




Usability Evaluation for Evidence-Based Psychosocial Interventions (USE-EBPI): A methodology for assessing complex intervention implementability

Implementation Research and Practice
Volume 1: Jan-Dec 2020 1–17
© The Author(s) 2020
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/2633489520932924
journals.sagepub.com/home/irp


Aaron R Lyon¹ , Kelly Koerner² and Julie Chung²

Abstract

Background: Most evidence-based practices in mental health are complex psychosocial interventions, but little research has focused on assessing and addressing the characteristics of these interventions, such as design quality and packaging, that serve as intra-intervention determinants (i.e., barriers and facilitators) of implementation outcomes. Usability—the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction—is a key indicator of design quality. Drawing from the field of human-centered design, this article presents a novel methodology for evaluating the usability of complex psychosocial interventions and describes an example “use case” application to an exposure protocol for the treatment of anxiety disorders with one user group.

Method: The Usability Evaluation for Evidence-Based Psychosocial Interventions (USE-EBPI) methodology comprises four steps: (1) identify users for testing; (2) define and prioritize EBPI components (i.e., tasks and packaging); (3) plan and conduct the evaluation; and (4) organize and prioritize usability issues. In the example, clinicians were selected for testing from among the identified user groups of the exposure protocol (e.g., clients, system administrators). Clinicians with differing levels of experience with exposure therapies (novice, $n=3$; intermediate, $n=4$; advanced, $n=3$) were sampled. Usability evaluation included Intervention Usability Scale (IUS) ratings and individual user testing sessions with clinicians, and heuristic evaluations conducted by design experts. After testing, discrete usability issues were organized within the User Action Framework (UAF) and prioritized via independent ratings (1–3 scale) by members of the research team.

Results: Average IUS ratings (80.5; $SD=9.56$ on a 100-point scale) indicated good usability and also room for improvement. Ratings for novice and intermediate participants were comparable (77.5), with higher ratings for advanced users (87.5). Heuristic evaluations suggested similar usability (mean overall rating = 7.33; $SD=0.58$ on a 10-point scale). Testing with individual users revealed 13 distinct usability issues, which reflected all four phases of the UAF and a range of priority levels.

Conclusion: Findings from the current study suggested the USE-EBPI is useful for evaluating the usability of complex psychosocial interventions and informing subsequent intervention redesign (in the context of broader development frameworks) to enhance implementation. Future research goals are discussed, which include applying USE-EBPI with a broader range of interventions and user groups (e.g., clients).

Plain language abstract: Characteristics of evidence-based psychosocial interventions (EBPIs) that impact the extent to which they can be implemented in real world mental health service settings have received far less attention than the characteristics of individuals (e.g., clinicians) or settings (e.g., community mental health centers), where EBPI implementation occurs. No methods exist to evaluate the usability of EBPIs, which can be a critical barrier or facilitator

¹University of Washington, Seattle, WA, USA

²Evidence Based Practice Institute, Seattle, WA, USA

Corresponding author:

Aaron R Lyon, University of Washington, 6200 NE 74th Street, Suite 100, Seattle, WA 98115, USA.
Email: lyona@uw.edu



of implementation success. The current article describes a new method, the Usability Evaluation for Evidence-Based Psychosocial Interventions (USE-EBPI), which uses techniques drawn from the field of human-centered design to evaluate EBPI usability. An example application to an intervention protocol for anxiety problems among adults is included to illustrate the value of the new approach.

Keywords

Human-centered design, user-centered design, usability, complex psychosocial interventions, evidence-based psychosocial interventions

Background

Complex interventions (i.e., those with several interacting components) are common in contemporary health care (Craig et al., 2013). In mental health, the majority of evidence-based practices are complex *psychosocial interventions*, involving interpersonal or informational activities, techniques, or strategies (England et al., 2015). Hundreds of evidence-based psychosocial interventions (EBPIs) have been developed, but are applied inconsistently in routine service delivery settings (Becker et al., 2013; Garland et al., 2008).

A wealth of research has focused on identifying multi-level determinants (i.e., barriers and facilitators) of implementation (Krause et al., 2014), most often specifying factors at the individual and organization/inner setting levels. Much less frequently targeted are intervention-level determinants (i.e., characteristics of EBPIs themselves; Dopp et al., 2019). This is surprising given long-standing recognition that intervention-level determinants are critical to successful implementation (Schloemer & Schroder-Back, 2018). Classic frameworks such as Rogers' (1962) *Diffusion of innovations* explicitly detail the importance of intervention determinants, including factors such as relative advantage and design quality and packaging. However, such frameworks have generally been too broad to articulate the specific intra-intervention characteristics that reflect good design quality. While some more recent work has articulated how characteristics of complex health innovations, such as clinical guidelines (Gagliardi et al., 2011) and genetic testing and consultation (Hamilton et al., 2014) may facilitate implementation, no efforts exist for EBPIs. Just as design problems can block uptake and use of electronic medical records and various decision support tools (Beuscart-Zephir et al., 2010), poor design is a major determinant of the extent to which EBPI users (clinicians, service recipients, others) adopt and sustain interventions (Lyon & Bruns, 2019).

Evaluation of intervention-level determinants

Attention to intervention-level determinants is most prominently reflected in research on intervention modification (Chambers & Norton, 2016). Extant frameworks tend to *describe or document* modifications (Rabin et al., 2018;

Stirman et al., 2019), but no methods exist to assess intra-intervention implementation determinants or to inform prospective adaptation. Lewis et al. (2015) conducted a systematic review of implementation instruments and found only 19 that addressed the intervention level. Most instruments focused on relative advantage ($n=7$), and none addressed design quality and packaging. Methods are needed to allow researchers and practitioners to more closely evaluate aspects of any EBPI—and especially intervention design—that impact implementation. Such methods are likely to be relevant to intervention developers (e.g., to inform iterative design of components of new interventions), implementation researchers (e.g., to test the degree to which intervention design is predictive of implementation), implementation practitioners (e.g., to determine which interventions are most likely to be fit the needs of consultees), and organizations interested in adopting EBPIs (e.g., to make adoption decisions).

Human-centered design and EBPI usability

We draw on methods from the field of human-centered design (HCD; also known as user-centered design). Most EBPIs have been developed independent from the HCD field, which has sought to clearly operationalize the concepts and metrics that reflect good design. As a result, EBPIs often have not been designed for typical end users and contexts of use, exacerbating the need for adaptations. As discussed later, EBPI users (i.e., the individuals who interact with a product) are often diverse, but primary users typically include both service providers and service recipients. HCD is focused on developing compelling and intuitive products, grounded in knowledge about the people and contexts where an innovation will be deployed (Courage & Baxter, 2005). Although the application of HCD methods has typically been limited to digital technologies, their potential for broader applications in health care is increasingly recognized (Roberts et al., 2016). Lyon and Koerner (2016) applied HCD principles to the tasks of EBPI development and redesign, suggesting that EBPI designs should demonstrate high learnability, efficiency, memorability, error reduction, a good reputation, low cognitive load, and should exploit natural constraints (i.e., incorporate or explicitly address the static properties of an intended destination

context that limit the ways a product can be used). Collectively, these design goals reflect key drivers of EBPI *usability*, or the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction (International Organization for Standardization [ISO], 1998).

Evaluation of usability is increasingly routine in digital health (National Cancer Institute [NCI], 2007), but systematic usability assessment procedures have never been applied to EBPIs. This is problematic given that EBPI usability strongly influences implementation outcomes that, in turn, drive clinical outcomes (Lyon & Bruns, 2019). Usability testing of EBPIs is critical because such assessments (1) allow for evaluation of intervention characteristics likely to be predictive of adoption (Rogers, 2010) and (2) uncover critical usability issues that could subsequently be addressed via prospective adaptation (i.e., redesign; Chambers et al., 2013). This information is relevant across multiple stages of intervention development, testing, and implementation by driving initial design, modification, and selection (e.g., of the most usable interventions) for research and practice applications. Presently, no methodologies exist to accomplish these goals.

