ORIGINAL ARTICLE

The Behavior Change Technique Taxonomy (v1) of 93 Hierarchically Clustered Techniques: Building an International Consensus for the Reporting of Behavior Change Interventions

Susan Michie, DPhil, CPsychol · Michelle Richardson, PhD · Marie Johnston, PhD, CPsychol · Charles Abraham, DPhil, CPsychol · Jill Francis, PhD, CPsychol · Wendy Hardeman, PhD · Martin P. Eccles, MD · James Cane, PhD · Caroline E. Wood, PhD

Abstract

Background CONSORT guidelines call for precise reporting of behavior change interventions: we need rigorous methods of characterizing active content of interventions with precision and specificity.

Objectives The objective of this study is to develop an extensive, consensually agreed hierarchically structured taxonomy of techniques [behavior change techniques (BCTs)] used in behavior change interventions.

Methods In a Delphi-type exercise, 14 experts rated labels and definitions of 124 BCTs from six published classification systems. Another 18 experts grouped BCTs according to similarity of active ingredients in an open-sort task. Inter-rater agreement amongst six researchers coding 85 intervention descriptions by BCTs was assessed.

Results This resulted in 93 BCTs clustered into 16 groups. Of the 26 BCTs occurring at least five times, 23 had adjusted kappas of 0.60 or above.

Conclusions "BCT taxonomy v1," an extensive taxonomy of 93 consensually agreed, distinct BCTs, offers a step change as a method for specifying interventions, but we anticipate further development and evaluation based on international, interdisciplinary consensus.

S. Michie (✉) · M. Johnston · C. E. Wood
Centre for Outcomes Research Effectiveness,
Research Department of Clinical, Educational and Health Psychology,
University College London, 1-19 Torrington Place,
London WC1E 7HB, UK
e-mail: s.michie@ucl.ac.uk

M. Richardson · C. Abraham
University of Exeter Medical School, University of Exeter,
St Luke's Campus, Heavitree Road,
Exeter EX1 2LU, UK

M. Johnston
Aberdeen Health Psychology Group, University of Aberdeen,
Institute of Applied Sciences,
College of Life Sciences and Medicine,
2nd floor, Health Sciences Building, Foresterhill,
Aberdeen AB25 2ZD, UK

J. Francis
Division of Health Services Research & Management,
City University London,
C332 Tait Building, City University London, Northampton Square,
London EC1V 0HB, UK

W. Hardeman
Behavioural Science Group, The Primary Care Unit,
University of Cambridge,
Cambridge Institute of Public Health, Robinson Way,
Cambridge CB2 0SR, UK

M. P. Eccles
Institute of Health and Society, Newcastle University,
21 Claremont Place,
Newcastle upon Tyne NE2 4AA, UK

J. Cane
School of Psychology, University of Kent,
Keynes College, Canterbury,
Kent CT2 7NP, UK

Introduction

Interventions to change behavior are typically complex, involving many interacting components [1]. This makes them challenging to replicate in research, to implement in practical applications, and to synthesize in systematic literature reviews. Complex interventions also present challenges for identifying the active, effective components within them. Replication, implementation, evidence synthesis, and identifying active components are all necessary if we are to better understand the effects and mechanisms of behavior change interventions and to accumulate knowledge to inform the development of more effective interventions. However, the poor description of interventions in research protocols and published reports presents a barrier to these essential scientific and translational processes [2, 3]. A well-specified intervention is essential before evaluation of effectiveness is worth undertaking: an under-specified intervention cannot be delivered with fidelity and, if evaluated, could not be replicated.

The CONSORT statement for randomized trials of "nonpharmacological" interventions recommends precise specification of trial processes, including some details of the delivery of interventions and "description of the different components of the interventions" [4]. As currently constituted, CONSORT gives no guidance as to what this description or components should be. Intervention components have been identified by Davidson et al. [3] as follows: who delivers the intervention, to whom, how often, for how long, in what format, in what context, and with what content. These are mainly procedures for delivery (often referred to as "mode" of delivery), except for the key intervention component, "content," i.e., the active ingredients that bring about behavior change (the "what" rather than the "how" of interventions).

The content, or active components of behavior change interventions, is often described in intervention protocols and published reports with different labels (e.g., "self-monitoring" may be labeled "daily diaries"); the same labels may be applied to different techniques (e.g., "behavioral counseling" may involve "educating patients" or "feedback, self-monitoring, and reinforcement" [5]). This may lead to uncertainty and confusion. For example, behavioral medicine researchers and practitioners have been found to report low confidence in their ability to replicate highly effective behavioral interventions for diabetes prevention [6]. The absence of standardized definitions and labels for intervention components means that systematic reviewers develop their own systems for classifying behavioral interventions and synthesizing study findings [7–10]. This proliferation of systems is likely to lead to duplication of effort and undermines the potential to accumulate

evidence across reviews. It also points to the urgent need for consensus. Consequently, the UK Medical Research Council's (MRC) guidance [1] for developing and evaluating complex interventions calls for improved methods of specifying and reporting intervention content in order to address the problems of lack of consistency and consensus.

A method recently developed for this purpose is the reliable characterization of interventions in terms of behavior change techniques (BCTs) [11]. By BCT, we mean an observable, replicable, and irreducible component of an intervention designed to alter or redirect causal processes that regulate behavior; that is, a technique is proposed to be an "active ingredient" (e.g., feedback, self-monitoring, and reinforcement) [12, 13]. BCTs can be used alone or in combination and in a variety of formats. Identifying the presence of BCTs in intervention descriptions included in systematic reviews and national datasets of outcomes has allowed the identification of BCTs associated with effective interventions. Effective BCTs have been identified for interventions to increase physical activity and healthy eating [14] and to support smoking cessation [10, 15], safe drinking [16], prevention of sexually transmitted infections [7, 17], and changing professional behavior [18].

Abraham and Michie [11] developed the first cross-behavior BCT taxonomy, building on previous intervention content analyses [7, 8]. These authors demonstrated reliability in identifying 22 BCTs and 4 BCT packages across 221 intervention descriptions in papers and manuals. This method has been widely used internationally to report interventions, synthesize evidence [14, 19–22], and design interventions [6, 23]. It has also enabled the specification of professional competences for delivering BCTs [24, 25] and as a basis for a national training program (see www.ncsct.co.uk). Guidance has also been developed for incorporating BCTs in text-based interventions [26].

Although the subsequent development of classification systems of defined and reliably identifiable BCTs has been accompanied by a progressive increase in their comprehensiveness and clarity, this work has been conducted by only a few research groups. For this method to maximize scientific advance, collaborative work is needed to develop agreed labels and definitions and reliable procedures for their identification and application across behaviors, disciplines, and countries.

Previous classification systems have either been in the form of an unstructured list or have been mapped to, or structured, according to categories (e.g., theory [7, 11] and theoretical mechanism [24, 25]) judged to be the most appropriate by the authors. In addition, they have mainly been developed for particular behavioral domains (e.g. physical activity, smoking, or safer sex). A comprehensive taxonomy will encompass a greater number of BCTs and therefore require structure to facilitate recall and access to the BCTs and thus to increase speed and accuracy of use. A true, i.e., hierarchically structured, taxonomy provides the advantage of making it more coherent to, and useable by, those applying it [27]. As the

number of identified BCTs has increased, so also has the need for such a structure, to improve the usability of the taxonomy.

