DIGITAL HEALTH

A quasi-experimental investigation of college students' ratings of two physical activity mobile apps with varied behavior change technique quantity

Ashlee Davis and **Rebecca Ellis**

Abstract

Background: An assessment of how users rate physical activity apps of varying behavior change technique content is necessary to understand if users recognize differences in an app's ability to promote physical activity.

Objective: The purpose of this study was to compare user ratings of an app with a lower behavior change technique count to an app with a higher behavior change technique count.

Method: Participants were randomly assigned to interact with either the high behavior change technique app or the low behavior change technique app using an iPad. Participants then completed a Mobile App Rating questionnaire.

Results: The final sample included 83 participants with an average age of 22.66 years (SD = 2.13; range = 20–29). Independent t-tests revealed significant group differences for perceived impact, $t(81) = 5.27$, $p < .001$, $g = 1.15$, 95% confidence interval (0.69, 1.62); engagement, $t(81) = 6.71$, $p < .001$, g = 1.15, 95% confidence interval (1.02, 1.87); aesthetics, $t(81) = 4.29$, $p < .001$, $g = 1.15$, 95% confidence interval (0.50, 1.38); and subjective quality, $t(81) = 6.46$, $p < .001$, $g = 1.15$, 95% confidence interval (0.75, 1.42), with participants from the high behavior change technique group scoring these qualities more positively than participants from the low behavior change technique group.

Conclusion: App users rated a physical activity app with higher behavior change technique content more favorably on aesthetics, engagement, subjective quality, and perceived impact than those with reduced behavior change technique content. Additional research is needed to understand how these perceptions influence users during the app selection process, as well as the efficacy of apps for promoting physical activity behavior change.

Introduction

Mobile health is a rapidly growing field, with mobile apps being examined for their ability to facilitate active lifestyles. In total, 77% of Americans own a smartphone and one in four of them use smartphones to look up health information, with about 19% using health-related mobile apps.[1,2] Mobile apps have the unique potential to provide health-related programs, such as physical activity interventions, to a large number of people for a relatively low

Department of Kinesiology and Health, Georgia State University, United States of America

Corresponding author:
Ashlee Davis, Department of Kinesiology and Health, Georgia State University, PO Box 3975, Atlanta, GA, 30302, United States of America.
Email: ahamilton16@student.gsu.edu
Twitter: Ashlee Davis – @AshleeSDavis; Rebecca Ellis – @RebeccaEllisPhD.

cost. As such, in 2017 there were over 325,000 commercially available mobile health apps in all major app stores.[3] However, research on the content and quality of these apps identified significant limitations, such as lack of theoretically based content.[4,5]

One way to incorporate theoretical content within mobile apps is the use of behavior change techniques (BCTs). As described by Michie et al., BCTs are the smallest active component designed to elicit a behavior change.[6] They can be used alone, but are typically used as part of a larger intervention to target a particular behavior, although most apps for physical activity (PA) use a combination of BCTs to promote PA uptake among users.[6] However, despite their popularity, mobile apps have not implemented enough BCTs to promote a significant change in PA behavior.[5] Additionally, there is limited evidence to suggest that mobile app users make app selections based on BCT content or that they are aware of differences in behavior change impact between apps with varying BCT content.

Several authors examined commercially available PA apps for BCT quality and quantity. In an examination of top PA and diet apps, Direito et al. examined the taxonomy of BCTs used in interventions and found free apps had an average of 6.6 BCTs (range 3–14) and paid apps had an average of 9.7 (range 2–18).[7] A similar study examined 100 PA apps and found that of the possible 93 BCTs (using BCT Taxonomy V1), only 39 were observed in the apps.[8] Further, they observed an average of 6.6 BCTs (SD = 3.3, median = 6) and did not find a significant difference in BCT content between free and paid apps, $t(98) = 1.43$, $p = 0.08$, $d = 0.29$.[8] These findings demonstrated that BCTs were under-utilized in PA apps. However, it should be noted that in general, the presence of more BCTs does not necessarily indicate a higher-quality intervention, as BCTs vary in their ability to promote behavior change. The least prevalent type of BCT reported in the study by Yang et al. was action planning (15%) and coping planning was not observed.[8] This is concerning because both are proposed to mediate the relationship between intention and behavior.[9] Carraro and Gaudreau found action planning and coping planning were significant predictors of PA, with spontaneous action planning and coping planning displaying moderate-to-large effects on PA.[9] Furthermore, apps must be correctly applied to improve effectiveness. Each BCT has a specific set of conditions needed for the technique to aid in promoting behavior change.[10] This suggests that although increasing the presence of BCTs may help increase the effectiveness of PA mobile apps, the BCTs utilized and **the parameters of use** should also be thoughtfully considered.

