

A Novel Method for Assessing Design Fidelity in Web-Based Behavioral Interventions

Jeffrey D. Lambert¹, Lewis R. Elliott², Adrian H. Taylor³, Paul Farrand⁴, Anne M. Haase⁵,
and Colin J. Greaves⁶

¹ Department for Health, University of Bath

² European Centre for Environment and Human Health, University of Exeter Medical School

³ Peninsula School of Medicine, Faculty of Health, University of Plymouth

⁴ Clinical Education, Development and Research (CEDAR), Psychology Program, University of Exeter

⁵ School of Health, Victoria University of Wellington

⁶ School of Sport, Exercise and Rehabilitation Sciences, University of Birmingham

Objective: Delivery is one of the most common ways of assessing fidelity in behavioral interventions. However, there is a lack of research reporting on how well an intervention protocol reflects its proposed theoretical principles (design fidelity). This study presents a systematic method for assessing design fidelity and applies it to the eMotion web-based intervention targeting physical activity and depression.

Method: The eMotion intervention comprises of 13 web-based modules, designed according to an underlying intervention map. An independent rater with expertise in behavior change coded the presence or absence of behavior change techniques (BCTs) in the content of eMotion. Results of coding were compared to the intervention designers' a priori specification for interrater reliability. **Results:** After discussion, the independent rater and the intervention designer had a high agreement for the presence of BCTs relating to behavioral activation ($AC1 = 0.91$) with "demonstration of behavior" and "monitoring of emotional consequences" having the lowest agreement ($AC1 < 0.4$). There was also high agreement for the presence of BCTs targeting physical activity ($AC1 = 0.88$) with "demonstration of behavior" and "monitoring of emotional consequences" having the lowest agreement ($AC1 < 0.4$). The eMotion description was then amended to align the interrater agreement. **Conclusions:** This study presents a novel method for assessing design fidelity. Developers of behavioral (and other multicomponent) interventions are encouraged to develop and refine this method and assess design fidelity in future interventions to ensure BCTs are operationalized as intended.

Keywords: fidelity, intervention, behavior change, depression, physical activity

Supplemental materials: <https://doi.org/10.1037/hea0001046.supp>

The National Institute for Health Behavior Change Consortium (NIH BCC) defines intervention fidelity as the reliability and validity of behavioral interventions (Bellg et al., 2004). Behavioral interventions can fail to represent their intended content and mechanisms of change at five levels: intervention design, provider training, intervention delivery, intervention

receipt, and enactment of the skills promoted (Bellg et al., 2004; Borrelli, 2011; Borrelli et al., 2005). Lack of consideration of intervention fidelity can lead to a false positive, where an intervention not faithful to its intended content yields a significant effect or a false negative where a potentially effective intervention is discarded because the intended con-

Jeffrey D. Lambert  <https://orcid.org/0000-0003-4774-9054>

Lewis R. Elliott  <https://orcid.org/0000-0003-3864-9465>

Adrian H. Taylor  <https://orcid.org/0000-0003-2701-9468>

Paul Farrand  <https://orcid.org/0000-0001-7898-5362>

Anne M. Haase  <https://orcid.org/0000-0001-8556-2165>

Colin J. Greaves  <https://orcid.org/0000-0003-4425-2691>

We thank Samantha van Beurden for her help pilot coding an earlier iteration of the coding manual. Jeffrey D. Lambert's time input was supported by the Economic and Social Research Council (ESRC; Grant: ES/J50015X/1). Colin J. Greaves' time input was supported by the U.K.'s National Institute for Health Research (NIHR; Career Development Fellowship CDF-2012–05-029). This report is independent research and the views expressed are those of the authors and not necessarily those of NIHR or the U.K. Department of Health.

Jeffrey D. Lambert served as lead for conceptualization, methodology, and writing (original draft, review, and editing). Lewis R. Elliott served in a supporting role for conceptualization and writing (original draft, review, and editing). Adrian H. Taylor served in a supporting role for supervision and writing (review and editing). Paul Farrand served in a supporting role for supervision and writing (review and editing). Anne M. Haase served in a supporting role for supervision and writing (review and editing). Colin J. Greaves served as lead for supervision and served in a supporting role for writing (original draft, review, and editing). All authors read and approved the final manuscript.

Correspondence concerning this article should be addressed to Jeffrey D. Lambert, Department for Health, University of Bath, 1 West 4.107, BA2 7AY, United Kingdom. Email: jl2426@bath.ac.uk

tent or mechanisms of change were not adequately operationalized or delivered (Bellg et al., 2004).

A recent systematic review of behavioral interventions promoting physical activity identified a range of methods used by researchers to assess fidelity across the domains of training, delivery, receipt, and enactment (Lambert, Greaves, Farrand, Cross, et al., 2017). These methods included: using checklists to assess provider competence before delivering an intervention (“training”), observing the presence of intended intervention components during delivery, verifying participants’ understanding of intervention components (“receipt”), and observing participant use of intervention components (e.g., goal setting) in participants’ day-to-day lives (“enactment”). However, the systematic review did not identify any examples of studies reporting the assessment of design fidelity.

Assessing design fidelity involves measuring whether behavioral interventions are consistent with stated components of their theoretical models. The NIH BCC proposes that intervention designers: (a) specify and incorporate the intervention components (or behavior change techniques [BCTs]) in the intervention, (b) use a “panel of experts” to assess intervention protocols to ensure they adequately incorporate the underlying theoretical and clinical guidelines, and (c) ensure that measures reflect the hypothesized theoretical constructs/mechanisms of action (Borrelli, 2011).

Several programs of work have sought to specify intervention components resulting in several taxonomies of BCTs (Borek et al., 2015; Hartmann-Boyce et al., 2016; Michie et al., 2011; Michie et al., 2013). The core aim of a taxonomy is to lay the foundation for the design of reliable and replicable behavior change interventions by enabling precise specification of the active ingredients of existing interventions (Michie et al., 2013).

Taxonomies help to specify the content of behavioral interventions. However, there is a lack of guidance concerning ways to ensure BCTs have been sufficiently translated into intervention protocols. Without a systematic assessment process, intervention designers could fail to operationalize intended BCTs into the actual intervention content (e.g., web-based interventions, manuals, protocols) or include additional unintended BCTs that target different processes to the underlying logic model. For example, a previous study compared the presence of BCTs in 13 published intervention descriptions to their corresponding intervention manuals and found a correspondence rate of only 74% (Abraham & Michie, 2008). This discrepancy was due to BCTs appearing in the intervention manuals but not in their corresponding published intervention descriptions and vice versa.