Simple, reliable grouping structures have previously been used by three groups of authors. Dixon and Johnston [25] grouped BCTs according to "routes to behavior change," "motivation," "action," and "prompts/cues"; Michie et al. [15] grouped according to "function" in changing behavior, "motivation," "self-regulation capacity/skills," "adjuvant," and "interaction"; and Abraham et al. [17] grouped according to "change target," that is, knowledge, awareness of own behavior, attitudes, social norms, self-efficacy, intention formation, action control, behavioral maintenance, and change facilitators. However, there is a need for a basic method of grouping which does not depend on a theoretical structure. We therefore adopted an empirical approach to developing an international consensus of BCT groupings.

Potential Benefits

There are at least five potential benefits of developing a cross-domain, internationally supported taxonomy. First, it will promote the *accurate replication* of interventions (and control conditions in comparative effectiveness research), a key activity in accumulating scientific knowledge and investigating generalizability across behaviors, populations, and settings. Second, specifying intervention content by BCT will facilitate *faithful implementation* of interventions found to be effective. Third, *systematic reviews* will be able to use a reliable method for extracting information about intervention content, thus identifying and synthesizing discrete, replicable, potentially active ingredients (or combinations of ingredients) associated with effectiveness. Earlier BCT classification systems, combined with the statistical technique of meta-regression, have allowed reviewers to synthesize evidence from complex, heterogeneous interventions to identify effective component BCTs [6, 14, 21, 28, 29]. Fourth, *intervention development* will be able to draw on a comprehensive list of BCTs (rather than relying on the limited set that can be brought to mind) to design interventions, and it will be possible to report the intervention content in well-defined and detailed ways. Fifth, linking BCTs with theories of behavior change has allowed the investigation of possible *mechanisms of action* [14, 29, 30].

The work reported here represents the first stages of a program of work to develop an international taxonomic classification system for BCTs, building on previous work. The aims of the work reported in this paper are as follows:

1. To generate a taxonomy that

    (a) Comprises an extensive hierarchical classification of clearly labeled, well-defined BCTs with a consensus that they are proposed active components of behavior change interventions, that they are distinct (non-overlapping and non-redundant) and

    precise, and that they can be used with confidence to describe interventions;

    (b) Has a breadth of international and disciplinary agreement

2. To assess and report the reliability of using BCT labels and definitions to code intervention descriptions

Method

Design

The work involved three main tasks. The first involved the rigorous development of a list of distinct BCT labels and definitions, using Delphi methods, with feedback from a multidisciplinary International Advisory Board and members of the study team. The inter-rater reliability of coding intervention descriptions using the list of BCTs was then assessed in two rounds of reliability testing. The third task was the development of a hierarchical structure.

Participants

Participants were international behavior change experts (i.e., active in their field and engaged in investigating, designing, and/or delivering behavior change interventions) who had agreed to take part in one or more of the study phases, members of the International Advisory Board, and the study team (including a "lay" person). All Board members, as leaders in their field, were eligible to take part as a behavior change expert. However, in light of their advisory role commitments, members were not routinely approached for further participation unless it would help widen participation in terms of country, discipline, and behavioral expertise.

For the Delphi exercise, 19 international behavior change experts were invited to take part. Experts were identified from a range of scientific networks on the basis of breadth of knowledge of BCTs, experience of designing and/or delivering behavior change interventions, and of being able to complete the study task in the allotted time. Recruitment was by email, with an offer of an honorarium of £140 (approximately US$200) on completing the task. Of the 19 originally approached, 14 agreed to take part (response rate of 74 %). Ten participants were female, with an age range of 37 to 62 years (M=50.57; SD=7.74). Expert participants were from the UK (eight), Australia (two), Netherlands (two), Canada (one), and New Zealand (one). Eleven were psychologists (six health psychologists, one clinical psychologist, three clinical and health psychologists, and one educational psychologist), one a cognitive behavior therapist, and two had backgrounds in health sciences or community health. Eleven were active practitioners in their discipline. Eleven had research or professional doctorates,

and two had registered psychologist status. There was a wide range of experience of using BCTs, with all having used at least six BCTs, more than half having used more than 30 BCTs, and four having used more than 50 BCTs for intervention design, delivery, and training (see Electronic Supplementary Materials Table 1 for more information).

For the international feedback phase, 16 of the 30 International Advisory Board members (see: http://www.ucl.ac.uk/health-psychology/BCTtaxonomy/collaborators.php) took part in discussions to comment on a prototype BCT classification system. Advisory Board members were identified by the study team as being leaders in their field within the key domains of interest (e.g., types of health-related behaviors, major disease types, and disciplines such as behavioral medicine) following consultation of websites, journals, and scientific and professional organizations. Advisory Board members were from the USA, Canada, Australia, UK, the Netherlands, Finland, and Germany. Feedback was also provided by members of the study team, who had backgrounds in psychology and/or implementation science and a "lay" person with a BA (Hons) in English but no background in psychology or behavior change.

Five members of the UK study team conducted the first round of reliability testing and six the second round. Eighteen of 19 participants approached from the pool of experts in behavior change interventions completed the open-sort grouping task. Eight were women, and ten were men with an age range of 27 to 67 years ($M = 43.94$); 16 were from the UK, and two were from Australia.

Procedure

Participants recruited for the Delphi exercise and open-sort grouping task provided written consent and were assured that their responses would remain confidential. All participants were asked to provide demographic information (i.e., age, gender, and nationality). Delphi exercise participants were also asked to provide their professional background (i.e., qualifications, registrations, job title, and area of work) and how many BCTs they had used professionally in intervention design, face-to-face delivery, and training (reported in increments of 5 up to 50+).

A prototype classification system was developed by the study team based on all known published classifications of BCTs following a literature review [27] (Step 1). An online Delphi-type exercise (see Pill [31]) with two "rounds" was used for initial evaluation and development of the classification system. Participants worked independently and rated the prototype BCT labels and definitions on a series of questions designed to assess omission, overlap, and redundancy (Step 2). The results of Step 2 subsequently informed the development of an improved BCT list, which was sent to the Delphi participants for round 2. They were asked to rate BCTs for clarity, precision, distinctiveness, and confidence of use (Step 3). The resulting list of BCTs was then scrutinized by the Advisory Board, who submitted verbal and written feedback, and was assessed by the lay and expert members of the study team (Step 4). Following each of Steps 2, 3, and 4, the results were synthesized by SM and MJ in preparation for the next step. Using the developed BCT list, members of the study team coded published descriptions of interventions, and inter-rater agreement for the presence of each BCT was calculated (Step 5). An open-sort grouping task was then carried out to generate reliable and stable groupings and create a hierarchical structure within the taxonomy (Step 6).

Step 1: Developing the Prototype Classification System

The labels and definitions of distinct BCTs were extracted from six BCT classification systems identified by a literature search (relevant papers [11, 14–16, 25, 26]). For BCTs with two or more labels ($n = 24$) and/or definitions ($n = 37$), five study team members rated their preferred labels and definitions. Where there was complete or majority agreement, the preferred label and/or definition was retained. Where there was some, little, or no agreement, new labels and definitions were developed by synthesizing the existing labels and definitions across classification systems. Definition wording was modified to include active verbs and to be non-directional (i.e., applicable to both the adoption of a new wanted behavior and the removal of an unwanted behavior).

Step 2: Delphi Exercise First Round

Participants were provided with the study definition of a BCT [13], i.e., having the following characteristics: (a) aim to change behavior, (b) are proposed "active ingredients" of interventions, (c) are the smallest components compatible with retaining the proposed active ingredients, (d) can be used alone or in combination with other BCTs, (e) are observable and replicable, (f) can have a measurable effect on a specified behavior/s, and (g) may or may not have an established empirical evidence base. It was explained that BCTs could be delivered by someone else or self-delivered.