In addition to understanding the BCTs present in mobile apps, it is necessary to recognize user preferences of app features. In an online cross-sectional study of young adults, researchers found "goal setting on outcome of behavior," "self-monitoring of behavior," and "self-monitoring on outcome of behavior" as the most preferred BCTs within a PA mobile app.[11] Interestingly, this study also examined associations among specific personality traits and preferred BCTs. Researchers found a positive relationship between "agreeableness" and "goal setting" (odds ratio (OR) 1.61, 95% confidence interval (CI) 1.06, 2.41), an inverse association between "neuroticism" and "feedback/self-monitoring" (OR 0.76, 95% CI 0.58, 1.00), and a positive relationship between "self-efficacy" and "feedback and self-monitoring" (OR 1.06, 95% CI 1.02, 1.11).[11] These findings suggested that although users preferred some BCTs to others, PA mobile apps were not "one size fits all." Similarly, Middleweerd et al. performed a qualitative study with Dutch university students in which participants used a PA app for 3 weeks then attended a focus group.[12] These researchers found participants favored the coaching, tailored feedback, and competition features of the app.[13] Finally, in another cross-sectional study on user perceptions of behavior change mechanisms, most respondents reported using a PA mobile app positively affected PA perceptions, attitudes, and beliefs.[14]

Based on the findings from previous studies, users can identify app features they find helpful. It is unclear, however, how these preferences impact their perceptions of the ability of the app to change behavior. Further, it has not been determined if it is merely the presence of a single preferred BCT that influences perception, or if users identify apps with a variety of BCTs as possibly having a greater impact on behavior. For example, if a higher BCT content app contains a single feature that a user prefers, will the user perceive another app of lower BCT content with that same feature to be equally as impactful? An assessment of how users rate PA apps of varying BCT content is necessary to help understand if users recognize potential differences in the PA app's ability to impact PA behavior. Stoyanov et al.[15] developed the Mobile App Rating Scale (MARS), a multidimensional scale for rating the quality of mobile apps, in response to this need. The MARS was originally designed for use by researchers and professionals, so the authors later developed an end-user version, the User Version of the MARS (uMARS). The uMARS scale consists of four objective quality subscales (engagement, functionality, aesthetics, and information quality), a subjective quality subscale, and a perceived impact subscale. Therefore, the purpose of this study was to examine user ratings of two PA mobile apps using an amended uMARS, specifically as it related to the engagement, functionality, aesthetics, subjective quality, and perceived impact of the apps. We compared user ratings

of an app with a lower BCT count to an app with a higher BCT count, primarily to identify potential differences in ratings of the perceived impact of the apps. The findings from this study will provide a quantitative assessment on whether users rate the impact of apps with varying BCT content differently. We hypothesized that participants in the higher BCT count app group would rate the app more positively on engagement, functionality, aesthetics, subjective quality, and perceived impact than those in the lower BCT count app group.

Methods

Participants

Eligible participants were Georgia State University students aged 18–29 years. We targeted this age group because a significantly greater percentage of smartphone users in this demographic use mobile apps.[1,16,17]

Measures

Personal history questionnaire. A demographics questionnaire assessed age, height, weight, gender, year in school or employment status, and race. This questionnaire also asked about participants' mobile app usage. Body mass index (BMI) was calculated from self-reported height and weight.

Stages-of-change (SOC) modified four-stage algorithm. This four-item self-report instrument assessed participants' SOC by providing a definition of "regular physical activity" (based on current guidelines) then asking participants if they were regularly active based on that definition and intended to continue to be (yes/no). If "yes," participants were categorized into stage four (action/maintenance). Participants who responded "no" selected one of three statements that categorized them into either stage one (precontemplation; less active than recommended and do not intend to become regularly active in next 6 months), stage two (contemplation; less active than recommended and intend to become regularly active in next 6 months), or stage three (preparation; less active than recommended and intend to become active in next month).[13] Construct validity was confirmed because the algorithm accurately distinguished PA levels across the stages.[13] We used this instrument to examine group differences for PA.

