

UNIVERSITY CAPE TOWN

DOCTORAL THESIS

Utilization of Personal Health Informatics Through Intermediary Users

Author:

Ntwa KATULE

Supervisors:

Dr. Melissa DENSMORE and

Prof. Ulrike RIVETT

*A thesis submitted in fulfilment of the requirements
for the degree of Doctor of Philosophy*

in the

Research Group Name

Department of Computer Science

June 2016

Declaration of Authorship

I, Ntwa KATULE, declare that this thesis titled, 'Utilization of Personal Health Informatics Through Intermediary Users ' and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

Date:

“Thanks to my solid academic training, today I can write hundreds of words on virtually any topic without possessing a shred of information, which is how I got a good job in journalism.”

Dave Barry

UNIVERSITY NAME (IN BLOCK CAPITALS)

Abstract

Faculty Name

Department of Computer Science

Doctor of Philosophy

Utilization of Personal Health Informatics Through Intermediary Users

by Ntwa KATULE

The Thesis Abstract is written here (and usually kept to just this page). The page is kept centered vertically so can expand into the blank space above the title too...

Acknowledgements

The acknowledgements and the people to thank go here, don't forget to include your project advisor...

Contents

Declaration of Authorship	i
Abstract	iii
Acknowledgements	iv
Contents	v
List of Figures	vi
List of Tables	vii
Abbreviations	viii
Physical Constants	ix
Symbols	x
1 Contextual Enquiry	1
1.1 Study Description	1
1.2 Data Collection Methods	1
1.3 Findings	1
2 Prototype I	2
2.1 Development of the Prototype	2
2.2 Prototype Evaluation Description	2
2.3 Prototype Evaluation Methods	2
2.4 Findings	2
3 Prototype II	3
3.1 Prototype Development Iteration II	3
3.2 Prototype Evaluation Description	3
3.3 Prototype Evaluation Methods	3
3.4 Findings	3

4	Summative Evaluation	4
4.1	Recruitment of Participants	4
4.2	Experiments	5
4.2.1	Experiment Design	5
4.2.2	Data Collection Methods	6
4.2.2.1	Family Wellness App Logs	6
4.2.2.2	Questionnaires	6
4.2.2.3	Interviews	8
4.3	Findings	8
4.3.1	Analysis of Usage Trend	8
4.3.2	User Experience of Intermediaries	16
4.3.3	Beneficiaries' IMI Scores on Health Self-Monitoring	22
4.3.3.1	IMI in Self-Monitoring of Diet	23
4.3.3.2	IMI in Self-Monitoring of Activity	25

Bibliography	28
---------------------	-----------

List of Figures

4.1	Total daily number of sessions from the two experimental conditions. . . .	9
4.2	Total clicks on feedback features for self-monitoring of wellness: “Logbook App” versus “Gamified App”.	13
4.3	Total clicks on feedback features in the two experimental conditions : self-monitoring features versus rewards features.	13
4.4	Average number of logbook sessions on 14 intermediaries: Age \geq median age(=15.5) versus Age $<$ median age.	14
4.5	Average number of gamification sessions on 14 intermediaries: Age \geq median age(=15.5) versus Age $<$ median age.	15
4.6	Average of total number of sessions of 14 intermediaries on both experimental conditions: Age \geq median age(=15.5) versus Age $<$ median. . . .	15
4.7	Average number of logbook sessions on 12 intermediaries: Age \geq median age(=15.5) versus Age $<$ median age.	16
4.8	Average number of gamification sessions on 10 intermediaries: Age \geq median age(=15.5) versus Age $<$ median age.	16
4.9	Intermediaries’ average perceived enjoyment to help others with cellphone tasks versus age group.	17
4.10	Intermediaries’ average perceived enjoyment in using the app versus age group.	18
4.11	Intermediaries’ average perceived enjoyment in using the app versus age group (Logbook and Gamification).	19
4.12	Trend on Average IMI Scores of Self-Monitoring of Diet at Baseline, Midline, and Endline.	24
4.13	Trend on Average IMI Scores of Self-Monitoring of Diet at Baseline, Logbook, and Gamification.	25
4.14	Trend on Average IMI Scores of Self-Monitoring of Activity at Baseline, Logbook, and Gamification.	26
4.15	Trend on Average IMI Scores of Self-Monitoring of Activity at Baseline, Logbook, and Gamification.	27

List of Tables

4.1	Daily usage comparison between Logbook and Gamified systems for 41 days	10
4.2	Usage comparison between Logbook and Gamified systems for 14 pairs of users	10
4.3	Pairs with usability/technical problems that hinder their participation . .	11
4.4	Usage on days comparison between Logbook and Gamified systems for 10 pairs of users	12
4.5	Usage on sessions comparison between Logbook and Gamified systems for 10 pairs of users	13
4.6	Age groups of intermediary participants	15
4.7	Age groups of intermediary participants	17
4.8	Pairs affected by poor design of gamification	21
4.9	Comparison of ten intermediaries' scores on sub-scales of competence, autonomy, enjoyment, and relatedness in using the "Family Wellness App	22
4.10	Comparison of ten beneficiaries' IMI scores in self-monitoring of diet at baseline, midline and endline	23
4.11	Pairwise comparisons of IMI scores in self-monitoring of diet: Baseline versus Midline	24
4.12	Pairwise comparisons of IMI scores in self-monitoring of diet: Baseline versus Endline	24
4.13	Pairwise comparisons of IMI scores in self-monitoring of diet: Midline versus Endline	24
4.14	Comparison of ten beneficiaries' IMI scores in self-monitoring of diet at baseline, after logbook, and after gamification conditions	25
4.15	Comparison of ten beneficiaries' IMI scores in self-monitoring of activity at baseline, midline and endline	26
4.16	Comparison of ten beneficiaries' IMI scores in self-monitoring of activity at baseline, logbook and gamification	27

Abbreviations

LAH List Abbreviations **Here**

Physical Constants

Speed of Light $c = 2.997\,924\,58 \times 10^8 \text{ ms}^{-\text{s}}$ (exact)

Symbols

a	distance	m
P	power	W (Js^{-1})
ω	angular frequency	rads^{-1}

For/Dedicated to/To my...

Chapter 1

Contextual Enquiry

1.1 Study Description

- Study Location:
- Recruitment of Participants:
- Participants Demography

1.2 Data Collection Methods

1.3 Findings

Chapter 2

Prototype I

2.1 Development of the Prototype

2.2 Prototype Evaluation Description

- Location:
- Recruitment of Participants:
- Participants Demography

2.3 Prototype Evaluation Methods

2.4 Findings

Chapter 3

Prototype II

3.1 Prototype Development Iteration II

3.2 Prototype Evaluation Description

- Location:
- Recruitment of Participants:
- Participants Demography

3.3 Prototype Evaluation Methods

3.4 Findings

Chapter 4

Summative Evaluation

4.1 Recruitment of Participants

With help of a research assistant who was a resident of Langa, we managed to recruit a total of fourteen adult participants (beneficiary users). We recruited these participants from two townships in Cape Town: Langa, and Athlone. In Langa there were six adult participants while in Athlone there were nine adult participants. The average age of these adult participants was 44.21 years with a standard deviation (S.D) of 9.99 years. The youngest adult was 26 years of age while the oldest was 60 years of age. Thirteen participants were females.