The BCTs (labels and definitions) from Step 1 were presented in a random order, and participants were asked five questions about each of them:

1. Does the definition contain what you would consider to be potentially active ingredients that could be tested empirically? Participants were asked to respond to this question using a five-point scale ("definitely no," "probably no," "not sure," "probably yes," and "definitely yes").
2. Please indicate whether you are satisfied that the BCT is conceptually unique or whether you consider that it is redundant or overlapping with other BCTs (with forced choice as to "whether it was conceptually unique, redundant, or overlapping").

3. If participants indicated that the BCT was "redundant," they were asked to state why they had come to this conclusion.
4. If they indicated that the BCT was "overlapping," they were asked to state (a) with which BCT(s) and (b) whether they can be separated ("yes" or "no").
5. If the BCTs were considered to be separate, participants were asked how the label or definition could be rephrased to reduce the amount of overlap or, if not separate, which label and which definition was better.

Participants were given an opportunity to make comments on the exercise and to detail any BCTs not included on the list. They were asked, "does the definition and/or label contain unnecessary characteristics and/or omitted characteristics?" This question item was open-ended. The exercise was designed to take 2 hour; follow-up reminders were sent to participants after 2 weeks, and all responses were submitted within 1 month of the initial request.

Frequencies, means, and/or modes of responses to questions (1) and (2) were considered for each BCT. Based on the distribution of responses, BCTs for which (a) more than a quarter of participants doubted that they contained active ingredients and/or (b) more than a third considered them to be overlapping or redundant were flagged as "requiring further consideration." These data, along with the responses to questions (3) to (5), guided the rewording of BCT labels and definitions and the identification of omitted BCTs. The BCTs for re-consideration and the newly identified BCTs were presented in the second Delphi exercise round.

Step 3: Delphi Exercise Second Round

The BCTs identified as requiring further consideration were presented; the rest of the BCTs were included for reference only, to assist judgments about distinctiveness. For each BCT, participants were asked three questions and asked to respond using a five-point scale ("definitely no," "probably no," "not sure," "probably yes," and "definitely yes"):

1. If you were asked to describe a behavior change intervention in terms of its component BCTs, would you think the following BCT was (a) clear, (b) precise, or (c) distinct?
2. Would you feel confident in using this BCT to describe the intervention?
3. Would you feel confident that two behavior change researchers or practitioners would agree in identifying this BCT?

If participants responded "probably no," "definitely no," or "not sure" to any question, they were asked to state their suggestions for improvement.

Frequencies, means, and/or modes were calculated for all questions for each BCT. BCTs for which more than a quarter

of participants responded "probably no" or "definitely no" or "not sure" to any question were flagged as needing to be given special attention. Using information on the distribution of ratings, the modal scores, and suggestions for improvement, SM and MJ amended the wording of definitions and labels. This included changes to make BCTs more distinct from each other where this had been identified as a problem and to standardize wording across BCTs. Where it was not obvious how to amend the BCT from the second round responses, other sources [32] were consulted for definitions of particular words or descriptions of BCTs.

Step 4: Feedback from the International Advisory Board

Sixteen of the 30 members of the Advisory Board took part in one of three 2-hour-long teleconferences to give advice to the study team, and the BCT list was refined based on their feedback.

Step 5: Reliability Testing Round 1

Five members of the study team coded 45 intervention descriptions. The descriptions were selected from *Implementation Science*, *BMC Public Health Services*, and *BMC Public Health* in 2009 and 2010 using quota sampling to ensure spread across preventive, illness management, and health professional behaviors. The study team then discussed reasons for discrepancies in round 1 and amended the BCT list as needed.

Step 6: Investigating Hierarchical Structure of the BCT List

An open-sort grouping task was delivered via an online computer program. Participants were asked to sort the developed list of BCTs into groups (up to a maximum of 24) of their choice and to label the groups. They were asked to "group together BCTs which have similar active ingredients, i.e., by the mechanism of change, NOT the mode of delivery." BCTs were presented to participants in a random order, and definitions for each BCT were made available.

For data analysis, a binary dissimilarity matrix containing all possible BCT×BCT combinations was produced for each participant, where a score of 1 indicated BCTs which were not sorted into the same group and a score of 0 indicated items which were sorted into the same group. Individual matrices were aggregated to produce a single dissimilarity matrix which could be used to identify the optimal grouping of BCTs using cluster analysis. Using hierarchical cluster analysis (HCA), the optimal number of groupings (2–20) were examined for suitability using measures of internal validity (Dunn's index) and stability (figure of merit, FOM) [33]. Bootstrap methods were used in conjunction with the HCA, whereby data were re-sampled 10,000 times, to identify which groupings were strongly supported by the data. The approximately unbiased (AU) p values yielded by this

method indicated the extent to which groupings were strongly supported by the data with higher AU values (e.g., 95 %) indicating stronger support for the grouping [34].

The words and phrases used in the labels given by participants were analyzed to identify any common themes and to help identify appropriate labels for the groupings. For each grouping, labels were created based on their content and, where applicable, based on the frequency of word labels given by participants. After the labels were assigned to relevant groupings, the fully labeled groups with the word frequency analysis were sent out to a subset of five of the original participants for refinement.

Step 7: Reliability Testing Round 2

An additional member of the study team was recruited for the second round of reliability testing. The team coded 40 intervention descriptions using the amended list. The six members each coded 9–14 intervention descriptions.

For both rounds, each intervention description was coded independently by two team members, and inter-rater agreement by BCT was assessed using kappas adjusted for prevalence and bias effects [35, 36]. Conventionally, a kappa of <0.60 is considered poor to fair agreement, 0.61–0.80 strong, and more than 0.80 near complete agreement [37]. The more frequent the BCTs, the greater the confidence that the kappa is a useful indicator of reliability of judging the BCT to be present. We therefore only report the kappa scores for BCTs which were observed at least five times by either coder in the 40 intervention descriptions.

Step 8: Feedback from Study Team Members

The BCT definitions were checked to ensure that they contained an active verb specifying the action required to deliver the intervention [38].The "lay" member of the study team (FR) read through the list to ensure syntactic consistency and general comprehensibility to those outside the field of behavioral science. Subsequently, the study team members made a final check of the resulting BCT labels and definitions.

Full details of the procedure are available in Electronic Supplementary Materials Table 2.

Results

The evolution of the taxonomy at the different steps of the procedure is summarized in Electronic Supplementary Materials Table 2.

Step 1: Developing the Prototype Classification System

Of the 124 BCTS in the prototype classification system, 32 were removed: five composite BCTs and 26 BCTs overlapping

with others were rated to have better definitions. One additional BCT was identified, given a label and definition informed by other sources and then added to the system. This produced a list of 94 BCTs.

Step 2: Delphi Exercise First Round

The means, modes, and frequencies of responses to the Delphi exercise first round questions are shown in Table 1. On the basis of these scores, 21 BCTs were judged to be "satisfactory" and 73 "requiring further consideration." Of the 73 reconsidered BCTs, four were removed, four were divided, and one BCT was added (see Electronic Supplementary Materials Table 2 for more details of changes at each step), giving 70 BCTs. During this process, one reason for overlap became evident: there was a hierarchical structure meaning that deleting overlapping BCTs would end up with only the superordinate BCT and a loss of specific variation (for example, adopting the higher order BCT "consequences" would have deleted "reward").