Amended uMARS questionnaire. This questionnaire was an amended version of uMARS (see supplemental file).[18] The amended uMARS was a 16-item questionnaire that asked participants to rate a mobile app on *engagement* ($n = 2$; e.g. "Does it use strategies to increase engagement by presenting its content in an interesting way?"), *functionality* ($n = 2$, e.g. "How easy is it to learn how to use the app?"), *aesthetics* ($n = 1$, e.g. "How good does the app look?"), *subjective quality* ($n = 4$; e.g. "What is your overall star rating of the app?"), and *perceived impact* ($n = 6$; e.g. "This app is likely to increase awareness of the importance of addressing physical activity."), with perceived impact our primary outcome measure. These 15 items were rated on a five-point Likert type scale. A 16th item asked if participants knew of the app they rated, but it was not part of the five subscales. Mean scores were calculated for each of the five subscales. The engagement, functionality, and aesthetics subscales are intended to objectively examine app quality in these areas. The subjective quality evaluated user opinions of the app's value, and the perceived impact subscale was designed to obtain information on the user's perception of the apps impact on the user's knowledge, attitudes, and intentions for a specific behavior. Stoyanov et al.[18] found high internal consistency and good test-retest reliability among all subscales of the uMARS in a study of adolescents and young adults using a mobile health app.

Mobile apps. To select the study apps, we searched the Google Play app store (United States) in May 2017 for PA apps. First, we used the search term "physical activity" to identify commercially available, free PA mobile apps with greater than 4.5 star ratings. This yielded 141 apps. The apps were screened to exclude irrelevant apps (e.g. not fitness related, geared towards children, etc.). Then 74 apps were reviewed for BCT content using the CALO-RE taxonomy based on their app store description and the techniques were quantified.[11,12] Michie et al. recommend using the CALO-RE taxonomy in interventions targeting physical activity.[19] Eight apps (four highest and four lowest) were downloaded on an Android device and the first author coded the BCTs present in each app. The second author reviewed the coded spreadsheet. In an effort to prevent bias, an attempt was made to identify the apps with the lowest and highest number of techniques that were comparable on visual attractiveness, so the apps were rated on aesthetics using the amended uMARS. Finally, the two apps that compared visually in the lowest BCT and the highest BCT categories were selected for use in this study. The Noom Walk Pedometer (NWP; two BCTs) was selected as the low BCT app and MapMyFitness (MMF; 11 BCTs) was selected as the high BCT app (see Figure 1).

Procedures

The study procedures were approved by Georgia State University's (GSU's) Institutional Review Board. Participants were recruited using flyers posted on the GSU Atlanta campus, as well as with classroom announcements. Interested volunteers emailed the
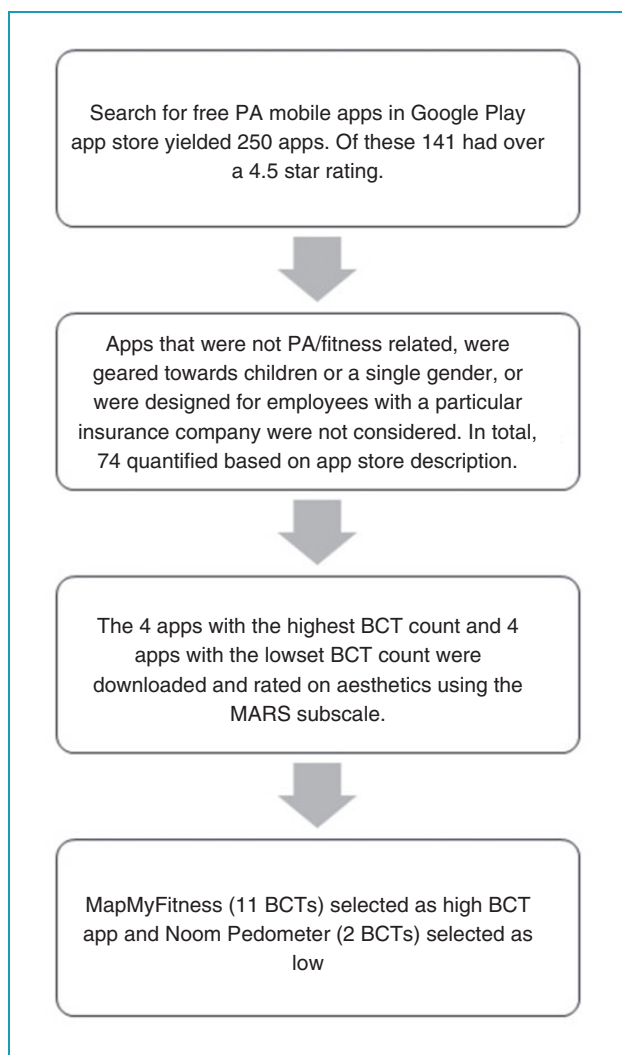
Figure 1. Mobile app selection process.