Each adult participant (beneficiary user) elected one of their children/grand children to become their intermediary user to form a pair of users. The two members of a pair were required to work together in using the “Family Wellness App” to self-monitor the wellness of one member of a pair (a beneficiary user). All beneficiary users were working with their children but one whom was working with her grand child. The average age of children participants (intermediary users) was 15.42 (S.D=2.06) years. The youngest intermediary user was 12 years of age while the oldest was 20 years of age. The number of females and males intermediary users were equal.

I gave detailed information of what the study was all about to both intermediary and beneficiary participants. I informed them about different modes of which I will collect data. All beneficiary participants signed informed consent forms agreeing to be part of the study. Since all intermediaries were under 21 years of age, they signed assent forms which were also signed by their parents/guardians who were part of the study.

I allocated one day to teach intermediary participants on how to use the “Family Wellness App”. In addition, each intermediary was given a user manual. After the training, I

gave out one Android phone (Samsung GT-S5300) to each pair of participants. These phones were installed with two natives apps. The first app was a pedometer and the second one was the main “Family Wellness App”. The “Family Wellness App” loaded all its content from a web application hosted remotely. The app was used for a total period of six weeks. Each pair of participants provided the service provider’s number of the SIM card that was inserted on their given Android phone. I allocated 1.3 GB of data to each SIM card. In addition each beneficiary participant was given a total of ZAR 240 as a compensation for transport and their time for the duration of the study. The details of the experiments are outlined on the next section.

4.2 Experiments

This phase of the study evaluated the effectiveness of gamification/rewards in motivating both intermediaries and beneficiaries to engage with the “Family Wellness App”. I was comparing two versions of the applications. The first version of the application was simply a logbook or journal that allows each pair of users to record and view wellness data of a beneficiary member of the pair. With the logbook app users could view physical activity feedbacks and recording and viewing summaries of nutrition components of food consumed by a beneficiary within a pair. The second version of the application was an extension of logbook to include rewards/gamified subsystem. I carried out this evaluation for a period of six weeks. I carried out the experiments from the mid-October 2015 to the end of November 2015. The details of how experiments were designed and how data were collected are presented on the next sub-sections.

4.2.1 Experiment Design

I used “within-group” design for the experiments. In within-group design, the same group of participants is exposed to different experimental conditions. This helps to minimize the number of groups needed to test hypotheses as only one group is used for both control and intervention. Another advantage of within-group design is that it minimizes the effect of confounding factors. The only problem with this approach is the learning effect and it lengthens the duration of the study. In order to minimize the impact of the learning effect on the outcome, I randomly assigned pairs of participants to two separate groups referred to as experimental sequences. The first experimental sequence started with the “Logbook App” and finished with the “Gamified App”. The second experimental sequence started with the “Gamified App” and finished with the “Logbook App”. I used the following abbreviations “LG” and “GL” to refer to the first and second experimental sequences respectively.

A total of seven pairs of participants were assigned to the LG group while the remaining seven pairs were assigned to the GL group. Both groups spent the first four weeks in their first experimental conditions of which the Logbook App for the LG group and the Gamified App for the GL group. After 27 days each group was switched to a different experimental condition. The LG group started using the Gamified App while the GL group started using the Logbook App. The second phase of the experiment lasted for a total of 14 days. In the next sub-section, I provide details of how data were collected during the duration of 41 days (6 weeks) of running the experiments.

4.2.2 Data Collection Methods

Data collection was a triangulation of application's logs, questionnaires and interviews.

4.2.2.1 Family Wellness App Logs

Application's logs consisted of information regarding the time when there were users' activities on the app, the pair that was accessing the app at that time, and the functionality that was being accessed by that pair. Logs were categorized to their respective experimental condition.

4.2.2.2 Questionnaires

I administered questionnaires at the baseline, mid-line (during switching of experimental conditions), and end-line. The list of questionnaires is provided below.

Baseline Questionnaires

At baseline both intermediary and beneficiary participants filled their respective questionnaires.

- **Intermediaries:** Intermediaries participants' baseline questionnaire had three sections. The first section captured demographic information such as age, gender, and number services/apps used on cellphones. The second section included an IMI (Intrinsic Motivation Inventory) questionnaire to assess participants' intrinsic motivation in using cellphones. The third section included an IMI questionnaire to assess participants' intrinsic motivation in helping their parents with cellphone based tasks.

- **Beneficiaries:** Beneficiary participants' baseline questionnaire had four sections. The first section captured demographic information such as age, gender, and number services/apps used on cellphones. The second section included an IMI questionnaire to assess participants' intrinsic motivation in using cellphones. The third section included an IMI questionnaire to assess participants' intrinsic motivation in self-monitoring of diet/nutrition. The fourth section included an IMI questionnaire to assess participants' intrinsic motivation in self-monitoring of physical activity.

Midline Questionnaires Also at midline both intermediary and beneficiary participants filled their respective questionnaires.

- **Intermediaries:** Intermediaries participants' midline questionnaire had only one section which included an IMI questionnaire to assess participants' intrinsic motivation in using the family wellness app.
- **Beneficiaries:** Beneficiary participants' baseline questionnaire had four sections. The first section included an IMI questionnaire to assess participants' intrinsic motivation in using the family wellness app. The third section included an IMI questionnaire to assess participants' intrinsic motivation in self-monitoring of diet/nutrition. The third section included an IMI questionnaire to assess participants' intrinsic motivation in self-monitoring of physical activity.

Endline Questionnaires At endline both intermediary and beneficiary participants filled their respective questionnaires.

- **Intermediaries:** Intermediaries participants' endline questionnaire had only one section which included an IMI questionnaire to assess participants' intrinsic motivation in using the family wellness app.
- **Beneficiaries:** Beneficiary participants' endline questionnaire had three sections. The first section included an IMI questionnaire to assess participants' intrinsic motivation in using the family wellness app. The third section included an IMI questionnaire to assess participants' intrinsic motivation in self-monitoring of diet/nutrition. The third section included an IMI questionnaire to assess participants' intrinsic motivation in self-monitoring of physical activity.

I developed the IMI questionnaires with guidance of materials found on a "Self-Determination Theory"¹ website which is maintained by researchers working on the theory including

¹<http://www.selfdeterminationtheory.org/intrinsic-motivation-inventory/>

Richard Ryan and Edward Deci (Deci and Ryan, 1985) whom were early pioneers in developing the theory. I pretested these questionnaires during the informative evaluation of prototype II in chapter 3. The most most important sub-scales for our theoretical construct were perceived competence and perceived autonomy which are part of the three basic psychological needs. The relatedness sub-scale is not yet validated but it was included in all questionnaires. Other sub-scale that was included in all questionnaires is perceived enjoyment. In each question from the IMI sub scales, respondents were supposed to rate there experience in a scale of 1 to 7 points which means that 1 implies the statement is "not true at all" and 7 means the statement is "very true". Perceived enjoyment is the only direct measure of intrinsic motivation while perceived competence and perceived autonomy are predictors of intrinsic motivation. Self-Determination theory suggests that a behaviour can be started as externally motivated and if external motivators support the three basic psychological needs which are relatedness, competitiveness, and autonomy then a behaviour that was once externally motivated can be internalized and users will start doing it because it is a good thing to do.

