

Outcomes for Implementation Research: Conceptual Distinctions, Measurement Challenges, and Research Agenda

Enola Proctor · Hiie Silmere · Ramesh Raghavan ·
Peter Hovmand · Greg Aarons · Alicia Bunger ·
Richard Griffey · Melissa Hensley

Published online: 19 October 2010

© The Author(s) 2010. This article is published with open access at Springerlink.com

Abstract An unresolved issue in the field of implementation research is how to conceptualize and evaluate successful implementation. This paper advances the concept of “implementation outcomes” distinct from service system and clinical treatment outcomes. This paper proposes a heuristic, working “taxonomy” of eight conceptually distinct implementation outcomes—acceptability, adoption, appropriateness, feasibility, fidelity, implementation cost, penetration, and sustainability—along with their nominal definitions. We propose a two-pronged agenda for research on implementation outcomes. Conceptualizing and measuring implementation outcomes will advance understanding of implementation processes, enhance efficiency in implementation research, and pave the way for studies of the comparative effectiveness of implementation strategies.

Keywords Implementation · Outcomes · Evaluation · Research methods

Background

A critical yet unresolved issue in the field of implementation science is how to conceptualize and evaluate success. Studies of implementation use widely varying approaches to measure how well a new mental health treatment, program, or service is implemented. Some infer implementation success by measuring clinical outcomes at the client or patient level while other studies measure the actual targets of the implementation, quantifying for example the desired provider behaviors associated with delivering the newly implemented treatment. While some studies of implementation strategies assess outcomes in terms of improvement in process of care, Grimshaw et al. (2006) report that meta-analyses of their effectiveness has been thwarted by lack of detailed information about outcomes, use of widely varying constructs, reliance on dichotomous rather than continuous measures, and unit of analysis errors.

This paper advances the concept of “implementation outcomes” distinct from service system outcomes and clinical treatment outcomes (Proctor et al. 2009; Fixsen et al. 2005; Glasgow 2007a). We define implementation outcomes as the effects of deliberate and purposive actions to implement new treatments, practices, and services. Implementation outcomes have three important functions. First, they serve as indicators of the implementation success. Second, they are proximal indicators of implementation processes. And third, they are key intermediate outcomes (Rosen and Proctor 1981) in relation to service system or clinical outcomes in treatment effectiveness and quality of care research. Because an intervention or

E. Proctor (✉) · R. Raghavan · P. Hovmand · A. Bunger ·
M. Hensley
George Warren Brown School of Social Work, Washington
University in St. Louis, Campus Box 1196, One Brookings
Drive, St. Louis, MO 63130, USA
e-mail: ekp@wustl.edu

H. Silmere
Division of Social Work, Roberts Wesleyan College,
Rochester, USA

R. Raghavan
Department of Psychiatry, Washington University School
of Medicine, St. Louis, MO, USA

G. Aarons
Department of Psychiatry, University of California,
San Diego, La Jolla, CA, USA

R. Griffey
Department of Emergency Medicine, Washington University
School of Medicine, St. Louis, MO, USA

treatment will not be effective if it is not implemented well, implementation outcomes serve as necessary preconditions for attaining subsequent desired changes in clinical or service outcomes.

Distinguishing implementation effectiveness from treatment effectiveness is critical for transporting interventions from laboratory settings to community health and mental health venues. When such efforts fail, as they often do, it is important to know if the failure occurred because the intervention was ineffective in the new setting (intervention failure), or if a good intervention was deployed incorrectly (implementation failure). Our current knowledge of implementation is thwarted by lack of theoretical understanding of the processes involved (Michie et al. 2009). Conceptualizing and measuring implementation outcomes will advance understanding of implementation processes, enable studies of the comparative effectiveness of implementation strategies, and enhance efficiency in implementation research.

This paper aims to advance the “vocabulary” of implementation science around implementation outcomes through four specific objectives: (1) to advance conceptualization of implementation outcomes by distinguishing implementation outcomes from service and clinical outcomes; (2) to advance clarity of terminology currently used in implementation science by nominating heuristic definitions of implementation outcomes, yielding a working “taxonomy” of implementation outcomes; (3) to reflect the field’s current language, conceptual definitions, and approaches to operationalizing implementation outcomes; and (4) to propose directions for further research to advance knowledge on these key constructs and their interrelationships.

Our objective of advancing a taxonomy of implementation outcomes is comparable to the work of Michie et al. (2005, 2009), Grimshaw et al. (2006), the Cochrane group, and others who are working to develop taxonomies and common nomenclature for implementation strategies. Our work is complementary to these efforts because implementation outcomes will provide researchers with a framework for evaluating implementation strategies.

Conceptual Framework for Implementation Outcomes

Our understanding of implementation outcomes is lodged within a previously published conceptual framework (Proctor et al. 2009) as shown in Fig. 1. The framework distinguishes between three distinct but interrelated types of outcomes—implementation, service, and client outcomes. Improvements in consumer well-being provide the most important criteria for evaluating both treatment and implementation strategies—for treatment research, improvements are examined at the individual client level whereas improvements at the population-level (within the providing

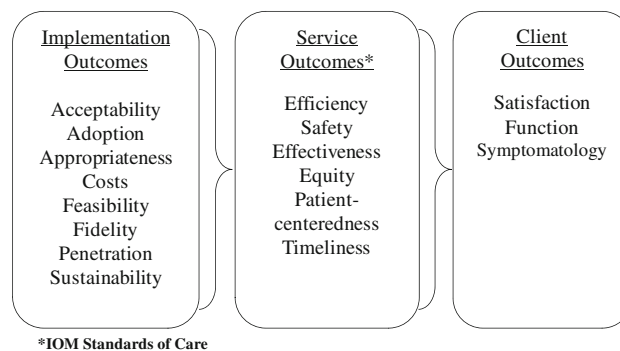


Fig. 1 Types of outcomes in implementation research

system) are examined in implementation research. However, as we argued above, implementation research requires outcomes that are conceptually and empirically distinct from those of service and clinical effectiveness.