student primary investigator (PI) who scheduled a face-to-face visit.

Participants who attended the face-to-face meeting consented to participate in the study then completed the participant history and the SOC questionnaires. The participants were randomly assigned to interact with either the MMF app (high BCT group) or the NWP app (low BCT group) using an iPad. Randomization was achieved by providing participants with an envelope from a randomly sorted stack, labeled either 1 or 2. The numbers corresponded with the iPad they were to use because only one PA app was loaded on each iPad (one had NWP, the other had MMF). Inside the envelope was a piece of paper with the name of the mobile app so participants knew which app to open. A student assistant was responsible for assigning a number to each iPad, labeling and shuffling the envelopes, and providing them to the researchers.

The researchers were unaware which app was on each iPad and were blinded to app assignment. Once a participant was provided with an envelope, the student entered a separate room to pick up the iPad corresponding with the envelope provided. Each participant was allowed 15 minutes to interact with the assigned mobile app in a secluded space. No identifying information was needed for the participant to interact with the app. Following the app use, participants were asked to complete the amended uMARS questionnaire. The entire visit took approximately 45 minutes to complete.

Statistical analyses

Tests of normality and outliers were performed before analyses. All variables were summarized using frequencies, means, and standard deviations. Independent *t*-tests and Chi-square tests were used to examine baseline differences on demographic variables (age, BMI, race, year in school) and SOC between the two app groups. Cronbach's alpha was calculated for the multi-item MAR questionnaire subscales. A Cronbach alpha of $\alpha > .70$ = acceptable and $\alpha > .80$ = good.[20] Group differences on the amended uMARS subscales were assessed using an independent *t*-test. Bonferroni corrections were used, so statistical calculations for the amended uMARS subscales were considered significant at an alpha level of $p < .01$. All other calculations were considered significant at an alpha level of $p < .05$. Analyses were conducted using SPSS version 23.

Results

The high BCT app (MMF) incorporated 11 BCTs and the low BCT app (NWP) incorporated two. Specifically, the high BCT app included the following: (a) provide instructions on how to perform behavior, (b) goal-setting behavior, (c) information about other's approval, (d) prompt review of behavioral goals, (e) facilitate social comparison, (f) set graded tasks, (g) provide information on when and where to perform behavior, (h) prompt self-monitoring of behavior, (i) prompt self-monitoring of behavioral outcomes, (j) teach to use prompts/cues, and (k) provide reward contingent on successful behavior. The low BCT app included: (a) information about others approval and (b) prompt self-monitoring of behavior.

In total, 89 individuals responded to recruitment announcements. Data for five were excluded because they were not 18–29 years of age ($n = 4$) or they did not report the app they evaluated ($n = 1$). One participant was removed due to non-response on too many items. The final sample included 83 participants with an average age of 22.66 years ($SD = 2.13$;

Table 1. Participant characteristics.

	MapMyFitness		Noom		Total	
Characteristics	M	SD	M	SD	M	SD
Age (years)	22.40	1.78	22.93	2.43	22.66	2.13
BMI	24.69	3.80	23.59	2.78	24.14	3.36
	n	%	*n*	%	*N*	%
Gender						
Female	24	57.14	29	70.73	53	63.9
Male	18	42.86	12	29.27	30	36.1
Race						
Asian	8	19.05	6	14.63	14	16.9
Black or African-American	19	45.24	21	51.22	40	48.2
Hispanic or Latino	4	9.52	3	7.32	7	8.2
Multiracial	2	4.76	4	9.76	6	7.2
White or Caucasian	8	19.05	7	17.07	15	18.1
Other	1	2.38	0	0.00	1	1.2
Year in school						
Junior	3	7.14	6	14.63	9	10.84
Senior	39	92.86	35	85.37	74	89.16
SOC						
Action/maintenance	34	80.95	32	78.05	66	79.5
Contemplation	1	2.38	6	14.63	7	8.43
Preparation	7	12.67	3	7.32	10	12.05
Previous use/knowledge of app	10	23.8	0	0	10	11.9