4.2.2.3 Interviews

I also conducted short unstructured interviews at midline and endline. I selected fewer intermediaries and beneficiaries for the interviews. Interviews responses were important in supplementing data collected through questionnaires and application's logs.

4.3 Findings

There were four primary outcomes in analysing the findings and these are: (1) usage trend of the app; (2) user experience/intrinsic motivation of both intermediaries and intermediaries in using the app; (3) intrinsic motivation of beneficiaries in self-monitoring of diet/nutrition; and (4) intrinsic motivation of beneficiaries in self-monitoring of physical activity.

4.3.1 Analysis of Usage Trend

The average number of days on which pairs used both versions of the application was 10.5 (SD=7.39) days. The most active usage was from a pair that utilized the app for a total of 26 days. The less active usage was from a pair that had used the app for only two days out of 41 days.

I contrasted usage in each day between the "Logbook App" and "Gamified App" by

computing the daily total number of sessions from all users in each of the two experimental conditions. The beginning of a new session is defined as the a continuous period of time where the end user interacts with the app without interruption. This period is whereby there is user’s activities detected on the app while there was no user activities for the past one hour. If delay between activities is less than one hour it is then assumed that the last session is still active. If a user comes back after one hour has elapsed since the last detected user’s activity then it is assumed they went away from the app and now they are coming back for a new session. There were 41 days of app usage in total. Therefore, each experimental condition had a total number of sessions for each day as shown on Figure 4.1. After the fourth week there was a decline in the daily total number of sessions in each of the two experimental conditions as the result of the learning effect. The decline was more apparent to users who where switched from gamification to logbook since the switching happened at the end of the fourth week of usage.

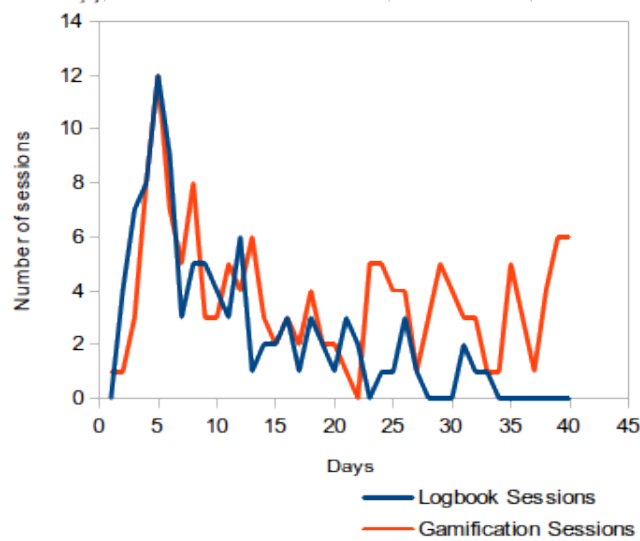


FIGURE 4.1: Total daily number of sessions from the two experimental conditions.

I expanded the comparison of Figure 4.1 by carrying out an inferential statistical test in order to determine if the daily total of number of sessions differs in between logbook and gamification conditions. Before deciding on which test to use, I applied “*Shapiro-Wilk Normality Test*”² to test for normal distribution in both daily logbook and gamification sessions. The distributions of the two independent samples were both not normal. An attempt to transform the data using transformation algorithms didn’t yield to normal distributions. Therefore, I made a decision to use a non-parametric test called Mann-Whitney U Test. The daily number of sessions in gamification condition were significantly higher than in logbook condition as shown on Table 4.1.

²<http://sdittami.altervista.org/shapirotest/ShapiroTest.html>

TABLE 4.1: Daily usage comparison between Logbook and Gamified systems for 41 days

Groups	N	Rank Average	Sum Ranks	U	Z	P
Daily logbook sessions	41	33.72	1701.5	1159.5	-2.9538	0.00318
Daily gamification sessions	41	49.28	1701.5			

I conducted another usage comparison which entailed contrasting ‘Logbook’ usage versus ‘Gamification’ usage for each pair as the aforementioned analyses were general not specifically examining on the performance of individual pairs and its results are shown on Table 4.2

TABLE 4.2: Usage comparison between Logbook and Gamified systems for 14 pairs of users

Mean	Logbook App	Gamified App
Ratio of days	M=0.187;SD=0.19	M=0.284 ;SD=0.234
	t(13)= 1.4714 ; p=0.1650 ; 95% CI= -0.239 to 0.0454	
Number of sessions/day	M=0.271 ;SD=0.3	M=0.51;SD=0.55
	t(13)=1.5768 ; p= 0.1388 ; 95% CI= - -0.565 to 0.088	

This is a step by step of how I arrived to results on Table 4.2. The first step was to count the followings for each pair of users in each experimental condition: (1) the total number of days; and (2) total number of sessions. Since the total number of days and sessions were relative to how many days users had a system with particular experimental condition available to them, I divided the number of sessions with total days on which users had a particular experimental condition available to them. The outcome is both the ratio of days with respect to the total number of days and number of sessions per day. Therefore, the comparison was based on a relative number and not an absolute number. For instance, lets say pair one is from a LG group which they were allowed to use a logbook system for 27 days and a gamified system for 14 days. Suppose the number of sessions in logbook was 50 and the number of sessions in Gamified system was 30 then the relative number of sessions per day spent on logbook is going to be “50 divide by 27 days” and the relative number of sessions spent on a gamified system is going to be “30 divide by 27. Before comparing the two versions of the system on usage, I had to test if the differences of both the ratio of days and number of sessions

per day between the two systems followed a normal distribution in order to identify a proper statistical test to use. The difference between logbook and gamification had a normal distribution in both the ratio of days and number of sessions per day, therefore I conducted a paired t-test, and its results are what I presented on the aforementioned Table 4.2 . On comparison of usage between a gamified system and logbook system, the former had a higher mean of ratio of days and mean of number of sessions per day compared to the latter without statistical significance.

There were four pairs who faced hurdles on utilizing the app and this affected both their motivation and ability to participate or access the system. These pairs are listed on Table 4.3.

TABLE 4.3: Pairs with usability/technical problems that hinder their participation

	Pair	Experimental Sequence	Problem
1	Pair A	GL group	App not loading
2	Pair B	GL group	Lack of data bundles.
3	Pair C	LG group.	Pedometer never transmitted data to the server.
4	Pair D	LG group.	Pedometer stopped transmitting data to the server.

