Unified Health Gamification can significantly improve Well-being in Corporate Environments*

Arash Shahrestani, Pieter Van Gorp, Pascale Le Blanc,
Fabrizio Greidanus, Kristel de Groot and Jelle Leermakers

Abstract— **There is a multitude of mHealth applications that aim to solve societal health problems by stimulating specific types of physical activities via gamification. However, physical health activities cover just one of the three World Health Organization (WHO) dimensions of health. This paper introduces the novel notion of *Unified Health Gamification* (UHG), which covers besides physical health also social and cognitive health and well-being. Instead of rewarding activities in the three WHO dimensions using different mHealth competitions, UHG combines the scores for such activities on unified leaderboards and lets people interact in social circles beyond personal interests. This approach is promising in corporate environments since UHG can connect the employees with intrinsic motivation for physical health with those who have quite different interests. In order to evaluate this approach, we realized an app prototype and we evaluated it in two corporate pilot studies. In total, eighteen pilot users participated voluntarily for six weeks. Half of the participants were recruited from an occupational health setting and the other half from a treatment setting. Our results suggest that the UHG principles are worth more investigation: various positive health effects were found based on a validated survey. The mean mental health improved significantly at one pilot location and at the level of individual pilot participants, multiple other effects were found to be significant: among others, significant mental health improvements were found for 28% of the participants. Most participants intended to use the app beyond the pilot, especially if it would be further developed.**

I. Introduction

In the area of participatory health and well-being, the multitude of mHealth apps and wearables focuses on promoting physical activities. There is a clear bias towards rewarding steps, runs or biking sessions. For example, various 5 to 8-week pilot studies have focused on goal setting and checking apps regarding the specific physical activities of walking and running (i.e., making steps) [1]–[3]. Such studies have achieved promising evidence in controlled groups, often recruited specifically via primary care physicians.

Unfortunately, daily steps only give a narrow view on activity levels. While this might not be problematic in cases where physicians prescribe specific activities for a controlled target group, it may be detrimental in less controlled settings such as corporate environments where people are recruited to join health games and competitions. In such a setting, people who perform frequent swimming (or biking, …) activities would find it quite unfair if less fit challenge participants would systematically win the corporate challenges simply because they happen to make more steps per day. According to a survey by Seaborn et al. [4], psychological frameworks such as the Self-Determination-Theory (SDT) should be considered. Against such a baseline, the exclusive rewarding of steps is detrimental for the feelings of competence of those who prefer different types of physical activities [5]. In fact, according to the WHO, it is even too restrictive to only consider physical activities: *"Health is a state of complete physical, mental and social well-being"* [6].

Aparicio et al. suggest to consider besides people's sense of competence also their sense of autonomy and relatedness in SDT-based game design [7]: not controlling the instructions given to people has been shown to improve their sense of autonomy and consequently, the intrinsic motivation of individuals [8]. In terms of relatedness, many mHealth apps already provide leaderboards and badges to relate to other people who enjoy performing one shared type of activity (e.g., running). Furthermore, individual mHealth apps may have team and communication features to discuss such activities. Both functionalities may support relatedness between people with the same type of health interests. However, they may fall short in corporate health settings, where individuals may have quite different interests.

Based on the above analysis of prior mHealth studies and their limitations in relation to psychosocial theories, we decided to explore a novel approach to mHealth games, called Unified Health Gamification (UHG). Based on the Know-Check-Move paradigm [1]–[3], UHG also aims to (1) raise awareness, (2) support goal-setting and (3) provide feedback on progress. The UHG novelty resides in its capability to engage people with different health interests and different capabilities into health competitions between families, colleagues, neighbors and/or similar social circles.

II. Methods

This paper investigates the health effects and the user perceptions of the first UHG app by considering health changes and user attitudes in two different target groups. This was done with pre- and post-tests during a six-week intervention period. *Fig. 1* shows the overall study design.

A. Unified Health Gamification and its Implementation

UHG is based on the following principles: (1) people should be capable to form teams with peers regardless of differences in health interests, (2) people should receive positive feedback at the personal level as well as on the level of their team, and (3) the implementation should be flexible with regards to the rules of health games.

