PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Gamification for health promotion: systematic review of behaviour change techniques in smartphone apps
AUTHORS	Edwards, Elizabeth; Lumsden, Jim; Rivas, Carol; Steed, Liz; Edwards, Lindsey; Thiyagarajan, Arun; Sohanpal, Ratna; Caton, Hope; Griffiths, Chris; Munafo, Marcus; Taylor, Stephanie; Walton, Robert

VERSION 1 - REVIEW

REVIEWER	Geraldine Naughton Australian Catholic University
REVIEW RETURNED	13-Jun-2016

GENERAL COMMENTS	This timely review of game-related smart phone apps for health promotion will generate much interest. The links to health related behavioural change is particularly salient and offers much relevance to health practitioners. The title is however, somewhat of a misnomer. Systematic reviews typically involve literature reviews. In this manuscript each smart phone app has been reviewed somewhat rigorously for behaviour change technology; but it is a review of the apps, not of the literature. In a sense, it is an original research project rather than a review, because each "app" was a participant. It is difficult to understand how it was accepted to be registered through Prospero. As the authors correctly identified, the impact of the apps involving behavior change challenges and game-based activities are yet to be known. However the topic is contemporary and provides a solid platform for future investigations in this area. The magnitude of technology available in health is well described. The rationale argues strongly for a scientific interpretation of what is available in health apps. The rigor of the Michie et al taxonomy provides a credible basis for categorizing gamification on health apps. Further, sound definitions of gaming characteristics were considered and applied within a contemporary timeframe. Results were clearly presented and described. Reporting could be improved by adding interquartile ranges (rather than whole ranges) to median values. Reporting percentages in whole numbers rather than to inconsistent numbers of decimal points may improve readability and recall of important findings. The discussion was well considered. The major limitation of advancing the understanding of the availability and nature of these apps rather than the effectiveness is well acknowledged The manuscript is strongly supported by relevant and clear supplementary material.

REVIEWER	Fabiana Lorencatto
	Centre for Health Services Research, City University London
REVIEW RETURNED	14-Jun-2016

GENERAL COMMENTS	Thank you for the opportunity to review this manuscript, which presents the results of a systematic review aiming to describe the behaviour change techniques commonly featured in smartphone apps that aim to change health related behaviours using gamification strategies.
	Overall, I am enthusiastic about the topic being addressed and methodology used in this manuscript.
	However, I have some concerns at this stage regarding the manuscript, particularly the background and methods section, which I feel warrant expansion and further clarification.
	Introduction
	1. Overall, the manuscript is clearly and succinctly written. However, I feel the introduction is lacking essential information.
	a. First, the manuscript as a whole assumes familiarity with behaviour change techniques (BCTs) and taxonomies of BCTs. These concepts are very loosely defined and described throughout. I am very familiar with the BCT literature and methodology and was thus able to interpret the methods and findings presented in this manuscript, but I am concerned that a reader less familiar with the relatively new BCT literature and methodology would be less able to readily interpret the methods and results presented. I would thus suggest defining BCTs and BCT taxonomies in the introduction, particularly the difference between behaviour change categories and BCTs. More importantly, I suggest highlighting and citing the rapidly increasing number of reviews similar to the current review, which aim to identify and characterise the content of interventions in terms of component BCTs, using taxonomies with established reliability. One such review is cited in the discussion (i.e. Dombrowski et al), but would be helpful to incorporate this in the introduction to help clarify the rationale/context for this study.
	b. On a related note, I was surprised to see that none of the literature that has specifically aimed to characterise the content of smartphone apps has been cited in the introduction. For instance, a highly relevant example is the work of Crane et al. (2015) who applied BCT taxonomies to code the content of alcohol reduction smartphone apps [i.e. Crane, D., Garnett, C., Brown, J., West, R., & Michie, S. (2015). Behavior change techniques in popular alcohol reduction apps: content analysis.Journal of medical Internet research, 17(5)]. Crane et al. also cite a number of reviews of the characteristics of smartphone apps targeting a range of health behaviours (refs 14-25 in Crane et al). There are plenty of such reviews in the current literature (a few further examples below). It would be helpful to summarise this existing literature in introduction to better contextualize the present review.
	• Direito, A., Dale, L. P., Shields, E., Dobson, R., Whittaker, R., & Maddison, R. (2014). Do physical activity and dietary smartphone applications incorporate evidence-based behaviour change techniques?. BMC Public Health, 14(1), 1.
	• Conroy, D. E., Yang, C. H., & Maher, J. P. (2014). Behavior

change techniques in top-ranked mobile apps for physical activity. American journal of preventive medicine, 46(6), 649-652.
• Hoeppner, B. B., Hoeppner, S. S., Seaboyer, L., Schick, M. R., Wu, G. W., Bergman, B. G., & Kelly, J. F. (2015). How smart are smartphone apps for smoking cessation? A content analysis. Nicotine & Tobacco Research, ntv117.