For **Pair A**, the app failed to load every time their intermediary user tried to use it. What I observed from the house where this pair lived in is that there was a poor Internet signal hence the app was always failing to load most of the time. The second pair (**Pair B**), data was allocated to the wrong phone number at the beginning of experiments but they never reported on time. These two pairs (Pair A and Pair B) had the lowest usage days which were 2 and 3 days respectively and they had used the app only in gamification condition. The last two pairs (**Pair C** and **Pair D**) on Table 4.3 had their pedometers not transmitting steps' data that were important in gamification for advancement of badges, and improvement of both the fish tank and the botanical garden. Pair C's pedometer never transmitted any readings to the server even in logbook condition. Pair D's pedometer stopped transmitting steps data on the fourth week of running the experiments and this was before this particular user was switched to gamification condition. The two intermediary users from Pair C and Pair D were close friends. Although the pedometer never transmitted data in Pair C, an intermediary user from this pair continued to use the wellness app because of the informal comparison with an intermediary user from Pair D while they were still in logbook condition. The usage in Pair C and Pair D were both eleven days. Their drop out started during gamification condition. Pair C used the app for only three days and Pair D used it for only one day.

Since gamification depended on transmission of steps to the server, pedometer problems affected Pair C and Pair D motivations to participate in gamification phase despite their efforts during logbook condition. The learning effect coupled with problems with their pedometers mediated their decrease in usage with the app. An intermediary user from another pair in "LG" group who happened to live close to the Pair D and Pair C, shared her concerns about the rewards from the gamified app. This was during the endline interviews. This particular intermediary user didn't appreciate her advancement in badges because she admitted that her peers (the two intermediary users from Pair C and Pair D) did more efforts than her but they were not getting anything so she didn't understand why she was ahead of them. She was referring to their usage during logbook condition as they were both using the "logbook App" at the same time. This proves that usability problems played a role to the some extent in demotivating participation in gamification condition of intermediary users from Pair B and Pair C.

I repeated the above analysis on usage (Table 4.2) without the pairs on Table 4.3. Therefore, the new analysis has only ten pairs of users. The differences on ratio of days had a normal distribution shape. A student t test demonstrated that mean ratio of days was significantly higher in the gamified system compared to the logbook system as shown on Table 4.4. The differences on number of sessions per day didn't have a normal distribution shape. I transformed the data using the natural log equation 4.1. The differences between transformed logbook and gamification data followed normal distribution. I performed a paired t test on the transformed data and the results showed that the Log mean of number of sessions per day was significantly higher on gamification condition when compared to logbook condition as shown on Table 4.5.

$$y = \log(x + 1) \quad (4.1)$$

TABLE 4.4: Usage on days comparison between Logbook and Gamified systems for 10 pairs of users

Mean	Logbook App	Gamified App
Ratio of days	M=0.18;SD=0.18	M=0.35 ;SD=0.25
	t(9)= 2.3490 ; p=0.0434 ; 95% CI= -0.334 to -0.0063	

Analysis from Table 4.5 suggests that there was an indication of an increase in frequency of daily usage when pairs were in gamification condition.

I further examined the utilization of different self-monitoring functionality as shown on Figure 4.2. One interesting trend is that during gamification condition, usage of the main self-monitoring features for steps and diet decreases compared to when users

TABLE 4.5: Usage on sessions comparison between Logbook and Gamified systems for 10 pairs of users

Log mean	Logbook App	Gamified App
Number of sessions/day	M=0.201 ;SD=0.196	M=0.459;SD=0.336
	t(9)= 2.6593 ; p= 0.0261 ; 95% CI= -0.477 to -0.039	

where in Logbook condition. This happens as users started to spend time on virtual rewards of where they were getting indirect feedbacks of steps and diet through badges, gardens etc. However, the trend in recording of diet/meals continued to remain the same between the two experimental conditions as the process of recording meals played an important role in earning some of the virtual rewards, therefore, intermediary users had to continue using that feature while in gamification condition. Figure 4.3 shows the distribution of clicks among feedback features of the "Gamified Wellness App" and "The Logbook App".

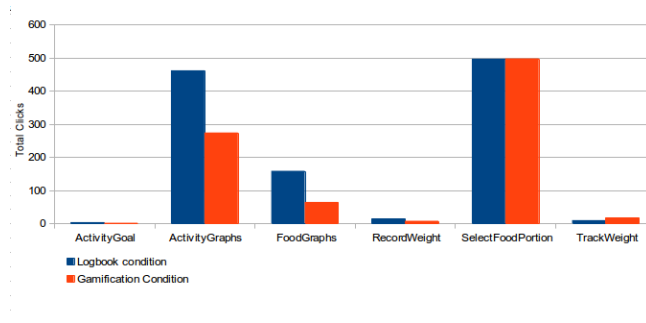


FIGURE 4.2: Total clicks on feedback features for self-monitoring of wellness: "Logbook App" versus "Gamified App".

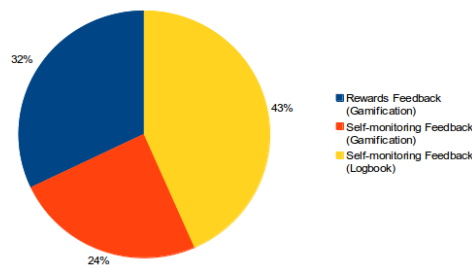


FIGURE 4.3: Total clicks on feedback features in the two experimental conditions : self-monitoring features versus rewards features.

From Figure 4.3, we can see that gamification condition's clicks on feedback features are distributed between the main self-monitoring and virtual rewards features. The number

of clicks on the main self-monitoring feedback features had decreased as users became interested to feedbacks provided by virtual rewards features more than the main self-monitoring features. This explanation also supplements the reason why usage on Figure 4.1 above shows the drop on Logbook condition as users who had been exposed gamification condition before logbook condition became less interested with the "Logbook App" after being switched to it. I expanded on this usage phenomenon by examining through baseline data. However, the baseline data such as demographic information were not statistically sufficient to explain the above usage but we can use that information to explain the aforementioned usage trend. The drop in the total number of logbook sessions appears slightly the same in both younger and older intermediaries as shown on Figure 4.4. But contrary to this is that younger intermediaries had slightly higher average on total number of sessions compared to the older ones while in gamification condition as shown on Figure 4.5. Also on the average of total number sessions in both experimental conditions, the trend on average shows that younger intermediaries to have more sessions in overall as shown on Figure as more sessions were from gamification condition. The term old and young are defined by the age \geq and $<$ median age (15.5 years old) respectively. Although the total number of sessions is relative to the number of days in which a pair of users had a particular experimental condition available to them, but young and old intermediaries were evenly distributed to both experimental sequences (LG and GL groups). This means young and old intermediaries were both present in almost the same number in both experimental sequences hence their differences in number sessions which is influenced by the number of days are expected to cancel each other. The distribution by age group is presented on Table 4.6.