Current aims

This article presents (1) a novel methodology for identifying, organizing, and prioritizing usability issues for psychosocial interventions and (2) an example application to an exposure procedure for anxiety disorders. Exposure is among the most effective interventions for disorders, such as obsessive compulsive disorder (Tryon, 2005). During exposure, clinicians support clients to approach fear-producing stimuli (exposure) while preventing fear-reducing behaviors, such as compulsions or other avoidance strategies (Himle & Franklin, 2009). Although the example used is specific to mental health and, for simplicity, incorporates only one user group (clinicians), the methodology is intended to be generalizable. Furthermore, while the feasibility of the usability testing techniques varies across settings (see Step 3, below), the methodology is intended for use by a range of professionals, including intervention developers, implementation researchers, implementation practitioners, and organizations interested in adopting EBPIs.

Methods

The *Usability Evaluation for Evidence-Based Psychosocial Interventions (USE-EBPI)* is a methodology for assessing the ease with which interventions are likely to be adopted and which components may impede implementation. It comprises four steps: (1) identify users/participants; (2) define and prioritize EBPI components; (3) plan and

conduct the test; and (4) organize and prioritize identified usability issues (Figure 1). All the steps and techniques described borrow from the extensive literature on HCD and usability testing (e.g., Albert & Tullis, 2013; Maguire, 2001), but have been adapted to ensure relevance to psychosocial interventions. Importantly, USE-EBPI is a prospective *usability evaluation* method and is not intended to retrospectively assess adaptations that have occurred or to be a comprehensive framework for EBPI redesign.

Step 1: identify users/participants

Explicit identification of representative end users is a basic tenet of HCD (Cooper et al., 2007). Product developers tend to underestimate user diversity and base designs on people like themselves (Cooper, 1999; Kujala & Mantyla, 2000), but explicit user identification produces more usable systems (Kujala & Kauppinen, 2004).

The USE-EBPI framework proposes a systematic user identification process (Table 1) drawn from the larger testing literature (Hackos & Redish, 1998; Kujala & Kauppinen, 2004). As indicated by the funnel shape for Step 1 (Figure 1), each stage of identification narrows the potential participant pool. The first sub-step is brainstorming an overly-inclusive, preliminary list of potential users (e.g., clinicians, clients, system administrators, etc.). For the exposure protocol, potential users included all behavioral health clinicians and clients who treat or experience exposure-relevant anxiety, as well as supervisors who support those clinicians. Other potential user groups (e.g., implementation intermediaries, service system administrators) were considered but determined to be too distal to the study aims.

Second, the most relevant subset of user characteristics are articulated, which may include personal (e.g., prior EBPI training or attitudes toward EBPIs [clinician], expectations, or prior treatment experiences [client]), task-related (e.g., experience with the specific EBPI, frequency of usage), and setting characteristics (e.g., intervention setting). User characteristics most relevant to the test of the exposure protocol included experience delivering or supervising exposure interventions (clinicians, supervisors) and anxiety severity (consumers).

Third, primary user groups (i.e., the core group[s] expected to use a product) are described and prioritized, with potential adjunctive input from secondary users (Cooper et al., 2007). Primary EBPI users often include clinicians and clients, while secondary users may include caregivers (for interventions that do not target them directly), system administrators (who often make adoption decisions), implementation intermediaries (who work to enhance EBPI adoption), and paraprofessionals (who may direct clients to interventions; Lyon & Koerner, 2016). Explicitly articulated negative or nonusers may be deprioritized. In our example, clinicians and clients were primary

users. However, only clinicians were selected for testing (Table 5) given the modest goals of the USE-EBPI pilot and because the exposure protocol materials were designed to be primarily clinician facing. Clinicians interested in exposure interventions were prioritized and disinterested clinicians were identified as non-users.

Fourth, typical and representative users are selected for testing. For tests involving more than a small number of users (e.g., $n = 6 +$), it is frequently advantageous to recruit participants into at least two different strata, defined the most critical characteristics. The sample size required for user tests is debated in the HCD literature. Although there is a classic assumption that, after five users, usability tests yield diminished returns (Hwang & Salvendy, 2010; Nielsen, 2000), it is likely sample sizes between 6 and 20 users (Beyer & Holtzblatt, 1998) are appropriate for complex EBPIs. For the exposure protocol evaluation, our team was interested in how existing expertise influenced user experiences. Novice, intermediate, and advanced clinicians were all included in the evaluation regardless of other characteristics. Ten users were determined to provide a sufficient testing sample, given findings that even seven participants can be sufficient when there is substantial complexity present (Turner et al., 2006).

Step 2: define and prioritize the EBPI's components

Because it is rarely feasible to conduct a usability test of the entirety of an EBPI's features, it is essential to constrain the scope of components included. The USE-EBPI framework delineates four types of EBPI components for testing, organized into two different categories, tasks and packaging (Table 2).

EBPI tasks. EBPIs include critical tasks that must be accomplished to have their intended effects. First, *content elements* (a.k.a., practice elements) are discrete clinical tasks or strategies used in the context of an intervention session (Chorpita et al., 2005). For behavioral health interventions, these may include techniques, such as cognitive restructuring or psychoeducation. In the exposure protocol, identified content elements are given in Table 2. The completion of an actual exposure was selected as the most important content to assess given (1) procedures clinicians use to assist clients in approaching and learning in feared contexts are widely considered the most critical core component for obtaining desired clinical outcomes and (2) clinicians tend to omit and drift from those critical elements (Waller & Turner, 2016).

Second, *structures* are dynamic processes that guide clinicians in selecting, organizing, delivering, maintaining, altering, or discontinuing content elements (Lyon et al., 2018). Structures differ from within-session

client–therapist processes (i.e., content elements) and include activities, such as measurement-based care (Scott & Lewis, 2015) and structured supervision (Dorsey et al., 2016). Structures identified in the example protocol are given in Table 2. Subjective units of distress (SUDs; a.k.a., “fear thermometer”) ratings were selected for testing as they are an integral component of most exposure protocols.

EBPI packaging. EBPI packaging refers to the static properties of how tasks are organized, communicated, or otherwise supported. Packaging includes both EBPI artifacts and parameters (Table 2). *Artifacts* reflect tangible, digital, or visual materials that support task completion (e.g., treatment manuals; Keenan et al., 1999). Identified exposure artifacts are provided in Table 2. Although all materials were provided to participants to review (see Step 3), it was determined that the brief exposure guide contained the most critical core content of the exposure procedures and would be feasible to test in its entirety.

EBPI *parameters* refer to any static aspect of an intervention that defines and constrains the intervention or service “space” within which tasks are completed, such as intervention modality (e.g., individual versus group). Although many parameters were embedded into the exposure protocol (e.g., sequencing fear hierarchy construction before exposure), none were explicitly selected for testing because our research team had no explicit research questions about parameters—such as testing in different practice settings or evaluating the role of language proficiency on usability—at this initial stage of evaluation.

Prioritizing EBPI components. Tasks and packaging can be prioritized for usability testing based on whether they represent core intervention components and whether there are known or suspected usability issues that may impact implementation. The actual exposure procedures in the example protocol above met both of these criteria. Although packaging is more likely to be the “adaptable periphery” of the EBPI, rather than a “core component” (Damschroder et al., 2009), key artifacts or parameters of an EBPI's packaging that are critical to effectiveness (e.g., brief exposure guide) also are likely to be core components. Core components may be identified based on (1) theory or logic models that specify causal pathways, (2) empirical unpacking studies that test the necessity of components, or (3) research evaluating the mechanisms through which interventions impact outcomes (Kazdin, 2007). Known or suspected usability problems with an EBPI's component tasks and packaging may also be prioritized (e.g., information from the literature about underuse of exposure procedures among community clinicians). Step 2 of USE-EBPI should

result in a prioritized list of components with which users most need to interact to achieve an EBPI's desired outcomes.

Step 3: plan and conduct the tests

Usability tests should systematically document usability problems, confirming those already suspected (e.g., derived from the literature) and eliciting new issues. USE-EBPI provides a standard set of user research questions to drive selection of testing techniques (Table 3). Categories of testing techniques include (a) quantitative instruments; (b) heuristic evaluation; (c) cognitive walkthroughs; (d) lab-based testing; and (e) in vivo testing. Only a subset will be relevant to any particular EBPI testing process. In USE-EBPI, we suggest triangulation using complementary methods (e.g., quantitative and lab-based).