2. Throughout the manuscript there is reference to 'theoretically consistent BCTs' (e.g. p. 4 line 25; p. 13 line 49, p. 12 line 17)- but no actual mention or discussion as to which theories these BCTs are consistent with. Could this please be clarified/ made more explicit? Such as in the Michie 2009 paper linking BCTs to Control Theory (reference 9 in the present manuscript)?

Methods

3. The search was conducted 1 year ago (i.e. June 2015). Given smartphone apps are a rapidly expanding and developing industry, this raises the question as to whether the search is now out of date and requires updating.

4. More detail is needed on the methods for BCT coding, particularly to enable a reader less familiar with this methodology to understand the analyses conducted. For instance, what content in the app was coded? Was it just text or images/video/other multimedia content as well? How were BCTs coded- in terms of a global present or absence rating (i.e. 1/0)? Did you also consider the frequency with which BCTs were present (i.e. BCTs delivered more than once within an app)? Was mode of BCT delivery considered/recorded? Was inter-rater reliability formally assessed (i.e. Kappa)? I am also not entirely clear what is meant by 'app content was assessed against criteria derived from a standard taxonomy of BCTs'- which criteria are these? It might be helpful to provide an example of app content coded into BCTs (e.g. a screen shot of an app page, and call-out boxes highlighting what content would be coded into what BCT; or some exemplary text/quote excerpt and associated BCTs).

5. A particular strength of this review is that combinations of BCTs were considered. An enduring limitation of many existing BCT reviews is that they only tend to look at individual BCTs and not combination of BCTs. I therefore commend the authors for addressing this gap. As such, could the authors provide more detail as to how combinations of BCTs were defined/assessed? As this will be of particular methodological interest to researchers conducting similar research.

Results

6. Overall results are clear. Minor suggestion- P. 9 lines 28-30, 34-36; consider moving the findings on correlation between number of BCTs and customer ratings/price later in the results section. These findings are currently presented before the results on BCTs identified, so the sequence in which the results is slightly confusing.

Results/Discussion

7. Results/discussion currently focused on the high frequency BCTs. This is, of course, of interest, but equally of interest are the low frequency BCTs- how do these align with relevant 'theory'- would we have expected any observed low frequency BCTs to have been

	included in apps? Are there any missed opportunities (Seeing as, if I understood correctly, 75% of the 93 BCTs were not identified in any apps)? Discussion 8. Overall, the discussion provides good consideration of the implications of the present findings, for both research and practice. However, as per comment 1b, the discussion would be strengthened by comparing the present findings to those of other reviews aiming to characterise the characteristics of smartphone apps (e.g. Crane et al. 2015, plus associated references). Currently comparisons are drawn with reviews/meta-analyses of internet based interventions, and other reviews that have identified similar combination of BCTs as the present review. Although this literature is relevant, even more focused/relevant comparisons could be drawn by comparing present findings with that of other reviews focused on smartphone apps specifically.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
Reviewer Name: Geraldine Naughton
Institution and Country: Australian Catholic University, Australia
Competing Interests: None declared

This timely review of game-related smartphone apps for health promotion will generate much interest. The links to health related behavioural change is particularly salient and offers much relevance to health practitioners.

The title is however, somewhat of a misnomer. Systematic reviews typically involve literature reviews. In this manuscript each smart phone app has been reviewed somewhat rigorously for behaviour change technology; but it is a review of the apps, not of the literature. In a sense, it is an original research project rather than a review, because each "app" was a participant.

We have clarified the title to avoid confusion with literature systematic reviews as follows: Gamification for health promotion: systematic review of behaviour change techniques in smartphone apps.