Thus, the first potential threat to intervention fidelity occurs earlier than delivery. If a BCT is not operationalized in an intervention protocol, a provider would be unable to deliver it in practice. By contrast, pharmacological interventions can use laboratory techniques to confirm whether the substance matches what is shown on the label, whether the amount of the substance is consistent with the labeled amount, whether there are process impurities in the substance, and whether the drug is absorbed in the body. The NIH BCC guidance attempts to mimic this pharmacological process in a behavioral context, by recommending the use of a “protocol review group” to ensure BCTs are accurately specified in intervention content. However, there is a lack of guidance regarding who this “protocol review group” should comprise of and how to effectively and reliably use the group to

determine whether the intervention protocol and materials accurately reflect the underlying BCTs in an unbiased way (design fidelity).

The main aims of this study were, therefore: (a) to propose a systematic method for assessing whether underlying BCTs are operationalized in behavioral interventions and (b) to apply this method to the development of the eMotion intervention (Lambert et al., 2018; Lambert, Greaves, Farrand, Haase, et al., 2017) as a case study.

Method

Before assessing design fidelity, intervention developers must first ensure they have specified the precise intended content of their intervention. This might be achieved from intervention mapping (Bartholomew et al., 2011), or a set of tables derived from the Behavior Change Wheel (Michie et al., 2011). To assess design fidelity of the eMotion intervention, an independent coder was asked to assess the presence and location of intended (and unintended) BCTs in the intervention content (e.g., web-pages, written information, worksheets, and behavior-tracking devices), session delivery plans, and other delivery materials (e.g., presentation slides, interactive game materials). The coder was independent of the study design team with experience in using the relevant BCTs and intervention development. The coding manual was developed before the beginning of the assessment. Intervention content was then compared against the intervention description to identify areas of concordance or discordance. The intervention developers and independent coder then resolved any discrepancies in coding through discussion. Discordant aspects of the intervention (areas where the intervention diverged from the intended design) were then addressed by either refining the intervention materials or intervention description. See Table 1 for an overview of the criteria for assessing design fidelity.

Brief Summary of eMotion

To test the assessment method described above, the design fidelity of the eMotion web-based intervention was assessed. The eMotion intervention is a web-based course, consisting of weekly modules, that provides people with access to evidence-based treatment based on behavioral activation (BA) techniques to promote physical activity and to reduce symptoms of depression. eMotion is based on self-determination theory (SDT; Deci & Ryan, 1985) and BA principles (Farrand et al., 2014; Hopko et al., 2003; Lejuez et al., 2001; Richards, 2010). The development of eMotion is described in detail elsewhere (Lambert, Greaves, Farrand, Haase, et al., 2017). In brief, eMotion was hosted on the online mental health platform “Living Life to the Full” (Williams et al., 2016). The eMotion intervention is comprised of audio-visual modules including an introduction, eight weekly modules, three unlockable modules, and a problem-solving module. eMotion delivers a staged approach to behavioral activation with a gradually increasing emphasis on incorporating physical activity into the process. It supports people in first building an understanding of how behavioral activation works. It then asks the user to identify routine, pleasurable, or necessary activities and rate them in terms of difficulty. The user then schedules and tries out the planned activities and is supported to review and problem-solve these activities over time. Throughout the intervention (ideally, as depression begins to lift following success with the initially planned

Table 1
Criteria for Assessing Presence/Absence of BCTs in Intervention Content

Domains	Criteria for assessing design fidelity
What should be assessed?	Presence and location of intended BCTs in intervention content (e.g. manuals, protocols, web-based delivery platforms) Presence and location of non-intended BCTs in intervention content (e.g. manuals, protocols, web-based delivery platforms)
How should data be collected?	Audit of intervention content coded according to criteria and protocols developed a priori (i.e. before beginning the audit of content)
Who should assess?	Coders should be versed in use of and coding of BCTs Coders should be independent of the intervention study/its design Coders should not be aware of where BCTs are located in the content
How should it be assessed?	Inter-rater reliability (e.g. by using agreement coefficients) between coders and intervention description (as specified by the intervention developers) Discrepancies resolved through discussion

Note. BCTs = behavior change techniques.

activities), the additional benefits of activities that include a physical activity component are reinforced/encouraged, and the process continues.

Specification of Intended Content of eMotion

Intervention mapping (Bartholomew et al., 2011) was used to specify the intervention content in eMotion. The behavior change taxonomy (v1; Michie et al., 2013) was used to label the intended BCTs wherever possible. The behavior change technique taxonomy (v1) is a taxonomy of 93 agreed-upon, distinct BCTs resulting from a Delphi-type exercise conducted by an international panel of 32 experts in behavior change interventions. This taxonomy was developed to increase the transparency and replicability of behavioral interventions (Michie et al., 2013). The eMotion intervention contained 17 BCTs, which in turn reflected key processes relating to SDT (Ryan & Deci, 2000), self-regulation theory (Bandura, 1991), and the principles of BA (Jacobson et al., 2001). Each BCT was intended to target the behavioral avoidance and inertia that accompanies depression and overcome sources of negative reinforcement (Farrand et al., 2014). Each BCT was therefore designed to target routine (e.g., doing the washing up), pleasurable (e.g., going to the cinema), or necessary activities (e.g., having a shower, paying a bill). Later in the program, a greater emphasis on routine, pleasurable, or necessary activities that incorporate physical activity were included (e.g., walking to the shops, walking the kids to school, playing tennis). BCTs were incorporated in each specific module of eMotion with a total of 65 occurrences (based on the presence/absence of a BCT in each module) of BCTs relating to routine, pleasurable, or necessary activities and 53 relating to physical activity across eMotion. The most frequently used techniques (targeting routine, pleasurable and necessary or physical activities) were “demonstration of behavior,” “self-monitoring of behavior,” “graded tasks,” “action planning,” “goal-setting behavior,” “problem-solving,” “review behavioral goal,” and “social reward.”