User research questions that drove the example application of USE-EBPI included: (1) What is overall level of usability for components of the exposure protocol and related materials for more experienced and less experienced users? (2) To what extent does the protocol align with established usability principles? (3) How effectively can users complete an exposure task? and (4) What specific usability issues do users experience when interacting with the protocol. Drawing from Table 3, testing methods selected to address these questions included the use of a quantitative instrument, a heuristic evaluation checklist, and lab-based usability testing (Table 4). All five USE-EBPI testing techniques are presented below to provide a comprehensive description of the USE-EBPI method.

Quantitative instruments. A wide variety of quantitative instruments exist to identify usability problems. Tools, such as the robust 10-item System Usability Scale (SUS [Brooke, 1996; Sauro, 2011]) are completed directly by users. Our research team has created an adapted version of the SUS for EBPIs (i.e., IUS—Figure 2). Nevertheless, USE-EBPI de-emphasizes quantitative measures as a first line approach. They efficiently identify the *presence* of a usability problem, but offer few details about the nature of the problem. We recommend the use of quantitative tools only (a) when combined with other qualitative usability assessment approaches or (b) to efficiently monitor usability improvements over time. In the example, the IUS was administered to participants to assess overall usability of the exposure protocol via a secure, web-based platform following participation in a user testing session (see below).

Heuristic evaluation. Heuristic evaluation involves expert review of a system or interface while applying a set of guidelines that reflect good design principles (Nielsen, 1994). Within USE-EBPI, heuristic evaluation involves

ratings from multiple individuals with expertise in EBPI design who independently review all relevant task and packaging components. Although these heuristics should be selected or adjusted according to the specific needs of the evaluation, the design goals articulated by Lyon and Koerner (2016) reflect USE-EBPI's default set (i.e., learnability, efficiency, memorability, error reduction, low cognitive load, and exploit natural constraints), with the exception of reputation (see Heuristic Evaluation Rubric for EBPIs [HERE], Figure 3). Evaluation is inherently mixed methods, with quantitative ratings as well as qualitative justification of those ratings for data complementarity and expansion (Palinkas et al., 2011). While an evaluator may spend multiple hours reviewing an EBPI manual and all associated materials, heuristic evaluation remains relatively efficient. Nevertheless, drawbacks include a risk of “lumping” different usability problems together, thus creating a list of problems with suboptimal specificity (Keenan et al., 1999; Khajouei et al., 2018). Heuristic analysis is also best applied by experts in design principles, the content area, or both (Nielsen, 1994), expertise that might not be available to all research teams.

HERE was selected to evaluate the exposure protocol in part because multiple members of the study team had adequate expertise in HCD. Three experts conducted independent HERE evaluations of all available artifacts (i.e., a how-to manual for the exposure protocol, brief exposure guide, a core fear map, and fear hierarchy examples). Raters reviewed all materials twice, once to understand the overall scope of the protocol and materials, and again to rate and log-specific usability issues.

Cognitive walkthroughs. Cognitive walkthroughs are more resource intensive than heuristic analyses largely because they require representative users. Although there are multiple approaches, walkthroughs generally focus on leading individual users or groups of users through key aspects of a product to identify the extent to which the product aligns with their expectations or internal cognitive models (Mahatody et al., 2010). Drawing from existing walkthrough procedures (Bligård & Osvalder, 2013), USE-EBPI presents users with common use scenarios and, using a sequential, mixed-methods data collection approach (Palinkas et al., 2011), asking them to rate whether they will be able to perform the correct actions (ranging from “A very good chance of success [5]” to “No/a very small chance of success [1]”) and then provide justifications. Average success ratings identify qualitative responses that require more in-depth review (e.g., via systematic content analysis [Hsieh & Shannon, 2005]). Despite their efficiency, cognitive walkthroughs tend to over-identify potential usability problems (Health and Human Services, n.d.). Although they were not applied in the exposure protocol example,

walkthroughs were considered as a lower-cost alternative to more intensive lab-based user testing.

Lab-based user testing. Individual, task-based user testing with observation is a hallmark of HCD because it captures direct interactions between users and features of a product. Typically, testing involves presenting a series of scenarios and observing how successfully and efficiently users complete a set of discrete tasks. EBPI usability tests build on established behavioral rehearsal methods (e.g., Beidas et al., 2014), but with the novel objective of evaluating the intervention instead of assessing user competence. First, participants are often instructed “*think aloud*” (Benbunan-Fich, 2001) when completing tasks to describe their processes and experiences as they navigate the EBPI tasks and packaging (qualitative). The pathway to completion (i.e., how the user completed the task and using which materials) is recorded for subsequent coding. Second, indicators of *task effectiveness* are drawn from the general usability testing literature (Hornbæk, 2006) and may include error rate (i.e., number of errors made across tasks), binary task success or failure (total percent of tasks completed), and help seeking (from the examiner) during tasks. Third, *task efficiency* (time to completion) may also be recorded. Lab-based testing can be done rapidly (Pawson & Greenberg, 2009), but is nevertheless a complex endeavor. Novice usability testers can struggle with categorizing and determining the severity of identified usability problems (Bruun, 2010).

In our example application, 10 representative users were recruited from an existing clinical practice network. Institutional review board approval was obtained by the second author from the Behavioral Health Research Collective. Clinicians were invited via recruiting emails. Interested participants ($n=16$) who had interest in exposure completed an online consent form and background questionnaire that included exposure familiarity (see Step 1). Based on the extent of their training in exposure, participants were sorted into novice, intermediate, and advanced groups and recruited from these strata to ensure equal representation. Prior to testing sessions, participants reviewed all artifacts. Testing was subsequently completed remotely, using a secure web conferencing platform and included (1) a “*think aloud*” review of the brief exposure guide and (2) a behavioral rehearsal role-play in which participants completed an exposure with the facilitator playing the role of a 20-year-old female client with contamination fears. Facilitators used a standardized testing guide that specified passive initial refusal to complete the exposure during the role-play, as well as multiple expressions of distress during which task effectiveness—operationalized as successful completion of the exposure—was tracked. Task effectiveness was prioritized given information that critical exposure elements are infrequently adequately delivered in community practice. Finally, participants completed (3) a

semi-structured interview to gather additional feedback about the exposure tasks and packaging. No incentive or compensation was provided. Two research team members were present for each session: a test facilitator and a scribe who took detailed notes for subsequent coding.

Following testing, the notes from each session were analyzed using an inductive qualitative content analysis procedure (Bradley et al., 2007; Hsieh & Shannon, 2005) in which two members of the research team reviewed all notes independently, generated codes for identified usability problems, rated task completion success (i.e., effectiveness; coded “failure,” “partial success” [with 1 + errors], and “successful”), and met to compare their coding and arrive at consensus judgments through consensus coding (Hill et al., 2005). Usability issues were defined as aspects of the intervention or its components and/or a demand on the user which make it unpleasant, inefficient, onerous, or impossible for the user to achieve their goals in typical usage situations (Lavery et al., 1997).

In vivo user testing. Unlike lab-based testing, in vivo testing involves more extensive applications of an EBPI in a destination context over longer periods of time, which allows for evaluation of the ways in which it interacts with contextual constraints. In vivo testing has the potential to expand the traditional acceptability and feasibility goals of pilot studies (Westlund & Stuart, 2017) with usability evaluation objectives. To be completed successfully, in vivo testing inherently requires some degree of intervention implementation and, as a result, is the most expensive—and also most externally valid—method of evaluating usability. If feasible to collect, real-world adherence data may be conceptualized as an indicator of EBPI task effectiveness. A/B testing, in which two designs are implemented simultaneously (e.g., an original design and a novel, alternative design) to determine whether one is superior (Albert & Tullis, 2013) is particularly useful during in vivo testing. Due to time and resource constraints, it was not feasible for the example application of USE-EBPI to conduct in vivo user testing. Tradeoffs between costs of time, money, and expertise versus quality of information require care in selecting and balancing which usability techniques are selected.

Step 4: organize and prioritize identified usability issues

Regardless of the techniques used, usability problems identified are classified and prioritized with a structured method within USE-EBPI. Usability taxonomies provide a means for the consistent and accurate classification, comparison, reporting, analysis, and prioritization of usability issues (Jeffries, 1994; Keenan et al., 1999). To organize and prioritize usability issues, USE-EBPI adapts an existing taxonomy for categorizing usability problems—the UAF (Khajouei et al., 2011). The UAF was selected because it is

theoretically driven and has demonstrated reliability among experts for categorizing usability problems (Andre et al., 2001).

