessment is its relationship with (i.e., prediction of) intended outcomes. For example, a positive or negative correlation of 0.70 or better would indicate that the essential components have been adequately identified and assessed, and they are effective [a correlation coefficient greater than 0.7 is considered strong; (14)]. A correlation of 0.70 or better explains 50% or more of the variance in outcomes. If there is a strong relationship, then the innovation would be “worth doing” with high fidelity so that socially significant outcomes could be achieved.

The fidelity-outcome relationship was tested in a study to develop a fidelity assessment for cognitive behavioral therapy for insomnia. The study found a 0.30 correlation between therapist fidelity scores and treatment outcomes, explaining about 10% of the variability in outcomes (15). In another example, the Washington State Institute for Public Policy (16) examined fidelity of the use of Functional Family Therapy (FFT) (17–19). The 12-month post treatment delinquency outcomes were assessed for referred youths in the 427 families treated by 25 FFT therapists. An analysis of the data found a  $-0.61$  correlation

between the fidelity of the therapists’ use of FFT and youth recidivism, explaining about 36% of the variability in delinquency outcomes. A quintile analysis found 8% recidivism among the youths in families treated by FFT therapists in the top 20% of fidelity scores, and 34% recidivism for the youths in families treated by FFT therapists in the bottom 20% of fidelity scores. When FFT was present and at sufficient strength (top 20%), outcomes for youths, families, and society were 4X better. The fidelity-outcome relationship provides evidence that the essential components have been (more or less) adequately specified and are effective—important information for potential users.

An, Dusing (4) recommend setting a standard for fidelity of the use of the independent variable (e.g., 80%) that must be met *before* an experiment begins. For example, Tofail, Fernald (20) conducted a study of individual and combined water, sanitation, handwashing and child nutrition interventions delivered by community health workers (CHWs) to pregnant women and their infants in 4,169 households in Bangladesh. Three months after initiating the experiment, assessments indicated fidelity was low (30%–60% range). Extra support for the CHWs was provided, fidelity scores improved (86%–93% range), and only then was the experiment conducted (21). Fidelity scores remained high (22) and ensured the continued presence and strength of the multifaceted independent variable. As a result, the researchers provided a true test of the combined water, sanitation, handwashing and child nutrition interventions—“it” was there, and outcomes could be evaluated and attributed to “it”.

In perhaps the most elegant implementation experiment ever conducted (23), a  $2 \times 2$  design was carried out in 14 rural Appalachian counties to test the effects of organization support on implementation success. The design included 2 factors: (a) the random assignment of delinquent youth within each county to a multisystemic therapy (MST) program or usual services and (b) the random assignment of counties to the ARC (Availability, Responsiveness, and Continuity) organizational intervention. MST teams were developed using established implementation protocols for therapist selection, training, and supervision, and therapist fidelity was regularly assessed. ARC specialists were trained and supervised by the ARC developers at the University of Tennessee and fidelity was monitored with on-site observation and activity logs. The combination of evidence-based treatment (MST) and facilitative organization support (ARC) produced the best outcomes for delinquent youths. Fidelity assessments provided assurance that each complex evidence-based innovation (MST and ARC) was there and at sufficient strength to conduct a credible test of their individual and combined effects in multiple counties over four years.

Fidelity assessment requires attention to the essential components, what “it” is and the key indicators of the presence and strength of “it.” With a required fidelity assessment, high fidelity [at least 80% according to An, Dusing (4)] ensures the independent variable was present and at sufficient strength to provide a valid test of its effects. If fidelity is not high, or not strongly related to outcomes, the time to correct the problem is during the original research. Have the essential components been identified adequately? Have the essential components been

measured adequately? If the identification and assessment of the essential components are adequate, do the essential components need to be improved (discarded, changed) to produce better outcomes? Deferring the solution to these fundamental questions puts the onus on potential users who want to benefit others. However, inadequate science plus local adaptation likely will not equal socially significant benefits for intended beneficiaries.

In science, fidelity is directly linked to the essential components of an innovation, fidelity provides indicators of the presence and strength of the essential components, and fidelity is highly correlated with outcomes. With a firm commitment to fidelity, “this” is defined, “that” is known, and “this” and “that” are improvable as the science evolves.

## Fidelity in implementation

Implementation is the planned process for putting something into effect (24–27). Thus, a “planned process” is the implementation independent variable (if this), and “putting something into effect” is the implementation dependent variable (then that). In a science of implementation, fidelity assessment is doubly important: (a) it provides an indication of the presence and strength of the implementation independent variable (the planned process), and (b) it provides an indication that the essential components of something have been put into effect (the implementation dependent variable).

What is the “something” (i.e., defined by the essential components) and how do we know it was “put into effect” (i.e., assessed with a measure of fidelity)? In a science of implementation, innovation fidelity always is a dependent variable, an outcome of effective implementation processes (25, 28). Thus, innovation fidelity has a dual role as an implementation dependent variable and an innovation independent variable. This is a common feature in nested systems where one component, simultaneously, is a singular unit and a part of a larger whole (29, 30). In effect, every independent variable at one level is a dependent variable at the next level (8, 31).