Coding Manual

The full eMotion intervention description was used to derive a set of checklist items that were used as a basis for coding (see [online supplementary material](#)). The key aim of the coding manual

was to facilitate the reliable identification of the BCTs in the audio-visual (participant-facing) content of eMotion. This coding manual was piloted (with another researcher) by applying the coding framework to a sample of eMotion content to ensure clarity of definitions and examples and to refine the coding manual where appropriate. The final coding manual contained definitions for 17 distinct BCTs, specifically targeting either a general behavior that can provide a sense of positive reinforcement (i.e. routine, pleasurable or necessary activities) or a behavior that included a physical activity component ([online supplementary material](#)).

Coding Procedures

An independent coder (with no knowledge of the intended location or prevalence of the BCTs and no involvement in the eMotion study) was asked to work through each module of the eMotion platform and rate the presence or absence of the specified BCTs using the coding manual. The coder was a doctoral-level student (at the time of coding) with expertise in health psychology. The coder also had experience in coding for BCTs in brochures encouraging walking in natural environments (Elliott et al., 2018). Abraham and Michie (2008) found that, after brief training, psychologists could identify BCTs in published intervention descriptions. The coder was thus provided with the coding manual and asked to become familiar with the BCT definitions before implementing the coding. After the first read, the coder was asked to work through the intervention again and rate the presence or absence of each BCT (specified in the coding manual) for each module of eMotion. Analysis of the coding responses was conducted for all intervention content. The coder was not instructed to identify the number of instances of a BCT, only whether it appeared at least once, in a module.

Analysis

To assess how well BCTs were operationalized, interrater reliability was used to calculate the agreement between BCTs identified by the independent coder and BCTs specified in the original intervention description. As the prevalence of some of the intended BCTs in the intervention description was low, the first-order agreement coefficient (AC1) statistic (Gwet, 2002) was preferred

to Kappa to estimate interrater reliability. The AC1 statistic corrects for the possibility of two coders agreeing on the presence of a BCT by chance. Interrater reliability was calculated for each BCT in eMotion and clustered by behavioral target (i.e. routine, pleasurable, and necessary vs. physical activities) to assess fidelity to the underlying theoretical model. Discussions were then held between the lead intervention developer and the independent coder to resolve any discrepancies. Any unresolved discrepancies were then discussed with another member of the design team (CG) and resolved through further discussion. Strength of design fidelity was defined in terms of established AC1 thresholds: < 0.2 = poor, $0.2 \leq 0.4$ = fairly poor, $> 0.4 \leq 0.6$ = moderate, $> 0.6 \leq 0.8$ good, and $> 0.8 \leq 1$ = very good (Landis & Koch, 2008).

Increasing Concordance With the Intended Design

After discussion, a contingency table (see Table 2) was applied to every instance where there was disagreement regarding the presence or absence of a BCT. This guided remedial action to refine the intervention or its description as needed. Disagreements occurred when the independent coder rated a BCT as present and the intervention designer rated the same BCT as absent or vice versa. Agreement occurred when both the independent coder and intervention designer rated the same BCT as present or absent.

This procedure was formative and was primarily used to refine and develop eMotion before its use in a pilot randomized controlled trial (Lambert et al., 2018). However, it could also provide a summative indication of design fidelity that could be used to provide evidence of how well specific BCTs have been operationalized in behavioral interventions (particularly in the final/refined versions). The contingency table, therefore, provided a systematic way of making decisions around BCTs specified in the intervention map that did not align with the intervention content.

Results

Agreement of Intended BCTs Targeting Behavioral Activation (the Promotion of Routine, Pleasurable, and Necessary Activities)

See online supplementary material for BCT definitions. There was good overall agreement for BCTs identified as present or absent in eMotion by the independent coder and the intervention designer for routine, pleasurable, and necessary activities ($AC1 = 0.72$) with a total of 37 (out of a possible 221) discrepancies (see

Figure 1). Good to perfect levels of agreement (>0.6) were found for the following BCTs: “credible source,” “information about health consequences,” “problem-solving,” “behavioral experiments,” “review behavioral goal,” “internal prompts/cues,” “social reward,” “framing/reframing,” “associative learning,” “instruction on how to perform a behavior” and “graded tasks.” “Information about monitoring of emotional consequences,” “goal setting,” and “self-monitoring of behavior” had a moderate agreement ($0.4-0.6$). Those with the poorest agreement (≤ 0.4) were “demonstration of behavior,” “monitoring of emotional consequences,” and “action planning.” Discussion between the independent coder and one of the intervention designers raised overall agreement considerably ($AC1 = 0.91$), yielding a total of 12 remaining discrepancies (see Figure 2). “Demonstration of behavior” and “monitoring of emotional consequences” still had poor design fidelity ($AC1 < 0.4$), with “monitoring of emotional consequences” being present when not intended and “demonstration of behavior” not being present when intended (according to the judgment of the independent coder). The remaining discrepancies were then resolved using the contingency table (see Table 2).

Agreement of Intended BCTs Targeting Physical Activity

There was good overall agreement with BCTs identified by the independent coder in eMotion for BCTs relating to physical activity ($AC1 = 0.67$), with a total of 44 (out of a possible 221) discrepancies (see Figure 1). Good to perfect levels of agreement (>0.6) were found for the following BCTs: “credible source,” “information about emotional consequences,” “information about health consequences,” “graded tasks,” “action planning,” “goal-setting behavior,” “problem-solving,” “behavioral experiments,” “internal prompts/cues,” “social reward,” “framing/reframing,” “associative learning,” and “instruction on how to perform a behavior.” “Demonstration of behavior,” “self-monitoring of behavior,” “monitoring of emotional consequences,” and “review behavioral goal” had the poorest agreement (≤ 0.4).

Discussion between the independent coder and one of the intervention designers raised agreement ($AC1 = 0.88$), with a residual total of 16 discrepancies (see Figure 2). The BCTs “demonstration of behavior” and “monitoring of emotional consequences” still had the poorest agreement, with “monitoring of emotional consequences” being present when not originally specified in the intervention content and “demonstration of behavior” not being present