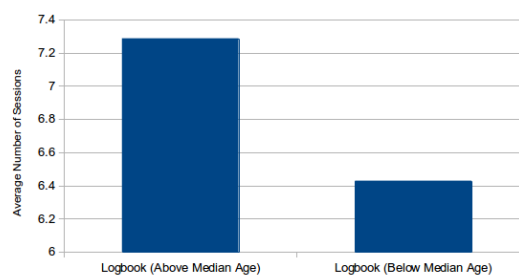


FIGURE 4.4: Average number of logbook sessions on 14 intermediaries: Age \geq median age(=15.5) versus Age $<$ median age.

If we recall about users who had usage problems on Table 4.3, three out of four intermediary users were above the median age and one below age, therefore one would argue

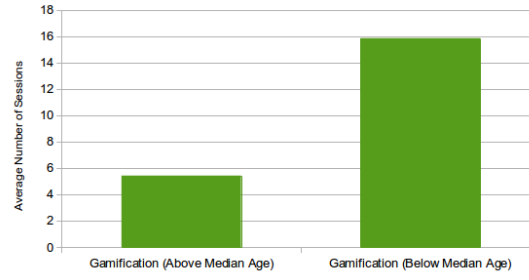


FIGURE 4.5: Average number of gamification sessions on 14 intermediaries: Age \geq median age(=15.5) versus Age $<$ median age.

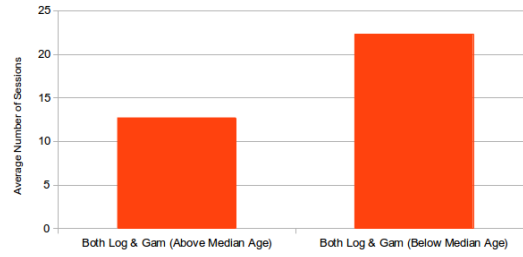


FIGURE 4.6: Average of total number of sessions of 14 intermediaries on both experimental conditions: Age \geq median age(=15.5) versus Age $<$ median.

TABLE 4.6: Age groups of intermediary participants

	Age Groups	Total users	No. of GL sequence	No. of LG sequence	No. of Females	No. of Males
1	Age ≥ 15.5 years	7	3	4	4	3
2	Age < 15.5 years	7	4	3	3	4

that the trend would change. So let's try to remove some pairs of users from the trends on averages from the aforementioned Figures 4.4 and 4.5. As we have seen on pairs A and B, their usage stopped early during the beginning of the experiments. As for Pair C and D, their motivation to use the app was only hindered during the gamification condition but they had used the app throughout logbook condition in which Pair C's pedometer was not working but since one of them had their pedometer working and the two intermediaries from these pairs were close, usage problems didn't affect their utilization of the app while they were in logbook condition. Therefore, in the logbook condition, let's remove pairs A and B only and the new trend on average number of sessions is shown on Figure 4.7. For the gamification condition, we remove Pairs A, B, C, and D and the new trend on average session is shown Figure 4.8. The same trends continue to hold as younger intermediaries appear to have higher average number sessions in gamification condition and lower average number of sessions in logbook condition. We

cannot confidently conclude any thing based on these trends but they can recommended to be considered in future work in the conclusion chapter. Koivisto and Hamari (2014) found that the usage of a gamified system is highly affected by the novelty effect which is inversely proportional to the age of participants, meaning that highly usage due to the novelty effect may be reported in much younger participants.

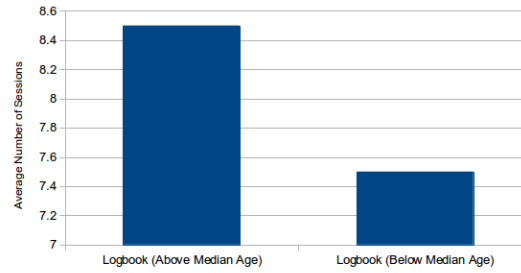


FIGURE 4.7: Average number of logbook sessions on 12 intermediaries: Age \geq median age(=15.5) versus Age $<$ median age.

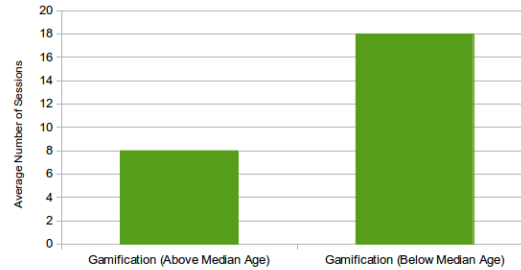


FIGURE 4.8: Average number of gamification sessions on 10 intermediaries: Age \geq median age(=15.5) versus Age $<$ median age.

On the next sub sections, user experiences of both intermediaries and beneficiaries are reported .

4.3.2 User Experience of Intermediaries

User experience was analysed through IMI(Intrinsic Motivation Inventory) questionnaires and interviews. The first step was to examine how baseline intrinsic motivation and demographic information such as age and gender influence user experience. From Figure 4.6 above, we saw that young intermediaries (age $<$ median=15.5 years) appeared to have more number of sessions per day on average, but contrary to this trend is

that, at baseline the average perceived enjoyment on helping with cellphone related tasks was higher in intermediaries with age \geq median compared to intermediaries with age $<$ median for the 12 intermediary users as shown on Figure 4.9.

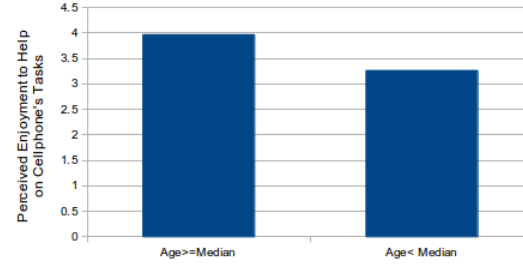


FIGURE 4.9: Intermediaries' average perceived enjoyment to help others with cellphone tasks versus age group.

I started by comparing the perceived enjoyment to use the family wellness app by age at midline and endline for the 12 intermediary users (excluding users from pairs A and B since they terminated their usage too soon). The purpose of this comparison was to understand why young intermediaries who appeared to be less enjoying on helping with cellphone related tasks, emerged as having more usage sessions than older intermediaries. In this comparison, six intermediary users were below the median age of 15.5 while the remaining six intermediary users were above the median age. More information on these two age groups is provided on Table 4.7. We can see that both age groups had representatives from both experimental sequences (GL, and LG) and genders. Figure 4.10 shows the averages of perceived enjoyment to use the family wellness app at midline and endline. The averages are higher on the intermediaries with age $<$ median for both midline and endline points.