Organize. The augmented version of the UAF is based on a theory of the interaction cycle (Norman, 1986) and states that, in any interaction, users begin with goals and intentions, and engage in (1) *planning*, which includes cognitive actions to determine *what to do* when interacting with a product to meet those goals; (2) *translating*, cognitive actions to determine *how to carry out their intentions*; (3) *actions*, executing behaviors to manipulate the product; and (4) *feedback*, understanding and interpreting information about the effects of actions. Using a consensus coding approach, usability problems in USE-EBPI are mapped to the interaction cycle to aid redesign. Table 4 displays the adapted UAF taxonomy with generic examples most relevant to classifying anticipated EBPI usability issues. For the exposure protocol, research team members assigned stages of the UAF to each usability issue. Because usability issues can often impact multiple stages of the interaction cycle, all relevant stages of the UAF were identified. Findings from the application of the UAF to the exposure protocol are given in the “Results” section.

Prioritize. Finally, most usability evaluation approaches include a process for determining the potential impact of each identified problem (Medlock et al., 2002; NCI, 2007). In the UAF, prioritizing based on severity and impact focuses redesign efforts on those problems that severely hinder key interactions, ensuring that only essential elements that need fixing receive attention. USE-EBPI employs revised UAF ratings in which priority (ranging from “low priority” [1] to “high priority” [3]) is assigned to each identified problem by two or more independent team members based on its (1) impact on users, (2) likelihood of a user experiencing it, and (3) criticality for an EBPI’s putative change mechanisms. Ratings are averaged across reviewers. For example, two research team members independently rated each usability problem. Mean scores were calculated (see the section “Results”). Although ratings inform prioritization, all decisions about which problems to address first in an EBPI redesign effort should be made by the design team when considering all available information.

Results

Quantitative ratings

IUS ratings (scale: 0–100) ranged from 65 to 85, with a mean of 80.5 ($SD=9.56$). Based on descriptors developed for the original SUS (Brooke, 1996), this range corresponds to descriptors between “below average” (2nd quartile) and “excellent” (4th quartile; Bangor et al., 2008). The mean was also in the “acceptable range” (3rd/4th

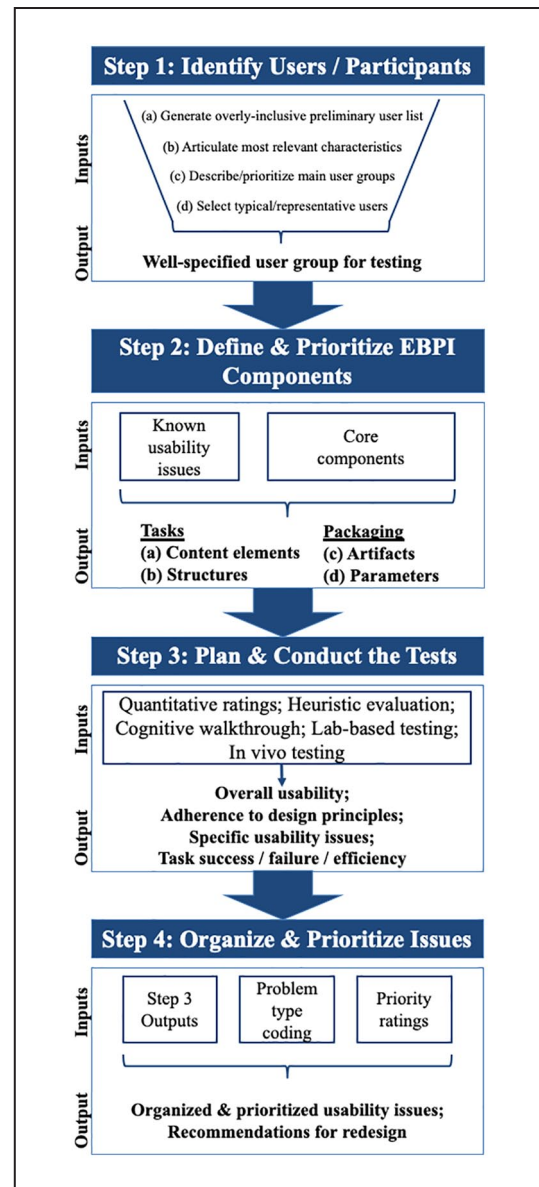


Figure 1. Steps of USE-EBPI methodology. A figure depicting the inputs, techniques, and outputs used across all phases of the method.

quartile). A 10-point difference was observed between advanced participants ($M=87.5$; $SD=8.66$) and both novice ($M=77.5$; $SD=10.90$) and intermediate participants ($M=77.5$; $SD=8.66$).

Heuristic evaluation

HERE ratings (scale: 1–10) indicated a mean overall assessment rating of 7.33 ($SD=0.58$; Table 6). The highest ratings were assigned for the efficiency of the exposure protocol ($M=8.33$; $SD=0.58$) and the lowest for its ability to exploit natural constraints ($M=5.00$; $SD=3.61$). *Exploit natural constraints* demonstrated the most variability

Table 1. EBPI usability test user/participant identification process (Step 1).

Substeps	Description	Exposure protocol example
1. Generate preliminary user list	<ul style="list-style-type: none"> • Generate an overly-inclusive list • Consider individuals in different roles 	<ul style="list-style-type: none"> • Behavioral health clinicians who treat anxiety • Supervisors who support those clinicians • Adult clients who experience anxiety
2. Articulate most relevant user characteristics	<ul style="list-style-type: none"> • Personal characteristics • Task-related characteristics • Geographic/social/setting characteristics 	<ul style="list-style-type: none"> • Experience delivering exposure interventions (clinicians) (supervisors) • Anxiety severity (consumers)
3. Describe and prioritize main user groups	<ul style="list-style-type: none"> • Articulate primary, secondary, and non-users 	<ul style="list-style-type: none"> • Clinicians (primary) • Supervisors (secondary) • Clinicians uninterested in providing exposure interventions (non-user)
4. Select typical and representative users for testing	<ul style="list-style-type: none"> • Sample into user subtype strata • Recruit ~6–20 users per test 	<ul style="list-style-type: none"> • Novice, intermediate, advanced in delivering exposure procedures • Recruited ten clinicians

EBPI: evidence-based psychosocial intervention.

Table 2. EBPI tasks and packaging components (Step 2).

Tasks	Definition	General examples	Exposure protocol components
Content elements	Discrete clinical techniques or strategies used in a session	Exposure; cognitive restricting; psychoeducation; agenda setting	Psychoeducation about anxiety/treatment; fear hierarchy construction; exposure procedures
Structures	Processes that guide the dynamic selection, organization, and maintenance of content	Team-based goal setting; measurement-based care; structured supervision; intervention algorithms	SUDS ratings/fear thermometer to guide exposure; post-session core fear hypothesis testing to drive decision-making
Artifacts	Tangible, digital, or visual materials that exist to support task completion	Intervention manuals; informational handouts; job aids; homework sheets	How-to manual; brief exposure guide (intended for pre- and post-session self-study and reflection by the provider); core fear map; example fear hierarchies
Parameters	Static properties that define and constrain the intervention or service “space”	Modality; prescriptive content sequencing; session length or length of stay/care episode; content delivery method; dosage; language	Individual delivery modality; English language; hierarchy construction-exposure sequencing

EBPI: evidence-based psychosocial intervention; SUDS: subjective units of distress.

Table 3. Testing techniques and user research questions (Step 3).