Logically, (a) implementation specialists engage in high fidelity implementation processes, (b) so that practitioners will provide high fidelity services, (c) so that recipients will benefit. Proctor, Bunger (32) found that studies related to this predicted relationship are not common. In their analysis of 400 studies of implementation outcomes, Proctor, Bunger (32) found 22 studies relating implementation outcomes with service outcomes, with 2 of those studies focusing on fidelity. Similarly, in a search for repeated measures of implementation variables, Fixsen, Van Dyke (33) found 17 articles that assessed innovation fidelity two or more times in the course of an experiment. Thus, although innovation fidelity is recognized as an implementation dependent variable, it is not studied frequently in implementation science.

Assessing innovation fidelity immediately directs attention to implementation processes. If fidelity is low, instead of adapting the innovation, what implementation processes can be used to prepare practitioners to use 80% or more of the innovation’s essential components consistently? What implementation

processes can be used to help an organization change to effectively support practitioners’ use on the innovation? What implementation processes can be used to help leaders and managers provide leadership for change, or change policies and procedures to support the continuing and effective use of an innovation? These are proximal implementation variables and have an immediate effect on the use (or nonuse) of any innovation (31, 34–37). If fidelity is high and outcomes are poor, the next right step is to conduct experiments to re-examine the putative essential components (try again, back to fidelity in science). If fidelity is not assessed at all, there are no next right steps and there is no prompt to improve the innovation or attend to implementation processes.

Fidelity is a necessary and critical link between implementation processes and ultimately achieving intended outcomes. In this process, potential users are not encouraged to adapt the very things (the essential components) that produce desired outcomes. Improved fit almost always requires changing practitioner behavior and organization and system behavior so that the essential components of an innovation can be used successfully (38–40). The processes for changing practitioner, organization, and system behavior are implementation independent variables. Practitioners, organizations, and systems that attempt to use innovations without making any changes in their ways of work—doing the same thing again and again and expecting different results—are certain to fail.

Implementation independent variables are “planned processes” that have an immediate and longer-term effect on the use (or nonuse) of any innovation. In a science of implementation, innovation fidelity is always a dependent variable for implementation independent variables, the test that “something has been put into effect”.

## Preparing for everyday use

An innovation should not be expected to be usable in general practice until it has undergone usability testing (39, 41). For multifaceted and complex interaction-based innovations in human services, well defined and operationalized essential components likely will be incomplete. And measurement of each component may not be feasible given the sometimes private or fast paced nature of human interactions. Even so, scientists must make every effort to define the essential components and find credible ways to assess their presence and strength. The test is that fidelity and outcomes are highly related.

Usability testing is a well-established, systematic approach to “working out the bugs” in any complex program or system intended for general use (39, 42–45). Usability testing is based on plan-do-study-act-cycle (PDSAC) logic (46–48). In usability testing, a small number of participants ( $n \sim 5$ ) attempt to use an innovation in each Cycle. The Plan is to use the essential components of the innovation. Each participant then Does the plan. The testing team Studies the results: did the individuals Do the Plan (fidelity) and to what extent were intended outcomes achieved? The testing team Acts on the information by changing the innovation and modifying the fidelity assessment to reflect

the changes. The participants in the next group then use the essential components of the improved innovation (the new Plan) in the next Cycle. This process is repeated until the innovation (the Plan) is improved to the point that intended outcomes are achieved reliably and are highly correlated with fidelity scores. Three to 5 cycles may be sufficient to detect and correct 80% or more of the errors in the original Plan (49).

In usability testing, with the evolving fidelity assessment as the standard, factors that negatively impact achieving the standard can be detected and corrected without compromising the function—if this, then that. Fidelity is the “bug detector,” an indication that something is getting in the way of doing what is required (i.e., the essential components) to produce desired outcomes. With each iteration in usability testing, “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” (8). The result is a set of robust and generalizable implementation methods that support the full and effective (high fidelity) use of the innovation so that desired outcomes can be achieved consistently. Of course, usability testing must be done with fidelity to be effective (47, 50).

Successive groups of 5 users to detect and correct errors is recommended by the developers of usability testing (42, 43). Barker, Reid (51) advocate usability testing with increasing numbers of users to ensure exposure to an increasing range of real-life situations. Barker, Reid (51) provide examples where “the rate of expansion can be exponential (i.e., not linear) by a multiple of 5... (e.g., 1–5–25–125–625, etc.).” At each level, external validity is strengthened as revised methods are established to resolve newly exposed problems before moving to the next level. Appropriate adaptations become a part of the definition of an innovation as exposure to new problems invites new solutions so that the problems are solved, and outcomes are achieved. In this way, “the pool of effective methods” is expanded to include constructive responses to variations related to culture, race, gender, socio-economic conditions, geography, seasons, territorial conflicts, local contexts, and so on. Usability testing also can detect the limits of the use of the innovation, the conditions under which the essential components do not produce the outcomes found under other conditions. For example, a well-defined innovation for youths adjudicated as serious delinquents does not produce similar outcomes for youths with severe mental health problems (52). Usability testing is work for the scientific community.