It is difficult to understand how it was accepted to be registered through Prospero. As the authors correctly identified, the impact of the apps involving behavior change challenges and game-based activities are yet to be known. However the topic is contemporary and provides a solid platform for future investigations in this area. The magnitude of technology available in health is well described. The rationale argues strongly for a scientific interpretation of what is available in health apps. The rigor of the Michie et al taxonomy provides a credible basis for categorizing gamification on health apps. Further, sound definitions of gaming characteristics were considered and applied within a contemporary timeframe.

We are grateful for confirmation of the methodological rigour of the work.

Results were clearly presented and described. Reporting could be improved by adding interquartile ranges (rather than whole ranges) to median values.

We prefer to use the range if possible rather than the interquartile range since this gives the reader

the exact number of behaviour change techniques in the apps. However we are happy to change the figures should the editors request this.

Reporting percentages in whole numbers rather than to inconsistent numbers of decimal points may improve readability and recall of important findings.

Thank you this was an error – we have now used whole number percentages throughout.

The discussion was well considered. The major limitation of advancing the understanding of the availability and nature of these apps rather than the effectiveness is well acknowledged. The manuscript is strongly supported by relevant and clear supplementary material.

We are grateful to the reviewer for their support and helpful comments.

Reviewer: 2
eviewer Name: Fabiana Lorencatto
Institution and Country: Centre for Health Services Research, City University London, UK
Competing Interests: none declared

Thank you for the opportunity to review this manuscript, which presents the results of a systematic review aiming to describe the behaviour change techniques commonly featured in smartphone apps that aim to change health related behaviours using gamification strategies.

Overall, I am enthusiastic about the topic being addressed and methodology used in this manuscript.

However, I have some concerns at this stage regarding the manuscript, particularly the background and methods section, which I feel warrant expansion and further clarification.

-Introduction
1. Overall, the manuscript is clearly and succinctly written. However, I feel the introduction is lacking essential information.

a. First, the manuscript as a whole assumes familiarity with behaviour change techniques (BCTs) and taxonomies of BCTs. These concepts are very loosely defined and described throughout. I am very familiar with the BCT literature and methodology and was thus able to interpret the methods and findings presented in this manuscript, but I am concerned that a reader less familiar with the relatively new BCT literature and methodology would be less able to readily interpret the methods and results presented. I would thus suggest defining BCTs and BCT taxonomies in the introduction, particularly the difference between behaviour change categories and BCTs.

Thank you – a paragraph has been added to the introduction to clarify this, as follows: A behaviour change technique is "an observable, replicable and irreducible component of an intervention designed to alter or redirect causal processes that regulate behaviour; that is, a technique is proposed to be an 'active ingredient' (e.g., feedback, self-monitoring, reinforcement)" [7]. These techniques have been clearly defined, linked with theories of behaviour change and classified into an internationally recognised taxonomy, comprising 93 individual techniques, grouped into 16 behaviour change categories [7].

More importantly, I suggest highlighting and citing the rapidly increasing number of reviews similar to the current review, which aim to identify and characterise the content of interventions in terms of component BCTs, using taxonomies with established reliability. One such review is cited in the discussion (i.e. Dombrowski et al), but would be helpful to incorporate this in the introduction to help

clarify the rationale/context for this study.

A paragraph has now been added to the introduction reviewing the literature in more detail, as follows: This taxonomy builds on previous work to identify the active components of complex interventions [8, 23, 24, 33-35]. For example, Dombrowski et al., coded behaviour change techniques for obese adults with obesity-related comorbidities in behavioural interventions applying a 26-category taxonomy developed by Abraham and Michie et al.,[34, 36] .

b. On a related note, I was surprised to see that none of the literature that has specifically aimed to characterise the content of smartphone apps has been cited in the introduction. For instance, a highly relevant example is the work of Crane et al. (2015) who applied BCT taxonomies to code the content of alcohol reduction smartphone apps [i.e. Crane, D., Garnett, C., Brown, J., West, R., & Michie, S. (2015). Behavior change techniques in popular alcohol reduction apps: content analysis.Journal of medical Internet research, 17(5)]. Crane et al. also cite a number of reviews of the characteristics of smartphone apps targeting a range of health behaviours (refs 14-25 in Crane et al). There are plenty of such reviews in the current literature (a few further examples below). It would be helpful to summarise this existing literature in introduction to better contextualize the present review.

• Direito, A., Dale, L. P., Shields, E., Dobson, R., Whittaker, R., & Maddison, R. (2014). Do physical activity and dietary smartphone applications incorporate evidence-based behaviour change techniques?. BMC Public Health, 14(1), 1.