For heuristic purposes, our model positions implementation outcomes as preceding both service outcomes and client outcomes, with the latter sets of outcomes being impacted by the implementation outcomes. As we discuss later in this paper, interrelationships among these outcomes require conceptual mapping and empirical tests. For example, one would expect to see a treatment’s strongest impact on client outcomes as an empirically supported treatment’s (EST) penetration increases in a service setting—but this hypothesis requires testing. Our model derives service outcomes from the six quality improvement aims set out in the reports on crossing the quality chasm: the extent to which services are safe, effective, patient-centered, timely, efficient, and equitable (Institute of Medicine Committee on Crossing the Quality Chasm 2006; Institute of Medicine Committee on Quality of Health Care in America 2001).

Methods

The paper’s methods were shaped around its overall aim: to advance clarity in the language used to describe outcomes of implementation. We convened a working group of implementation researchers to identify concepts for labeling and assessing outcomes of implementation processes. One member of the group was a doctoral student RA who coordinated, conducted, and reported on the literature search and constructed tables reflecting various iterations of the heuristic taxonomy. The RA conducted literature searches using key words and search programs to identify literature on the current state of conceptualization and measurement of these outcomes, primarily in the health and behavioral sciences. We searched in a number of databases with a particular focus on MEDLINE, CINAHL Plus, and PsycINFO. Key search terms included the name

of the implementation outcome (e.g., “acceptability,” “sustainability,” etc.) along with relevant synonyms combined with any of the following: innovation, EBP, evidence based practice, and EST. We scanned the titles and abstracts of the identified sources and read the methods and background sections of the studies that measured or attempted to measure implementation outcomes. We also included information from relevant conceptual articles in the development of nominal definitions. Whereas our primary focus was on the implementation of evidence based practices in the health and behavioral sciences, the keyword “innovation” broadened this scope by also identifying studies that focused on other areas such as physical health that may inform implementation of mental health treatments. Because terminology in this field currently reflects widespread inconsistency, we followed leads beyond what our keyword searches “hit” upon. Thus we read additional articles that we found cited by authors whose work we found through our electronic searches. We also conducted searches of CRISP, TAGG, and NIH reporter and studies to identify funded mental health research studies with “implementation” in their titles or abstracts, to identify examples of outcomes pursued in current research.

We used a narrative review approach (Educational Research Review), which is appropriate for summarizing different primary studies and drawing conclusions and interpretation about “what we know,” informed by reviewers’ experiences and existing theories (McPheeters et al. 2006; Kirkevoid 1997). Narrative reviews yield qualitative results, with strengths in capturing diversities and pluralities of understanding (Jones 1997). According to McPheeters et al. (2006), narrative reviews are best conducted by a team. Members of the working group read and reviewed conceptual and theoretical pieces as well as published reports of implementation research. As a team, we convened recurring meetings to discuss the similarities and dissimilarities. We audio-taped and transcribed meeting discussions, and a designated individual took thorough notes. Transcriptions and notes were posted on a shared computer file for member review, revision, and correction.

Group processes included iterative discussion, checking additional literature for clarification, and subsequent discussion. The aim was to collect and portray, from extant literature, the similarities and differences across investigators’ use of various implementation outcomes and definitions for those outcomes. Discussions often led us to preserve distinctions between terms by maintaining in our “nominated” taxonomy two different implementation outcomes because the literature or our own research revealed possible conceptual distinctions. We assembled the identified constructs in the proposed heuristic taxonomy to portray the current state of vocabulary and

conceptualization of terms used to assess implementation outcomes.

Taxonomy of Implementation Outcomes

Through our process of iterative reading and discussion of the literature, we worked to nominate definitions that (1) achieve as much consistency as possible with any existing definitions (including multiple definitions we found for a single construct), yet (2) serve to sharpen distinctions between constructs that might be similar. For several of the outcomes, the literature did not offer one clear nominal definition.

Table 1 depicts the resultant working taxonomy of implementation outcomes. For each implementation outcome, the table nominates a level of analysis, identifies the theoretical basis to the construct from implementation literature, shows different terms that are used for the construct in the literature, suggests the point or stage within implementation processes at which the outcome may be most salient, and lists the types of existing measures for the construct that our search identified. The implementation outcomes listed in Table 1 are probably only the “more obvious,” and we expect that other concepts may emerge from further analysis of the literature and from the kind of empirical work we call for in our discussion below. Many of the implementation outcomes can be inferred or measured in terms of expressed attitudes and opinions, intentions, or reported or observed behaviors. We now list and discuss our nominated conceptual definitions for each implementation outcome in our proposed taxonomy. We reference similar definitions from the literature, and also comment on marked differences between our definitions and others proposed for the term.

Acceptability is the perception among implementation stakeholders that a given treatment, service, practice, or innovation is agreeable, palatable, or satisfactory. Lack of acceptability has long been noted as a challenge in implementation (Davis 1993). The referent of the implementation outcome “acceptability” (or the “*what is acceptable*”) may be a specific intervention, practice, technology, or service within a particular setting of care. Acceptability should be assessed based on the stakeholder’s knowledge of or direct experience with various dimensions of the treatment to be implemented, such as its content, complexity, or comfort. Acceptability is different from the larger construct of service satisfaction, as typically measured through consumer surveys. Acceptability is more specific, referencing a particular treatment or set of treatments, while satisfaction typically references the general service experience, including such features as waiting times, scheduling, and office environment. Acceptability may be measured from the perspective of