Step 3: Delphi Exercise Second Round

The means, modes, and frequencies of responses to the five Delphi exercise second round questions are shown in Table 2. On the basis of these scores, 38 BCTs were judged to be "satisfactory" and 32 "requiring further consideration." Of the reconsidered BCTs, seven labels and 35 definitions were amended, and seven BCTs were removed (see Electronic Supplementary Materials Table 2 for more details), giving 63 BCTs. Together with the 21 BCTs judged to be "satisfactory" in the first round, there were 84 BCTs at the end of the Delphi exercise. Some further standardization of wording across all BCTs was made by study team members (e.g., specifying "unwanted" or "wanted" behaviors rather than the more generic "target" behaviors and ensuring that all definitions included active verbs).

Step 4: Feedback from the International Advisory Board

The Advisory Board members made two general recommendations: first, to make the taxonomy more usable by empirically grouping the BCTs, and secondly, to consider publishing a sequence of versions of the taxonomy (with each version clearly labeled) that would achieve a balance between stability/usability and change/evolution. Feedback from members led to the addition of two and the removal of four BCTs. Further refinement of labels and definitions resulted in a list of 82 BCTs.

Step 5 and 7: Reliability Testing Round 1 and 2

Inter-rater agreement for BCTs is shown in Table 3. For the first round of reliability testing, 22 BCTs were observed five

Table 1 BCTs judged to be satisfactory and those requiring further consideration in Delphi exercise round 1 (step 2): means, modes, and frequencies of responses to questions

Question	Satisfactory (BCTs=21)			Reconsider (BCTs=73)		
	Range of means	Range of SDs	Frequency of modes, % (N)	Range of means	Range of SDs	Frequency of modes, % (N)
(1) Does the definition contain what you would consider to be potentially active ingredients that could be tested empirically?: 1 "definitely yes," 2 "probably yes," 3 "not sure," 4 "probably no," and 5 "definitely no"	1.21 (SD=0.47) to 1.93 (SD=0.92)	0.43 to 1.17	1=71 % (15) 2=29 % (6) 3=0 % (0) 4=0 % (0) 5=0 % (0)	1.29 (SD=0.47) to 4.79 (SD=0.43)	0.43 to 1.34	1=44 % (32) 2=51 % (37) 3=4 % (3) 4=1 % (1) 5=0 % (0)
(2) Please indicate whether you are satisfied that it is conceptually unique or whether you consider that it is redundant or overlapping with other BCTs (1 conceptually unique; 2 redundant; 3 overlapping)	N/A	N/A	1=100 % (21)	N/A	N/A	1=60 % (44) 2=0 % (0) 3=40 % (29)

or more times and therefore could be assessed. Adjusted kappa scores ranged from 0.38 to 0.85, with three scores below 0.60. Results from the first round of reliability testing led to the addition of five and the removal of one BCT resulting in a list of 86 BCTs.

For the second round of reliability testing, 15 BCTs were observed five or more times. Adjusted kappa scores ranged from 0.60 to 0.90. In all, 26 BCTs were tested for reliability, 23 of which achieved kappa scores of 0.60 or above and met our criteria of a BCT (see Table 3).

Step 6: Investigating Hierarchical Structure of the BCT List

Participants created an average of 15.11 groups (SD=6.11; range, 5–24 groups). Measures of internal validity indicated that the maximum internal validity Dunn index value (.57) was for a 16-cluster solution using hierarchical cluster analysis (see Fig. 1), with no increase in internal validity on subsequent cluster solutions (>16). Similarly, FOM values showed greater stability in the 16-cluster solution compared to the 2–15 cluster solutions, and there was negligible increase in stability over cluster solutions 17–20. Therefore, hierarchical clustering methods identified the 16-cluster solution as the optimal solution. The frequency of the words and phrases used in the labels given by participants is shown in Table 4. On the basis of participant responses, the groups were assigned the following labels (number of component BCTs in brackets): reinforcement (10), reward and threat (7), repetition and substitution (7), antecedents (4), associative learning (8), covert learning (3), consequences (6), feedback and monitoring (5), goals and planning (9), social support (3), comparison of behavior (3), self-belief (4), comparison of outcomes (3), identity (5), shaping knowledge (4), and regulation (4). Three of these labels were modified to facilitate comprehensibility across disciplines: "reinforcement" was changed to "scheduled consequences," and "associative learning" was changed to "associations." "Consequences" was then changed to "natural consequences" to distinguish it from "scheduled consequences."

The final results of the cluster analysis are shown in Table 5. Seven of the 16 clusters (clusters 3, 4, 5, 8, 10, 15, and 16) showed AU values greater than 95 %, indicating that these groupings were strongly supported by the data. Clusters 1, 2, 9, 12, and 13 had AU values between 90 % and 95 %, and clusters 6, 7, 11, and 14 had AU values less than 90 %; these were 73 %, 85 %, 83 %, and 86 %, respectively. The standard errors (SE) of AU values for all clusters were less than 0.009.

Step 8: Feedback from Study Team Members

Feedback from study team members led to the addition of three BCTs, the division of one BCT, and further refinement of labels and definitions. This resulted in a taxonomy of 93 BCTs.

Table 2 BCTs judged to be satisfactory and those requiring further consideration in Delphi exercise round 2 (step 3): means, modes, and frequencies of responses to questions

Question	Satisfactory (BCTs=38)			Reconsider (BCTs=32)		
	Range of means	Range of SDs	Frequency of modes, % (N)	Range of means	Range of SDs	Frequency of modes, % (N)
(1) If you were asked to describe a behavior change intervention in terms of its component BCTs, would you think the following BCT was						
(a) Clear?	1.07 (0.27) to 1.79 (1.05)	0.27 to 1.15	1=95 % (36) 2=5 % (2) 3=0 % (0) 4=0 % (0) 5=0 % (0)	1.13 (0.36) to 3.36 (1.15)	0.36 to 1.76	1=56 % (18) 2=38 % (12) 3=3 % (1) 4=3 % (1) 5=0 % (0)
(b) Precise?	1.07 (0.27) to 2.14 (1.29)	0.27 to 1.29	1=87 % (33) 2=13 % (5) 3=0 % (0) 4=0 % (0) 5=0 % (0)	1.14 (0.36) to 3.07 (1.33)	0.36 to 1.33	1=56 % (18) 2=41 % (13) 3=0 % (0) 4=3 % (1) 5=0 % (0)
(c) Distinct?	1.14 (0.36) to 2.14 (1.17)	0.36 to 1.38	1=89 % (34) 2=11 % (4) 3=0 % (0) 4=0 % (0) 5=0 % (0)	1.64 (0.93) to 3.21 (1.05)	0.78 to 1.66	1=59 % (19) 2=31 % (10) 3=9 % (3) 4=0 % (0) 5=0 % (0)
(2) Confidence in identifying BCT: would you feel confident in using this BCT to describe the intervention? (1 "definitely yes," 2 "probably yes," 3 "not sure," 4 "probably no," and 5 "definitely no")	1.21 (0.43) to 1.93 (1.07)	0.43 to 1.12	1=87 % (33) 2=13 % (5) 3=0 % (0) 4=0 % (0) 5=0 % (0)	1.14 (0.36) to 3.07 (1.00)	0.36 to 1.46	1=47 % (15) 2=47 % (15) 3=3 % (1) 4=3 % (1) 5=0 % (0)
(3) Confidence in others identifying BCT: would you feel confident that two behavior change researchers or practitioners would agree in identifying this BCT? (1 "definitely yes," 2 "probably yes," 3 "not sure," 4 "probably no," and 5 "definitely no")	1.21 (0.43) to 2.14 (0.95)	0.43 to 1.20	1=76 % (29) 2=24 % (9) 3=0 % (0) 4=0 % (0) 5=0 % (0)	1.36 (0.63) to 3.29 (1.27)	0.51 to 1.46	1=46 % (15) 2=41 % (13) 3=6 % (2) 4=6 % (2) 5=0 % (0)