* Research supported by EIT Digital project GameBus (ID 15024).
A. Shahrestani, P. Van Gorp and P. Le Blanc are with Eindhoven University of Technology (corresponding author: p.m.e.v.gorp@tue.nl). F. Greidanus is with ZuidZorg, K. de Groot is with GGzE and J. Leermakers is with Finaps.

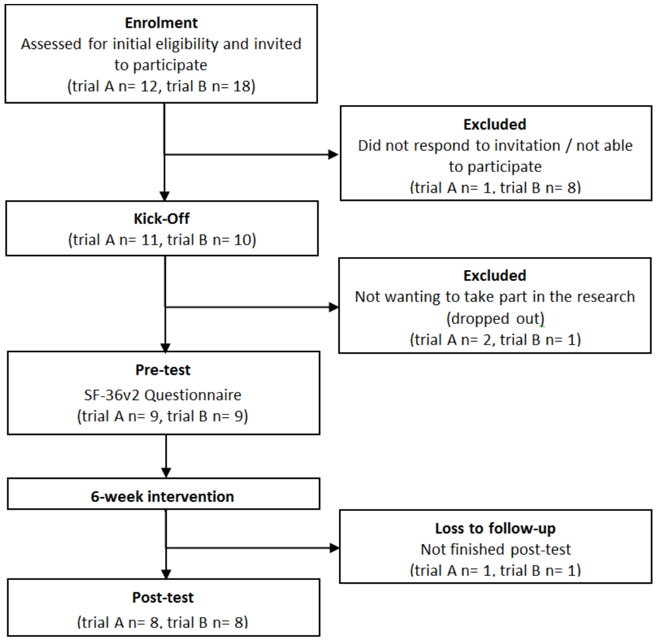

Figure 1: Study Design Graph

Significant investments were made to realize GameBus[1], the first prototype of an UHG app with integrations to third party activity apps and wearables. Furthermore, various concept validations were carried out during app development, in collaboration with intended sponsors. The first GameBus version was released late 2015. This paper presents the results of the first pilots which aimed at testing the effectiveness of the approach at that point in time.

Fig. 2 and *Fig. 3* show two key screens of the UHG implementation: the left figure shows a leaderboard in which different teams compete for a specific health challenge. The top-three teams, displayed with a green bullet, are at positions for a reward. The current app user was in the team with the orange bar, at position 6. This team required 1662 more points to reach a winning position, as indicated by the extended grey bar and numerical label alongside it. The right figure shows the details of the example health challenge. In this example, only physical activities are shown but they do cover besides steps also other distance based sports, where

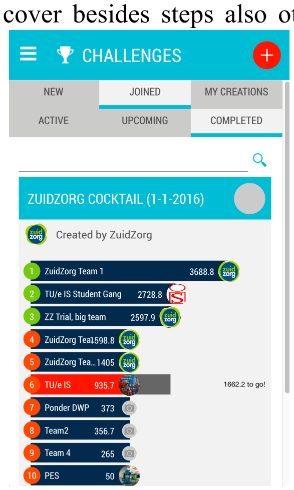

Figure 2: UHG Leaderboard

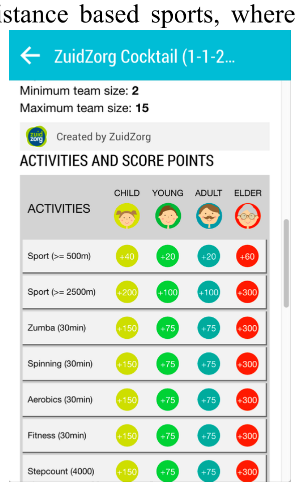

Figure 3: UHG Points Table

[1] See http://www.gamebus.eu/ and https://app.gamebus.eu/ in particular.

any activity covering more than 500 meters is already worth some points. Intense sports activities are worth more points and furthermore the user's age is considered in the points system. The latter is based on SDT (i.e., taking account individual competences). The right figure shows at the top that the example challenge has a minimum team size of 2, thereby making it impossible to join the challenge individually. Team scores are based on averaging team member scores. Challenge design configurations can be changed from within the app, without programming effort. The app provides various other screens, providing an overview of badges, a newsfeed, a chat, and integrations to apps like Runkeeper (and to various brain training apps).

B. *Semi-Controlled Setup: Pilot Studies in Corporations*

In order to have short feedback loops, the mobile application was tested inside corporate environments rather than within society at large. Organizations embracing innovative solutions to well-being were recruited and two pilot locations were found in the region of Eindhoven.