Testing technique	Questions it can answer	Strengths/weaknesses	Exposure protocol application
Quantitative instruments	How significant overall are an EBPI's usability issues? Does the severity of EBPI usability issues vary across groups?	<p><u>Strengths:</u></p> <ul style="list-style-type: none"> • Rapid comparison of alternatives <p><u>Weaknesses:</u></p> <ul style="list-style-type: none"> • Only indicates presence of a usability problem 	IUS to assess overall usability
Heuristic evaluation	How well does an EBPI align with established usability principles? What is the likelihood that an EBPI user will encounter common usability issues?	<p><u>Strengths:</u></p> <ul style="list-style-type: none"> • No user recruitment <p><u>Weaknesses:</u></p> <ul style="list-style-type: none"> • Ratings require design expertise • Risk of excessive problem "lumping" 	HERE to assess alignment with usability principles
Cognitive walkthroughs	How learnable is the EBPI for new or infrequent users? How well does the basic structure and process of an EBPI align with users' goals, expectations, and internal mental models? Which EBPI activities are likely to be most problematic for users? What major EBPI usability issues are readily detectable?	<p><u>Strengths:</u></p> <ul style="list-style-type: none"> • More efficient than traditional testing <p><u>Weaknesses:</u></p> <ul style="list-style-type: none"> • May over-identify usability problems 	Not applied
Lab-based testing	What specific usability issues do users (new or experienced) encounter when completing targeted EBPI activities? How effectively and efficiently do users complete EBPI activities? How frequently do users make errors when completing EBPI activities? To what extent does a user's experience interacting with an EBPI match their expectations?	<p><u>Strengths:</u></p> <ul style="list-style-type: none"> • Gold standard usability assessment • Observational identification of usability problems <p><u>Weaknesses:</u></p> <ul style="list-style-type: none"> • Time and resource intensive • Requires trained usability tester 	Testing exposure procedures using: Think aloud technique Behavioral rehearsal/ role-play Evaluation of task effectiveness Not applied
In-vivo testing	How successfully can an EBPI operate in an intended destination context? How do users interact with an EBPI over time? What contextual variables are most influential on an EBPI's functioning? In what ways do secondary and tertiary users interact with an EBPI? What aspects of an EBPI are commonly omitted from delivery and why? Which of multiple competing EBPI design options can best be delivered in the destination context? What aspects of an EBPI are most related to its adoption or discontinuation?	<p><u>Strengths:</u></p> <ul style="list-style-type: none"> • Augments standard pilot tests • Greatest external validity <p><u>Weaknesses:</u></p> <ul style="list-style-type: none"> • Requires some real-world implementation • Most expensive 	Not applied



Highest cost

EBPI: evidence-based psychosocial intervention; IUS: Intervention Usability Scale; HERE: Heuristic Evaluation Rubric for EBPI.

Table 4. Adapted User Action Framework (UAF) for organizing EBPI usability issues (Step 4).

Step of interaction cycle	Generic descriptors of usability problems
<p>Planning What interferes with user understanding what to do and/or determining what to do? Example: An EBPI uses many specialized, complex terms, and specifies rigid sequence. User finds it difficult to determine a best sequence to accomplish clinical goals</p>	<p>Problems with: A. User's model of system (e.g., fit with beliefs and expectations; awareness and understanding of model and metaphors; idiom has to be learned to do task planning) B. Goal decomposition (e.g., ability to establish sequence of tasks to accomplish goals and determine what to do next) C. Memory limitations not supported D. User's knowledge of system state, modalities</p>
<p>Translation Can the user translate plans into actions to achieve their goals during EBPI interactions? Example: A mandatory procedure requires additional assessment when a patient reports symptoms above a cutoff, but the measure is difficult to score by hand</p>	<p>Problems with: A. Existence of visual cues B. Conceptual and procedural clarity of core tasks (content and structures) C. Preferences (e.g., appropriate default values, aesthetics, poor attempts at humor; locus of control) D. Presentation of cognitive affordance (e.g., perceptual issues, including noticeability, legibility, quality of graphics, layout, and grouping) E. Efficiency (e.g., number of steps, short cuts, anticipating most likely next tasks)</p>
<p>Actions Can the user successfully perform actions with respect to the elements (tasks and packaging components) within typical within clinical workflow and inner setting or daily life? Example: A depressed patient is to complete between session assignments but no prompts or feedback is provided to remember or support new behavior</p>	<p>Problems with: A. Perceiving and manipulating affordances B. Timing of presentation of opportunities for actions C. Feedback during actions D. Accommodating user differences (e.g., new and experienced users; differently abled users) E. Complexity or excessive workload F. Inconsistency</p>
<p>Feedback (assessment) Can user understand effect of actions on the system/ interaction/ intervention; perceive, aggregate, and interpret feedback generated by EBPI applications? Example: Provider and patient collect no patient-reported outcomes and therefore lack data about improvement in symptoms</p>	<p>Problems with: A. Information collection (e.g., efficiency; accessibility) B. Feedback (e.g., presence; comparison of present state to goal state; interpretability) C. Presentation of feedback (e.g., noticeability, legibility, timing, presentation medium) D. Content/meaning of feedback (e.g., clarity, completeness, correctness, relevance) preferences and efficiency E. Information displays for results of task</p>

EBPI: evidence-based psychosocial intervention.

Intervention Usability Scale (IUS)
(Adapted from the System Usability Scale)

	Strongly Disagree				Strongly Agree
1. I think I would like to use this intervention frequently	1	2	3	4	5
2. I found the intervention unnecessarily complex	1	2	3	4	5
3. I thought the intervention was easy to use	1	2	3	4	5
4. I think that I would need the support of an expert consultant to be able to use this intervention	1	2	3	4	5
5. I found the various components of this intervention were well integrated	1	2	3	4	5
6. I thought there was too much inconsistency in this intervention	1	2	3	4	5
7. I would imagine that most people would learn to use this intervention very quickly	1	2	3	4	5
8. I found this intervention very cumbersome to use	1	2	3	4	5
9. I felt very confident using this intervention	1	2	3	4	5
10. I needed to learn a lot of things before I could get going with this intervention	1	2	3	4	5

Adapted from the System Usability Scale:
Brooke, J. (1996). SUS-A quick and dirty usability scale. *Usability evaluation in industry*, 189(194), 4-7.

Figure 2. The IUS, as applied in the current project.

across raters. Qualitative reasons given for low ratings for that domain indicated that, aside from some references to what types of exposures can be accomplished in a clinician's office, the exposure protocol materials did not speak to the context of use in any identifiable way.

Lab-based testing

Task effectiveness. Successful task completion during the behavioral rehearsal was coded for nine of the 10 participants (one participant did not attempt it). Two novice

participants (66%) and one intermediate participant (25%) failed the exposure task. No advanced participants failed the task. Reasons for failure included engaging in contraindicated behaviors, such as providing reassurance to the client during the exposure and unilaterally selecting the easiest trigger from a fear hierarchy (rather than collaboratively choosing something mid-range). No participants were coded as achieving partial success.

Usability problem prioritization. Consensus coding yielded 13 distinct usability problems. In Table 7, usability problems

Heuristic Evaluation Rubric for EBPIs (HERE)	
Criteria:	Scale (1-10; 1=not at all; 10=extremely)
<p>1. Learnability</p> <p>The EBPI provides users with opportunities to rapidly build understanding of, or facility in, its use.</p>	1 2 3 4 5 6 7 8 9 10
<p>2. Efficiency</p> <p>The EBPI can be applied by users to resolve identified problems with minimal time, effort, and cost.</p>	1 2 3 4 5 6 7 8 9 10
<p>3. Memorability</p> <p>Users of the EBPI can remember and successfully apply important elements of the EBPI protocol without many added supports.</p>	1 2 3 4 5 6 7 8 9 10
<p>4. Error reduction</p> <p>The EBPI explicitly prevents or allows rapid recovery from errors or misapplications of content.</p>	1 2 3 4 5 6 7 8 9 10
<p>5. Low cognitive load</p> <p>The EBPI task structure is sufficiently simple so that amount of thinking required to complete a task minimized.</p>	1 2 3 4 5 6 7 8 9 10
<p>6. Exploit natural constraints</p> <p>The EBPI incorporates or explicitly addresses the static properties of the intended destination context, which may affect the ways it can be used.</p>	1 2 3 4 5 6 7 8 9 10
<p>7. Overall assessment</p>	1 2 3 4 5 6 7 8 9 10
Notes / explanation of ratings:	
Guidelines for application of HERE:	
<ul style="list-style-type: none"> • Always use more than one evaluator • Evaluators should ideally have “double expertise” (usability + subject domain) • Evaluators should review all relevant/available EBPI materials (including training materials and other implementation supports) 	
<p>Developed by Aaron Lyon & Kelly Koerner. Based on design goals for EBPIs articulated in: Lyon, A. R., & Koerner, K. (2016). User-centered design for psychosocial intervention development and implementation. <i>Clinical Psychology: Science and Practice</i>, 23(2), 180-200.</p>	

Figure 3. The HERE, as applied in the current project.

are organized based on priority scores, as these account for both likelihood of occurrence and anticipated impact. Usability problem priority scores from the UAF across the two raters were correlated at $r = .65$. Problems receiving the highest average priority ratings included ambiguity about contraindicated behaviors listed in the brief exposure guide ($M = 3.0$) and the procedures failing to block the use of these contraindicated behaviors during the behavioral rehearsal ($M = 3.0$). In general, usability problems receiving the highest priority scores were also experienced by the greatest number of users ($r = .66$).