Table 3 Inter-rater agreement for each BCT: adjusted kappas for two rounds of reliability testing

Round 1 and 2	Adjusted kappa	Round 1 only	Adjusted kappa	Round 2 only	Adjusted kappa
Pharmacological support [11.1]	0.87, 0.85	Non-specific encouragement[a]	0.82	Social comparison [6.2]	0.90
Self-monitoring of behavior [2.3]	0.82, 0.75	Review of outcome goal [1.7]	0.78	Material reward [10.2 and 10.10]	0.85
Restructuring of the physical environment [30]	0.82, 0.85	Discrepancy between current behavior [1.6]	0.73	Incentive [10.1]	0.80
Social support (practical) [3.2]	0.78, 0.70	Self-monitoring of outcome of behavior [2.4]	0.73	Monitoring outcome of behavior by others without feedback [2.5]	0.70
Behavioral practice/rehearsal [8.1]	0.78, 0.70	Health consequences [5.1]	0.69		
Problem solving/coping planning [1.2]	0.73, 0.75	Feedback on behavior [2.2]	0.69		
Persuasive argument [9.1]	0.73, 0.60	Action planning (including implementation intentions) [1.4]	0.64		
Review behavior goal(s) [1.5]	0.69, 0.75	Social support (general) [3.1]	0.60		
Goal setting (outcome) [1.3]	0.64, 0.85	Goal setting behavior [1.1]	0.56		
Prompts/cues [7.1]	0.42, 0.70	Tailored personalized message[a]	0.50		
Demonstration of the behavior [6.1]	0.87, 0.75	Instruction on how to perform the behavior [4.1]	0.38		

Some BCT labels differ as a result of the consensus exercises; number in brackets indicates related BCT in Electronic Supplementary Materials Table 3

Reliability shown for BCTs observed at least five times

[a] BCT not in Taxonomy v1

Discussion

An extensive hierarchically organized taxonomy of 93 distinct BCTs has been developed in a series of consensus exercises involving 54 experts in delivering and/or designing behavior change interventions. These experts were drawn from a variety of disciplines including psychology, behavioral medicine, and health promotion and from seven countries. The resulting BCTs therefore have relevance among experts from varied behavioral domains, disciplines, and countries and potential relevance to the populations from which they were drawn. The extent to which we can generalize our findings across behaviors, disciplines, and countries is an important question for future research. Building on a preliminary list generated from six published BCT classification systems, BCTs were added, divided, and removed, and their labels and definitions refined to capture the smallest components compatible with retaining the proposed active ingredients with the minimum of overlap. This resulted in 93 clearly defined, non-redundant BCTs, grouped into 16 clusters, for use in specifying the detailed content of a wide range of behavior change interventions. Of the 26 BCTs which could be assessed for inter-rater reliability, 23 had kappa scores of 0.60 or above and met our

definition of a reliable BCT. BCT Taxonomy v1 is the first consensus-based, cross-domain taxonomy of distinct BCTs to be published, with reliability data for the most frequent BCTs. The process of building a shareable consensus language and methodology is necessarily collaborative and will be an ongoing cumulative and iterative process, involving an international network of advisors and collaborators (see www.ucl.ac.uk/health-psychology/BCTtaxonomy/).

The methodologies used here represent an attempt to get a basic version of a taxonomy, a foundation on which to build future improvements. Like other classificatory systems, e.g., Linnaeus' classification of plants, or even systems based on consensus such as DSM [39] or ICD [40], we anticipate and plan to continue to work on improvements. There is no agreed methodology for this work, and there are limitations to the methods we have used. The purpose of the Delphi exercise was to develop a prototype taxonomy on which to build. It was one of a series of exercises adapted to develop the taxonomy. Our Delphi-type methods involved 14 individuals, an appropriate number for these methods [31], but a number that makes the choice of participants important. We attempted to ensure adequate coverage of behavior change experts (see Electronic Supplementary Materials Table 1). While we had some diversity of expertise,

Fig. 1 Results of hierarchical cluster analysis (step 6): dendrogram for 85 behavior change techniques (BCTs) partitioned across 16 clusters
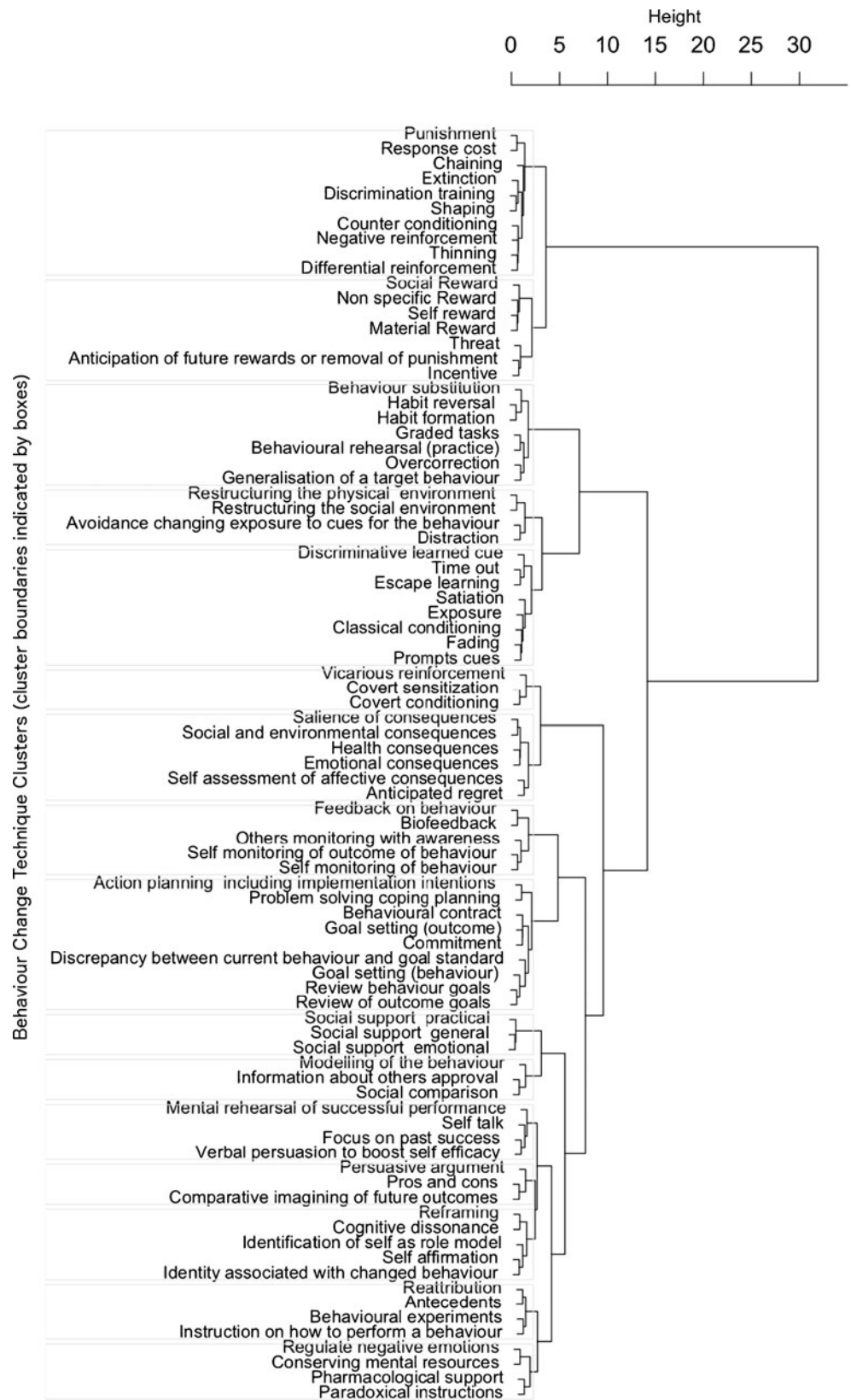
we acknowledge the predominance of European experts from a psychological background within our sample. At various stages, we made arbitrary decisions such as the cut-offs for amending BCT labels and descriptions and the minimum frequency of occurrence of BCTs for reporting reliability. In the absence of agreed standards for such