Unfortunately, usability testing is not common. Instead of doing the work required to establish evidence-based innovations and evidence-based implementation processes, the challenges are passed to potential users. The notion of tailoring asks thousands of potential users of an innovation to do the work the developers were unwilling or unable to do—specify the essential components and provide indicators to assess their presence and strength. Increasing uncertainty and variability by tailoring, and shifting responsibility to local users will not solve the problems confronting human services (53–58).

For researchers and program developers, usability testing is an extra step to establish the internal validity and external validity of

an innovation before it is released for general use. Extra steps are not unusual in science. Early in the evidence-based innovation movement, concerns about the “evidence” led to CONSORT guidelines related to the quality of randomized controlled trials (RCTs) [described by Altman, Schulz (59)]. As Eldridge, Ashby (60) stated, internal validity can be strengthened with “good design, conduct, and analysis of the trial, with minimal bias, ... and sufficient sample size.” In the seminal CONSORT paper, there is no mention of fidelity, no mention of an independent variable, and only 3 uses of the word “intervention” where researchers were encouraged to “suggest a plausible explanation for how the intervention under investigation might work” (p. 667). With regard to evidence, Altman, Schulz (59) summarized CONSORT by stating, “Reports of RCTs should be written with... close attention to minimizing bias. Readers should not have to speculate; the methods used should be transparent, so that readers can readily differentiate trials with unbiased results from those with questionable results. Sound science encompasses adequate reporting, and the conduct of ethical trials rests on the footing of sound science”.

Currently, the growing science to service gap has led to the guidelines outlined in this paper regarding the internal and external validity of the innovation itself. For evidence-based innovations, it is not enough to have rigorously derived “evidence,” the “innovation” also must be well defined. The CONSORT statement can be paraphrased: Readers should not have to speculate; with a usable innovation readers can readily differentiate innovations with clearly specified essential components from those with questionable components. Sound science encompasses adequate description and measurement of essential components, and the conduct of ethical usability testing rests on the footing of sound science.

Avoiding fidelity assessment is avoiding learning what we need to know to create evidence-based implementation processes. Encouraging users to adapt methods may increase their acceptability to users but not their benefits to recipients. In science and in practice, changing methods changes outcomes.

## Fidelity in practice

Fidelity sets a minimum standard, the least that users need to do in order to say they are “using” an innovation (61–64). Taking anything to a useful scale requires increasing standardization of innovations, implementation supports, and operating environments to reduce unwanted (potentially harmful) sources of variability (5, 65–68). Achieving a useful standard requires greater attention to innovation development and implementation methods so that thousands of practitioners can use evidence-based innovations with fidelity to provide benefits to whole populations.

For example, smallpox was eradicated globally using surveillance teams and containment teams (69). The teams were the innovation independent variables in the efforts to eradicate smallpox for the population on Earth. Foege (70) and colleagues, working in rural Africa in the 1960s, developed containment

teams to isolate and treat the infected and inoculate the exposed. In their early work (i.e., usability testing) they relied on local networks to identify newly infected people. Foege and colleagues found that local networks often missed new outbreaks of smallpox. The response was to establish surveillance teams to systematically and reliably find those who were infected. As they gained experience and collected more data, they established protocols (i.e., operationalized essential components) for each team and “increased the pool of effective methods” based on effectiveness and efficiency outcomes. Foege (70) recounts the challenges faced in India in the 1970s (population over 600 million in 27 states). National, state, regional, and local implementation supports were established so that surveillance teams and containment teams could be created and sustained in every state and each “block” of 100,000 people so that they could reach all the people in each urban neighborhood and each village. Foege’s story is about increasing the specificity of protocols, increasing the frequency of fidelity, process, and outcome monitoring, and increasing the reliance on data so that effectiveness and efficiency of the surveillance teams and containment teams were immediately and continually improved.

“The strategy for smallpox eradication did not change from country to country, but the local culture determined which tactics were most useful. Only the specific locality can provide information on who is sick, who is hiding from the vaccinators, when people are available for vaccination, how to hire watch guards, or how to secure the cooperation of the community. In all cultures, an approach of respect for local customs is needed” (70). Thus, “methods to respect local customs” was one of the essential components of the standard protocol for a surveillance team.

Surveillance teams and containment teams were not adapted to fit local contexts. Instead, the protocols for implementation processes, teams, fidelity assessments, and outcome assessments were standardized to reduce variability and error and improve outcomes. In one month, surveillance teams searched 140,000 villages in one state using standard protocols. The surveillance teams did not stop 140,000 times to figure out how the essential components of surveillance teams should be adapted. Instead, the surveillance teams used the standard protocols 140,000 times with high fidelity to accomplish the intended outcomes.

In human services, standard fidelity assessments for evidence-based programs have been developed and used for many years across many contexts. For example, the Teaching-Family Model began in 1967 as a group home residential treatment program for delinquent youths. After the first replication attempt failed in 1971, a fidelity assessment was developed, tested, and refined (71–73). Early work (i.e., usability testing) provided evidence that practitioner development was insufficient for sustainability (74). This led to developing Teaching-Family organizations with implementation teams built into each organization, and sustainability (5+ years) improved from 15% ( $n=84$  group homes) to 83% ( $n=219$  group homes) (75). The Teaching-Family fidelity assessment has been used in every Teaching-Family treatment service setting (i.e., group home, foster family, homebased, or school-based) for over 50 years (75–77).