Table 1 Taxonomy of implementation outcomes

Implementation outcome	Level of analysis	Theoretical basis	Other terms in literature	Saliency by implementation stage	Available measurement
Acceptability	Individual provider Individual consumer	Rogers: “complexity” and to a certain extent “relative advantage”	Satisfaction with various aspects of the innovation (e.g. content, complexity, comfort, delivery, and credibility)	Early for adoption Ongoing for penetration Late for sustainability	Survey Qualitative or semi-structured interviews Administrative data Refused/blank Administrative data Observation Qualitative or semi-structured interviews Survey
Adoption	Individual provider Organization or setting	RE-AIM: “adoption” Rogers: “trialability” (particularly for early adopters)	Uptake; utilization; initial implementation; intention to try	Early to mid	Observation Qualitative or semi-structured interviews Survey
Appropriateness	Individual provider Individual consumer Organization or setting	Rogers: “compatibility”	Perceived fit; relevance; compatibility; suitability; usefulness; practicability	Early (prior to adoption)	Survey Qualitative or semi-structured interviews Focus groups Survey
Feasibility	Individual providers Organization or setting	Rogers: “compatibility” and “trialability”	Actual fit or utility; suitability for everyday use; practicability	Early (during adoption)	Survey Administrative data
Fidelity	Individual provider	RE-AIM: part of “implementation”	Delivered as intended; adherence; integrity; quality of program delivery	Early to mid	Observation Checklists Self-report
Implementation Cost	Provider or providing institution	TCU Program Change Model: “costs” and “resources”	Marginal cost; cost-effectiveness; cost-benefit	Early for adoption and feasibility Mid for penetration Late for sustainability	Administrative data
Penetration	Organization or setting	RE-AIM: necessary for “reach”	<i>Level of institutionalization? Spread? Service access?</i>	Mid to late	Case audit Checklists
Sustainability	Administrators Organization or setting	RE-AIM: “maintenance” Rogers: “confirmation”	Maintenance; continuation; durability; incorporation; integration; institutionalization; sustained use; routinization;	Late	Case audit Semi-structured interviews Questionnaires Checklists

various stakeholders, such as administrators, payers, providers, and consumers. We presume rated acceptability to be dynamic, changing with experience. Thus ratings of acceptability may be different when taken, for example, pre-implementation and later throughout various stages of implementation. The literature reflects several examples of measuring provider and patient acceptability. Aarons' Evidence-Based Practice Attitude Scale (EBPAS) captures the acceptability of evidence-based mental health treatments among mental health providers (Aarons 2004). Aarons and Palinkas (2007) used semi-structured interviews to assess case managers' acceptance of evidence-based practices in a child welfare setting. Karlsson and Bendtsen (2005) measured patients' acceptance of alcohol screening in an emergency department setting using a 12-item questionnaire.

Adoption is defined as the intention, initial decision, or action to try or employ an innovation or evidence-based practice. Adoption also may be referred to as "uptake." Our definition is consistent with those proposed by Rabin et al. (2008) and Rye and Kimberly (2007). Adoption could be measured from the perspective of provider or organization. Haug et al. (2008) used pre-post items to capture substance abuse providers' adoption of evidence-based practices, while Henggeler et al. (2008) report interview techniques to measure therapists' adoption of contingency management.

Appropriateness is the perceived fit, relevance, or compatibility of the innovation or evidence based practice for a given practice setting, provider, or consumer; and/or perceived fit of the innovation to address a particular issue or problem. "Appropriateness" is conceptually similar to "acceptability," and the literature reflects overlapping and sometimes inconsistent terms when discussing these constructs. We preserve a distinction because a given treatment may be perceived as appropriate but not acceptable, and vice versa. For example, a treatment might be considered a good fit for treating a given condition but its features (for example, rigid protocol) may render it unacceptable to the provider. The construct "appropriateness" is deemed important for its potential to capture some "pushback" to implementation efforts, as is seen when providers feel a new program is a "stretch" from the mission of the health care setting, or is not consistent with providers' skill set, role, or job expectations. For example, providers may vary in their perceptions of the appropriateness of programs that co-locate mental health services within primary medical, social service, or school settings. Again, a variety of stakeholders will likely have perceptions about a new treatment's or program's appropriateness to a particular service setting, mission, providers, and clientele. These perceptions may be function of the organization's culture or climate (Klein and Sorra 1996).

Bartholomew et al. (2007) describe a rating scale for capturing appropriateness of training among substance abuse counselors who attended training in dual diagnosis and therapeutic alliance.

Cost (incremental or implementation cost) is defined as the cost impact of an implementation effort. Implementation costs vary according to three components. First, because treatments vary widely in their complexity, the costs of delivering them will also vary. Second, the costs of implementation will vary depending upon the complexity of the particular implementation strategy used. Finally, because treatments are delivered in settings of varying complexity and overheads (ranging from a solo practitioner's office to a tertiary care facility), the overall costs of delivery will vary by the setting. The true cost of implementing a treatment, therefore, depends upon the costs of the particular intervention, the implementation strategy used, and the location of service delivery.

Much of the work to date has focused on quantifying intervention costs, e.g., identifying the components of a community-based heart health program and attaching costs to these components (Ronckers et al. 2006). These cost estimations are combined with patient outcomes and used in cost-effectiveness studies (McHugh et al. 2007). A review of literature on guideline implementation in professions allied to medicine notes that few studies report anything about the costs of guideline implementation (Callum et al. 2010). Implementing processes that do not require ongoing supervision or consultation, such as computerized medical record systems, may carry lower costs than implementing new psychosocial treatments. Direct measures of implementation cost are essential for studies comparing the costs of implementing alternative treatments and of various implementation strategies.

Feasibility is defined as the extent to which a new treatment, or an innovation, can be successfully used or carried out within a given agency or setting (Karsh 2004). Typically, the concept of feasibility is invoked retrospectively as a potential explanation of an initiative's success or failure, as reflected in poor recruitment, retention, or participation rates. While feasibility is related to appropriateness, the two constructs are conceptually distinct. For example, a program may be appropriate for a service setting—in that it is compatible with the setting's mission or service mandate, but may not be feasible due to resource or training requirements. Hides et al. (2007) tapped aspects of feasibility of using a screening tool for co-occurring mental health and substance use disorders.