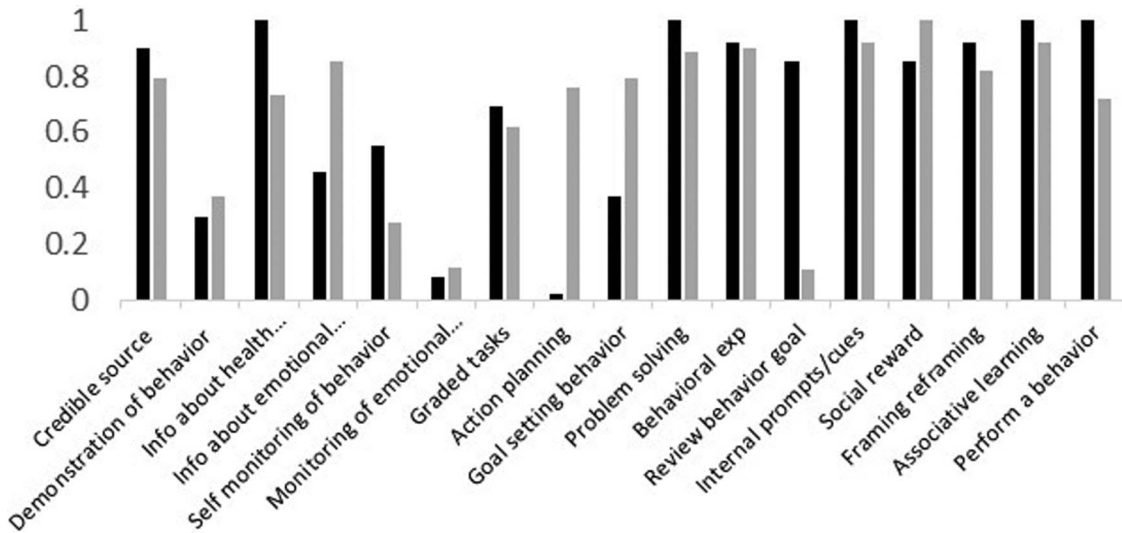
Table 2
Contingency Table Guiding Remedial Action for Discrepant BCTs

Intervention	Independent coder	
	Present	Absent
Specification		
Present	Good design fidelity in terms of intended techniques (no action required)	Designer adds technique to intervention content or amends intervention description to reflect its absence
Absent	Designer removes technique from intervention content or amends intervention description to reflect its presence	Good design fidelity in terms of non-intended techniques (no action required)

Note. BCTs = behavior change techniques.

Figure 1

Agreement (AC1) About the Presence or Absence of Techniques in the Behavioral Activation Components (Black) and Physical Activity Components (Gray) of the Intervention (Before Discussion)



Note. AC1 = first-order agreement coefficient.

when intended (according to the judgment of the independent coder). The remaining discrepancies were then resolved using the contingency table (see Table 3).

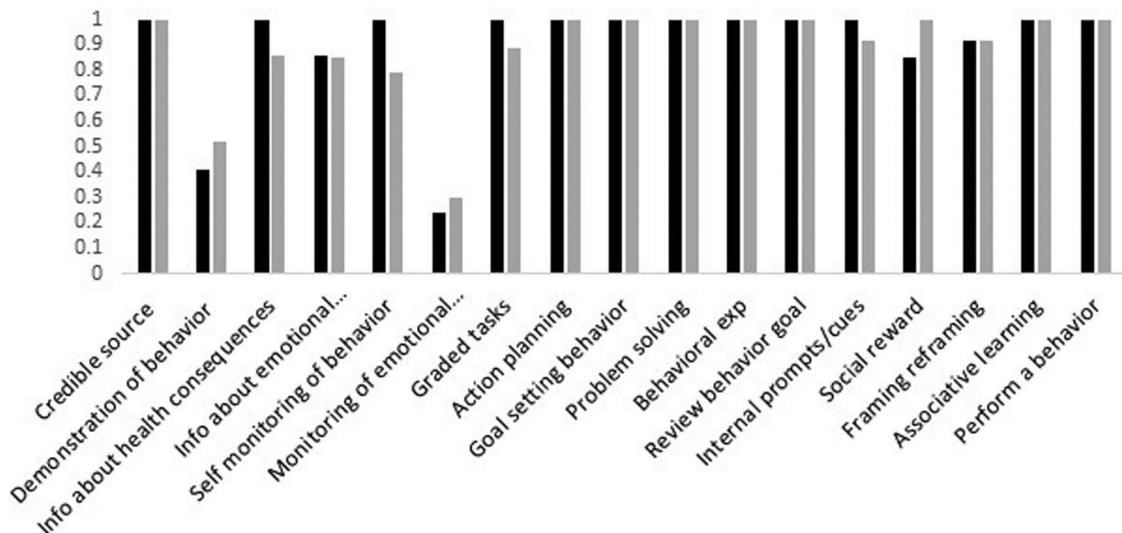
Applying the Contingency Table

For BCTs targeting routine, pleasurable, or necessary activities, after application of the contingency table (see Table 2), one instance of “social reward,” one instance of “framing/

reframing,” and four instances of “demonstration of behavior” were removed from the intervention description as they were not adequately reflected in the content of eMotion as intended. One instance of “information about emotional consequences” and five instances of “monitoring of emotional consequences” were added to the intervention description. These changes resulted in a complete agreement between the intervention designers’ specifications and the independent coder (AC1 = 1.0).

Figure 2

Agreement (AC1) About the Presence or Absence of Techniques in the Behavioral Activation Components (Black) and Physical Activity Components (Gray) of the Intervention (After Discussion)



Note. AC1 = first-order agreement coefficient.