1) *ZuidZorg*

Participants in pilot A were employees of ZuidZorg. ZuidZorg is an extramural healthcare provider in The Netherlands in the area of Home Care, Maternity and Child Health Care, stimulating health, independence and self-reliance of its clients in the neighbourhood. In the context of this paper, the innovation department of this organization was interested specifically in the health and well-being of its own employees, especially during office hours.

2) *GGzE*

Participants in pilot B were forensic psychiatric clients from GGzE. GGzE is a mental healthcare institution in the southern part of the Netherlands. It provides one of The Netherlands' 13 forensic psychiatric clinics and it has a strategic focus on the well-being of its clients. This pilot focused on comorbid clients with addiction problems, next to psychiatric disorders.

C. *Participant Recruitment*

This study aimed at pilot sizes of 10 participants (per pilot), being the lower boundary mentioned by several studies [9]–[13]. Additionally, studies suggest that a pilot study sample should be 9-10% of the projected sample of the main pilot; which is indeed the case of aforementioned pilot locations [14], [15]. Both pilots started with a group session where the functioning of the GameBus app was explained. Participants also received instructions about the study during this kick-off session lasting 1.5 hours.

Equal teams of three participants were formed using the app during a testing period of six weeks. Users were stimulated to complete 11 challenges and compete against each other (within their own team and against other teams) by scoring points for completing different activities using sports apps (e.g., Runkeeper) and brain trainers. Points could also be earned by uploading to the app pictures that demonstrated other specific healthy activities (e.g., drinking water, having a healthy lunch with a colleague, playing table soccer).

At the time of the pilot, GameBus was an Android-based Application which could only be installed on Android devices. Participants who did not have an own Android device were offered a loan Android mobile (smart) phone or an Android Tablet. Participants were supported with the installation process and were given the opportunity to practice and ask questions to dedicated pilot managers. Moreover, pilot A participants were offered an activity wristband tracker (Fitbit Charge HR) to use during the pilot period. All participants were joining voluntarily and they were required to provide informed consent prior to the study.

A random sample of men and women was recruited over a two week period from ZuidZorg Veldhoven and GGzE during the autumn of 2015. A message regarding GameBus and the collaboration of the organization with GameBus was posted publically on the central messaging systems of the pilot organization. Potential participants were approached in order to raise awareness. Eligibility criteria of the participants included (1) the ability to use a smartphone and applications, (2) the ability to do physical, cognitive and social activities, and (3) the willingness to use GameBus. At this stage, one in four of the initially eligible participants was excluded from pilot A and one in two from pilot B.

D. Measurements

In order to measure user perceptions and health effects, we combined focus groups, surveys and personal interviews. The main focus group session was a pilot kick-off session. During this session, we demonstrated the app such that users could already drop out if they would never use it. Personal interviews took place after the pilot, for qualitative feedback.

For scoring a participant's health status, an empirically validated and commonly used 36-item patient-reported short-form (SF-36v2) survey was used [16]–[18]. SF-36 was created to use in clinical practice and research, health policy evaluations, and general population surveys. It is widely recognized and used for measuring an individual's health status [19]–[21]. The items and scales presented in the survey are specifically constructed for scoring using a Likert scale.

SF-36 divides health in physical (PCS) and mental health (MCS). Within those categories, it covers eight health dimensions: physical function (PF), role limitations due to physical problems (RP), bodily pain (BP), general health (GH), vitality (VT), social function (SF), role limitations due to emotional problems (RE) and mental health (MH). The measurements of the first four dimensions directly contribute to the result of physical health, while the latter four are related to mental health [18]. Note that the scores for overall health status and dimensions are scored on a 0 to 100 scale, such that the lowest and highest possible scores are respectively 0 and 100; therefore higher scores correspond to a better health state.

SF-36v2 is an improvement of the 1990 SF-36 version in several ways, such as: shortened items, simplified wording and most importantly providing the possibility to compare own results with the US General Population Norms. We extended the SF-36v2 questionnaire further for the post-survey: it was supplemented with questions regarding usefulness, usability, and intentions to use. For these questions, we used five-point Likert scale items and open items. Furthermore, personal interviews were organized to gain deeper understanding about participants' quality of life, perceived app usability and perceived app usefulness.