Fidelity is defined as the degree to which an intervention was implemented as it was prescribed in the original protocol or as it was intended by the program developers (Dusenbury et al. 2003; Rabin et al. 2008). Fidelity has been measured more often than the other implementation

outcomes, typically by comparing the original evidence-based intervention and the disseminated/implemented intervention in terms of (1) adherence to the program protocol, (2) dose or amount of program delivered, and (3) quality of program delivery. Fidelity has been the overriding concern of treatment researchers who strive to move their treatments from the clinical lab (efficacy studies) to real-world delivery systems. The literature identifies five implementation fidelity dimensions including adherence, quality of delivery, program component differentiation, exposure to the intervention, and participant responsiveness or involvement (Mihalic 2004; Dane and Schneider 1998). Adherence, or the extent to which the therapy occurred as intended, is frequently examined in psychotherapy process and outcomes research and is distinguished from other potentially pertinent implementation factors such as provider skill or competence (Hogue et al. 1996). Fidelity is measured through self-report, ratings, and direct observation and coding of audio- and videotapes of actual encounters, or provider-client/patient interaction. Achieving and measuring fidelity in usual care is beset by a number of challenges (Proctor et al. 2009; Mihalic 2004; Schoenwald et al. 2005). The foremost challenge may be measuring implementation fidelity quickly and efficiently (Hayes 1998).

Schoenwald and colleagues (2005) have developed three 26–45-item measures of adherence at the therapist, supervisor and consultant level of implementation (available from the MST Institute www.mstinstitute.org). Ratings are obtained at regular intervals, enabling examination of the provider, clinical supervisor, and consultant. Other examples from the mental health literature include Bond et al. (2008) 15-item Supported Employment Fidelity Scale (SE Fidelity Scale) and Hogue et al. (2008) Therapist Behavior Rating Scale–Competence (TBRS-C), an observational measure of fidelity in evidence based practices for adolescent substance abuse treatment.

Penetration is defined as the integration of a practice within a service setting and its subsystems. This definition is similar to (Stiles et al. 2002) notion of service penetration and to Rabin et al.'s (2008) notion of niche saturation. Studying services for persons with severe mental illness, Stiles et al. (2002) apply the concept of service penetration to service recipients (the number of eligible persons who use a service, divided by the total number of persons eligible for the service). Penetration also can be calculated in terms of the number of providers who deliver a given service or treatment, divided by the total number of providers trained in or expected to deliver the service. From a service system perspective, the construct is also similar to “reach” in the RE-AIM framework (Glasgow 2007b). We found infrequent use of the term penetration in the implementation literature; though studies seemed to tap into this

construct with terms such as a given treatment's level of institutionalization.

Sustainability is defined as the extent to which a newly implemented treatment is maintained or institutionalized within a service setting's ongoing, stable operations. The literature reflects quite varied uses of the term “sustainability,” but our proposed definition incorporates aspects of those offered by Johnson et al. (2004), Turner and Sanders (2006), Glasgow et al. (1999), Goodman et al. (1993), and Rabin et al. (2008). Rabin et al. (2008) emphasizes the integration of a given program within an organization's culture through policies and practices, and distinguishes three stages that determine institutionalization: (1) passage (a single event such as transition from temporary to permanent funding), (2) cycle or routine (i.e., repetitive reinforcement of the importance of the evidence-based intervention through including it into organizational or community procedures and behaviors, such as the annual budget and evaluation criteria), and (3) niche saturation (the extent to which an evidence-based intervention is integrated into all subsystems of an organization). Thus the outcomes of “penetration” and “sustainability” may be related conceptually and empirically, in that higher penetration may contribute to long-term sustainability. Such relationships require empirical test, as we elaborate below. Indeed Steckler et al. (1992) emphasize sustainability in terms of attaining long-term viability, as the final stage of the diffusion process during which innovations settle into organizations. To date, the term sustainability appears more frequently in conceptual papers than actual empirical articles measuring sustainability of innovations. As we discuss below, the literature often uses the same term (niche saturation, for example) to reference multiple implementation outcomes, underscoring the need for conceptual clarity as we seek to advance in this paper.

Research Agenda to Advance Implementation Outcomes

Advancing the conceptualization, measurement, and empirical understanding of implementation outcomes requires research on several critical issues. We propose two major themes for this research—(1) conceptualization and measurement, and (2) theory building—and identify important issues within each of these themes.

Research on Conceptualization and Measurement of Implementation Outcomes

Research on several fronts is required to advance the conceptual and measurement properties of implementation outcomes, five of which we identify and discuss.

Consistency of Terminology

For each outcome listed in Table 1, we found literature using different and sometimes inconsistent terminology. Sometimes studies used different labels for what appear to be the same construct. In other cases, studies used one term for a label or nominal definition but a different term for operationalizing or measuring the same construct. This problem was pronounced for three implementation outcomes—acceptability, appropriateness, and feasibility. These constructs were frequently used interchangeably or measured under the common generic label as client or provider perceptions, reactions, and attitudes toward, or satisfaction with various aspects of the innovation, EST, or clinical practice guidelines. For example, Graham et al. (2007) assessed doctors' attitudes and perceptions toward clinical practice guidelines with a survey that tapped all three of these outcomes, although none of them were explicitly labeled as such: acceptability (e.g. perceived quality of and confidence in guidelines), appropriateness (e.g. perceived usefulness of guidelines), and feasibility (e.g. these guidelines provide recommendations that are implementable). Other studies interchanged the terms for acceptability and feasibility within the same article. For example, Wilkie et al. (2003) begin by describing the measurement of "usability" (of a computerized innovation), including its "acceptability" to clients but later use the findings to conclude that the innovation was feasible.

While language inconsistency is typical in most still-developing fields, implementation research may be particularly susceptible to this problem. No one discipline is "home" to implementation research. Studies are conducted across a broad range of disciplines, published in a scattered set of journals, and consequently are rarely cross referenced. Beyond mental health, we found articles referencing these implementation outcomes in physical health, smoking cessation, cancer, and substance abuse literatures, addressing a wide variety of topics.

Clearly, the field of implementation science now has only the beginnings of a common language to characterize implementation outcomes, a situation that thwarts the conceptual and empirical advancement of the field but could be overcome by use of a common lexicon. Just as Michie et al. (2009) state the "imperative that there be a consensual, common language" (p. 4) to describe behavior change techniques, so is common language needed for implementation outcomes.