Table 3
Amended Module Content in eMotion After Applying Contingency Table

Module	Updated BCTs in eMotion description
Introduction	Credible source (BA); Information about emotional consequences (BA/PA); Demonstration of behavior (BA)
Week 1	Credible source (BA/PA); Information about emotional consequences (BA/PA); Information about health consequences (PA); Demonstration of behavior (PA); Self-monitoring of behavior (BA); Monitoring of emotional consequences (BA); <u>Framing/reframing (BA)</u>
Week 2	Credible source (PA); Information about emotional consequences (BA/PA); Demonstration of the behavior (BA/PA); Monitoring of emotional consequences (BA); Graded tasks (BA/PA); Action planning (BA); Goal setting (behavior; BA); Problem-solving (BA); Review behavioral goal (BA/PA)
Week 3	Information about emotional consequences (BA/PA); Monitoring of emotional consequences (BA); Graded tasks (BA) Graded tasks (PA); Action planning (BA); Goal setting (behavior; BA); Problem-solving (BA); Review behavioral goal (BA); Social reward (BA); <u>Demonstration of the behavior (BA/PA)</u>
Week 4	<u>Demonstration of the behavior (BA)</u> ; Demonstration of the behavior (PA); Self-monitoring of behavior (BA); Monitoring of emotional consequences (BA/PA) ; Graded tasks (BA/PA); Action planning (BA); Goal setting (behavior) (BA); Problem-solving (BA); Review behavioral goal (BA); Social reward (BA/PA)
Week 5	<u>Demonstration of the behavior (BA)</u> ; Demonstration of the behavior (PA); Self-monitoring of behavior (BA); Monitoring of emotional consequences (BA/PA) ; Graded tasks (BA/PA); Action planning (BA); Goal setting (behavior; BA); Problem-solving (BA); Review behavioral goal (BA); Social reward (BA/PA)
Week 6	<u>Demonstration of the behavior (PA)</u> ; Demonstration of the behavior (BA); Self-monitoring of behavior (BA); Monitoring of emotional consequences (BA/PA) ; Graded tasks (BA/PA); Action planning (BA); Goal setting (behavior; BA); Problem-solving (BA); Review behavioral goal (BA); Social reward (BA/PA)
Week 7	<u>Demonstration of the behavior (BA)</u> ; Self-monitoring of behavior (BA); Monitoring of emotional consequences (BA/PA) ; Graded tasks (BA/PA); Action planning (BA); Goal setting (behavior; BA); Problem-solving (BA); Review behavioral goal (BA); Social reward (BA/PA)
Week 8	Information about emotional consequences (BA/PA); Information about health consequences (PA); Self-monitoring of behavior (BA); <u>Self-monitoring of behavior (PA)</u> ; Monitoring of emotional consequences (BA/PA); Graded tasks (PA); Problem-solving (BA); Review behavioral goal (BA); Internal prompts/cues (BA); <u>Internal prompts/cues (PA)</u> ; Social reward (BA/PA)
Moving on with physical activity	Credible source (PA); Information about emotional consequences (PA); Information about health consequences (PA); Demonstration of the behavior (PA); Monitoring of emotional consequences (PA) ; Graded tasks (PA); Action planning (PA); Goal setting (behavior; PA); Problem-solving (PA); Instruction on how to perform a behavior (PA); <u>Framing/reframing (PA)</u>
Monitoring your physical activity	Information about emotional consequences (PA); Demonstration of the behavior (PA); Self-monitoring of behavior (PA); Monitoring of emotional consequences (PA); Graded tasks (PA); Action planning (PA); Behavioral experiments (PA)
Increasing your physical activity	Information about health consequences (PA) ; Demonstration of the behavior (PA); Self-monitoring of behavior (PA) ; Graded tasks (PA); Instruction on how to perform a behavior (PA)
Problem-solving	Demonstration of the behavior (BA); Graded tasks (BA/PA); Problem-solving (BA/PA); <u>Demonstration of the behavior (PA)</u> ; Social reward (BA)

Note. BCTs = behavior change techniques; BA = behavioral activation; PA = physical activity. Bold = BCT added, Underline = BCT removed.

For BCTs targeting physical activity, one instance of “self-monitoring,” one instance of “internal prompts/cues,” one instance of “framing/reframing,” and three instances of “demonstration of behavior” were removed from the intervention description as they were not adequately reflected in the content of eMotion as intended. One instance of “information about emotional consequences,” one instance of “information about health consequences,” one instance of “demonstration of behavior,” one instance of “graded tasks,” one instance of “self-monitoring of behavior,” and five instances of “monitoring of emotional consequences” were added to the intervention description as they were present in the intervention content and considered to be relevant (although not originally being present in the intervention description). These changes resulted in a complete agreement between the intervention specification and the independent coder for all BCTs targeting physical activity (AC1 = 1.0; see Table 3).

Discussion

Summary of Findings

The current study presents a novel methodology for checking intervention design fidelity, using independent coding to check the presence or absence of intended and unintended BCTs systematically. An application of the method to the development of the eMotion intervention identified a wide range of discrepancies relating to both the BA and physical activity components. The intervention, or its description, was then amended to remove all discrepancies and generate a more theoretically consistent intervention. Calculating interrater reliability enabled greater transparency of the coding process, offering an auditable approach and likely to increase rigor and attention to detail compared with simple peer-review. It also allowed us to prioritize BCTs which showed the poorest agreement. This approach goes substantially beyond proofread-

ing, resembling qualitative methods such as framework analysis, whereby text is coded against a prespecified framework of codes (Gale et al., 2013).

The proposed method can be used to check that intervention content fully represents its intended design in terms of both component BCTs and its theoretical integrity. Intervention designers could report the reliability statistic to help other researchers to better understand the content and theoretical underpinnings of the intervention and also to appraise any future evidence relating to the intervention. First, it could hasten systematic reviews exploring the effects of specific BCTs on various outcomes. Second, it provides a better indicator of BCTs than the coding of published intervention descriptions, which are often limited in detail (Dombrowski et al., 2007; Hoffmann et al., 2014). This method can also be used to help improve design fidelity at the development stage (as exemplified here). In any case, it is recommended that the assessment of design fidelity (quantitative or otherwise) be reported in articles describing behavioral interventions and added as an optional item to guidance on reporting interventions (e.g., the TiDier checklist (Hoffmann et al., 2014)).

Much work has already been conducted to specify a process for mapping BCTs to theoretical determinants (Cane et al., 2012) and describe BCTs using a common language (Abraham & Michie, 2008; Michie et al., 2013). Guidelines have also been developed for intervention reporting (Bellg et al., 2004; Borek et al., 2015; Borrelli, 2011; Borrelli et al., 2005; Hoffmann et al., 2014). The present study seeks to complement these existing approaches by presenting a method for assessing whether or not specific BCTs are sufficiently operationalized (i.e. present as intended) in behavior change interventions. The basic method could be extended to other complex behavioral interventions combining multiple behavioral targets (Taylor et al., 2014; Thompson et al., 2018; Ussher et al., 2015).

Relation to Other Literature

The low initial agreement for some specific BCTs (e.g., “demonstration of behavior” and “monitoring of emotional consequences”) may have been due to ambiguity around the definition of these techniques, rather than differences in interpretation between the coders. In a recent study, coders found it hard to distinguish between “demonstration of behavior” and “instruction on how to perform a behavior,” despite achieving modest reliability results (Abraham et al., 2015). The purpose of the discussion is to clarify any discrepancies that arise due to ambiguities in the definitions. For example, in the present study, the discussion improved the reliability of most BCTs. However, the discussion revealed that the intervention designer used the BCT “review behavioral goal” as a broader category to include “monitoring of emotional consequences.” Five instances of the BCT “monitoring of emotional consequences” were then added to the intervention description to reconcile this discrepancy, initially overlooked by the intervention designer.