Table 4 Hierarchical structure labeling (step 6): frequency of words from labels given by participants (conjunctions removed)

Rank	Word/phrase	Frequency
1	Behavior/behavioral	24
2	Monitoring	10
2	Emotional/emotion/emotions/emotional regulation	10
3	Environment/environmental	9
4	Consequences	8
4	Self-efficacy	8
5	Feedback	7
5	Motivation	7
5	Reinforcement/reinforcing	7
6	Change	6
6	Conditioning	6
6	Identity	6
6	Planning	6
7	Antecedents	5
7	Goal-setting	5
7	Information	5
7	Learning	5
7	Manipulate	5
7	Other	5
7	Persuasion	5
7	Punishment	5
7	Self-regulation	5
7	Social	5
7	Social-support	5
8	Cognitions	4
8	Goals	4
8	Outcome expectancies	4
9	Resources	4
9	Restructuring	4
9	Reward	4
10	Commitment	3
10	Contingencies	3
10	Cues	3
10	Factors	3
10	Increase	3
10	Influence	3
10	Knowledge	3
10	Modeling	3
10	Physical	3
10	Practice	3
10	Prompts	3

Table shows most frequently used words ranked from 1 to 10

decisions, we were guided by the urgent need to develop an initial taxonomy which was fit for purpose and would therefore form a basis for future development. Our amendments of the BCT labels and definitions also depend on the expertise available, and we therefore based our amendments on a wide range of inputs: the data we collected from Delphi participants and coders, expert modification, international advice, and lay user improvements.

Compared with many of the previous "taxonomies" which are more accurately described as "nomenclatures," BCT Taxonomy v1 is not only a list of reliable, distinct BCTs but it also has a hierarchical structure. Such a structure has been shown to improve processing of large quantities of information by organizing it into "chunks" [41] that compensates for human memory limitations. In turn, this enables the user to attend to and recall the full range of BCTs available when reporting and designing interventions.

Use of an open-sort grouping task is an improvement on previous efforts to develop hierarchical structure in that it allowed for the individual groupings defined by participants to hold equal weight within the final solution, rather than using consensus approaches amongst small groups of participants. Second, the groupings increase the practical use of BCTs by aiding recall. Distinct sets of individual items with semantic similarity can be more easily recalled than a single list of individual items both in the short-term and in the long-term, particularly when the semantic category is cued [42–44]. The hierarchical structure shown in the dendrogram (see Fig. 1) gives an indication of the distance between clusters of BCTs and can be used as a starting point to compare the conceptual similarities and differences between BCTs. Sixteen clusters are too many for easy recall, and a higher-level cluster would be desirable. A simpler, higher-level structure of grouping BCTs has been used by Dixon and Johnston [25] and Michie et al. [24]. However, such a structure was not apparent in our data and points to the need for further research to refine the hierarchical structure of this taxonomy.

Other advantages of v1 are that it is relevant to a wide range of behavior rather than being restricted to a single behavioral domain; it provides examples of how the BCTs can be implemented and gives users enough detail to operationalize BCTs.

The results indicated that using the taxonomy to code intervention descriptions was generally reliable for those BCTs occurring relatively frequently. However, it was not possible to assess reliability for the 62 BCTs occurring with low frequency in the 85 coded intervention descriptions. Of the BCTs which could be assessed, three had kappa scores below 0.60 ["instruction on how to perform the behavior," "tailored personal message," and "goal setting (behavior)"]. Exploring reasons for discrepancies between coders may help to identify where further refinement of BCT labels/definitions and training may be required. For example, users reported difficulties distinguishing between "goal setting (behavior)" (i.e., when goal is unspecified, the most general BCT in the grouping should be coded) and other

Table 5 Results of hierarchical cluster analysis of behavior change techniques (step 6): grouping within the 16 cluster solution, approximately unbiased *p* values (AU), and standard errors

Cluster label and component BCTs	AU, % (SE)
(1) Scheduled consequences	91 (.004)
Punishment [14.2]	
Response cost [14.1]	
Chaining [14.5]	
Extinction [14.3]	
Discrimination training [14.6]	
Shaping [14.4]	
Negative reinforcement [14.10]	
Counter-conditioning [14.7]	
Thinning [14.9]	
Differential reinforcement [14.8]	
(2) Reward and threat	90 (.005)
Social reward [10.4]	
Material reward [10.2]	
Self-reward [10.9]	
Non-specific reward [10.3]	
Threat [10.11]	
Anticipation of future rewards or removal of punishment [14.10]	
Incentive [10.1]	
(3) Repetition and substitution	97 (.002)
Behavior substitution [8.2]	
Habit reversal [8.4]	
Habit formation [8.3]	
Graded tasks [8.7]	
Overcorrection [8.5]	
Behavioral rehearsal/practice [8.1]	
Generalization of a target behavior [8.6]	
(4) Antecedents	96 (.002)
Restructuring the physical environment [12.1]	
Restructuring the social environment [12.2]	
Avoidance/changing exposure to cues for the behavior [12.3]	
Distraction [12.4]	
(5) Associations	97 (.002)
Discriminative (learned) cue [7.2]	
Time out [7.4]	
Escape learning [7.5]	
Satiation [7.6]	
Exposure [7.7]	
Classical conditioning [7.8]	
Fading [7.3]	
Prompts/cues [7.1]	
(6) Covert learning	73 (.008)
Vicarious reinforcement [16.3]	
Covert sensitization [16.1]	
Covert conditioning [16.2]	
(7) Natural consequences	85 (.006)

Table 5 (continued)