Referent for Rating the Outcome

Several of the proposed implementation outcomes could be used to rate (1) a specific treatment; (2) the implementation strategy used to introduce that treatment into the care

setting; or (3) a broad effort to implement several new treatments at once. A lingering issue for the field is whether implementation processes should be tackled and studied specifically (one new treatment) or in a more generalized way (the extent to which a system's care is evidence-based or guideline congruent). Understanding the optimal specificity of the referent for a given implementation outcome is critical for measurement. As a beginning step, researchers should report the referent for all implementation outcomes measured.

Level of Analysis for Outcomes

Implementation of new treatments is an inherently multi-level enterprise, involving provider behavior, care organization, and policy (Proctor et al. 2009; Raghavan et al. 2008). Implementation outcomes are important at each level of change, but the research has yet to determine which level or unit of analysis is most appropriate for particular implementation outcomes. Certain outcomes, such as acceptability, may be most appropriate for individual level analysis (for example, providers, consumers), while others, such as penetration may be more appropriate for aggregate analysis, at the level of the health care organization. Currently, very few studies reporting implementation outcomes specify the level of measurement, nor do they address issues of aggregation within or across levels.

Construct validity. The constructs reflected in Table 1 and the terms employed in our taxonomy of implementation outcomes derive largely from the research literature. Yet it is important to also understand outcome perceptions and preferences through the voice of those who design and deliver health care. Qualitative data, reflecting language used by various stakeholders as they think and talk about implementation processes, is important for validating implementation outcome constructs. Through in-depth interviews, stakeholders' cognitive representations and mental models of outcomes can be analyzed through such methods as cultural domain analysis (CDA). A "cultural domain" refers to a set of words, phrases, and/or concepts that link together to form a single conceptual subject (Luke 2004; Bates and Sarkar 2007), and methods for CDA, such as free-listing and pile-sorting, have been used since the 1970s (Bates and Sarkar 2007). While primarily used in anthropology, CDA is aptly suited for health services research that endeavors to understand how stakeholders conceptualize implementation outcomes, informing the generation of definitions of implementation outcomes. The actual words used by stakeholders may or may not reflect the terms used in academic literature and reflected in our proposed taxonomy (acceptability, appropriateness, feasibility, adoption, fidelity, penetration, sustainability and costs). But

such research can identify the terms and distinctions that are meaningful to implementation stakeholders.

Measurement Properties of Implementation Outcomes

The literature reflects a wide array of approaches for measuring implementation outcomes, ranging from qualitative, quantitative survey, and record archival. Michie et al. (2007) studied perceived difficulties implementing a mental health guideline, coding respondent descriptions of implementation difficulties as 0, 0.5, or 1. Much measurement has been “home-grown,” with virtually no work on the psychometric properties or measurement rigor. Measurement development is needed to enhance the portability and usefulness of implementation outcomes in real world settings of care. Measures used in efficacy research will likely prove too cumbersome for real-world studies of implementation. For example, detailed assessment of fidelity through coding of encounter videotapes would be too time-intensive for a multi-agency study assessing fidelity of treatment implementation.

Theory-Building Research

Research is also needed to advance our theoretical understanding of the implementation process. Empirical studies of the five issues we list here will inform theory, illuminate the “black box” of implementation processes, and help shape models for developing and testing implementation strategies.

Salience of Implementation Outcomes to Stakeholders

Any effort to implement change in care involves a range of stakeholders, including the treatment developers who design and test the effectiveness of ESTs, policy makers who design and pay for service, administrators who shape program direction, providers and supervisors, patients/clients/consumers and their family members, and interested community members and advocates. The success of efforts to implement evidence-based treatment may rest on their congruence with the preferences and priorities of those who shape, deliver, and participate in care. Implementation outcomes may be differentially salient to various stakeholders, just as the salience of clinical outcomes varies across stakeholders (Shumway et al. 2003). For example, implementation cost may be most important to policy makers and program directors, feasibility may be most important to direct service providers, and fidelity may be most important to treatment developers. To ensure applicability of implementation outcomes across a range of settings and to maximize their external validity, all stakeholder groups and priorities should be represented in this research.

Salience of Implementation Outcomes by Point in the Implementation Process

The implementation of any new treatment or service is widely recognized as a process, involving a sequence of activities, beginning with initial considerations of what and how to change current care. Chamberlain has identified ten steps for the implementation of an evidence-based treatment, Multidimensional Treatment Foster Care (MTFC), beginning with consideration of adopting MTFC and concluding when a service site meets certification criteria for delivering the treatment (Chamberlain et al. 2008). As we suggest in Table 1, certain implementation outcomes may be more important at some phases of implementation process than at other phases. For example, feasibility may be most important once organizations and providers try new treatments. Later, it may be a “moot point,” once the treatment—initially considered novel or unknown—has become part of normal routine.

The literature suggests that studies usually capture fidelity during initial implementation, while adoption is often assessed at 6 (Waldorff et al. 2008), 12 (Adily et al. 2004; Fischer et al. 2008), or 18 months (Cooke et al. 2001) after initial implementation. But most studies fail to specify a timeframe or are inconsistent in choice of a time point in the implementation process for measuring outcomes. Research is needed to explore these issues, particularly longitudinal studies that measure multiple implementation outcomes before, during, and after implementation of a new treatment. Such research may reveal “leading” and “lagging” indicators of implementation success. For example, if acceptability increases for several months, following which penetration increases, then we may view acceptability as a leading indicator of penetration. Leading indicators can be useful for managing the implementation process as they signal future trends.

Where leading indicators may identify future trends, lagging indicators reflect delays between when changes happen and when they can be observed. For example, sustainability may be observed only well into, or even after the implementation process. Being aware of lagging indicators of implementation success may help managers avoid over-reacting to slow change and wait for evidence of what may soon prove to be successful implementation.