E. Data Analysis

All scores for health status and dimensions were determined via the 'QualityMetric Health Outcomes Scoring Software 4.5' software package. One participant did not fill out the post-survey. We have compensated that by setting the participant's score to the average score of other participants [22]. Each score of a SF-36v2 scale with five points or more above or below a previous score is indicated as a significant change. This is based on the fact that the comparison group of the underlying SF-36v2 studies had a mean health score of 50 and a standard deviation of 10 [23], while medium effects in trials typically have the size of half a standard deviation. Looking at minimal clinically important difference (MCID, indicating how big an effect should be for significant health change) this difference is also near 5 points for SF-36v2 [24].

We also calculated MCID values based on "average change"-based anchoring: MCID values were calculated as the averages of the changes in MCS and PCS scores for those participants who had shown Health Transition Item (HTI) progression in the SF-36 pre- and post-surveys. HTI is an appropriate independent anchor since it is not used in the scoring of the SF-36 MCS or PCS elements [25].

Further descriptive and statistical analyses of the SF-36v2 pre- and post-intervention data were performed with SPSS Statistics 22, while Microsoft Excel was used to analyse other quantitative data of non-SF36v2 survey items and qualitative data of performed interviews.

III. RESULTS

A. Participant Recruitment and Survey Completion

1) ZuidZorg

In total, twelve participants were approached to participate in the ZuidZorg pilot. Of these twelve, one did not respond to the invitation to take part in the kick-off session. All eleven other agreed initially to be in the pilot. Two participants (18%) had to drop out of the study and they were not taken into account in the results of this study. This resulted eventually in a remaining of nine participants; male (56%) and female (44%) with an average of 40 (SD= 11.3) and 36 (SD= 10.9) years of age respectively [26].

All participants filled out the survey of the pre-test resulting in a response rate of 100%. This dropped to 89% since one participant did not respond to the post-test survey. Item response rates were perfect in the case of pilot A since all participants answered all questions, resulting in an average item response rate of 100% for the pre- and post-test.

2) GGzE

In total, eighteen participants were approached to participate in the GGzE pilot. Of these eighteen, five declined the opportunity to be in the study; since they were very isolated individuals and did not have the desire to participate in a group activity. Another three were automatically excluded and were not present at the kick-off session due to the status of their treatment. Of the remaining ten participants, all agreed initially to be in the pilot. This eventually resulted in a remaining of nine participants; male (77.8%) and female (22.2%) with an average age of 35.1 (SD= 12.2) for male and 27 (SD= 2.8) for female respectively [26].

Pilot B also resulted in a response rate of 100% for the pre-test and 89% for the post-test of the survey. The average item response rate for both measurements was 100%.

B. Health Effects, User Perceptions, and Intentions

1) Health Status

The sample showed overall statistically significant improvement in the mean mental health (MSC) scale score on aggregate level at pilot B [27]. Physical health (PCS) also increased at both locations on pilot sample level, although not significantly.

Another statistically significant improvement in health status on this level was observed at the Role-Emotional (RE) dimension: decrease of limitations in usual role activities due to emotional problems. Box plotted scales for pre and post-test scores with 95% confidence intervals (CI) further strengthen this finding, since intervals are not overlapping. Due to this result and equal sample sizes there is strong evidence that the difference between the two medians is statistically significant at the 0.05 level [28]. Other changes in normative values of other SF-36v2 scales are presented in the tabular overview of the online supplement [27].

On the individual participant level significant effects were also observable for both pilots. PCS scores increased for 17% of total participants, ranging from a minimum of 13 to maximum of 22 points. In 28% of the cases the MCS score increased significantly ranging from 10 to 19 points (see *Table 1*). The HTI anchor-based MCID values for PCS and MCS turned out to be 1.99 and -0.66 for pilot A and 10.77 and 1.46 for pilot B, respectively. These values confirmed the participant-level improvements that were identified by our statistical analyses. Conversely however, MCS improvement cases based on the MCID value for MCS at pilot A were rejected since they were disqualified by statistical analyses.

Internal consistency reliability estimates (Cronbach's coefficient alpha) for the eight SF-36v2 scales were determined. Most of the scales that contribute to the scoring of the eight different health concepts met or exceeded the recommended 0.70. Several scored excellent, being close to or higher than 0.90 (e.g., for RE: $\alpha = 0.970$ and for RP: $\alpha = 0.938$) [29]. Other scales – RP and BP (pilot A) and SF (pilot B – post) – were questionable. Their low alpha values may require more related items in the survey than the currently 4, 2 and 2 respectively to test the same concept [30].