Fidelity assessment has been used on a large scale by Positive Behavior Interventions and Supports (PBIS). PBIS is an

evidence-based multifaceted whole school intervention developed to reduce student discipline problems and improve academic outcomes (78, 79). Fidelity assessments were developed and tested (80), revised (81), and used as PBIS expanded to over 30,000 of the 100,000 schools in the US. McIntosh, Mercer (82) analyzed PBIS fidelity data each year for 5 years for 5,331 schools located in 1,420 school districts in 37 states. They found evidence of shifting patterns of fidelity they categorized as sustainers, slow starters, late abandoners, and rapid abandoners. Similarly, Motivational Interviewing (83) is in widescale use with standard assessments of fidelity and supervision (84–86).

Forgatch, Patterson (87) used elements of two previously developed observation methods and, “through an iterative process” (i.e., usability testing), established new protocols for observation of the Parent Management Training-Oregon (PMTO) program essential components. The PMTO fidelity assessment has been used in the national scale up of PMTO in Norway (88). The first “generations” of PMTO practitioners learned from the original researchers, became certified PMTO therapists, and then learned to carry out the implementation supports with fidelity in Norway and beyond (89). A 10-year assessment of implementation capacity development (i.e., recruitment, training, coaching, fidelity assessments, administration, leadership, etc.) for scale up in Norway found continued high fidelity use of essential implementation components, high fidelity use of essential innovation components by “generations” of practitioners, and sustained benefits for children and families (5, 90).

These examples of long-term and functional fidelity are not typical at this stage of the evidence-based movement. Reviews of the literature over the past several decades consistently find that fidelity is not measured in the majority of outcome studies, and repeated use of any fidelity measure is even less common (91, 92). When it is present, fidelity assessment has focused on form (e.g., frequency, duration, dosage, participation) rather than function (i.e., fidelity scores are highly correlated with desired outcomes). Eventually, with a high correlation as a benchmark, the wide variety of “fidelity measures” will be replaced by functional ones that can be relied on by potential users. In the meantime, potential users will continue to cope with innovations with uncertain essential components and questionable fidelity assessments, and scaling to achieve socially significant benefits will remain an aspirational goal.

Fortunately, there are fidelity assessments in everyday use in many sectors to ensure the expanded and sustained use of evidence-based innovations (21, 31, 52, 93–99). In everyday use, the essential components of an innovation must be present (used with fidelity) so that their outcomes can be produced.

## Summary

To close the science to service gap and produce benefits to recipients of those services (i.e., the goals of implementation in human services), we need to get the science right, right from the beginning. For any innovation or implementation independent variable, scientists must specify the essential components, provide indicators (fidelity measures) of the presence and strength of



those essential components, and provide evidence that outcomes are strongly associated with the strength of those essential components. For any use of an innovation or implementation independent variable (practice, program, or policy), fidelity is the standard to achieve so that desired outcomes can be realized. With a firm commitment to fidelity, “this” is defined, “that” is known, and “this” and “that” are improvable as the science evolves.

Once a usable innovation is established, potential users can vary non-essential components to suit their circumstances. The realities of human services often require modifications in the delivery of services, and practitioners introduce their personality into service delivery. Fidelity data provide evidence to determine when modifications have “gone too far” and have compromised the essential components. As variations occur, maintaining the fidelity-outcome correlation is paramount so that benefits accrue to the intended population.

Fidelity is thoroughly embedded in the definition of any intervention that is evidence-based or scalable. Bond and Drake (100), pioneers in implementation research, note that, “Fidelity specification and measurement confer multifarious benefits to funders, program managers, clinicians, researchers, and patients. 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. On the other hand, fidelity measurement provides clarity regarding the intervention model, its differentiation from other models, and its degree of implementation”.

At present fidelity assessments are missing or inadequate and, therefore, there is a notable lack of evidence that an independent variable is there. Consequently, the “science” in implementation science is not progressing as it might, and potential users are left wondering what to put in place to reliably produce promised benefits to people.

## 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.

## References

1. Albers B, Verweij L, Blum K, Oesch S, Schultes MT, Clack L, et al. Firm, yet flexible: a fidelity debate paper with two case examples. *Implement Sci.* (2024) 19(1):79. doi: 10.1186/s13012-024-01406-3
2. Chambers DA. Advancing adaptation of evidence-based interventions through implementation science: progress and opportunities. *Front Health Serv.* (2023) 3:1204138. doi: 10.3389/frhs.2023.1204138
3. Schoenwald SK, Garland AF. A review of treatment adherence measurement methods. *Psychol Assess.* (2013) 25(1):146–56. doi: 10.1037/a0029715
4. An M, Dusing SC, Harbourne RT, Sheridan SM. What really works in intervention? Using fidelity measures to support optimal outcomes. *Phys Ther.* (2020) 100(5):757–65. doi: 10.1093/ptj/pzaa006
5. Tommeraaas T, Ogden T. Is there a scale-up penalty? Testing behavioral change in the scaling up of parent management training in Norway. *Adm Policy Ment Health.* (2016) 44:203–16. doi: 10.1007/s10488-015-0712-3
6. Beard MT. Theory construction and testing: an introduction and overview. In: Beard MT, editor. *Theory Construction and Testing*. Lisle, IL: Tucker Publications, Inc. (1995). p. 1–18.
7. Popper K. *Conjectures and Refutations: The Growth of Scientific Knowledge*. New York: Harper Torchbooks (1963).
8. Fixsen DL, Van Dyke MK, Blase KA. Is implementation science a science? Not yet. *Front Public Health.* (2024) 12:1454268. doi: 10.3389/fpubh.2024.1454268
9. Carroll C, Patterson M, Wood S, Booth A, Rick J, Balain S. A conceptual framework for implementation fidelity. *Implement Sci.* (2007) 2(1):40. doi: 10.1186/1748-5908-2-40
10. Dusenbury L, Brannigan R, Falco M, Hansen WB. A review of research on fidelity of implementation: implications for drug abuse prevention in school settings. *Health Educ Res.* (2003) 18(2):237–56. doi: 10.1093/her/18.2.237