Modeling Interrelationships Among Implementation Outcomes

Our team’s observations of implementation suggest that implementation outcomes are themselves interrelated in dynamic and complex ways (Woolf 2008; Repenning 2002; Hovmand and Gillespie 2010; Klein and Knight 2005) and are likely to change throughout an agency’s process to

adopt and implement ESTs. For example, the perceived appropriateness, feasibility, and implementation cost associated with an intervention will likely bear on ratings of the intervention's acceptability. Acceptability, in turn, will likely affect adoption, penetration, and sustainability. Similarly, consistent with Rogers' theory of the diffusion of innovation, the ability to adopt or adapt an innovation for local use may increase its acceptability (Rogers 1995). This suggests that when providers believe they do not have to implement a treatment "by the book" (or with precise fidelity), they may rate the treatment as more acceptable.

Modeling the interrelationships between implementation outcomes will also inform their definitional boundaries and thus shape the taxonomy. For example, if two outcomes which we now define as distinct concepts are shown through research to always occur together, the empirical evidence would suggest that the concepts are really the same thing and should be combined. Similarly, if two of the outcomes are shown to have different empirical patterns, evidence would confirm their conceptual distinction.

Modeling Attainment of Implementation Outcomes

Once researchers have advanced consistent, valid, and efficient measures for implementation outcomes, the field will be equipped to conduct important research treating these constructs as dependent variables, in order to identify correlates or predictors of their attainment. Their measurement will enable research to determine which features of a treatment itself or which implementation strategies help make new treatments acceptable, feasible to implement, or sustainable over time. The diffusion of innovation literature posits that the implementation outcome, adoption of an EST, is a function of such factors as perceived need to do things differently (Rogers 1995) perception of the new treatment's comparative advantage (Frambach and Schillewaert 2002; Henggeler et al. 2002) and as easy to understand (Berwick 2003). Such suppositions require empirical test using measures of implementation outcomes.

Using Implementation Outcomes to Model Implementation Success

Reliable, valid measures of implementation outcomes will enable empirical testing of the success of efforts to implement new treatments, and pave the way for comparative effectiveness research on implementation strategies. In most current initiatives to move evidence-based treatments into community care settings, the success of the implementation is assumed and evaluated from data on clinical outcomes. We believe that an exclusive focus on clinical outcomes thwarts understanding the process of

implementation, as well as the effects of contextual factors that must be addressed and that are captured in implementation outcomes.

Established evidence for a "proven" treatment does not ensure successful implementation. Implementation also requires addressing a number of important contextual factors, such as provider attitudes, professional behavior, and the service system. Constructs in the proposed taxonomy of implementation outcomes have potential to capture those provider attitudes (acceptability) and behaviors (adoption, uptake) as well as contextual factors (system penetration, appropriateness, implementation cost).

For purposes of stimulating debate and future research, we suggest that successful implementation be considered in light of a "portfolio" of factors, including the effectiveness of the treatment to be implemented and implementation outcomes such as included in our taxonomy. For example, implementation success (I , in the equation below) could be modeled to reflect (1) the effectiveness (E) of the treatment being implemented, plus (2) implementation factors (IO's), which heretofore have been insufficiently conceptualized, distinguished, and measured and rarely used to guide implementation decisions.

$$I = fE + \text{IO's}$$

For example, in situation "A", an evidence-based treatment may be highly effective but given its high cost, only mildly acceptable to key stakeholders and low in sustainability. The overall potential success of implementation in this case might be modeled as follows:

$$\begin{aligned} \text{Implementation success} = & f \text{ of effectiveness (= high)} \\ & + \text{acceptability (= moderate)} \\ & + \text{sustainability (low)}. \end{aligned}$$

In situation "B", a given treatment might be only moderately effective but highly acceptable to stakeholders because current care is poor, the treatment is inexpensive, and current training protocols ensure high penetration through providers. This treatment's potential might be modeled in the following equation:

$$\begin{aligned} \text{Implementation success} = & \\ & f \text{ of treatment effectiveness (moderate)} \\ & + \text{acceptability (high) + potential to improve care (high)} \\ & + \text{penetration (high)}. \end{aligned}$$

Thus using implementation outcomes, the success of implementation may be modeled and tested, thereby making decisions about what to implement more explicit and transparent.

To increase the success of implementation, implementation strategies need to be employed strategically. For

example, implementation strategies could be employed to increase provider acceptance, improve penetration, reduce implementation costs, and achieve sustainability of the treatment being implemented. Understanding how to achieve implementation outcomes requires the kind of work now underway by Michie et al. (2009) advance a taxonomy of implementation strategies and reflect their demonstrated effects.

Summary and Implications

The science of implementation cannot be advanced without attention to implementation outcomes. All studies of implementation should explicate and measure implementation outcomes. Given the rudimentary state of the field, we chose a narrative approach to reviewing the literature and constructing a taxonomy. Our purpose is to advance the clarity of language, provoke debate, and stimulate more systematic work toward the aims of advancing the conceptual, linguistic, and methodological clarity in the field. A taxonomy of implementation outcomes can help organize the key variables and frame research questions required to advance implementation science. Their measurement and empirical test can help specify the mechanisms and causal relationships within implementation processes and advance an evidence base around successful implementation.