2) Usefulness of Concept

After the kick-off session, one participant dropped out of pilot A due to unwillingness to perform physical activities and due to not perceiving the application as sufficiently useful. Furthermore, one participant dropped out of pilot B due to lacking affinity with technology in general and electronic devices in particular. All remaining participants perceived GameBus useful at this stage.

Perceived usefulness measurement during the post-test survey resulted in 100% positive feedback for pilot A, while 50% plus 38% of pilot B participants were still positive and moderately positive regarding the concept, respectively. 50% of pilot A participants felt that the app motivated them extremely to exercise more and work on mental health, while the other 50% evaluated the app to be slightly motivational.

Additionally, qualitative analyses of pilot A interviews resulted in a further explanation of the high aforementioned score [31]: all participants reported to be more aware of their physical and mental health, and they acknowledged changes in their behavior. This was illustrated by an increase in physical activity during working hours: various participants took the stairs instead of elevators, they walked to colleagues instead of calling them and they monitored their fluid intake consciously. Most participants reduced their coffee intake during working hours while others switched to water completely. After work, participants reported to go, on average, more often to the gym per week and they kept up with cognitive activities by solving several puzzles each day.

3) Usability of Concept

Conducted interviews and the post-test survey resulted in suggestions for improvement of the application usability. Both groups reported the same items as the most important ones [31]: two out of eight pilot A participants and five out of eight pilot B participants experienced unpleasant app loading times which had a slight impact on their motivation. Other participants did not find this to be an issue for decreased usability. One participant of each pilot (12.5%) mentioned the need for a manual, since the app was mentioned to be non-intuitive at times and it lacked an in-app tutorial. Compatibility was also discussed: being solely an Android application, availability on Apple hardware and office computers was requested. This was not necessarily seen as a decrease in usability, but rather an opportunity to increase it.

4) Intentions to Use

After the six week intervention 75% of pilot A' sample was willing and interested to continue using GameBus during and after office hours. The remainder was not sure due to the status of the app at that time. However, 63% of this group still recommended GameBus to other colleagues, friends and family. Others would consider it after solutions to usability limitations are implemented.

TABLE I. INDIVIDUAL NORM BASED SF-36V2 SCORES

#	Individual Norm based SF-36v2 scores [a]							
	Trial A				*Trial B*			
	Pre: PCS	*Post: PCS*	*Pre: MCS*	*Post: MCS*	*Pre: PCS*	*Post: PCS*	*Pre: MCS*	*Post: MCS*
1	56.27	60.45	58.71	57.57	**44.29**	**58.80**	56.00	50.25
2	47.91	48.87	**41.88**	**52.13**	51.26	39.64	50.26	41.47
3	54.65	54.64	44.58	45.97	**33.48**	**56.36**	**36.83**	**55.98**
4	56.95	56.95	48.58	48.58	49.63	48.27	**41.49**	**59.26**
5	56.08	55.58	28.61	28.86	40.50	44.01	**46.37**	**60.91**
6	57.08	57.67	57.00	53.84	60.24	52.67	51.93	51.56
7	57.21	57.81	58.27	58.24	61.82	54.98	**25.59**	**44.97**
8	52.71	55.91	51.79	52.07	**47.52**	**60.79**	55.68	48.48
9	51.06	55.99	54.13	49.66	59.44	51.94	50.97	51. 61
Mean	54.44	55.99	49.28	49.66	49.80	51.94	46.12	51.61
SD	3.24	3.16	9.77	8.75	9.60	6.91	9.95	6.35

a. Dark colored cells indicate significant improvement on individual level

IV. Conclusions and Outlook

This study reported on the first scientific study that encompassed the unified gamification of activities from all health dimensions (physical, social, cognitive). The mobile application encourages smartphone users to regularly participate in challenges and gradually increase their health status, especially when it comes to one's mental health (MH) and emotional role health dimension (RE). Since both pilots involved small sample sizes, the significance of these results inspires follow-up research. Preliminary results in controlled corporate environments demonstrate the usefulness of such an approach, with a positive rate of 100% in pilot A and 88% in pilot B, measured after the use of GameBus.