Usability problem organization. Application of the UAF interaction cycle to the usability issues indicated that most

impacted more than one step of that cycle (Table 7). Seven of the usability issues interfered with the *planning* phase, seven negatively impacted *translation* of plans into actions, five issues interfered with performance of *actions*, and seven problems related *feedback*. Only one usability issue—confusing, non-intuitive formatting or labeling in the brief exposure guide—was determined to impact all four steps of the UAF interaction cycle.

Discussion

Complex psychosocial interventions are common in contemporary health care services. Their usability is a critical, but understudied, determinant of implementation outcomes.

Evaluation of usability provides insights to drive adoption decisions as well as proactive adaptation to improve intervention implementability (Lyon et al., 2019). USE-EBPI is the first method developed to directly assess the usability of complex psychosocial interventions.

USE-EBPI application to exposure protocol components

IUS results indicated that overall clinician-rated usability of the exposure protocol components tested was good,

Table 5. Demographics of participants.

Characteristic	N	%
Gender		
Male	4	40
Female	6	60
Race		
Aboriginal (First Nations, Metis, Inuit)	0	
Asian	0	
Native Hawaiian or other Pacific Islander	0	
Black or African American	1	10
White/non-Hispanic	8	80
More than one race or other	1	10
Highest degree earned		
MA	2	20
MSW	4	40
PhD	3	30
Other	1	10
Age		
20–29	1	10
30–39	3	30
50–59	2	20
60–69	4	40

Table 6. HERE evaluation ratings.

Item	M	SD
Learnability <i>The EBPI provides users with opportunities to rapidly build understanding of, or facility in, its use</i>	7.33	1.155
Efficiency <i>The EBPI can be applied by users to resolve identified problems with minimal time, effort, and cost</i>	8.33	0.577
Memorability <i>Users of the EBPI can remember and successfully apply important elements of the EBPI protocol without many added supports</i>	6.33	0.577
Error reduction <i>The EBPI explicitly prevents or allows rapid recovery from errors or misapplications of content</i>	7.67	0.577
Low cognitive load <i>The EBPI task structure is sufficiently simple so that the amount of thinking required to complete a task is minimized</i>	6.33	0.577
Exploit natural constraints <i>The EBPI incorporates or explicitly addresses the static properties of the intended destination context, which may affect the ways it can be used</i>	5.00	3.606
Overall assessment	7.33	0.577

HERE: Heuristic Evaluation Rubric for EBPI; EBPI: evidence-based psychosocial intervention.

based on established SUS norms. For comparison, mean ratings were comparable with SUS ratings of the iPhone, but lower than a typical microwave oven (Kortum & Bangor, 2013). This indicates that, while the materials could be improved, the current state is likely acceptable for many users. Nevertheless, differences in IUS ratings by clinician experience level illustrate the value in stratifying by experience. Advanced clinicians viewed the materials near the “excellent” range, whereas practitioners less experienced with exposure were more impacted by usability issues. IUS ratings were largely consistent with HERE rubric ratings by experts, which independently suggested moderate to good usability. However, HERE evaluation also yielded unique information about the protocol’s difficulty exploiting natural constraints, which provides potential direction for subsequent intervention adaptations.

Lab-based testing revealed additional detail about the specific usability problems and further underscored the utility of including users with varying experience levels. Interestingly, novice users attempted and correctly performed many aspects of the exposure, but were not able to rapidly identify and avoid *proscribed* behaviors, and three ultimately failed the exposure task. These findings signal one critical usability problem experienced by novice users (i.e., that the intervention failed to block contraindicated behaviors) that is ripe to be addressed in future proactive adaptations (see below). This, and the 12 additional usability problems, provide insight into potential reasons for lower ratings or task failure and can be used to identify redesign directions.

Implications for adaptation and redesign

Although it is beyond the scope of this article to detail the full EBPI adaptation or redesign processes that may result

Table 7. Prioritization and categorization of usability problems (Steps 3 and 4 results).

Average UAF priority rating	% encountering problem	Impact by clinician experience level	Usability problem	Usability problem organization: step of UAF impacted (planning translation action feedback)			
3.0	20		Failure to provide needed feedback to block contraindicated behaviors	P	T	A	F
3.0	50		Contraindicated behaviors are ambiguous (layout and meaning hinders error avoidance)	P	T	A	F
2.5	70		Confusing/non-intuitive presentation of cognitive affordances (e.g., legibility, layout, grouping)	P	T	A	F
2.5	70		Unclear idiom “Processing” causes confusion	P	T	A	F
2.5	60		Lack of feedback on accuracy of hierarchy level	P	T	A	F
2.0	30		Absence of labeling and instructions interferes with user understanding purpose/rationale	P	T	A	F
2.0	30		Omission of content some users expect in exposure (method of action beyond habituation)	P	T	A	F
2.0	30		Insufficient support of exposure planning	P	T	A	F
1.5	10		Insufficient feedback to reinforce success	P	T	A	F
1.5	30		Lack of needed content—troubleshooting for family/system issues	P	T	A	F
1.5	10		Confusing/non-intuitive grouping (blend therapist and client barriers in same list)	P	T	A	F
1.0	20		Unclear idiom-“Habituation” is unclear	P	T	A	F
1.0	10		Omission of content-user expects defined developmental level of patient to be specified	P	T	A	F

UAF: User Action Framework.
 Clinician experience level (1 circle= 1 participant; darkened if impacted; gray if not impacted):
 Novice ○
 Intermediate ◐
 Advanced ●

from the assessment described, results suggested how the intervention’s implementability may be enhanced. While usability testing is necessarily problem-focused, redesign decisions should be sure to retain known strengths and positive aspects of the intervention. Focusing redesign on the highest priority problems is intended to help avoid excessive adaptations that may not be critical to ensuring implementability.

Although any adaptation must ultimately be codified in the written intervention protocol (an artifact), adaptations

informed by USE-EBPI might include those made to any aspects of the intervention (e.g., content, structures). As indicated above, the most critical usability issues had to do with clearer and understandable signaling about behaviors that undermine the purported active mechanism of exposure (e.g., excessive reassurance). The exposure protocol may benefit from the following proactive adaptations: First, overall usability of the intervention might be improved by clearer labeling in the brief exposure guide. Second, the intervention could provide strategic and timely

delivery of instruction to clinicians on how to use the guide before, during, and after a session for self-supervision. Third more novice-friendly idioms and additional supports (e.g., example scripts) in ambiguous areas (e.g., “exposure processing”) would reduce confusion, especially for less experienced clinicians. Fourth, clearer visual grouping of content presented in select artifacts (e.g., the brief exposure guide) may enhance ease of comprehension. In addition, assignment UAF interaction cycle steps to each usability problem further facilitates redesign decisions. Top priority issues (i.e., those receiving a 2.5 or above) were least likely to involve the planning step, suggesting that appropriate adaptations might be focused less on communicating concepts in understandable ways, and more on their applications.

Limitations

The current application of USE-EBPI has a number of limitations. First, the method was only applied with one intervention. Future research should broaden its applications to a wider range of evidence-based programs and practices. Second, the study applied the method to only one of the identified primary user groups (clinicians). As described previously, USE-EBPI is intended to be applicable to a wide range of primary and secondary users. Future research should examine the extent to which EBPI usability testing with clinicians and clients surfaces unique problems to be considered during redesign. Finally, we did not collect explicit information about the extent to which the various USE-EBPI techniques were feasible for use by different stakeholders. Nevertheless, we expect that lower-cost approaches (e.g., quantitative instruments) may be readily applied by community-based stakeholders, whereas more detailed and time-intensive techniques will require expert support and/or more resources (see Table 3).

Conclusion

Intervention-level determinants of successful implementation are understudied in contemporary implementation research and few methods exist to identify EBPI components for prospective adaptation (Lyon & Bruns, 2019). The USE-EBPI methodology allows for evaluation of a critical intra-intervention determinant—intervention usability—for complex psychosocial interventions in health care. The current study provides preliminary evidence for its utility in generating information about the implementability of specific interventions as well as informing subsequent intervention redesign.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This project was supported, in part, by grants K08MH095939, R34MH109605, and P50MH115837, awarded by the National Institute of Mental Health.