## Author contributions

DF: Conceptualization, Investigation, Writing – original draft, Writing – review & editing.

## Funding

The authors declare that no financial support was received for the research and/or publication of this article.

## Acknowledgments

The content of this paper benefitted considerably from the insightful comments and suggestions made by the two reviewers.

## Conflict of interest

DF was employed by Active Implementation Research Network, Inc.

## Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

## 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.

11. 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
12. Durlak JA, DuPre EP. Implementation matters: a review of research on the influence of implementation on program outcomes and the factors affecting implementation. *Am J Community Psychol.* (2008) 41:327–50. doi: 10.1007/s10464-008-9165-0
13. Lemire C, Rousseau M, Dionne C. A comparison of fidelity implementation frameworks used in the field of early intervention. *Am J Eval.* (2023) 44(2):236–52. doi: 10.1177/10982140211008978
14. U.S. Government Accounting Office. Program evaluation: A variety of rigorous methods can help identify effective interventions (2009). Available at: <http://gao.gov/products/GAO-10-30> (Accessed November 24, 2010).
15. Cross WF, McCarten J, Funderburk JS, Crean HF, Lockman J, Titus CE, et al. Measuring fidelity of brief cognitive behavior therapy for insomnia: development, reliability and validity. *Eval Program Plann.* (2024) 109:102531. doi: 10.1016/j.evalproplan.2024.102531
16. Washington State Institute for Public Policy. *Washington State's Implementation of Functional Family Therapy for Juvenile Offenders: Preliminary Findings.* Olympia, WA: Washington State Institute for Public Policy (2002). Report No.: 02-08-1201.
17. Sexton TL, Alexander JF. Functional family therapy. *Juv Justice Bull.* (2000):1–7.
18. Alexander JF, Pugh C, Parsons B, Sexton TL. Functional family therapy. In: Elliott DS, editor. *Book Three: Blueprints for Violence Prevention.* 2nd ed. Golden, CO: Venture (2000). p. 3–79.
19. Alexander JF, Parsons B. Short-term family intervention: a therapy outcome study. *J Consult Clin Psychol.* (1973) 2:195–201.
20. 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 Adolesc Health.* (2018) 2(4):255–68. doi: 10.1016/S2352-4642(18)30031-2
21. 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(1):360. doi: 10.1186/s13063-018-2708-2
22. Masud Parvez S, Azad R, Rahman MM, Unicomb L, Ram P, Naser AM, et al. Achieving optimal technology and behavioral uptake of single and combined interventions of water, sanitation hygiene and nutrition, in an efficacy trial (WASH benefits) in rural Bangladesh. *Trials.* (2018) 19(358):1–16. doi: 10.1186/s13063-018-2710-8
23. Glisson C, Schoenwald SK, Hemmelgarn A, Green P, Dukes D, Armstrong KS, et al. Randomized trial of MST and ARC in a two-level evidence-based treatment implementation strategy. *J Consult Clin Psychol.* (2010) 78(4):537–50. doi: 10.1037/a0019160
24. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. *Implement Sci.* (2009) 4(50):1–15. doi: 10.1186/1748-5908-4-50
25. 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). p. iii-119.
26. Rycroft-Malone J. The PARiHS framework: a framework for guiding the implementation of evidence-based practice. *J Nurs Care Qual.* (2004) 19(4):297–305. doi: 10.1097/00001786-200410000-00002
27. Greenhalgh T, Robert G, MacFarlane F, Bate P, Kyriakidou O. Diffusion of innovations in service organizations: systematic review and recommendations. *Milbank Q.* (2004) 82(4):581–629. doi: 10.1111/j.0887-378X.2004.00325.x
28. Wiltsey Stirman S, Gutner CA, Crits-Christoph P, Edmunds J, Evans AC, Beidas RS. Relationships between clinician-level attributes and fidelity-consistent and fidelity-inconsistent modifications to an evidence-based psychotherapy. *Implement Sci.* (2015) 10(115):1–10. doi: 10.1186/s13012-015-0308-z
29. Koestler A. *The Ghost in the Machine.* 1990 reprint ed. NY: Penguin Group (1967).
30. Fixsen DL, Blase KA, Van Dyke MK. *Implementation Practice and Science.* 1st ed. Chapel Hill, NC: Active Implementation Research Network, Inc. (2019). p. 378.
31. Schoenwald SK, Sheidow AJ, Letourneau EJ. Toward effective quality assurance in evidence-based practice: links between expert consultation, therapist fidelity, and child outcomes. *J Clin Child Adolesc Psychol.* (2004) 33(1):94–104. doi: 10.1207/S15374424JCCP3301\_10
32. Proctor EK, Bunger AC, Lengnick-Hall R, Gerke DR, Martin JK, Phillips RJ, et al. Ten years of implementation outcomes research: a scoping review. *Implement Sci.* (2023) 18(1):31. doi: 10.1186/s13012-023-01286-z