TABLE 4.7: Age groups of intermediary participants

	Age Groups	Total users	No. of GL sequence	No. of LG sequence	No. of Females	No. of Males
1	Age ≥ 15.5 years	6	3	3	2	4
2	Age < 15.5 years	6	2	4	4	2

There several factors that influence intermediaries to use the app and these are:

1. **Self-monitoring task:** The task itself of self-monitoring without rewards sparked interests of some intermediary users. This might be as the result of the novelty effect of visualization mechanisms. For instance one intermediary user who happened to be the youngest among all intermediaries reported the highest perceived

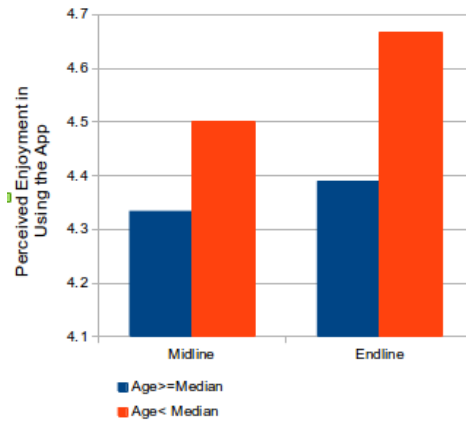


FIGURE 4.10: Intermediaries' average perceived enjoyment in using the app versus age group.

enjoyment during logbook condition compared to other intermediary users. In addition it seems the phone also had an impact on triggering her interests.

2. **Informal comparisons:** Comparison on steps graphs can be another source motivation. For intermediaries who were close, they did this form of informal comparison. In the absence of gamification, still intermediaries thought that they were competing with other therefore those who were close had face to face interactions of where they did comparison with one another.
3. **Gamification comparison:** This kind of comparison increased the number of times intermediary user checked the app. The “Gamified App” was designed in such a way that a pair will earn rewards based on usage and the average number of steps walked by a beneficiary participant who is a member of the pair. The purpose of rewards was to foster users' intrinsic experiences such as competitiveness and a sense of autonomy which are predictors of intrinsic motivation. Rewards depended on four parameters and these were the number of steps walked by a beneficiary user, the number of days the app has been used by an intermediary to either to record meals or to view feedback on meals, points, steps, gardens, etc. During interviews, I discovered that in one pair not only the beneficiary was using the pedometer, an intermediary was also taking turns to use the pedometer, therefore they were collaborating in accumulating steps. Both an intermediary user and a beneficiary user had discussions of whether the person whose turn it was had walk enough steps. They did this to accumulate more steps than other pairs.
4. **Requests from beneficiary users:** There were times where intermediary users engaged with the app only upon receiving requests from beneficiaries.

Intermediaries were interested to fulfil requests from beneficiaries provided an intermediary was interested with the self-monitoring task; steps comparison; or gamification comparison. The trend on average perceived enjoyment in both logbook and gamified conditions appeared to be slightly higher in younger intermediaries compared to older intermediaries as shown on Figure 4.11.

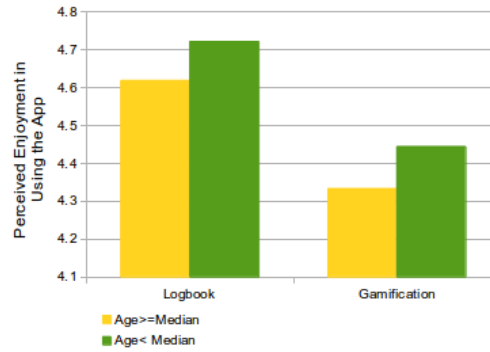


FIGURE 4.11: Intermediaries' average perceived enjoyment in using the app versus age group (Logbook and Gamification).

Figures 4.10 and 4.11 suggest that the task of helping was more interesting to younger intermediaries due to the existence of the aforementioned motivational affordances. However, not all intermediaries had positive user experience on utilizing gamification, that is why the average perceived enjoyment appears to be lower for both age groups while in gamification condition compared to the logbook condition. If we go back we can see that four pairs had usage problems as mentioned on Table 4.3. From those four pairs, two pairs had severe usability problems and these were pair A and Pair C. Pair B tried to use the app a few times but since they didn't have data bundles usage was terminated immediately, however, they continued to receive SMS feedback throughout the experimental conditions of which some of it was positive as they had managed to attain some badges. User from pair D also appeared to be more interested during gamification even though he terminated usage after his first week on gamification. We can see that problems in this pair only started on the fourth week of logbook condition (one week before switching to gamification condition), therefore, the pedometer didn't affect much the motivation of the intermediary user from Pair D as he enjoyed the gamification condition despite the fact that he logged in once throughout this experimental condition and the pedometer was not transmitting data to the server. The two users from Pair B and Pair D, had their motivation not affected by usability problems as for Pairs A and C. SMS feedbacks and reminders were sent to all fourteen pairs and this somehow engaged users from Pair B and Pair D. During gamification, if a particular pair managed

to acquire a new badge, all other pairs were notified through SMS including the pair that had attained a badge. In addition intermediaries were reminded of what has been done and what remains to be done in order to attain higher badges. During logbook condition, intermediaries were notified with SMS of how many steps have been walked by their intermediary so far and they were also reminded that to assist their beneficiaries in monitoring their wellness. Therefore since, usability problems in Pair B and D were not severe as for pairs A and C, and combination of text messages feedbacks, motivation of users in pairs D, and C appeared to be slightly higher in gamification condition compared to logbook condition. From here we can conclude that Pairs A and C motivations during gamification condition were negatively affected by the severe usability problems. Gamification condition didn't harm motivation of only users with severe usability problems. There was a very intriguing phenomenon from two other intermediary users from **Pair E** and **Pair F** as mentioned on Table 4.8. The two intermediary users had used the app more often in gamification condition compared to when they were in logbook condition but had reported both lower scores in perceived enjoyment, and perceived competence when in gamified condition compared to when they were in logbook condition. Two of these intermediary users were in LG and GL groups respectively. I examined the performance of these two users in gamification rewards and discovered that two users never managed to make any progress in attaining a single reward despite their efforts in using the "Gamified App". This harmed both their perceived enjoyment and perceived competence to use the "Family Wellness App" while in gamification condition. One of these two intermediary participants sent an SMS to the researcher asking what he was supposed to do in order to advance in badges and he was informed that all the information was specified in their given user manual. In addition, SMS reminders were sent out to all intermediaries to inform them of how far they have gone in achieving rewards and what is remaining in terms of usage and steps in order for them to reach the next badge. Their beneficiary participants were not walking enough steps despite the fact that these intermediaries had put more efforts in using the App during gamification condition. Badges were earned in combination of both the app usage and average number of steps walked by a beneficiary user. One beneficiary who was working with one of these intermediary users also reported in interviews that there was always a contention with her intermediary when this beneficiary wanted to see what was going in the app as the intermediary was not voluntarily willing to help some of the time.

Therefore the conclusion was that, the two intermediary users on Table 4.8 had a negative experience as the result of failure of our gamification design to match challenges with abilities. i.e. efforts of beneficiaries differed hence challenges should have matched with individual abilities of beneficiaries within pairs. When challenges are too difficult as they don't match users' skills, end users can become demotivated (Zhang, 2008). As

TABLE 4.8: Pairs affected by poor design of gamification

	Pair	Experimental Sequence	No. of logbook sessions/day
1	Pair E	LG group	The beneficiary participant was not walking enough steps hen
2	Pair D	LG group.	Same as above.

result only eight intermediary users were considered in the main sub scales of intrinsic motivation (autonomy, competence, and enjoyment).