Cluster label and component BCTs	AU, % (SE)
Health consequences [5.1]	
Social and environmental consequences [5.3]	
Salience of consequences [5.2]	
Emotional consequences [5.6]	
Self-assessment of affective consequences [5.4]	
Anticipated regret [5.5]	
(8) Feedback and monitoring	97 (.002)
Feedback on behavior [2.2]	
Biofeedback [2.6]	
Other(s) monitoring with awareness [2.1 and 2.5]	
Self-monitoring of outcome of behavior [2.4]	
Self-monitoring of behavior [2.3]	
(9) Goals and planning	90 (.002)
Action planning (including implementation intentions) [1.4]	
Problem solving/coping planning [1.2]	
Commitment [1.9]	
Goal setting (outcome) [1.3]	
Behavioral contract [1.8]	
Discrepancy between current behavior and goal standard [1.6]	
Goal setting (behavior) [1.1]	
Review behavior goal(s) [1.5]	
Review of outcome goal(s) [1.7]	
(10) Social support	100 (.001)
Social support (practical) [3.2]	
Social support (general) [3.1]	
Social support (emotional) [3.3]	
(11) Comparison of behavior	83 (.006)
Modeling of the behavior [6.1]	
Information about others' approval [6.3]	
Social comparison [6.2]	
(12) Self-belief	92 (.005)
Mental rehearsal of successful performance [15.2]	
Self-talk [15.4]	
Focus on past success [15.3]	
Verbal persuasion to boost self-efficacy [15.1]	
(13) Comparison of outcomes	90 (.005)
Persuasive argument [9.1]	
Pros and cons [9.2]	
Comparative imagining of future outcomes [9.3]	
(14) Identity	86 (.006)
Identification of self as role model [13.1]	
Self-affirmation [13.4]	
Identity associated with changed behavior [13.5]	
Reframing [13.2]	
Cognitive dissonance [13.3]	
(15) Shaping knowledge	95 (.003)
Reattribution [4.3]	
Antecedents [4.2]	

Table 5 (continued)

Cluster label and component BCTs	AU, % (SE)
Behavioral experiments [4.4]	
Instruction on how to perform a behavior [4.1]	
(16) Regulation	98 (.001)
Regulate negative emotions [11.2]	
Conserving mental resources [11.3]	
Pharmacological support [11.1]	
Paradoxical instructions [11.4]	

Some BCT labels differ as a result of the consensus exercises; number in brackets indicates related BCT in Electronic Supplementary Materials Table 3

AU adjusted unbiased *p* value, *SE* standard error of AU

goal-related BCTs, and between "instruction on how to perform the behavior" and "demonstration of the behavior." In considering reasons for discrepancies, we agreed that "tailored personal message" was a mode of delivery rather than a BCT and therefore removed it from the taxonomy. Since high reliability depends on both the content of the taxonomy and the training of the user to use it, we are currently evaluating methods of BCT user training and conducting more detailed analyses of reliability of application of the v1 classification system.

Future Developments

This is a fast-moving field: the first reliable taxonomy of BCTs was published only 4 years before the current one [11]; while widely cited and influential, this "taxonomy" included only 22 BCTs and 4 BCT packages so limiting the intervention content that could be classified. We anticipate that further refinement and development of BCT Taxonomy v1 will occur as a result of its use and feedback from primary researchers, systematic reviewers, and practitioners (e.g., the BCT, "increase positive emotions" appended as a footnote to Electronic Supplementary Materials Table 3 has been identified and will be included in future revisions of the taxonomy). In order to continue the development of the taxonomy and to further improve the accuracy and reliability of its use, training courses and workshops involving researchers and practitioners from five countries, with varying scientific and professional backgrounds and level of expertise, are being coordinated internationally. This will facilitate the comparison of reliability across different populations (e.g., disciplinary background, behavior, and continent). A web-based users resource, including the most recent version of the taxonomy, guidance on its use, and a discussion board for questions, comments, and feedback, has been developed to facilitate collaboration and synthesis of feedback (see www.ucl.ac.uk/health-psychology/BCTtaxonomy/).

Research is needed to link BCTs to theories of behavior change, for both designing and evaluating theory-based interventions. Preliminary attempts have been made to link BCTs to domains of theoretical constructs [17, 45], and this is part of an ongoing program of research. Guidance on developing interventions informed by considering theoretical determinants of behavior can be found in Kok et al. [46] and used in combination with the taxonomy. Work has also begun to link BCTs to a framework of behavior change interventions designed for use by policymakers, organizational change consultants, and systems scientists [47]. While some of the BCTs such as those dealing with incentives or environmental changes might be used in large-scale interventions, including health policy interventions, many would only be appropriate for smaller scale, personally delivered interventions. The current BCT Taxonomy v1 is a methodological tool for specifying intervention content and does not, itself, make links with theory.

The aim is to produce a consensual "core" BCT Taxonomy that may be extended and/or modified according to context, e.g., target behavior, country, specific setting. The BCT Taxonomy project will encourage authors to report how they have amended the core taxonomy so that other researchers can identify the links between the version used and the core taxonomy. Future work that increases the diversity of expertise and the geographical and cultural contexts in which BCTs are used would help to elucidate the extent to which BCT Taxonomy v1 is relevant across contexts, countries, and cultures and the extent to which specific adaptations will be needed. To date, the taxonomy and coded interventions have predominantly focused on interventions delivered to the individuals whose behavior change is targeted. Further work needs to be done to extend it to the BCTs relevant to community and population-level interventions [47].

BCT Taxonomy v1 thus lays the foundation for the reliable and systematic specification of behavior change interventions. This significantly increases the possibility of identifying the active ingredients within interventions components and the conditions under which they are effective, and of replicating and implementing effective interventions, thus advancing the science of behavior change. Historically, it has often been concluded that how BCTs are delivered may have as great or larger impact on outcomes as the techniques themselves [48]. Dimensions of behavior change interventions other than content, such as mode and context of delivery [5], and competence of those delivering the intervention [24, 25] would thus also benefit from being specified by detailed taxonomies. Elucidation of how content, mode, and context of delivery interact in their impact on outcomes is a key research goal for the field of behavioral science.

In summary, the work reported in this paper is foundational for our long-term goals of developing a comprehensive, hierarchical, reliable, and generalizable BCT Taxonomy as a method for specifying, evaluating, and implementing behavior change interventions that can be applied to many different types of intervention, including organizational and community interventions, and that has multidisciplinary and international acceptance and use. The work reported here is a step toward the objective of developing agreed methods that permit and facilitate the aims of CONSORT and UK MRC guidance of precise reporting of complex behavioral interventions. The next steps underway are to test the reliability and usability of BCT Taxonomy v1 across different behaviors and populations and to set up a system for its continuous development guided by an international, multidisciplinary team.

Conflicts of Interest The authors have no conflicts of interest to disclose.