33. Fixsen DL, Van Dyke MK, Blase KA. Repeated measures of implementation variables. *Front Health Serv.* (2023) 3:1–9. doi: 10.3389/frhs.2023.1085859
34. Jacobs C, Michelo C, Moshabela M. Implementation of a community-based intervention in the most rural and remote districts of Zambia: a process evaluation of safe motherhood action groups. *Implement Sci.* (2018) 13(1):74. doi: 10.1186/s13012-018-0766-1
35. Caron EB, Dozier M. Effects of fidelity-focused consultation on clinicians' implementation: an exploratory multiple baseline design. *Adm Policy Ment Health.* (2019) 46:445–57. doi: 10.1007/s10488-019-00924-3
36. Skogøy BE, Sørgaard K, Maybery D, Ruud T, Stavnes K, Kufås E, et al. Hospitals implementing changes in law to protect children of ill parents: a cross-sectional study. *BMC Health Serv Res.* (2018) 18(1):609. doi: 10.1186/s12913-018-3393-2
37. Joyce B, Showers B. *Student Achievement Through Staff Development.* 3rd ed. Alexandria, VA: Association for Supervision and Curriculum Development (2002). p. 217.
38. Nord WR, Tucker S. *Implementing Routine and Radical Innovations.* Lexington, MA: D. C. Heath and Company (1987).
39. Hirschhorn LR, Semraul K, Kodkany B, Churchill R, Kapoor A, Spector J, et al. Learning before leaping: integration of an adaptive study design process prior to initiation of BetterBirth, a large-scale randomized controlled trial in Uttar Pradesh, India. *Implement Sci.* (2015) 10(117):1–9. doi: 10.1186/s13012-015-0309-y
40. Webster-Stratton CH, Reid JM, Marsenich L. Improving therapist fidelity during implementation of evidence-based practices: incredible years program. *Psychiatr Serv.* (2014) 65(6):789–95. doi: 10.1176/appi.ps.201200177
41. Epstein D, Klerman JA. When is a program ready for rigorous impact evaluation? The role of a falsifiable logic model. *Eval Rev.* (2013) 36:375–401. doi: 10.1177/0193841X12474275
42. Genov A. Iterative usability testing as continuous feedback: a control systems perspective. *J Usability Stud.* (2005) 1(1):18–27.
43. Nielsen J. Usability for the masses. *J Usability Stud.* (2005) 1(1):2–3.
44. Akin BA, Bryson SA, Testa MF, Blase KA, McDonald T, Melz H. Usability testing, initial implementation, and formative evaluation of an evidence-based intervention: lessons from a demonstration project to reduce long-term foster care. *Eval Program Plann.* (2013) 41(0):19–30. doi: 10.1016/j.evalproplan.2013.06.003
45. 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. *Crit Care Nurs Clin North Am.* (2001) 13(4):497–509. doi: 10.1016/S0899-5885(18)30017-0
46. Speroff T, Connor O, T G. Study designs for PDSA quality improvement research. *Qual Manag Health Care.* (2004) 13(1):17–32. doi: 10.1097/00019514-200401000-00002
47. 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. *BMJ Qual Saf.* (2014) 23:290–8. doi: 10.1136/bmjqs-2013-001862
48. Shewhart WA. *Statistical Method from the Viewpoint of Quality Control.* Dover Publications (1939).
49. Nielsen J. Why you only need to test with 5 users. Alertbox, March 19, 2000. (2000). Available at: <https://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/> (Accessed April 22, 2007).
50. 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. *BMJ Qual Saf.* (2019) 28(5):356–65. doi: 10.1136/bmjqs-2017-007605
51. 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(1):12. doi: 10.1186/s13012-016-0374-x
52. Schoenwald SK, Brown TL, Henggeler SW. Inside multisystemic therapy: therapist, supervisory, and program practices. *J Emot Behav Disord.* (2000) 8(2):113–27. doi: 10.1177/10634266000800207
53. U.S. Public Health Service. *Report of the Surgeon General's Conference on Children's Mental Health: A National Action Agenda.* Washington DC: US Public Health Service (2000).
54. U.S. Department of Education Institute of Education Sciences. *The Condition of Education.* Washington, DC: U.S. Department of Education (2010). Available online at: [http://nces.ed.gov/ipeds/data/ipedsdatatools/2010/ipeds\\_data\\_tool\\_tables/](http://nces.ed.gov/ipeds/data/ipedsdatatools/2010/ipeds_data_tool_tables/)
55. Kruk ME, Gage AD, Arseneault 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 Glob Health.* (2018) 6:e1196–252. doi: 10.1016/S2214-109X(18)30386-3
56. McGinty EE, Alegria M, Beidas RS, Braithwaite J, Kola L, Leslie DL, et al. The lancet psychiatry commission: transforming mental health implementation research. *Lancet Psychiatry.* (2024) 11:368–96. doi: 10.1016/S2215-0366(24)00040-3
57. World Health Organization. *Health in 2015: From MDGs, Millennium Development Goals to SDGs, Sustainable Development Goals.* Geneva, Switzerland: World Health Organization (2015).
58. United Nations. *The Sustainable Development Goals Report.* New York: United Nations Publications (2023).

59. Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D, et al. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Ann Intern Med.* (2001) 134(8):663–94. doi: 10.7326/0003-4819-134-8-200104170-00012
60. Eldridge S, Ashby D, Bennett C, Wakelin M, Feder G. Internal and external validity of cluster randomised trials: systematic review of recent trials. *Br Med J.* (2008) 336(7649):876–80. doi: 10.1136/bmj.39517.495764.25
61. Blase KA, Fixsen DL. *Core Intervention Components: Identifying and Operationalizing What Makes Programs Work.* Washington, DC: Office of the Assistant Secretary for Planning and Evaluation, Office of Human Services Policy, U.S. Department of Health and Human Services (2013).
62. Hall G, Loucks SF. A developmental model for determining whether the treatment is actually implemented. *Am Educ Res J.* (1977) 14(3):263–76. doi: 10.3102/00028312014003263
63. Hall G, Hord SM. *Implementing Change: Patterns, Principles and Potholes.* 4th ed. Boston: Allyn and Bacon (2011).
64. Bond GR, Salyers MP. Prediction of outcome from the dartmouth assertive community treatment fidelity scale. *CNS Spectr.* (2004) 9(12):937–42. doi: 10.1017/S1092852900009792
65. Winter SG, Szulanski G. Replication as strategy. *Organ Sci.* (2001) 12(6):730–43. doi: 10.1287/orsc.12.6.730.10084
66. Huaynoca S, Chandra-Mouli V, Jr Y, Denno N, M D. Scaling up comprehensive sexuality education in Nigeria: from national policy to nationwide application. *Sex Educ.* (2013) 14(2):191–209. doi: 10.1080/14681811.2013.856292
67. Fixsen DL, Blase KA, Fixsen AAM. Scaling effective innovations. *Criminol Public Policy.* (2017) 16(2):487–99. doi: 10.1111/1745-9133.12288
68. Strain P, Fox L, Barton EE. On expanding the definition and use of procedural fidelity. *Res Pract Persons Severe Disabl.* (2021) 46(3):173–83. doi: 10.1177/15407969211036911
69. Fenner F, Henderson DA, Arita I, JeZek Z, Ladnyi ID. *Smallpox and its Eradication.* Geneva, Switzerland: World Health Organization (1988).
70. Foege WH. *House on Fire: The Fight to Eradicate Smallpox.* Berkeley and Los Angeles, CA: University of California Press, Ltd. (2011).
71. Braukmann CJ, Fixsen DL, Kirigin KA, Phillips EA, Phillips EL, Wolf MM. Achievement place: the training and certification of teaching-parents. In: Wood WS, editor. *Issues in Evaluating Behavior Modification.* Champaign, IL: Research Press (1975). p. 131–52.
72. Bedlington MM, Braukmann CJ, Ramp KA, Wolf MM. A comparison of treatment environments in community-based group homes for adolescent offenders. *Crim Justice Behav.* (1988) 15:349–63. doi: 10.1177/0093854888015003007
73. Bedlington MM, Solnick JV, Schumaker JB, Braukmann CJ, Kirigin KA, Wolf MM, editors. *Evaluating Group Homes: The Relationship Between Parenting Behaviors and Delinquency.* Toronto, Canada: American Psychological Association Convention (1978). p. 1–12.
74. Blase KA, Fixsen DL, Phillips EL. Residential treatment for troubled children: developing service delivery systems. In: Paine SC, Bellamy GT, Wilcox B, editors. *Human Services That Work: From Innovation to Standard Practice.* Baltimore, MD: Paul H. Brookes Publishing (1984). p. 149–65.
75. Fixsen DL, Blase KA, Timbers GD, Wolf MM. In search of program implementation: 792 replications of the teaching-family model. *Behav Anal Today.* (2001/2007) 8(1):96–110. doi: 10.1037/h0100104
76. Wolf MM, Kirigin KA, Fixsen DL, Blase KA, Braukmann CJ. The teaching-family model: a case study in data-based program development and refinement (and dragon wrestling). *J Organ Behav Manage.* (1995) 15:11–68. doi: 10.1300/J075v15n01\_04
77. Fixsen DL, Blase KA. The teaching-family model: the first 50 years. *Perspect Behav Sci.* (2018) 42(2):189–211. doi: 10.1007/s40614-018-0168-3
78. Horner RH, Sugai G, Horner HF. A school-wide approach to student discipline. *School Admin.* (2000) 57(2):20–4.
79. Sugai G, Sprague J, Horner R, Walker H. Preventing school violence: the use of office discipline referrals to assess and monitor school-wide discipline interventions. *J Emot Behav Disord.* (2000) 8(2):94–101. doi: 10.1177/10634266000800205