I compared between the logbook and gamification scores from an IMI questionnaire that assessed intrinsic motivation of intermediaries at midline and endline in using the “The Family Wellness App”. This questionnaire had four sub-scales which were perceived competence, perceived autonomy, perceived enjoyment, and perceived relatedness. For the reason stated above, I decided to exclude users from Pair A and C on Table 4.3 , and Pairs E and F on Table 4.8. A total of ten pairs were considered for analysis of on the aforementioned intrinsic motivation sub-scales.

The hypotheses of interest for intermediaries were:

1. Hypothesis 1

- H_0 : There is no difference in perceived competence in using the app between a logbook app and gamified app
- H_A : There is a difference in perceived competence in using the app between a logbook app and gamified app

2. Hypothesis 1

- H_0 : There is no difference in perceived autonomy in using the app between a logbook app and gamified app
- H_A : There is a difference in perceived autonomy in using the app between a logbook app and gamified app

3. Hypothesis 3

- H_0 : There is no difference in perceived enjoyment in using the app between a logbook app and gamified app
- H_A : There is a difference in perceived enjoyment in using the app between a logbook app and gamified app

4. Hypothesis 4

- H_0 : There is no difference in perceived relatedness in using the app between a logbook app and gamified app

- H_A : There is a difference in perceived relatedness in using the app between a logbook app and gamified app

The results on perceived competence, perceived autonomy, perceived enjoyment, and perceived relatedness of between the “Logbook App” and the “Gamified App” are shown on Table 4.9. The distribution of the differences of all the scores from the four aforementioned sub-scales, followed a normal distribution hence met a condition for using a paired student t-test.

TABLE 4.9: Comparison of ten intermediaries’ scores on sub-scales of competence, autonomy, enjoyment, and relatedness in using the “Family Wellness App

Mean	Logbook App	Gamified App
Perceived competence	M=5.23; SD=1.02	M=5.96; SD=0.66
	t(9)=-3.4949; p=0.0068 ; 95% CI= -1.204 to -0.258	
Perceived autonomy	M=3.95; SD=0.86	M=3.96; SD=0.94
	t(9)= -0.0269; p= 0.9792; 95% CI= -0.596 to 0.582	
Perceived enjoyment	M=4.18; SD=1.11	M=4.75; SD=1.32
	t(9)=-1.6930; p=0.1247; ; 95% CI= -1.343 to 0.193	
Perceived relatedness	M=4.22; SD=0.63	M=4.37; SD=0.9
	t(9)= -0.7193; p=0.4902; 95% CI= -0.622 to 0.322	

Perceived competence of intermediaries in using the “Family Wellness App” was significantly higher in the gamified condition than in the logbook condition in the ten intermediaries that were analysed. Perceived autonomy, perceived enjoyment, and perceived relatedness were all not different between gamification condition and logbook condition. For perceived enjoyment and autonomy, there are cases where the self-monitoring task was interesting enough on itself or when combined with informal comparison of steps among intermediary users. Therefore, there some intermediaries who felt more enjoyment while they were in logbook condition, which is also the same case for perceived autonomy. For perceived relatedness, the app brought users together regardless of an experimental condition especially the ones that already knew each other before.

4.3.3 Beneficiaries’ IMI Scores on Health Self-Monitoring

I used the IMI (intrinsic motivation inventory) questionnaire to assess motivation of intermediaries in self-monitoring of Diet. I excluded pairs (A,B, C, and D) from Table 4.3 problems that led discontinuation of usage. In total only ten out of fourteen beneficiaries had their results included for analysis. We have seen from the sub section (4.2.2.2) that this questionnaire was administered at baseline, midline and endline.

In the first comparison, I compared the IMI score of each participant at baseline, midline, and endline regardless of the experimental condition. In the second comparison I compared scores at baseline, logbook, and gamification condition. The IMI score was computed from the average of all scores from sub-scales of perceived competence, perceived autonomy, perceived relatedness, perceived enjoyment, perceived effort, and perceived usefulness. I used one way ANOVA with repeated measures to test if there a difference between scores at: (1) baseline, midline, and endline, and (2) baseline, logbook and gamification. I used Mauchy's test³ to check if baseline, midline and endline scores had the same covariance in each ANOVA test I carried out and this helped in deciding of what correction to use on SPSS results among "Sphericity Assumed", "Greenhouse-Geisser", and "Huynh-Feldt" as reported on the next sub-sections.

4.3.3.1 IMI in Self-Monitoring of Diet

The results on self-monitoring of diet (baseline, midline, and endline) are shown on Table 4.10. The Mauchly's test indicated that the assumption of sphericity was not violated with $\chi^2(2)=3.76$, $p=0.152$.

TABLE 4.10: Comparison of ten beneficiaries' IMI scores in self-monitoring of diet at baseline, midline and endline

Mean IMI Score	Baseline	Midline	Endline
Self-monitoring of Diet	M=4.48; SD=1.24	M=5.07; SD=1.19;	M=5.55; SD=0.95
	F(2,18)=3.787; p=0.042		

Therefore, the results (N=10) on "Self-monitoring of Diet" shown on Table 4.10 were from "Sphericity Assumed" output. ANOVA showed that there was a significant difference of average IMI scores on self-monitoring of diet measured at baseline, midline and endline.

I further conducted a student t-test to see if there was difference in IMI scores to self-monitor diet by contrasting the following: (1) baseline versus midline; (2) baseline versus endline; and (3) midline versus endline. The results are shown on Tables 4.11, 4.12, and 4.13. There was a significant difference between IMI scores on baseline versus endline but not on baseline versus midline and midline versus endline. Motivation to self-monitor diet appeared to increase with time as shown on Figure 4.12. The interpretation of the above findings are that the wellness app appears to have a significant effect with time on motivation of beneficiaries to self-monitor their diet.

³Read more on how Mauchy's test is used from <http://Business analysis attempts to study the existing market. /www.statisticshell.com/docs/repeatedmeasures.pdf>

TABLE 4.11: Pairwise comparisons of IMI scores in self-monitoring of diet: Baseline versus Midline

Mean	Baseline	Midline
IMI Score	M=4.48; SD=1.24	M=5.07; SD=1.19
	t(9)=-1.298; p=0.227 ; 95% CI= -1.621 to 0.439	

TABLE 4.12: Pairwise comparisons of IMI scores in self-monitoring of diet: Baseline versus Endline

Mean	Baseline	Endline
IMI Score	M=4.48; SD=1.24	M=5.55; SD=0.95
	t(9)=-2.457; p=0.036 ; 95% CI= -2.06083 to -0.08517	

TABLE 4.13: Pairwise comparisons of IMI scores in self-monitoring of diet: Midline versus Endline

Mean	Midline	Endline
IMI Score	M=5.07; SD=1.19	M=5.55; SD=0.95
	t(9)=-1.975; p=0.08 ; 95% CI= -1.0342 to 0.07017	

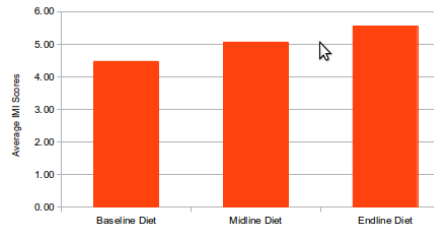


FIGURE 4.12: Trend on Average IMI Scores of Self-Monitoring of Diet at Baseline, Midline, and Endline.