BMI: body mass index; SOC: states of change.

range = 20–29) who were mostly female (63.9%) and African-American (48.2%; see Table 1). There were no significant differences between the high and low BCT app groups on demographic variables or SOC. The amended uMARS subscales had acceptable-to-good internal consistency (see Table 2). The independent *t*-tests revealed significant group differences for engagement, $t(81) = 6.71$, $p < .001$, $g = 1.15$, 95% CI (1.02, 1.87), aesthetics $t(81) = 4.29$, $p < .001$, $g = 1.15$, 95% CI (0.50, 1.38), subjective quality $t(81) = 6.46$, $p < .001$, $g = 1.15$, 95% CI (0.75, 1.42), and perceived impact $t(81) = 5.27$, $p < .001$, $g = 1.15$, 95% CI (0.69, 1.62). Specifically, participants rated MMF (high BCT group) as having greater engagement, aesthetics, subjective quality, and perceived impact than NWP (low BCT group; see Table 2). No significant group difference was observed for functionality, $t(81) = 2.09$, $p = .04$, $g = 1.15$, 95% CI (0.02, 1.00).

Table 2. Group differences for MARS subscales.

	MapMyFitness		Noom		
	M	SD	M	SD	Cronbach's α
Aesthetics[a]	3.45	1.11	2.51	0.87	–
Engagement[a]	3.29	1.08	1.84	0.86	0.89
Functionality[a]	4.07	0.83	3.56	1.34	0.75
Perceived impact[a]	3.30	1.02	2.07	1.09	0.95
Subjective quality[a]	2.88	0.82	1.80	0.70	0.85

[a]Significant group differences found at $p \leq .008$.
MARS: Mobile App Rating Scale.

Discussion

Mobile apps are being researched as an option to deliver behavioral interventions; however, limited research has focused on user ratings of apps as it relates to app quality and how users perceive an app's ability to impact behavior. Therefore, the purpose of this study was to examine user perceptions of two PA mobile apps using an amended uMARS. Participants in the higher BCT app group rated MMF significantly higher on four of the five amended uMARS subscales. This was the anticipated outcome because previous research indicated app users had preferences related to the BCTs utilized within an app.[11,12] The findings from the present study suggested that app users might have also related the presence of more BCTs to app effectiveness due to the group difference found for perceived impact. Several studies cited low theoretical content as a potential limitation of apps targeting health behavior.[21,22] However, real-world implications should be considered. Before an app could be examined for efficacy, app users would first have to identify an app they believe would help promote PA. Previous research did not provide clear evidence to suggest app users would identify an app with more theoretical content as being more effective. Our results supported the argument that the amount of theoretical content in an app might influence user ratings of the quality and perceived impact of the app.

Two previous cross-sectional studies examined PA app users' perceptions of behavior change mechanisms and app effectiveness.[14,23] The studies found PA app use led participants to be more motivated to change their behavior and that a greater percentage of users perceived PA apps effectively affected their behavior. Their results supported the concept that PA apps can be used to positively influence user behavior. Our findings extended their results by demonstrating

users may understand that some apps can support their behavior change efforts better than others. Taken together, these findings are promising because they provide further guidance on how to best design apps that will lead to PA behavior change. App developers and researchers should consider using a BCT taxonomy to increase the BCT quantity and quality offered by PA apps. The results can also impact how apps are marketed to consumers. Although our study did not test how users make app selections, the results may support emphasizing the presence of BCTs in the app description to encourage app use. However, as previously mentioned, it is also important to remember that all BCTs are not equally effective for promoting behavior change. As such, additional research is needed to understand whether perceived app effectiveness is related to actual PA behavior promotion.