Acknowledgments Preparation of this manuscript was supported by grants P30 MH068579 and UL1RR024992.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Aarons, G. A. (2004). Mental health provider attitudes toward adoption of evidence-based practice: The evidence-based practice attitude scale (EBPAS). *Mental Health Service Research, 6*(2), 61–74.
- Aarons, G. A., & Palinkas, L. A. (2007). Implementation of evidence-based practice in child welfare: Service provider perspectives. *Mental Health Services Research, 34*, 411–419.
- Adily, A., Westbrook, J., Coiera, E., & Ward, J. (2004). Use of online evidence databases by Australian public health practitioners. *Medical Informatics & the Internet in Medicine, 29*(2), 127–136.
- Bartholomew, N. G., Joe, G. W., Rowan-Szal, G. A., & Simpson, D. D. (2007). Counselor assessments of training and adoption barriers. *Journal of Substance Abuse Treatment, 33*, 193–199.
- Bates, D. M., & Sarkar, D. (2007). lme4: Linear mixed-effects models using S4 classes. *R Package Version, 0*, 99875–99876.
- Berwick, D. M. (2003). Disseminating innovations in health care. *Journal of the American Medical Association, 289*(15), 1969–1975.
- Bond, G. R., McHugo, G. J., Becker, D. R., Rapp, C. A., & Whitley, R. (2008). Fidelity of supported employment: Lessons learned from the national evidence-based practice project. *Psychiatric Rehabilitation Journal, 31*(4), 300–305.
- Callum, P., Iger, J., Ray, M., Sims, C. A., & Falk, R. E. (2010). Outcome and experience of implementing spinal muscular atrophy carrier screening on sperm donors. *Fertility and Sterility, 94*(5), 1912–1914.
- Chamberlain, P., Brown, C. H., Saldana, L., Reid, J., Wang, W., Marsenich, L., et al. (2008). Engaging and recruiting counties in an experiment on implementing evidence-based practice in California. *Administration and Policy In Mental Health, 35*, 250–260.
- Cooke, M., Mattick, R. P., & Walsh, R. A. (2001). Implementation of the ‘fresh start’ smoking cessation programme to 23 antenatal clinics: A randomized controlled trial investigating two methods of dissemination. *Drug and Alcohol Review, 20*, 19–28.
- Dane, A. V., & Schneider, B. H. (1998). Program integrity in primary and early secondary prevention: Are implementation effects out of control? *Clinical Psychology Review, 18*, 23–45.
- Davis, F. (1993). User acceptance of information technology: System characteristics, user perceptions and behavioral impacts. *International Journal of Man-Machine Studies, 38*, 475–487.
- Dusenbury, L., Brannigan, R., Falco, M., & Hansen, W. B. (2003). A review of research on fidelity of implementation: Implications for drug abuse prevention in school settings. *Health Education Research, 18*(2), 237–256.
- Fischer, M. A., Vogeli, C., Stedman, M. R., Ferris, T. G., & Weissman, J. S. (2008). Uptake of electronic prescribing in community-based practices. *Journal of General Internal Medicine, 23*(4), 358–363.
- Fixsen, D. L., Naoom, S. F., Blasé, K. A., Friedman, R. M., & Wallace, F. (2005). *Implementation research: A synthesis of the literature* (No. FMHI Publication #231). Tampa, FL: University of South Florida, Louis de la Parte Florida Mental Health Institute, National Implementation Research Network.
- Frambach, R. T., & Schillewaert, N. (2002). Organizational innovation adoption: A multi-level framework of determinants and opportunities for future research. *Journal of Business Research Special Issue: Marketing Theory in the Next Millennium, 55*, 163–176.
- Glasgow, R. E. (2007a). eHealth evaluation and dissemination research. *American Journal of Preventive Medicine, 32*(5S), S119–S126.
- Glasgow, R. E. (2007b). *The RE-AIM model for planning, evaluation and reporting on implementation and dissemination research*. Paper presented at the NIH Conference on Building the Science of Dissemination and Implementation in the Service of Public Health: Bethesda, MD.
- Glasgow, R. E., Vogt, T. M., & Boles, S. M. (1999). Evaluating the public health impact of health promotion interventions: The RE-AIM framework. *American Journal of Public Health, 89*, 1322–1327.
- Goodman, R. M., McLeroy, K. R., Steckler, A. B., & Hoyle, R. H. (1993). Development of level of institutionalization scales for health promotion programs. *Health Education Quarterly, 20*(2), 161–179.
- Graham, I. D., Brouwers, M., Davies, C., & Tetroe, J. (2007). Ontario doctors’ attitudes toward and use of clinical practice guidelines in oncology. *Journal of Evaluation in Clinical Practice, 13*(4), 607–615.
- Grimshaw, J., Eccles, M., Thomas, R., MacLennan, G., Ramsay, C., Fraser, C., et al. (2006). Toward evidence-based quality improvement: Evidence (and its limitations) of the effectiveness of guideline dissemination and implementation strategies 1966–1998. *Journal of General Internal Medicine, 21*, S14–S20.