ORCID iD

Aaron R Lyon  <https://orcid.org/0000-0003-3657-5060>

References

- Albert, W., & Tullis, T. (2013). *Measuring the user experience: Collecting, analyzing, and presenting usability metrics* (2nd ed.). Morgan Kaufmann.
- Andre, T. S., Hartson, H. R., Belz, S. M., & McCreary, F. A. (2001). The user action framework: A reliable foundation for usability engineering support tools. *International Journal of Human-Computer Studies*, 54(1), 107–136. <https://doi.org/10.1006/ijhc.2000.0441>
- Bangor, A., Kortum, P. T., & Miller, J. T. (2008). An empirical evaluation of the System Usability Scale. *International Journal of Human-Computer Interaction*, 24(6), 574–594. <https://doi.org/10.1080/10447310802205776>
- Becker, E. M., Smith, A. M., & Jensen-Doss, A. (2013). Who’s using treatment manuals? A national survey of practicing therapists. *Behaviour Research and Therapy*, 51(10), 706–710. <https://doi.org/10.1016/j.brat.2013.07.008>
- Beidas, R. S., Cross, W., & Dorsey, S. (2014). Show me, don’t tell me: Behavioral rehearsal as a training and analogue fidelity tool. *Cognitive and Behavioral Practice*, 21(1), 1–11. <https://doi.org/10.1016/j.cbpra.2013.04.002>
- Benbunan-Fich, R. (2001). Using protocol analysis to evaluate the usability of a commercial web site. *Information & Management*, 39(2), 151–163. [https://doi.org/10.1016/S0378-7206\(01\)00085-4](https://doi.org/10.1016/S0378-7206(01)00085-4)
- Beuscart-Zephir, M. C., Pelayo, S., & Bernonville, S. (2010). Example of a Human Factors Engineering approach to a medication administration work system: Potential impact on patient safety. *International Journal of Medical Informatics*, 79(4), e43–e57. <https://doi.org/10.1016/j.ijmedinf.2009.07.002>
- Beyer, H., & Holtzblatt, K. (1998). *Contextual design: Defining customer-centered systems*. Morgan Kaufmann.
- Bligård, L., & Osvalder, A. (2013). Enhanced cognitive walkthrough: Development of the cognitive walkthrough method to better predict, identify, and present usability problems. *Advances in Human-Computer Interaction*, 2013, Article 931698. <https://doi.org/10.1155/2013/931698>
- Bradley, E. H., Curry, L. A., & Devers, K. J. (2007). Qualitative data analysis for health services research: Developing taxonomy, themes, and theory. *Health Services Research*, 42(4), 1758–1772. <https://doi.org/10.1111/j.1475-6773.2006.00684.x>
- Brooke, J. (1996). SUS: A quick and dirty usability scale. In P. W. Jordan, B. Thomas, I. L. McClelland, & B. Weerdmeester (Eds.), *Usability evaluation in industry* (pp. 189–194). CRC Press.
- Bruun, A. (2010, October). *Training software developers in usability engineering: A literature review* [Conference session].

- 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries, Reykjavik, Iceland.
- Chambers, D. A., Glasgow, R. E., & Stange, K. C. (2013). The dynamic sustainability framework: Addressing the paradox of sustainment amid ongoing change. *Implementation Science*, 8, Article 117. <https://doi.org/10.1186/1748-5908-8-117>
- Chambers, D. A., & Norton, W. E. (2016). The adaptome: Advancing the science of intervention adaptation. *American Journal of Preventive Medicine*, 51(4, Suppl. 2), S124–S131. <https://doi.org/10.1016/j.amepre.2016.05.011>
- Chorpita, B. F., Daleiden, E. L., & Weisz, J. R. (2005). Identifying and selecting the common elements of evidence based interventions: A distillation and matching model. *Mental Health Services Research*, 7(1), 5–20.
- Cooper, A. (1999). *The inmates are running the asylum*. Macmillan Publishing.
- Cooper, A., Reimann, R., & Cronin, D. (2007). *About face 3: The essentials of interaction design*. John Wiley.
- Courage, C., & Baxter, K. (2005). *Understanding your users: A practical guide to user requirements methods, tools, and techniques*. Morgan Kaufmann.
- Craig, P., Dieppe, P., Macintyre, S., Michie, S., Nazareth, I., & Petticrew, M. (2013). Developing and evaluating complex interventions: The new Medical Research Council guidance. *International Journal of Nursing Studies*, 50(5), 587–592. <https://doi.org/10.1016/j.ijnurstu.2012.09.010>
- Damschroder, L. J., Aron, D. C., Keith, R. E., Kirsh, S. R., Alexander, J. A., & Lowery, J. C. (2009). Fostering implementation of health services research findings into practice: A consolidated framework for advancing implementation science. *Implementation Science*, 4, Article 50. <https://doi.org/10.1186/1748-5908-4-50>
- Dopp, A., Parisi, K. E., Munson, S. A., & Lyon, A. R. (2019). Integrating implementation and user-centred design strategies to enhance the impact of health services: Protocol from a concept mapping study. *Health Research Policy and Systems*, 17, Article 1.
- Dorsey, S., Pullmann, M. D., Kerns, S. E. U., Jungbluth, N., Berliner, L., Thompson, K., & Segell, E. (2016). Efficient sustainability: Existing community-based supervisors as evidence-based treatment supports. *Implementation Science*, 11(Suppl. 1), Article 85.
- England, M. J., Butler, A. S., & Gonzales, M. L. (2015). *Psychosocial interventions for mental and substance use disorders: A framework for establishing evidence-based standards*. National Academies Press.
- Gagliardi, A. R., Brouwers, M. C., Palda, V. A., Lemieux-Charles, L., & Grimshaw, J. M. (2011). How can we improve guideline use? A conceptual framework of implementability. *Implementation Science*, 6(1), Article 26.
- Garland, A. F., Hawley, K. M., Brookman-Frazee, L., & Hurlburt, M. S. (2008). Identifying common elements of evidence-based psychosocial treatments for children's disruptive behavior problems. *Journal of the American Academy of Child and Adolescent Psychiatry*, 47(5), 505–514. <https://doi.org/10.1097/CHI.0b013e31816765c2>
- Hackos, J. T., & Redish, J. C. (1998). *User and task analysis for interface design*. John Wiley.
- Hamilton, A. B., Oishi, S., Yano, E. M., Gammage, C. E., Marshall, N. J., & Scheuner, M. T. (2014). Factors influencing organizational adoption and implementation of clinical genetic services. *Genetics in Medicine*, 16(3), 238–245.
- Health and Human Services. (n.d.). *Use cognitive walkthroughs cautiously usability guidelines*. <https://webstandards.hhs.gov/guidelines/204>
- Hill, C. E., Knox, S., Thompson, B. J., Williams, E. N., Hess, S. A., & Ladany, N. (2005). Consensual qualitative research: An update. *Journal of Counseling Psychology*, 52, 196–205.
- Himle, M. B., & Franklin, M. E. (2009). The more you do it, the easier it gets: Exposure and response prevention for OCD. *Cognitive and Behavioral Practice*, 16(1), 29–39. <https://doi.org/10.1016/j.cbpra.2008.03.002>
- Hornbæk, K. (2006). Current practice in measuring usability: Challenges to usability studies and research. *International Journal of Human-Computer Studies*, 64(2), 79–102. <https://doi.org/10.1016/j.ijhcs.2005.06.002>
- Hsieh, H. F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research*, 15(9), 1277–1288. <https://doi.org/10.1177/1049732305276687>
- Hwang, W., & Salvendy, G. (2010). Number of people required for usability evaluation: The 10±2 rule. *Communications of the ACM*, 53(5), 130–133. <https://doi.org/10.1145/1735223.1735255>
- International Organization for Standardization. (1998). *Ergonomic requirements for office work with visual display terminals (VDTs)—Part 11: Guidance on usability*.
- Jeffries, R. (1994). Usability problem reports: Helping evaluators communicate effectively with developers. In J. Nielsen & R. L. Mack (Eds.), *Usability inspection methods* (pp. 273–294). John Wiley.
- Kazdin, A. E. (2007). Mediators and mechanisms of change in psychotherapy research. *Annual Review of Clinical Psychology*, 3, 1–27. <https://doi.org/10.1146/annurev.clinpsy.3.022806.091432>
- Keenan, S. L., Hartson, H. R., Kafura, D. G., & Schulman, R. S. (1999). The usability problem taxonomy: A framework for classification and analysis. *Empirical Software Engineering*, 4(1), 71–104.
- Khajouei, R., Ameri, A., & Jahani, Y. (2018). Evaluating the agreement of users with usability problems identified by heuristic evaluation. *International Journal of Medical Informatics*, 117, 13–18. <https://doi.org/10.1016/j.ijmedinf.2018.05.012>
- Khajouei, R., Peute, L. W., Hasman, A., & Jaspers, M. W. (2011). Classification and prioritization of usability problems using an augmented classification scheme. *Journal of Biomedical Informatics*, 44(6), 948–957. <https://doi.org/10.1016/j.jbi.2011.07.002>
- Kortum, P. T., & Bangor, A. (2013). Usability ratings for everyday products measured with the System Usability Scale. *International Journal of Human-Computer Interaction*, 29(2), 67–76. <https://doi.org/10.1080/10447318.2012.681221>
- Krause, J., Van Lieshout, J., Klomp, R., Huntink, E., Aakhus, E., Flottorp, S., . . . Baker, R. (2014). Identifying determinants of care for tailoring implementation in chronic diseases: An evaluation of different methods. *Implementation Science*, 9, Article 102. <https://doi.org/10.1186/s13012-014-0102-3>
- Kujala, S., & Kauppinen, M. (2004, October 23–27). *Identifying and selecting users for user-centered design* [Paper presentation]. Nordic Conference on Computer Human Interaction, Tampere, Finland.