Regarding the other subscales, for engagement, aesthetics, and subjective quality, participants also rated MMF (high BCT group) significantly higher than NWP (low BCT group). However, no group differences were observed for functionality. According to Stoyanov et al., when scored separately, the objective quality items (engagement, aesthetics, functionality) can be used to evaluate strengths and weaknesses in these areas.[18] The lack of group differences for functionality was ideal, as it supported the expectation that the group differences observed were not simply due to differences in how well each app worked. Although efforts were made to select apps that were similar in visual appeal, the group differences for aesthetics may have introduced bias in the participants' ratings of the other scales. According to the 2017 U.S. Mobile App Report, 21% of app users between the ages of 18–24 reported deleting an app because they did not like the logo.[16] Further, Singh suggested that color can influence perceptions of a product.[24] It is possible that an app rated as less attractive is

perceived as lower quality and, thus, less effective. Regarding the engagement and subjective quality sub-scales, our results supported findings from the Hoj et al. study. They found that app users who reported more frequent app use perceived PA apps to have a greater impact on behavior.[14] In our study, the app that users found to be more engaging and of higher subjective quality was also rated to have a greater perceived impact on behavior. Neither study explored the specific mechanisms, but the relationship between engagement and perceived impact should be further explored to better influence app design.

Other study limitations to consider when interpreting these results include the study population was limited to college students aged 18–29 years, which means the findings are not generalizable to other populations. These findings still provide valuable information as this demographic uses mobile apps at a higher rate.[1,16] Another limitation is that most participants reported being in the action/maintenance SOC. This may have introduced bias due to potential similarities in BCT preferences among regularly active individuals. Also, the present study has limited external validity, as participants used the mobile apps in a laboratory setting for a limited period. This improves the internal validity of the study but limits our ability to generalize these findings to how participants would use the app in a real-world setting. Future research should attempt to replicate these findings in a more naturalistic setting. Further, 10 study participants reported use or knowledge of the app before study participation. We did not analyze this variable as a potential moderator because all 10 participants were in the high BCT group, but it should be noted that familiarity (or lack thereof) with the app could influence ratings. As it relates to the apps, the high BCT app was still relatively low in BCT quantity, considering there are 40 BCTs listed in the CALO-RE taxonomy.[19] Still, the high BCT app had more BCTs than the average of 6.6 observed in previous studies that quantified app BCT content.[7,8] Lastly, this study did not consider the quality of the BCTs being used in each app, but rather focused on quantity of BCTs. This study focused on BCT quantity because previous literature cites low theoretical (BCT) content as a potential factor in why PA apps do not result in significant changes in PA behavior.[25] However, additional studies should be conducted to examine how BCT quality impacts user ratings of the quality and perceived impact of an app.

Conclusion

In summary, the results of this study indicated app users rated a PA app with higher BCT content more favorably on aesthetics, engagement, subjective quality,

and perceived impact than those with reduced BCT content. These findings supported previous literature that suggested theoretical content might influence app effectiveness. The results from this study should encourage app designers and researchers to consider the quantity and quality of BCTs included in PA mobile apps. Also, app developers might utilize this information to influence how to market PA apps to consumers. Future researchers should include app users from a wider range of age groups and all PA levels. In addition, researchers should experimentally test apps with higher BCT counts in comparison to apps with lower BCT counts on their ability to promote PA.

ORCID iDs
Ashlee Davis ⓘ https://orcid.org/0000-0002-5240-7481
Rebecca Ellis ⓘ https://orcid.org/0000-0002-4134-2053

References

1. Fox S and Duggan M. Mobile Health [Internet]. Pew Research Center (2012, accessed 28 May 2019). Available from: https://www.pewinternet.org/2012/11/08/mobile-health-2012/.
2. Perrin A. 10 facts about smartphones [Internet]. Pew Research Center (2017, accessed 31 May 2019). Available from: https://www.pewresearch.org/fact-tank/2017/06/28/10-facts-about-smartphones/.
3. mHealth Economics. Report: Status and trends in digital health [Internet]. Research2Guidance; (2017, accessed 31 May 2019). Available from: https://research2guidance.

com/product/mhealth-economics-2017-current-status-and-future-trends-in-mobile-health/.

4. Armstrong S. Which app should I use? *BMJ* 2015 Sep 9; h4597.

5. Conroy DE, Yang C-H and Maher JP. Behavior change techniques in top-ranked mobile apps for physical activity. *Am J Prev Med* 2014 Jun; 46(6): 649–652.