Having clear and checked operationalizations of BCTs in the intervention content alongside the training materials (e.g., delivery protocol) can ensure that correct training, delivery, receipt, and enactment takes place. In a previous study investigating the delivery of a behavioral intervention for physical activity (Proactive), it was found that only 44% of prespecified BCTs were delivered by facilitators (Hardeman et al., 2008). Poor delivery may have been due to the facilitator’s own biases or insufficient training. However, it may also have been a result of inadequate operationalization of techniques in the manuals and protocols that informed the training and subsequent delivery. This highlights the many stages at

which intervention fidelity can fail (design, training, delivery, receipt, enactment; Borrelli, 2011). It is essential, therefore, to have robust methods for assessing fidelity at each step, so that process evaluations can determine what factors may have caused an intervention to fail (or succeed) and what aspects could be improved for future implementation or research.

The NIH BCC recommends a list of strategies to enhance (e.g., by providing precise information about intervention dose) and assess fidelity, which was applied in a review of 10 years of health behavior research (Borrelli et al., 2005). In this study, 80%, 22%, 35%, 49%, and 57% of the 342 included articles showed evidence of adhering to strategies relating to design, training, delivery, receipt, and enactment, respectively (Borrelli et al., 2005). Although 80% was reported for design (which appears high), only strategies to enhance (rather than assess) design fidelity were recommended (e.g., provide a manual; Borrelli, 2011). This study progresses this field of research by clearly articulating a method for assessing design fidelity.

Strengths and Limitations

The main strength of this study is the development and testing of a novel methodology for enhancing confidence that BCTs present in behavioral intervention descriptions are operationalized in intervention materials as intended. An independent coder was used with no involvement in the development of eMotion who was experienced in coding behavior change materials (Elliott et al., 2018). This reduces the possible bias of finding BCTs due to previous knowledge of the intervention development process. A robust, chance-corrected, statistical approach was also used to check interrater reliability.

However, several limitations should be acknowledged. First, each researcher interprets the data according to their subjective interpretation and prior experience, which could lead to disagreements regarding either the presence or absence of BCTs or regarding the BCT definitions themselves (interrater reliability for coding of BCTs is far from being perfectly reliable for most BCTs; Abraham & Michie, 2008). In the present study, the independent coder was versed in the behavior change taxonomy but not specifically with the BA model. Hence, a more substantial “panel of experts” might have been used to enhance: (a) assessment reliability and (b) the design fidelity of the eMotion intervention. The choice of coders is an important potential source of bias in using this methodology, so it needs careful consideration. Second, the independent coder could have been subjected to a priming effect (Tipper, 1985) as they worked their way through eMotion. The eMotion intervention is modular, with some modules looking very similar. Awareness of whether specific BCTs are supposed to occur could make the coder more likely to highlight them as present, even if they are not. Third, although coding was applied at a modular level, the “dose” of BCTs could have been mapped at smaller units of analysis (e.g., slide by slide). However, coding at smaller levels of analysis would be far more resource-intensive, and a compromise between rigor and pragmatism was adopted in the present study. Fourth, we treated disagreements on the presence or omission of BCTs with equal valence. This is because they could (if not rectified) lead to “false negative” or “false positive” findings, which are both threats to reliability and replicability. Nonetheless, in different types of behavior change interventions, authors may judge these to not be equally benign and thus adjust, statistically or otherwise, for this in their assessments of design fidelity. Fifth, while our approach worked well with a web-based intervention, it may not apply equally well to face-to-face interventions, which are typically less modular. However, this approach is intended to

qualitatively code the underlying protocols and training materials rather than the delivery itself. Coding for complex person-centered, individually tailored interventions is within the domain of delivery fidelity, for which different approaches are needed (e.g., the MITI or OPTION coding frameworks; Elwyn et al., 2001; Moyers et al., 2003).

Future Research and Implications for Practice

The eMotion intervention was informed by intervention mapping (Bartholomew et al., 2011), which could explain the fact that there were relatively few discrepancies after discussion. Further research could test this empirically by assessing and comparing design fidelity for interventions that did or did not use intervention mapping or other intervention design frameworks (Bartholomew et al., 2011; Michie et al., 2011). This approach could also be used in randomized controlled trials to confirm differentiation of intervention and comparator arms (i.e. how well an intervention differs from a comparator or control group concerning critical domains). A similar idea has been proposed in a recent protocol by Lorencatto et al. (2016).

Current taxonomies may lack the range of ideas needed to code interventions that target both behavior change and mental health. First, some techniques may be used differently to induce psychological as opposed to behavioral change. For example, in CBT, “behavioral experiments” is a technique used to help people test the validity of persistent negative beliefs or assumptions to improve their mood (Beck, 1979). In the field of health behavior change, it is used to “try out” a new behavior to see if it is enjoyable or acceptable (or to see whether perceived barriers can be overcome). Second, current taxonomies focus on the antecedents of behavior rather than the use of the behavior itself. For example, a core technique in behavioral activation involves supporting people to engage in achievable, self-selected behaviors to improve their mood (Lejuez et al., 2001). Future research should aim to develop a wider taxonomy of “change techniques” that include ways to influence not just behavior but also psychological change. A broader taxonomy could, therefore, cover different applications of similar techniques (as per the examples here), but also additional techniques that are not currently described. It could also provide examples of the use of specific techniques in both mental health and behavioral interventions. In developing taxonomies for public health interventions, it might also be useful to incorporate techniques for inducing social and systems-level change (Rutter et al., 2017).

Finally, future research could test whether increasing the number of raters or the quality of training impacts the reliability ascertained, for example, by comparing the results of having an intervention manual coded by two raters with the use of multiple raters.

Conclusion

The present study has developed and tested a new method for assessing design fidelity in behavioral interventions. As illustrated by its application to the eMotion intervention, this method can also be used to improve design fidelity at the intervention development stage. Developers of behavioral interventions (and other multicomponent, theory-based interventions) are encouraged to develop and refine this method and assess design fidelity in future interventions to ensure behavior change techniques are operationalized as intended.