The aforementioned ANOVA comparison among baseline, midline, and endline doesn't discern between different experimental conditions of which pairs of users were exposed to. The second task was to examine the comparison on IMI scores to self-monitor diet, among baseline, logbook, and gamification conditions. The ANOVA results (N=10) are shown on Table 4.14. I chose the "Sphericity Assumed" output of the ANOVA test since the Mauchly's test indicated that the assumption of sphericity was not violated with $\chi^2(2)=2.19$, $p=0.335$. Therefore, the results on "Self-monitoring of Diet" shown on Table 4.14 were from "Sphericity Assumed" output. ANOVA showed that there was no significant difference of average IMI scores on self-monitoring of diet measured during baseline, logbook and gamification conditions. The trend on averages shows both logbook and gamification to be slightly higher than baseline as shown on Figure 4.13. The conclusion from this finding is that is that both versions of the prototype have show an indication of increasing motivation of beneficiaries to self-monitor diet. However, the

logbook condition achieved a significant result. In the informative evaluation that was conducted before this evaluation, it was found out that beneficiary users tended to give more value to the main feedback features for self-monitoring of steps and diet (Katule et al., 2016), while in the Figure 4.2 above we saw that the trend on utilization of the main feedback features for self-monitoring of steps and diet decreased during gamification condition. This could be one of the reasons of why we achieved a significant result only on logbook when compared to baseline and not on gamification when compared to baseline. Another explanation is attributed to the our smaller sample sizes hence the power of the effect size is reduced.

TABLE 4.14: Comparison of ten beneficiaries' IMI scores in self-monitoring of diet at baseline, after logbook, and after gamification conditions

Mean IMI Score	Baseline	Logbook	Gamification
Self-monitoring of Diet	M=4.48; SD=1.241	M=5.28; SD=1.05	M=5.34; SD=1.16
	F(2,18)=3.787; p=0.087		

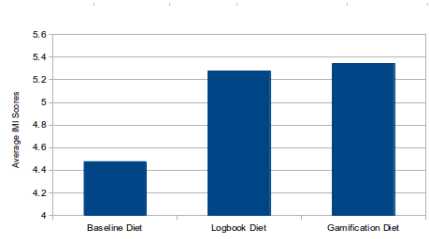


FIGURE 4.13: Trend on Average IMI Scores of Self-Monitoring of Diet at Baseline, Logbook, and Gamification.

4.3.3.2 IMI in Self-Monitoring of Activity

The results (N=9) on self-monitoring of activity are shown on Table 4.15. The results are based on a sample of nine beneficiary users as one participant didn't complete this part of the questionnaire at baseline. The Mauchly's test indicated that the assumption of sphericity was violated with $\chi^2(2)=8.248$, $p=0.016$. The value ϵ on Greenhouse Geisser was " <0.75 ", therefore, the results on "Self-monitoring of Diet" shown on Table 4.15 were selected from "Greenhouse-Geisser" output. ANOVA showed that there was no significant difference of average IMI scores on self-monitoring of activity measured at

baseline, midline and endline. The trend of means appears to increase from baseline to endline as shown on Figure 4.14. There are several factors that could have contributed to results not being significant among baseline, midline, endline points,. The first hypothesized reason is tracking of physical activity appeared to be easy in majority of the participants even without tracking devices as people can estimate the distance they walk daily and they consider this as tracking even though they might have means to record this information, hence their motivation was high at baseline unlike diet self-monitoring which they consider it to be cumbersome due to external barriers such as health food being expensive, therefore at baseline participants felt more motivated to track their activity. The second hypothesized reason is that the sample size was small hence there was a smaller power in detecting significant difference. But we have seen that the trend in motivation increases with time.

TABLE 4.15: Comparison of ten beneficiaries' IMI scores in self-monitoring of activity at baseline, midline and endline

Mean IMI Score	Baseline	Midline	Endline
Self-monitoring of activity	M=4.82; SD=1.002	M=5.28; SD=1.003	M=5.41; SD=0.894
	F(1.182, 9.455)=2.936; p=0.116		

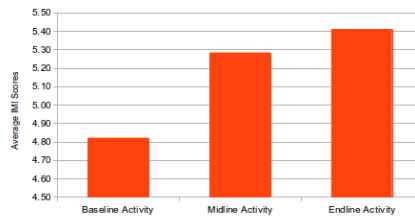


FIGURE 4.14: Trend on Average IMI Scores of Self-Monitoring of Activity at Baseline, Logbook, and Gamification.

I conducted another analysis (N=9) to examine if there is a difference among baseline, logbook, and gamification in self-monitoring of activity. The Mauchly's test indicated that the assumption of sphericity was violated with $\chi^2(2)=6.788$, $p=0.034$. The value of ϵ on Greenhouse Geisser was " <0.75 ", therefore, the results on "Self-monitoring of Activity" shown on Table 4.16 were selected from "Greenhouse-Geisser" output. ANOVA showed that there was no significant difference of average IMI scores on self-monitoring of activity measured at baseline, logbook and gamification. The same reasons on Table 4.15 for not achieving statistical significance are also valid in this case. Also the trend in motivation increases in both logbook and gamification as shown on

Figure 4.15

TABLE 4.16: Comparison of ten beneficiaries' IMI scores in self-monitoring of activity at baseline, logbook and gamification

Mean IMI Score	Baseline	Logbook	Gamification
Self-monitoring of activity	M=4.82; SD=1.002	M=5.33; SD=0.9762	M=5.37; SD=0.9276
	F(1.234, 9.872)=2.783; p=0.123		

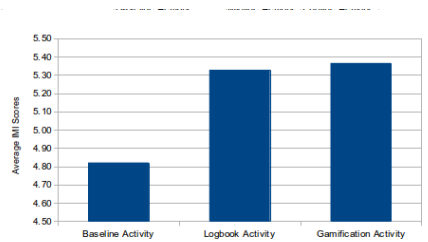


FIGURE 4.15: Trend on Average IMI Scores of Self-Monitoring of Activity at Baseline, Logbook, and Gamification.

Bibliography

Edward L Deci and Richard M Ryan. *Intrinsic motivation and self-determination in human behavior*. Springer Science & Business Media, 1985.

Jonna Koivisto and Juho Hamari. Demographic differences in perceived benefits from gamification. *Computers in Human Behavior*, 35:179–188, June 2014.

Ping Zhang. Technical opinion motivational affordances: reasons for ict design and use. *Communications of the ACM*, 51(11):145–147, 2008.

Ntwa Katule, Melissa Densmore, and Ulrike Rivett. Leveraging intermediated interactions to support utilization of persuasive personal health informatics. In *Proceedings of the Eighth International Conference on Information and Communication Technologies and Development*, page 19. ACM, 2016.