80. Horner RH, Todd AW, Lewis-Palmer T, Irvin LK, Sugai G, Bolland JB. The school-wide evaluation tool (SET): a research instrument for assessing school-wide positive behavior support. *J Posit Behav Interv.* (2004) 6(1):3–12. doi: 10.1177/10983007040060010201
81. Kim J, McIntosh K. Empirically deriving cut scores in the positive behavioral interventions and supports (PBIS) tiered fidelity inventory (TFI) through a bookmarking process. *J Posit Behav Interv.* (2025) 27(2):94–106. doi: 10.1177/10983007241276536
82. 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(8):992–1001. doi: 10.1007/s11121-016-0700-1
83. Miller WR, Yahne CE, Moyers TB, Martinez J, Pirritano M. A randomized trial of methods to help clinicians learn motivational interviewing. *J Consult Clin Psychol.* (2004) 72(6):1050–62. doi: 10.1037/0022-006X.72.6.1050
84. Martino S, Ball SA, Nich C, Frankforter TL, Carroll KM. Community program therapist adherence and competence in motivational enhancement therapy. *Drug Alcohol Depend.* (2008) 96(1):37–48. doi: 10.1016/j.drugalcdep.2008.01.020
85. Martino S, Gallon S, Ball SA, Carroll KM. A step forward in teaching addiction counselors how to supervise motivational interviewing using a clinical trials training approach. *J Teach Addict.* (2008) 6(2):39–67. doi: 10.1080/15332700802127946
86. Miller WR. Guidelines from the international motivational interviewing network of trainers (MINT). *Motivational Interviewing Newsletter: Updates, Education, and Training.* (2000).
87. Forgatch MS, Patterson GR, DeGarmo DS. Evaluating fidelity: predictive validity for a measure of competent adherence to the Oregon model of parent management training. *Behav Ther.* (2005) 36(1):3–13. doi: 10.1016/S0005-7894(05)80049-8
88. Forgatch MS, DeGarmo DS. Sustaining fidelity following the nationwide PMTO implementation in Norway. *Prev Sci.* (2011) 12(3):235–46. doi: 10.1007/s11121-011-0225-6
89. Sigmarsdóttir M, Forgatch M, Vikar Guðmundsdóttir E, Thorlacius Ö, Thorn Svendsen G, Tjaden J, et al. Implementing an evidence-based intervention for children in Europe: evaluating the full-transfer approach. *J Clin Child Adolesc Psychol.* (2018) 48:1–14. doi: 10.1080/15374416.2018.1466305
90. Ogden T, Bjørnebekk G, Kjøbli J, Patras J, Christiansen T, Taraldsen K, et al. Measurement of implementation components ten years after a nationwide introduction of empirically supported programs—a pilot study. *Implement Sci.* (2012) 7:49. doi: 10.1186/1748-5908-7-49
91. Moncher FJ, Prinz RJ. Treatment fidelity in outcome studies. *Clin Psychol Rev.* (1991) 11:247–66. doi: 10.1016/0272-7358(91)90103-2
92. Naleppa MJ, Cagle JG. Treatment fidelity in social work intervention research: a review of published studies. *Res Soc Work Pract.* (2010) 20(6):674–81. doi: 10.1117/1049731509352088
93. Harrison J, Spybrook J, Curtis A, Cousins L. Integrated dual disorder treatment: fidelity and implementation over time. *Soc Work Res.* (2017) 41(2):111–20. doi: 10.1093/swr/svx002
94. Harrison J, Taylor H. Ten years of implementation: assertive community treatment and integrated dual disorder treatment in a statewide system. *Arch Psychol.* (2019) 3(2):1–15. doi: 10.31296/aop.v3i2.92
95. 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(2):270–80. doi: 10.1037/a0013223
96. Bond GR, Becker DR, Drake RE, Vogler KM. A fidelity scale for the individual placement and support model of supported employment. *Rehabil Couns Bull.* (1997) 40:265–84.
97. Bond GR, Evans L, Salyers MP, Williams J, Kim H-W. Measurement of fidelity in psychiatric rehabilitation. *Ment Health Serv Res.* (2000) 2(2):75–87. doi: 10.1023/A:1010153020697
98. McGrew JH, Bond GR, Dietzen L, Salyers MP. Measuring the fidelity of implementation of a mental health program model. *J Consult Clin Psychol.* (1994) 62(4):670–8. doi: 10.1037/0022-006X.62.4.670
99. McHugo GJ, Drake RE, Teague GB, Xie H. Fidelity to assertive community treatment and client outcomes in the New Hampshire dual disorders study. *Psychiatr Serv.* (1999) 50(6):818–24. doi: 10.1176/ps.50.6.818
100. Bond GR, Drake RE. Assessing the fidelity of evidence-based practices: history and current status of a standardized measurement methodology. *Adm Policy Ment Health.* (2020) 47(6):874–84. doi: 10.1007/s10488-019-00991-6