References

Abraham, C., & Michie, S. (2008). A taxonomy of behavior change techniques used in interventions. *Health Psychology, 27*(3), 379–387. <https://doi.org/10.1037/0278-6133.27.3.379>

Abraham, C., Wood, C. E., Johnston, M., Francis, J., Hardeman, W., Richardson, M., & Michie, S. (2015). Reliability of identification of behavior change techniques in intervention descriptions. *Annals of Behavioral Medicine, 49*(6), 885–900. <https://doi.org/10.1007/s12160-015-9727-y>

Bandura, A. (1991). Social cognitive theory of self-regulation. *Organizational Behavior and Human Decision Processes, 50*(2), 248–287. [https://doi.org/10.1016/0749-5978\(91\)90022-L](https://doi.org/10.1016/0749-5978(91)90022-L)

Bartholomew, L. K., Parcel, G. S., Kok, G., Gottlieb, N. H., & Fernandez, M. E. (2011). *Planning health promotion programs: An intervention mapping approach* (3rd ed.). Wiley, Ltd.

Beck, A. (1979). *Cognitive therapy of depression*. Guilford Press.

Bellg, A. J., Borrelli, B., Resnick, B., Hecht, J., Minicucci, D. S., Ory, M., Ogedegbe, G., Orwig, D., Ernst, D., & Czajkowski, S. (2004). Enhancing treatment fidelity in health behavior change studies: Best practices and recommendations from the NIH behavior change consortium. *Health Psychology, 23*(5), 443–451. <https://doi.org/10.1037/0278-6133.23.5.443>

Borek, A. J., Abraham, C., Smith, J. R., Greaves, C. J., & Tarrant, M. (2015). A checklist to improve reporting of group-based behaviour-change interventions. *BMC Public Health, 15*(1), Article 963. <https://doi.org/10.1186/s12889-015-2300-6>

Borrelli, B. (2011). The assessment, monitoring, and enhancement of treatment fidelity in public health clinical trials. *Journal of Public Health Dentistry, 71*(s1), S52–S63. <https://doi.org/10.1111/j.1752-7325.2011.00233.x>

Borrelli, B., Sepinwall, D., Ernst, D., Bellg, A. J., Czajkowski, S., Breger, R., DeFrancesco, C., Levesque, C., Sharp, D. L., Ogedegbe, G., Resnick, B., & Orwig, D. (2005). A new tool to assess treatment fidelity and evaluation of treatment fidelity across 10 years of health behavior research. *Journal of Consulting and Clinical Psychology, 73*(5), 852–860. <https://doi.org/10.1037/0022-006X.73.5.852>

Cane, J., O'Connor, D., & Michie, S. (2012). Validation of the theoretical domains framework for use in behaviour change and implementation research. *Implementation Science, 7*(1), Article 37. <https://doi.org/10.1186/1748-5908-7-37>

Deci, E. L., & Ryan, R. M. (Eds.). (1985). Conceptualizations of intrinsic motivation and self-determination. *Intrinsic motivation and self-determination in human behavior* (pp. 11–40). Springer. https://doi.org/10.1007/978-1-4899-2271-7_2

Dombrowski, S. U., Sniehotta, F. F., Avenell, A., & Coyne, J. C. (2007). Current issues and future directions in psychology and health: Towards a cumulative science of behaviour change: Do current conduct and reporting of behavioural interventions fall short of best practice? *Psychology & Health, 22*(8), 869–874. <https://doi.org/10.1080/08870440701520973>

Elliott, L. R., White, M. P., Taylor, A. H., & Abraham, C. (2018). How do brochures encourage walking in natural environments in the U. K.? A content analysis. *Health Promotion International, 33*(2), 299–310. <https://doi.org/10.1093/heapro/daw083>

Elwyn, G., Edwards, A., Mowle, S., Wensing, M., Wilkinson, C., Kinnerley, P., & Grol, R. (2001). Measuring the involvement of patients in shared decision-making: A systematic review of instruments. *Patient Education and Counseling, 43*(1), 5–22. [https://doi.org/10.1016/S0738-3991\(00\)00149-X](https://doi.org/10.1016/S0738-3991(00)00149-X)

Farrand, P., Pentecost, C., Greaves, C., Taylor, R. S., Warren, F., Green, C., Hillsdon, M., Evans, P., Welsman, J., & Taylor, A. H. (2014). A written self-help intervention for depressed adults comparing behavioural activation combined with physical activity promotion with a self-help intervention based upon behavioural activation alone: Study protocol for a parallel group pilot randomised controlled trial (BacPac). *Trials, 15*(1), 1–12. <https://doi.org/10.1186/1745-6215-15-196>

Gale, N. K., Heath, G., Cameron, E., Rashid, S., & Redwood, S. (2013). Using the framework method for the analysis of qualitative data in