Since the completion of the pilots, various improvements have been made in order to increase the usability of the app. 25% of pilot A and 50% of pilot B participants saw the same opportunities for changes. Since prototype app loading times were due to network latencies, they have been addressed by means of caching. Furthermore, the app was released for iOS devices and for all mainstream browsers.

In our ongoing work, we are handling more recent feedback related to usability details and we are preparing larger pilots with a target sample size of 500 in order to increase the statistical power and generate deeper insights. We are also investigating detailed challenge design trade-offs. For example, we are investigating what is the preferred duration of a challenge and what are preferable team sizes. Additionally, additional data sources are being connected via the app's Open API. Finally, we are preparing tutorials to enable the use of the app in less controlled settings.

V. References

[1] L. G. Glynn, P. S. Hayes, M. Casey, F. Glynn, A. Alvarez-Iglesias, J. Newell, G. Ólaighin, D. Heaney, M. O'Donnell, and A. W. Murphy, "Effectiveness of a smartphone application to promote physical activity in primary care: The SMART MOVE randomised controlled trial," *Br. J. Gen. Pract.*, vol. 64, no. 624, pp. 384–391, 2014.

[2] T. Harries, P. Eslambolchilar, R. Rettie, C. Stride, S. Walton, and H. C. van Woerden, "Effectiveness of a smartphone app in increasing physical activity amongst male adults: a randomised controlled trial," *BMC Public Health*, vol. 16, no. 1, p. 925, 2016.

[3] J. C. Walsh, T. Corbett, M. Hogan, J. Duggan, and A. McNamara, "An mHealth Intervention Using a Smartphone App to Increase Walking Behavior in Young Adults: A Pilot Study.," *JMIR mHealth uHealth*, vol. 4, no. 3, p. e109, 2016.

[4] K. Seaborn and D. I. Fels, "Gamification in theory and action: A survey," *Int. J. Hum. Comput. Stud.*, vol. 74, pp. 14–31, 2015.

[5] E. L. Deci, "Effects of externally mediated rewards on intrinsic motivation.," *J. Pers. Soc. Psychol.*, vol. 18, no. 1, pp. 105–115, 1971.

[6] F. P. Grad, "Premable of the Constitution of the World Health Organization," *Bull. World Health Organ.*, vol. 80, no. 12, pp. 981–982, 2002.

[7] A. F. Aparicio, F. L. G. Vela, J. L. G. Sánchez, and J. L. I. Montes, "Analysis and application of gamification," in *Proceedings of the 13th International Conference on Interacción Persona-Ordenador - INTERACCION '12*, 2012, pp. 1–2.

[8] R. M. Ryan, C. S. Rigby, and A. Przybylski, "The motivational pull of video games: A self-determination theory approach," *Motiv. Emot.*, vol. 30, no. 4, pp. 347–363, 2006.

[9] M. A. Hertzog, "Considerations in determining sample size for pilot studies," *Res. Nurs. Heal.*, vol. 31, no. 2, pp. 180–191, 2008.

[10] G. van Belle, *Statistical Rules of Thumb*, 2nd ed. New York: Wiley-Interscience, 2002.

[11] R. Hill, "What Sample Size is 'Enough' in Internet Survey Research," *Interpers. Comput. Technol. An Electron. J. 21st Century*, vol. 6, pp. 3–4, 1998.

[12] S. Isaac and W. Michael, *Handbook in Research and Evaluation*, 3rd editio. Edits, 1995.

[13] S. Julious, "Sample size of 12 per group rule of thumb for a pilot study," *Pharm. Stat.*, vol. 4, no. 4, pp. 287–291, 2005.

[14] K. Cocks and D. J. Torgerson, "Sample size calculations for pilot randomized trials: A confidence interval approach," *J. Clin. Epidemiol.*, vol. 66, no. 2, pp. 197–201, 2013.

[15] L. M. Connelly, "Pilot studies," *MedSurg Nurs.*, vol. 17, no. 6, pp. 411–412, 2008.

[16] N. K. Aaronson, M. Muller, P. D. a Cohen, M. L. Essink-Bot, M. Fekkes, R. Sanderman, M. a G. Sprangers, A. Te Velde, and E. Verrips, "Translation, validation, and norming of the Dutch language version of the SF-36 Health Survey in community and chronic disease populations," *J. Clin. Epidemiol.*, vol. 51, no. 11, pp. 1055–1068, 1998.