References

1. Craig P, Dieppe P, Macintyre S, et al. Developing and evaluating complex interventions: The new Medical Research Council guidance. *BMJ*. 2008:337.
2. Michie S, Fixsen D, Grimshaw JM, Eccles MP. Specifying and reporting complex behavior change interventions: The need for a scientific method. *Implement Sci*. 2009;40:1-6.
3. Davidson KW, Goldstein M, Kaplan RM, et al. Evidence-based behavioral medicine: What is it and how do we achieve it? *Ann Behav Med*. 2003;26:161-171.
4. Boutron I, Moher D, Altman DG, et al. Extending the CONSORT statement to randomized trials of non-pharmacologic treatment: Explanation and elaboration. *Ann Intern Med*. 2008;148:295-309.
5. Michie S, Johnston M, Francis J, Hardeman W, Eccles M. From theory to intervention: Mapping theoretically derived behavioral determinants to behavior change techniques. *Appl Psychol* 2008;57:660-680.
6. Michie S, Hardeman W, Fanshawe T, Provost TA. Investigating theoretical explanations for behavior change: The case study of ProActive. *Psychol Health*. 2008;23:25-39.
7. Albarracin D, Gillette J, Earl AN, Glasman LR. A test of major assumptions about behavior change: A comprehensive look at the effects of passive and active HIV-prevention interventions since the beginning of the epidemic. *Psychol Bull*. 2005;131:856-897.
8. Hardeman W, Griffin S, Johnston M, Kinmonth AL, Wareham NJ. Interventions to prevent weight gain: A systematic review of psychological models and behavior change methods. *Int J Obesity*. 2000;24:131-143.
9. Mischel W. *Presidential address*. Washington: Association for Psychological Science Annual Convention; 2012.
10. West R, Walia A, Hyder N, Shahab L, Michie S. Behavior change techniques used by the English Stop Smoking Services and their associations with short-term quit outcomes. *Nicotine Tob Res*. 2010;12:742-747.
11. Abraham C, Michie S. A taxonomy of behavior change techniques used in interventions. *Health Psychol*. 2008;27:379-387.
12. Michie S, Abraham C, Eccles MP, et al. Strengthening evaluation and implementation by specifying components of behavior change interventions: A study protocol. *Implement Sci*. 2011;6:10.
13. Michie S, Johnston M. Behavior change techniques. In: Gellman MD, Turner JR, eds. *Encyclopedia of behavioral medicine*. New York: Springer; 2011.
14. Michie S, Abraham C, Whittington C, McAteer J, Gupta S. Effective techniques in healthy eating and physical activity interventions: A meta-regression. *Health Psychol* 2009;28:690-701.
15. Michie S, Hyder N, Walia A, West R. Development of a taxonomy of behavior change techniques used in individual behavioral support for smoking cessation. *Addict Behav*. 2011;36:315-319.
16. Michie S, Whittington C, Hamoudi Z, et al. Identification of behavior change techniques to reduce excessive alcohol consumption. *Addiction*. 2012;107:1431-1440.
17. Abraham C, Good A, Warren MR, Huedo-Medina T, Johnson B. Developing and testing a SHARP taxonomy of behavior change techniques included in condom promotion interventions. *Psychol Health*. 2011;26(Supplement 2):299.
18. Ivers N, Jamtvedt G, Flottorp S, Young JM, et al. Audit and feedback: Effects on professional practice and patient outcomes. *Cochrane Database Syst Rev*. 2012; (6): CD000259.
19. Araujo-Soares V, MacIntyre T, MacLennan G, Sniehotta FF. Development and exploratory cluster-randomized opportunistic trial of a theory-based intervention to enhance physical activity among adolescents. *Psychol Health* 2009;24:805-822.
20. Gardner B, Whittington C, McAteer J, Eccles MP, Michie S. Using theory to synthesize evidence from behavior change interventions: The example of audit and feedback. *Soc Sci Med*. 2010;70:1618-1625.
21. Michie S, Jochelson K, Markham WA, Bridle C. Low-income groups and behavior change interventions: A review of intervention content, effectiveness and theoretical frameworks. *J Epidemiol Community Health* 2009;63:610-622.
22. Quinn F. *On integrating biomedical and behavioral approaches to activity limitation with chronic pain: Testing integrated models between and within persons*. Aberdeen: University of Aberdeen; 2010.
23. Cahill K, Moher M, Lancaster T. Workplace interventions for smoking cessation. *Cochrane Database Syst Rev*. 2008; (4): CD003440.
24. Michie S, Churchill S, West R. Identifying evidence-based competences required to deliver behavioral support for smoking cessation. *Ann Behav Med*. 2011;41:59-70.
25. Dixon D, Johnston M. *Health behavior change competency framework: Competences to deliver interventions to change lifestyle behaviors that affect health*. Edinburgh: Scottish Government; 2012.
26. Abraham C. Mapping change mechanisms and behaviour change techniques: A systematic approach to promoting behaviour change through text. In: Abraham C, Kools M, eds. *Writing Health Communication: An Evidence-Based Guide for Professionals*. London: SAGE Publications; 2011.
27. Stavri Z, Michie S. Classification systems in behavioral science: Current systems and lessons from the natural, medical and social sciences. *Health Psychol Rev*. 2012;6:113-140.
28. de Bruin M, Viechtbauer W, Hospers HJ, Schaalma HP, Kok G. Standard care quality determines treatment outcomes in control groups of HAART-adherence intervention studies: Implications for the interpretation and comparison of intervention effects. *Health Psychol*. 2009;28:668-674.
29. Dombrowski SU, Sniehotta FF, Avenell A, et al. Identifying active ingredients in complex behavioral interventions for obese adults with obesity-related co-morbidities or additional risk factors for co-morbidities: A systematic review. *Health Psychol Rev* 2012;6:7-32.

30. Michie S, Johnston M. Theories and techniques of behavior change: Developing a cumulative science of behavior change. *Health Psychol Rev* 2012;6:1-6.

31. Pill J. The Delphi method: Substance context, a critique and the annotated bibliography. *Socioecon Planning Sci*. 1991;5:57-71.

32. Vandenbos GR. *APA dictionary of psychology*. Washington, DC: American Psychological Association; 2006.

33. Brock G, Pihur V, Datta S, Datta S. Package 'clvalid': Validation of clustering results. *J Statistical Software*. 2008;25:1-22.

34. Suzuki R, Shimodaira H. Pvclust: An R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics*. 2006;22:1540-1542.

35. Byrt T, Bishop J, Carlin JB. Bias, prevalence and kappa. *J Clin Epidemiol*. 1993;46:423-429.

36. Lantz CA, Nebenzahl E. Behavior and interpretation of the kappa statistic: Resolution of the two paradoxes. *J Clin Epidemiol*. 1996;49:431-434.

37. Landis JR, Koch GG. Measurement of observer agreement for categorical data. *Biometrics*. 1977;33:159-174.

38. Michie S, Johnston M. Changing clinical behavior by making guidelines specific. *BMJ*. 2004;328:343-345.

39. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. 4th ed. Text Revision. Washington, DC: American Psychiatric Association; 2000.

40. World Health Organisation. *ICD-10 international statistical classification of diseases and related health problems*. Geneva, Switzerland: Illu; 1992.

41. Miller GA. *The magical number seven, plus or minus two: Some limits on our capacity for processing information. Essential Sources in the Scientific Study of Consciousness*. Cambridge: A Bradford Book; 2003:357-372.

42. Baddeley A. Short-term memory for word sequences as a function of acoustic, semantic and formal similarity. *Q J Exp Psychol*. 1966;18:362-365.

43. Polyn SM, Erlikhman G, Kahana MJ. Semantic cuing and the scale insensitivity of recency and contiguity. *J Exp Psychol Learn Mem Cogn*. 2011;37:766-775.

44. Tulving E, Pearlsto Z. Availability versus accessibility of information in memory for words. *J Verb Learn Verb Behav*. 1966;5:381-391.

45. Michie S, Johnston M, Abraham C, et al. Making psychological theory useful for implementing evidence based practice: A consensus approach. *Qual Saf Health Care*. 2005;14:26-33.

46. Kok G, Schaalma H, Ruiter RAC, Van Empelen P, Brug J. Intervention mapping: A protocol for applying health psychology theory to prevention programmes. *J Health Psychol*. 2004;9:85-98.

47. Michie S, van Stralen MM, West R. The behavior change wheel: A new method for characterising and designing behavior change interventions. *Implement Sci*. 2011;6:42.

48. Kolehmainen N, Francis JJ. Specifying content and mechanisms of change in interventions to change professionals' practice: An illustration from the Good Goals study in occupational therapy. *Implement Sci*. 2012;7:100.