• Conroy, D. E., Yang, C. H., & Maher, J. P. (2014). Behavior change techniques in top-ranked mobile apps for physical activity. American journal of preventive medicine, 46(6), 649-652.

• Hoeppner, B. B., Hoeppner, S. S., Seaboyer, L., Schick, M. R., Wu, G. W., Bergman, B. G., & Kelly, J. F. (2015). How smart are smartphone apps for smoking cessation? A content analysis. Nicotine & Tobacco Research, ntv117.

A paragraph specifically about smartphone apps has now been added to the introduction with references as suggested, as follows: Although apps have proliferated, work aiming to characterise the use of behaviour change techniques in smartphone apps and smartphone games is relatively novel. Two reviews include Direito et al., who used a 26-category taxonomy developed by Abraham and Michie et al.,[36, 37] and Conroy et al., who used the Coventry, Aberdeen and London- Revised (CALO-RE) developed also by Michie et al., and found limited use of behaviour change techniques in diet and physical activity apps[8, 38]. Crane et al., examined use of behaviour change techniques in alcohol reduction apps using the BCT Taxonomy v1[7]. Findings again found limited use of behaviour change techniques.

2. Throughout the manuscript there is reference to 'theoretically consistent BCTs' (e.g. p. 4 line 25; p. 13 line 49, p. 12 line 17)- but no actual mention or discussion as to which theories these BCTs are consistent with. Could this please be clarified/ made more explicit? Such as in the Michie 2009 paper linking BCTs to Control Theory (reference 9 in the present manuscript)?

The relationship linking BCTs and control theory has now been clarified in the discussion, as follows: Control theory suggests that setting goals, monitoring of behaviour, receiving feedback and reviewing relevant goals In the light of feedback may be effective in changing behaviour[41] and is one of a broader group of theories involving feedback loops and self-regulation[42].

Methods
3. The search was conducted 1 year ago (i.e. June 2015). Given smartphone apps are a rapidly expanding and developing industry, this raises the question as to whether the search is now out of

date and requires updating.

The search was performed one year ago and represents a snapshot in time at that date. The process of coding BCTs for apps is particularly time consuming and if we repeated the search then we would most likely have a similar delay in the finished paper. We were transparent in specifying the date used so that future reviewers can build on our work.

4. More detail is needed on the methods for BCT coding, particularly to enable a reader less familiar with this methodology to understand the analyses conducted. For instance, what content in the app was coded? Was it just text or images/video/other multimedia content as well?

The entire app content was coded as each app was downloaded onto a test phone – this has now been clarified in the paper, as follows: The entire app content was coded for behaviour change techniques including text, images, video and other multimedia content.

How were BCTs coded- in terms of a global present or absence rating (i.e. 1/0)? Did you also consider the frequency with which BCTs were present (i.e. BCTs delivered more than once within an app)?

Two raters (EE and JL) used the app independently on the test phone and then met to discuss the presence or absence. Disagreements were resolved by reference to a third reviewer who is a health psychologist familiar with BCT coding (LS). This has now been clarified in the text of the methods section. BCT's were classified as either present or absent – there was no count of the number of times each BCT was used. This is now clarified in the methods section and mentioned in the limitations section of the discussion, as follows: Two researchers trained in behaviour change technique coding (EE, JL) coded apps independently. App content was coded using the BCT taxonomy (v1)[7]. Techniques were classified as either present or absent. An example of the coding process and application of behaviour change techniques to app content is provided (Supplementary Figure 2). The number of individual behaviour change techniques included in each app was counted. There was no count of the frequency in which techniques were used in each individual app. Similarly, we did not examine the frequency with which behaviour change techniques were used in each individual app and the mode of delivery of each behaviour change technique.

Was mode of BCT delivery considered/recorded?

We did not record the mode of delivery systematically – although method of delivery is often similar between apps eg reward tables. This has now been addressed in the limitations section of the discussion, as follows: Similarly, we did not examine the frequency with which behaviour change techniques were used in each individual app and the mode of delivery of each behaviour change technique.

Was inter-rater reliability formally assessed (i.e. Kappa)?