- Kujala, S., & Mantyla, M. (2000). How effective are user studies? In S. McDonal, Y. Waern, & G. Cockton (Eds.), *People and computers XIV—Usability or else!* (pp. 61–71). Springer London.
- Lavery, D., Cockton, G., & Atkinson, M. P. (1997). Comparison of evaluation methods using structured usability problem reports. *Behavior & Information Technology*, *16*(4–5), 246–266. <https://doi.org/10.1080/014492997119824>
- Lewis, C. C., Stanick, C. F., Martinez, R. G., Weiner, B. J., Kim, M., Barwick, M., & Comtois, K. A. (2015). The society for implementation research collaboration instrument review project: A methodology to promote rigorous evaluation. *Implementation Science*, *10*(1), Article 2. <https://doi.org/10.1186/s13012-014-0193-x>
- Lyon, A. R., & Bruns, E. J. (2019). User-centered redesign of evidence-based psychosocial interventions to enhance implementation—Hospitable soil or better seeds? *JAMA Psychiatry* *76*, 3–4.
- Lyon, A. R., & Koerner, K. (2016). User-centered design for psychosocial intervention development and implementation. *Clinical Psychology*, *23*, 180–200. <https://doi.org/10.1111/cpsp.12154>
- Lyon, A. R., Munson, S. A., Renn, B. N., Atkins, D. A., Pullmann, M. D., Friedman, E., & Areán, P. A. (2019). Use of human-centered design to improve implementation of evidence-based psychotherapies in low-resource communities: Protocol for studies applying a framework to assess usability. *JMIR Research Protocols*, *8*(10), Article e14990.
- Lyon, A. R., Stanick, C., & Pullmann, M. D. (2018). Toward high-fidelity treatment as usual: Evidence-based intervention structures to improve usual care psychotherapy. *Clinical Psychology: Science & Practice*, *25*, e12265.
- Lyon, A. R., Wasse, J. K., Ludwig, K., Zachry, M., Bruns, E. J., Unutzer, J., & McCauley, E. (2016). The Contextualized Technology Adaptation Process (CTAP): Optimizing health information technology to improve mental health systems. *Administration and Policy in Mental Health*, *43*(3), 394–409. <https://doi.org/10.1007/s10488-015-0637-x>
- Maguire, M. (2001). Methods to support human-centred design. *International Journal of Human-Computer Studies*, *55*(4), 587–634.
- Mahatody, T., Sagar, M., & Kolski, C. (2010). State of the art on the cognitive walkthrough method, its variants and evolutions. *International Journal of Human-Computer Interaction*, *26*(8), 741–785. <https://doi.org/10.1080/10447311003781409>
- Medlock, M. C., Wixon, D., Terrano, M., Romero, R., & Fulton, B. (2002). Using the RITE method to improve products: A definition and a case study. *Usability Professionals Association*, *51*. <https://pdfs.semanticscholar.org/5340/ef8a91900840263a4036b0433a389b7097b2.pdf>
- National Cancer Institute. (2007). *NCI evaluation guidelines for Phase II eHealth SBIR/STTR grantees and contractors*.
- Nielsen, J. (1994). Heuristic evaluation. In J. Nielsen & R. L. Mack (Eds.), *Usability inspection methods* (pp. 25–64). John Wiley.
- Nielsen, J. (2000). *Why you only need to test with 5 users*. <https://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/>
- Norman, D. A. (1986). Cognitive engineering. In D. A. Norman & S. W. Draper (Eds.), *User centered system design: New perspectives on human-computer interaction* (pp. 31–61). Lawrence Erlbaum.
- Palinkas, L. A., Aarons, G. A., Horwitz, S. M., Chamberlain, P., Hurlburt, M. S., & Landsverk, J. (2011). Mixed method designs in implementation research. *Administration and Policy in Mental Health and Mental Health Services Research*, *38*(1), 44–53.
- Pawson, M., & Greenberg, S. (2009). Extremely rapid usability testing. *Journal of Usability Studies*, *4*(3), 124–135.
- Rabin, B. A., McCreight, M., Battaglia, C., Ayele, R., Burke, R. E., Hess, P. L., . . . Glasgow, R. E. (2018). Systematic, multimethod assessment of adaptations across four diverse health systems interventions. *Frontiers in Public Health*, *6*, Article 102. <https://doi.org/10.3389/fpubh.2018.00102>
- Roberts, J. P., Fisher, T. R., Trowbridge, M. J., & Bent, C. (2016). A design thinking framework for healthcare management and innovation. *Healthcare*, *4*(1), 11–14. <https://doi.org/10.1016/j.hjdsi.2015.12.002>
- Rogers, E. M. (1962). *Diffusion of innovations*. Free Press of Glencoe.
- Rogers, E. M. (2010). *Diffusion of innovations* (4th ed.). Simon & Schuster.
- Sauro, J. (2011). *Measuring usability with the System Usability Scale (SUS)*. <https://measuringu.com/sus/>
- Schloemer, T., & Schroder-Back, P. (2018). Criteria for evaluating transferability of health interventions: A systematic review and thematic synthesis. *Implementation Science*, *13*(1), Article 88. <https://doi.org/10.1186/s13012-018-0751-8>
- Scott, K., & Lewis, C. C. (2015). Using measurement-based care to enhance any treatment. *Cognitive and Behavioral Practice*, *22*(1), 49–59. <https://doi.org/10.1016/j.cbpra.2014.01.010>
- Stirman, S. W., Baumann, A. A., & Miller, C. J. (2019). The FRAME: An expanded framework for reporting adaptations and modifications to evidence-based interventions. *Implementation Science*, *14*(1), Article 58. <https://doi.org/10.1186/s13012-019-0898-y>
- Tryon, W. W. (2005). Possible mechanisms for why desensitization and exposure therapy work. *Clinical Psychology Review*, *25*(1), 67–95. <https://doi.org/10.1016/j.cpr.2004.08.005>
- Turner, C. W., Lewis, J. R., & Nielsen, J. (2006). Determining usability test sample size. In W. Karwowski & B. Raton (Eds.), *International encyclopedia of ergonomics and human factors* (2nd ed., Vol. 3) (pp. 3084–3088). CRC Press.
- Waller, G., & Turner, H. (2016). Therapist drift redux: Why well-meaning clinicians fail to deliver evidence-based therapy, and how to get back on track. *Behaviour Research and Therapy*, *77*, 129–137. <https://doi.org/10.1016/j.brat.2015.12.005>
- Westlund, E., & Stuart, E. A. (2017). The nonuse, misuse, and proper use of pilot studies in experimental evaluation research. *American Journal of Evaluation*, *38*(2), 246–261.