- Haug, N. A., Shopshire, M., Tajima, B., Gruber, V., & Guydish, J. (2008). Adoption of evidence-based practices among substance abuse treatment providers. *Journal of Drug Education, 38*(2), 181–192.
- Hayes, S. C. (1998). President's message: Dissemination research now. *The Behavior Therapist, 21*(9), 166–169.
- Henggeler, S. W., Lee, T., & Burns, J. A. (2002). What happens after the innovation is identified? *Clinical Psychology: Science & Practice, 9*(2), 191–194.
- Henggeler, S. W., Chapman, J. E., Rowland, M. D., Halliday-Boykins, C. A., Randall, J., Shackelford, J., et al. (2008). Statewide adoption and initial implementation of contingency management for substance-abusing adolescents. *Journal of Consulting and Clinical Psychology, 76*(4), 556–567.
- Hides, L., Lubman, D. I., Elkins, K., Catania, L. S., & Rogers, N. (2007). Feasibility and acceptability of a mental health screening tool and training programme in the youth alcohol and other drug (AOD) sector. *Drug and Alcohol Review, 26*, 509–515.
- Hogue, A., Liddle, H. A., & Rowe, C. (1996). Treatment adherence process research in family therapy: A rationale and some practical guidelines. *Psychotherapy, 33*(2), 332–345.
- Hogue, A., Dauber, S., Chincilla, P., Fried, A., Henderson, C., Inclan, J., et al. (2008). Assessing fidelity in individual and family therapy for adolescent substance abuse. *Journal of Substance Abuse Treatment, 35*, 137–147.
- Hovmand, P. S., & Gillespie, D. F. (2010). Implementation of evidence based practice and organizational performance. *Journal of Behavioral Health Services & Research, 37*(1), 79–94.
- Institute of Medicine Committee on Crossing the Quality Chasm. (2006). *Adaption to mental health and addictive disorder: Improving the quality of health care for mental and substance-use conditions*. Washington, DC: Institute of Medicine, National Academies Press.
- Institute of Medicine Committee on Quality of Health Care in America. (2001). *Crossing the quality chasm: A new health system for the 21st century*. Washington, DC: Institute of Medicine, National Academy Press.
- Johnson, K., Hays, C., Center, H., & Daley, C. (2004). Building capacity and sustainable prevention innovations: A sustainability planning model. *Evaluation Program Planning, 27*, 135–149.
- Jones, K. (1997). Mission drift in qualitative research, or moving toward a systematic review of qualitative studies. *The Qualitative Report, 9*(1), 95–112.
- Karlsson, A., & Bendtsen, P. (2005). Acceptability of a computerized alcohol screening and advice routine in an emergency department setting: a patient perspective. *Addictive Behaviors, 30*, 767–776.
- Karsh, B. T. (2004). Beyond usability: Designing effective technology implementation systems to promote patient safety. *Quality and Safety in Health Care, 13*, 388–394.
- Kirkevoid, M. (1997). Integrative nursing research—an important strategy to further the development of nursing science and practice. *Journal of Advanced Nursing, 25*, 977–984.
- Klein, K. J., & Knight, A. P. (2005). Innovation implementation: Overcoming the challenge. *Current Directions in Psychological Science, 14*(5), 243–246.
- Klein, K. J., & Sorra, J. S. (1996). The challenge of innovation implementation. (Special topic forum on the management of innovation). *Academy of Management Review, 21*(4), 1055–1081.
- Luke, D. A. (2004). *Multilevel modeling*. Sage University Papers Series on Quantitative Applications in the Social Sciences. Thousand Oaks, CA: Sage.
- McHugh, R. K., Otto, M. W., Barlow, D. H., Gorman, J. M., Shear, M. K., & Woods, S. W. (2007). Cost-efficacy of individual and combined treatments for panic disorder. *The Journal of Clinical Psychiatry, 68*(7), 1038–1044.
- McPheeters, M. L., Briss, P., Teutsch, S. J., & Truman, B. (2006). Systematic reviews in public health. In R. C. Brownson & D. B. Petitti (Eds.), *Applied epidemiology: Theory to practice* (2nd ed., pp. 99–124). New York: Oxford University Press.
- Michie, S., Johnson, M., Abraham, C., Lawton, R., Parker, D., & Walker, A. (2005). Making psychological theory useful for implementing evidence based practice: A consensus approach. *Quality and Safety in Health Care, 14*(1), 26–33.
- Michie, S., Pilling, S., Garety, P., Whitty, P., Eccles, M. P., Johnston, M., et al. (2007). Difficulties implementing a mental health guideline: An exploratory investigation using psychosocial theory. *Implementation Science, 2*(8), 1–8.
- Michie, S., Fixsen, D., Grimshaw, J., & Eccles, M. (2009). Specifying and reporting complex behavior change interventions: The need for a scientific method. *Implementation Science, 4*(1), 40.
- Mihalic, S. (2004). The importance of implementation fidelity. *Emotional and Behavioral Disorders in Youth, 4*, 83–86.
- Proctor, E. K., Landsverk, J., Aarons, G., Chambers, D., Glisson, C., & Mittman, B. (2009). Implementation research in mental health services: An emerging science with conceptual, methodological, and training challenges. *Administration and Policy in Mental Health and Mental Health Services Research, 36*(1), 24–34.
- Rabin, B. A., Brownson, R. C., Haire-Joshu, D., Kreuter, M. W., & Weaver, N. L. (2008). A glossary for dissemination and implementation research in health. *Journal of Public Health Management and Practice, 14*(2), 117–123.
- Raghavan, R., Bright, C., & Shadoin, A. (2008). Toward a policy ecology of implementation of evidence-based practices in public mental health settings. *Implementation Science, 3*, 26.
- Repenning, N. P. (2002). A simulation-based approach to understanding the dynamics of innovation implementation. *Organization Science, 13*(2), 109–127.
- Rogers, E. M. (1995). *Diffusion of innovations* (4th ed.). New York: The Free Press.
- Ronckers, E., Groot, W., Steenbakkens, M., Ruland, E., & Ament, A. (2006). Costs of the 'hartslag limburg' community heart health intervention. *BMC Public Health, 6*(1), 51.
- Rosen, A., & Proctor, E. K. (1981). Distinctions between treatment outcomes and their implications for treatment evaluation. *Journal of Consulting and Clinical Psychology, 49*(3), 418–425.
- Rye, C. B., & Kimberly, J. R. (2007). The adoption of innovations by provider organizations in health care. *Medical Care Research and Review, 64*(3), 235–278.
- Schoenwald, S. K., Letourneau, E. J., & Halliday-Boykins, C. A. (2005). Predicting therapist adherence to a transported family-based treatment for youth. *Journal of Clinical Child & Adolescent Psychology, 34*, 658–670.
- Shumway, M., Saunders, T., Shern, D., Pines, E., Downs, A., Burbine, T., et al. (2003). Preferences for schizophrenia treatment outcomes among public policy makers, consumers, families, and providers. *Psychiatric Services, 54*, 1124–1128.
- Steckler, A., Goodman, R., & McElory, K. (1992). Measuring the diffusion of innovative health promotion programs. *American Journal of Health Promotion, 6*, 214–224.
- Stiles, P. G., Boothroyd, R. A., Snyder, K., & Zong, X. (2002). Service penetration by persons with severe mental illness: How should it be measured? *Journal of Behavioral Health Services & Research, 29*(2), 198.
- Turner, K. M. T., & Sanders, M. R. (2006). Dissemination of evidence-based parenting and family support strategies: Learning from the triple p-positive parenting program system approach. *Aggression and Violent Behavior, 11*(2), 176–193.
- Waldorff, F., Steenstrup, A., Nielsen, B., Rubak, J., & Bro, F. (2008). Diffusion of an e-learning programme among Danish general practitioners: A nation-wide prospective survey. *BMC Family Practice, 9*(1), 24.

- Wilkie, D. J., Judge, M. K. M., Berry, D. L., Dell, J., Zong, S., Gillespie, R., et al. (2003). Usability of a computerized PAIN-ReportIt in the general public with pain and people with cancer pain. *Journal of Pain and Symptom Management*, 25(3), 213–224.
- Woolf, S. H. (2008). The meaning of translational research and why it matters. *Journal of the American Medical Association*, 299(2), 211–213.