Kappa scores have been added to the methods section as follows: We used Cohen's kappa to assess inter-rater reliability of BCT coding at the initial stage of review. There was substantial agreement between the two reviewers ($\kappa = .79$, 95% CI, .76 to .81). All discrepancies in reviewer coding were then resolved through discussion with a third trained reviewer (LS), a health psychologist.

I am also not entirely clear what is meant by 'app content was assessed against criteria derived from a standard taxonomy of BCTs'- which criteria are these?

The wording has been clarified throughout – the standard taxonomy was used.

It might be helpful to provide an example of app content coded into BCTs (e.g. a screen shot of an app page, and call-out boxes highlighting what content would be coded into what BCT; or some exemplary text/quote excerpt and associated BCTs).

We have included an example of app content coding as suggested. This is Supplementary Figure 2.

5. A particular strength of this review is that combinations of BCTs were considered. An enduring limitation of many existing BCT reviews is that they only tend to look at individual BCTs and not combination of BCTs. I therefore commend the authors for addressing this gap. As such, could the authors provide more detail as to how combinations of BCTs were defined/assessed? As this will be of particular methodological interest to researchers conducting similar research.

This part of the paper was descriptive only, seeking to identify the combinations of BCTs that smartphone game developers are currently using. We had insufficient power to examine effects against other parameters such as rating or price. We agree that this is an interesting area of work, ideally, we would look at pre-defined, theoretically consistent groups of techniques and then examine whether these were present in the games, however, the theoretical framework linking to the latest taxonomy has not yet been fully developed so we were unable to progress further.

Results

6. Overall results are clear. Minor suggestion- P. 9 lines 28-30, 34-36; consider moving the findings on correlation between number of BCTs and customer ratings/price later in the results section. These findings are currently presented before the results on BCTs identified, so the sequence in which the results is slightly confusing.

Thank you this adjustment has been made to make the paper clearer.

Results/Discussion

7. Results/discussion currently focused on the high frequency BCTs. This is, of course, of interest, but equally of interest are the low frequency BCTs- how do these align with relevant 'theory'- would we have expected any observed low frequency BCTs to have been included in apps? Are there any missed opportunities (Seeing as, if I understood correctly, 75% of the 93 BCTs were not identified in any apps)?

Thank you. This is an important point – the aim of our paper was simply to describe common BCTs and combinations. Low frequency BCTs may be equally important. This has now been addressed in the limitations section of the discussion as follows: In this review, we focussed on commonly used behaviour change techniques. It would be interesting to examine behaviour change techniques that were not used or had a low frequency of use, to determine how these aligned with relevant behavioural and cognitive theories and hence identify potential opportunities for app developers.

Discussion

8. Overall, the discussion provides good consideration of the implications of the present findings, for both research and practice. However, as per comment 1b, the discussion would be strengthened by comparing the present findings to those of other reviews aiming to characterise the characteristics of smartphone apps (e.g. Crane et al. 2015, plus associated references).

The discussion has been extended to cover this point as follows: Social support as a behaviour

change technique is also common in physical activity apps[38]. Other reviews have found that the behaviour change technique providing instruction on how to perform behaviour has featured highly amongst physical activity apps (n=33, 83% of apps) [37] (n=111, 66% of apps)[38], however, this technique was found in relatively few apps in our review (n=25, 39% of apps). It is possible that this technique may be more suited to physical activity apps since it was not found in apps to reduce alcohol consumption [40]. Alcohol reduction apps also featured a range of techniques not found in smartphone games: facilitate self-recording; provide information on consequences; give options for additional and later support; offer/direct towards appropriate written materials[40]. Whilst, these techniques may be more suited to alcohol reduction apps it is also possible that they do not lend themselves to use on the gaming platform.

Currently comparisons are drawn with reviews/meta-analyses of internet based interventions, and other reviews that have identified similar combination of BCTs as the present review. Although this literature is relevant, even more focused/relevant comparisons could be drawn by comparing present findings with that of other reviews focused on smartphone apps specifically.

The discussion has now been extended now to include more comparisons specifically with other apps as follows: One study examining technology-based delivery found that popularity and user ratings were only weakly associated with behaviour change technique content [40].

VERSION 2 – REVIEW

REVIEWER	Geraldine Naughton Australian Catholic University Australia
REVIEW RETURNED	01-Aug-2016

GENERAL COMMENTS	As stated previously I would prefer NOT to see the word SYSTEMATIC in the title because it is misleading. Otherwise it appears that my concerns have been addressed.