- multi-disciplinary health research. *BMC Medical Research Methodology*, 13(1), Article 117. <https://doi.org/10.1186/1471-2288-13-117>
- Gwet, K. (2002). Kappa statistic is not satisfactory for assessing the extent of agreement between raters. *Statistical Methods for Inter-Rater Reliability Assessment*, 1(6), 1–6.
- Hardeman, W., Michie, S., Fanshawe, T., Prevost, A. T., Mcloughlin, K., & Kinmonth, A. L. (2008). Fidelity of delivery of a physical activity intervention: Predictors and consequences. *Psychology & Health*, 23(1), 11–24. <https://doi.org/10.1080/08870440701615948>
- Hartmann-Boyce, J., Aveyard, P., Koshiaris, C., & Jebb, S. A. (2016). Development of tools to study personal weight control strategies: Ox-FAB taxonomy. *Obesity*, 24(2), 314–320. <https://doi.org/10.1002/oby.21341>
- Hoffmann, T. C., Glasziou, P. P., Boutron, I., Milne, R., Perera, R., Moher, D., Altman, D. G., Barbour, V., Macdonald, H., Johnston, M., Lamb, S. A., Dixon-Woods, M., McCulloch, P., Wyatt, J. C., Chan, A.-W., & Michie, S. (2014). Better reporting of interventions: Template for intervention description and replication (TIDieR) checklist and guide. *British Medical Journal*, 348, Article g1687. <https://doi.org/10.1136/bmj.g1687>
- Hopko, D. R., Lejuez, C. W., Ruggiero, K. J., & Eifert, G. H. (2003). Contemporary behavioral activation treatments for depression: Procedures, principles, and progress. *Clinical Psychology Review*, 23(5), 699–717. [https://doi.org/10.1016/S0272-7358\(03\)00070-9](https://doi.org/10.1016/S0272-7358(03)00070-9)
- Jacobson, N. S., Martell, C. R., & Dimidjian, S. (2001). Behavioral activation treatment for depression: Returning to contextual roots. *Clinical Psychology: Science and Practice*, 8(3), 255–270. <https://doi.org/10.1093/clipsy.8.3.255>
- Lambert, J. D., Greaves, C. J., Farrand, P., Cross, R., Haase, A. M., & Taylor, A. H. (2017). Assessment of fidelity in individual level behaviour change interventions promoting physical activity among adults: A systematic review. *BMC Public Health*, 17(1), 765. <https://doi.org/10.1186/s12889-017-4778-6>
- Lambert, J. D., Greaves, C. J., Farrand, P., Haase, A. M., & Taylor, A. H. (2017). Development of a web-based intervention (eMotion) based on behavioural activation to promote physical activity in people with depression. *Mental Health and Physical Activity*, 13, 120–136. <https://doi.org/10.1016/j.mhpa.2017.10.003>
- Lambert, J. D., Greaves, C. J., Farrand, P., Price, L., Haase, A. M., & Taylor, A. H. (2018). Web-based intervention using behavioral activation and physical activity for adults with depression (the eMotion study): Pilot randomized controlled trial. *Journal of Medical Internet Research*, 20(7), Article e10112. <https://doi.org/10.2196/10112>
- Landis, J. R., & Koch, G. G. (2008). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159–174. <https://doi.org/10.2307/2529310>
- Lejuez, C. W., Hopko, D. R., & Hopko, S. D. (2001). A brief behavioral activation treatment for depression. *Behavior Modification*, 25(2), 255–286. <https://doi.org/10.1177/0145445501252005>
- Lorenccatto, F., Gould, N. J., McIntyre, S. A., During, C., Bird, J., Walwyn, R., Cicero, R., Glidewell, L., Hartley, S., Stanworth, S. J., Foy, R., Grimshaw, J. M., Michie, S., & Francis, J. J. (2016). A multidimensional approach to assessing intervention fidelity in a process evaluation of audit and feedback interventions to reduce unnecessary blood transfusions: A study protocol. *Implementation Science*, 11(1), Article 163. <https://doi.org/10.1186/s13012-016-0528-x>
- Michie, S., Ashford, S., Sniehotta, F. F., Dombrowski, S. U., Bishop, A., & French, D. P. (2011). A refined taxonomy of behaviour change techniques to help people change their physical activity and healthy eating behaviours: The CALO-RE taxonomy. *Psychology & Health*, 26(11), 1479–1498. <https://doi.org/10.1080/08870446.2010.540664>
- Michie, S., Richardson, M., Johnston, M., Abraham, C., Francis, J., Hardeman, W., Eccles, M. P., Cane, J., & Wood, C. E. (2013). The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. *Annals of Behavioral Medicine*, 46(1), 81–95. <https://doi.org/10.1007/s12160-013-9486-6>
- Michie, S., van Stralen, M. M., & West, R. (2011). The behaviour change wheel: A new method for characterising and designing behaviour change interventions. *Implementation Science*, 6(1), Article 42. <https://doi.org/10.1186/1748-5908-6-42>
- Moyers, T., Martin, T., Catley, D., Harris, K. J., & Ahluwalia, J. S. (2003). Assessing the integrity of motivational interviewing interventions: Reliability of the motivational interviewing skills code. *Behavioural and Cognitive Psychotherapy*, 31(2), 177–184. <https://doi.org/10.1017/S1352465803002054>
- Richards, D. (2010). Behavioural Activation. In J. Bennett-Levy, D. Richards, P. Farrand, H. Christensen, K. Griffiths, D. Kavanagh, B. Klein, M. A. Lau, J. Proudfoot, L. Ritterband, J. White, & C. Williams (Eds.), *Oxford guide to low intensity CBT interventions* (pp. 141–150). Oxford University Press. <https://doi.org/10.1093/med:psych/9780199590117.003.0012>
- Rutter, H., Savona, N., Glonti, K., Bibby, J., Cummins, S., Finegood, D. T., Greaves, F., Harper, L., Hawe, P., Moore, L., Petticrew, M., Rehfuss, E., Shiell, A., Thomas, J., & White, M. (2017). The need for a complex systems model of evidence for public health. *Lancet*, 390(10112), 2602–2604. [https://doi.org/10.1016/S0140-6736\(17\)31267-9](https://doi.org/10.1016/S0140-6736(17)31267-9)
- Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, 55(1), 68–78. <https://doi.org/10.1037/0003-066X.55.1.68>
- Taylor, A. H., Thompson, T. P., Greaves, C. J., Taylor, R. S., Green, C., Warren, F. C., Kandiyali, R., Aveyard, P., Ayres, R., Campbell, J., Ussher, M., Michie, S., & West, R. (2014). A pilot randomised trial to assess the methods and procedures for evaluating the clinical effectiveness and cost-effectiveness of exercise assisted reduction then stop (EARS) among disadvantaged smokers. *Health Technology Assessment*, 18(4), 1–324. <https://doi.org/10.3310/hta18040>
- Thompson, T. P., Lambert, J. D., Greaves, C. J., & Taylor, A. H. (2018). Intervention delivery fidelity assessment of a counseling-based intervention for promoting smoking reduction and increasing physical activity. *Health Psychology*, 37(7), 627–637. <https://doi.org/10.1037/hea0000613>
- Tipper, S. P. (1985). The negative priming effect: Inhibitory priming by ignored objects. *The Quarterly Journal of Experimental Psychology*, 37(4), 571–590. <https://doi.org/10.1080/14640748508400920>
- Ussher, M., Lewis, S., Aveyard, P., Manyonda, I., West, R., Lewis, B., Marcus, B., Riaz, M., Taylor, A. H., Barton, P., Daley, A., Essex, H., Eslinger, D., & Coleman, T. (2015). The London Exercise And Pregnant smokers (LEAP) trial: A randomised controlled trial of physical activity for smoking cessation in pregnancy with an economic evaluation. *Health Technology Assessment*, 19(84), 1–136. <https://doi.org/10.3310/hta19840>
- Williams, C., Mcclay, C., Martinez, R., Morrison, J., Haig, C., Jones, R., & Farrand, P. (2016). Online CBT life skills programme for low mood and anxiety: Study protocol for a pilot randomized controlled trial. *Trials*, 17(1), Article 220. <https://doi.org/10.1186/s13063-016-1336-y>

Received October 1, 2019

Revision received June 26, 2020

Accepted September 11, 2020 ■