6. Michie S, Wood CE, Johnston M, et al. Behaviour change techniques: The development and evaluation of a taxonomic method for reporting and describing behaviour change interventions (a suite of five studies involving consensus methods, randomised controlled trials and analysis of qualitative data). *Health Technol Assess* 2015 Nov; 19(99): 1–188.

7. Direito A, Pfaeffli Dale L, Shields E, et al. Do physical activity and dietary smartphone applications incorporate evidence-based behaviour change techniques? *BMC Public Health* 2014 Dec; 14(1): 646.

8. Yang C-H, Maher JP and Conroy DE. Implementation of behavior change techniques in mobile applications for physical activity. *Am J Prev Med* 2015 Apr; 48(4): 452–455.

9. Carraro N and Gaudreau P. Spontaneous and experimentally induced action planning and coping planning for physical activity: A meta-analysis. *Psychol Sport Exerc* 2013 Mar; 14(2): 228–248.

10. Peters G-JY, de Bruin M and Crutzen R. Everything should be as simple as possible, but no simpler: Towards a protocol for accumulating evidence regarding the active content of health behaviour change interventions. *Health Psychol Rev* 2015 Jan 1; 9(1): 1–14.

11. Belmon LS, Middelweerd A and te Velde SJ and Brug J. Dutch young adults ratings of behavior change techniques applied in mobile phone apps to promote physical activity: A cross-sectional survey. *JMIR MHealth UHealth* 2015 Nov 12; 3(4): e103.

12. Middelweerd A, van der Laan DM, van Stralen MM, et al. What features do Dutch university students prefer in a smartphone application for promotion of physical activity? A qualitative approach. *Int J Behav Nutr Phys Act* 2015; 12(1): 31.

13. Kosma M and Ellis R. Establishing construct validity of a stages-of-change algorithm for physical activity. *Am J Health Promot* 2010 Nov 1; 25(2): e11–20.

14. Hoj TH, Covey EL, Jones AC, et al. How do apps work? An analysis of physical activity app users' perceptions of behavior change mechanisms. *JMIR MHealth UHealth* 2017 Aug 3; 5(8): e114.

15. Stoyanov SR, Hides L, Kavanagh DJ, et al. Mobile app rating scale: A new tool for assessing the quality of health mobile apps. *JMIR MHealth UHealth* 2015 Mar 11; 3(1): e27.

16. Pedotto K and Chen V (2017) The 2017 U.S. Mobile App Report. Avaialbale from : https://www.comscore.com/Insights/Presentations-and-Whitepapers/2017/The-2017-US-Mobile-App-Report

17. Krebs P and Duncan DT. Health app use among us mobile phone owners: A national survey. *JMIR MHealth UHealth* [Internet]. 2015 Nov 4; 3(4). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4704953/.

18. Stoyanov SR, Hides L, Kavanagh DJ, et al. Development and validation of the User Version of the Mobile Application Rating Scale (uMARS). *JMIR MHealth UHealth* 2016 Jun 10; 4(2): e72.

19. Michie S, Ashford S, Sniehotta FF, et al. A refined taxonomy of behaviour change techniques to help people change their physical activity and healthy eating behaviours: The CALO-RE taxonomy. *Psychol Health* 2011; 26(11): 1479–1498.

20. George D, and Mallery P. *IBM SPSS Statistics 21 Step by Step: A simple guide and reference*. 13th ed. Boston: Pearson, 2014.

21. Pagoto S, Schneider K, Jojic M, et al. Evidence-based strategies in weight-loss mobile apps. *Am J Prev Med* 2013 Nov 1; 45(5): 576–582.

22. Morrissey EC, Corbett TK, Walsh JC, et al. Behavior change techniques in apps for medication adherence: A content analysis. *Am J Prev Med* 2016 May 1; 50(5): e143–6.

23. Wang Q, Egelandsdal B, Amdam GV, et al. Diet and physical activity apps: Perceived effectiveness by app users. *JMIR MHealth UHealth* 2016 Apr 7; 4(2): e33.

24. Singh S. Impact of color on marketing. Manag Decis [Internet] (2012, accessed 28 May 2019) Available from: https://www.emeraldinsight.com/doi/abs/10.1108/00251740610673332.

25. Flores Mateo G, Granado-Font E, Ferré-Grau C, et al. mobile phone apps to promote weight loss and increase physical activity: A systematic review and meta-analysis. *J Med Internet Res* 2015 Nov 10; 17(11): e253.