[17] J. F. Scoggins and D. L. Patrick, "The use of patient-reported outcomes instruments in registered clinical trials: Evidence from ClinicalTrials.gov," *Contemp. Clin. Trials*, vol. 30, no. 4, pp. 289–292, 2009.

[18] J. E. Ware and C. D. Sherbourne, "The MOS 36-Item Short-Form Health Survey (SF-36) I. Conceptual Framework and Item Selection," *Source Med. Care Med. CARE*, vol. 30, no. 6, pp. 473–483, 1992.

[19] C. Macias, T. Panch, Y. M. Hicks, J. S. Scolnick, D. L. Weene, D. Öngür, and B. M. Cohen, "Using Smartphone Apps to Promote Psychiatric and Physical Well-Being," *Psychiatr. Q.*, vol. 86, no. 4, pp. 505–519, 2015.

[20] R. Wang, C. Wu, Y. Zhao, X. Yan, X. Ma, M. Wu, W. Liu, Z. Gu, J. Zhao, and J. He, "Health related quality of life measured by SF-36: a population-based study in Shanghai, China.," *BMC Public Health*, vol. 8, p. 292, 2008.

[21] G. Kolt, R. Rosenkranz, T. Savage, A. Maeder, C. Vandelanotte, M. J. Duncan, C. M. Caperchione, R. Tague, C. Hooker, and W. Mummery, "WALK 2.0 - using Web 2.0 applications to promote health-related physical activity: a randomised controlled trial protocol.," *BMC Public Health*, vol. 13, no. 1, p. 436, 2013.

[22] J. E. J. Ware, M. Kosinski, and J. E. Dewey, *How to score version Two of the SF-36 health survey*. Lincoln, RI: QualityMetric Incorporated, 2000.

[23] S. S. Farivar, H. Liu, and R. D. Hays, "Half standard deviation estimate of the minimally important difference in HRQOL scores?," *Expert Rev. Pharmacoecon. Outcomes Res.*, vol. 4, no. 5, pp. 515–523, 2004.

[24] S. L. Parker, S. K. Mendenhall, D. N. Shau, O. Adogwa, W. N. Anderson, C. J. Devin, and M. J. McGirt, "Minimum clinically important difference in pain, disability, and quality of life after neural decompression and fusion for same-level recurrent lumbar stenosis: understanding clinical versus statistical significance.," *J. Neurosurg. Spine*, vol. 16, no. 5, pp. 471–478, 2012.

[25] B. M. Auffinger, R. R. Lall, N. S. Dahdaleh, A. P. Wong, S. K. Lam, T. Koski, R. G. Fessler, and Z. A. Smith, "Measuring Surgical Outcomes in Cervical Spondylotic Myelopathy Patients Undergoing Anterior Cervical Discectomy and Fusion: Assessment of Minimum Clinically Important Difference," *PLoS One*, vol. 8, no. 6, pp. 8–13, 2013.

[26] A. Shahrestani, "GameBus Pilot Studies (2): Characteristics of study participants in both trials," 2016. [Online]. Available: http://bit.ly/2ktOijq. [Accessed: 20-Oct-2016].

[27] A. Shahrestani, "GameBus Pilot Studies (4): Normative values of the SF-36v2 dimension scores," 2016. [Online]. Available: http://bit.ly/2kvRQRT. [Accessed: 20-Oct-2016].

[28] R. Mcgill, J. W. Tukey, and W. A. Larsen, "Variations of Box Plots," *Am. Stat.*, vol. 32, no. 1, pp. 12–16, 1978.

[29] A. Shahrestani, "GameBus Pilot Studies (6): Internal Consistency Reliability - Chronbach's Alpha," 2016. [Online]. Available: http://bit.ly/2kZOPGk. [Accessed: 20-Oct-2016].

[30] D. L. Streiner, "Starting at the Beginning: An Introduction to Coefficient Alpha and Internal Consistency," *J. Pers. Assess.*, vol. 80, no. 1, pp. 99–103, 2003.

[31] A. Shahrestani, "GameBus Pilot Studies (3): Coding of Participant Interviews," 2016. [Online]. Available: http://bit.ly/2kZVD6T. [Accessed: 20-Oct-2016].