

Improving Large Number Comprehension with an Infinite Runner Game

Thijs van den Hout

June 27, 2018

Abstract

Large numbers are increasingly important in our society. We encounter large numbers in the news, research, macroeconomics and they are the backbone of big data. Despite their importance, people often misjudge large numbers, leading to misconceptions and ungrounded discussions about the aforementioned topics. Training can improve comprehension of large numbers. For example by embedding a large number in an understandable context (Barrio, Goldstein, & Hofman, 2016), or pointing out a often wrongly predicted value on a number line (Guay, Davis, DeLaunay, Charlesworth, & Landy, Under review). This research proposes a serious game to train users on the comprehension of large magnitudes and exponential growth. The game is an endless runner-type game and is based on the notion of walking on an extended number line with varying speeds to give the player an impression of the magnitude and scale of numbers. An experiment was conducted regarding the effectiveness of this training method. Scores extracted from questionnaires filled in before and after playing the game revealed that the game is indeed effective in the training of large number comprehension. Especially notable is the change in response to price-difference questions, one of the used measurements for large number comprehension. These results imply the proposed game is a viable training method for large number comprehension and may be developed further for real-life use.

1 Introduction

1.1 Project background

Even though large numbers are increasingly important in political and societal matters, for most people large numbers and exponential growth are hard to understand (Landy, Charlesworth, & Ottmar, 2014). Therefore, discussions on issues such as global warming, income inequality, and population growth are often baseless. Hence, it is important to improve large number comprehension.

This research project proposes an educational game which trains its users on the comprehension of large numbers. Towards that end, it is important to know how the brain represents numbers, how currently numbers are trained, and how digital games can help achieve a better understanding of large numbers.

1.2 Cognitive representation of large numbers

Much research has been done on how humans reason with numbers across many orders of magnitude. The biggest share of research in this field suggests that humans represent number magnitudes in a logarithmic model, which becomes increasingly more linear with age (Siegler & Opfer, 2003; Barth & Paladino, 2011). When estimating values on a number line, a logarithmic model would place one million in the middle between one thousand and one billion, while in a linear scale the value of one million would only be 1/1000th of the way from the thousand mark. Furthermore, people with a more linear sense of (large) numbers

perform better in answering numerically based questions (Landy, Silbert, & Goldin, 2013; Petrova, Pligt, & Garcia-Retamero, 2014). Suitably, subjects that were classified as linear performed better on placing the number one million on a number line from one thousand to one billion than those classified otherwise (Landy et al., 2013).

Recent studies show there is overlap between both logarithmic and linear representations (Landy et al., 2013). On a number line between one thousand and one billion, about half of the participants placed the value for one million about half way on the number line, indicating a log-scale representation. Interestingly, the participants that displayed a log-scale representation did place numbers linearly on either side of this central division (Landy et al., 2013). A subsequent study showed a more detailed distinction of the uniform spacing between numbers words (e.g. thousand, million, billion) (David, Arthur, & Erin, 2017). Discontinuities were observed near the borders of these categories. Participants that were accurate on the number line task highlight contrasts within categories, while inaccurate participants adapt their responses toward the center of categories.

An analogous problem is one of the psychophysics of price. This entails the phenomenon of consumers making less effort to save a set amount off a purchase as the price increases (Grewal & Marmorstein, 1994), which, like the logarithmic intuition of numerical cognition, sprouted from the Weber-Fechner law in psychophysics (Fechner, 1860). For example, someone will likely make a large effort to save €50 on a groceries bill, while they would not worry to save €50 on a new car or house. Even though the impact may feel different, the actual value is the same in both scenarios (Caryn, 1989). This is called the distance effect in psychophysics, which claims that it becomes increasingly difficult to distinguish between two numbers as the difference between them decreases (Moyer & Landauer, 1967). The distance effect is the most prominent manifestation of the Weber-Fechner law in numerical cognition.

1.3 Existing training methods

Training of small numbers is commonplace in elementary schools. Children learn to understand numbers by various means, such as familiarizing themselves with number lines. For example, in a study by Moeller et al. (2012), children had to estimate numbers by positioning themselves on a number line that was taped on the floor with only numbers at the boundaries (Moeller et al., 2012).

Training the understanding of large numbers, however, is more complicated since these numbers are hardly imaginable, and therefore less intuitive. Additionally, training using a stationary number line is infeasible as large numbers become more intractable. No training of large number numeracy is actively given to people in relevant sectors.

Previous studies prove understanding of large numbers can still be trained in various ways. Providing subjects with a perspective of the relevant number in a more understandable context, improved the understanding of numbers in the news (Barrio et al., 2016). It is also shown that merely pointing out the actual value of an often-misinterpreted value on a number line can increase understanding of large numbers (Guay et al., Under review). Interestingly, such training showed to selectively improve political judgements of numerical nature such as "Rate the impact of reducing a \$6.9 billion National Science Foundation budget by \$370 million." (Guay et al., Under review). Half of the participants underwent a short training of number line placements while the other half did not. Results showed the group that received training gave more accurate, lower impact ratings than the group that did not receive training.

1.4 Educational games

There exist abundant tools and number games for children to master numbers under 1000, and they are proven to be effective (Noemí & Máximo, 2014; Laurillard, 2016). Many of these games consist of number orderings, addition training, or multiplication training. Often a number line is presented to guide the child towards the correct answer.

At the time of writing, no serious game regarding improvement of understanding of large numbers has been proposed.

1.5 Aim of the project

This project hopes to build on the promising findings in the previous studies by creating an accessible game to tutor users in dealing with the understanding of large numbers.

The current study introduces an interactive and accessible game to train people on the comprehension of large numbers and relative number magnitudes.

The research question of this project is:

How can an infinite runner game on an extended number line be improve large number comprehension?

The effect of the game is measured by comparing participants' scores on questionnaires containing numerical questions before and after playing the proposed game. As control parameter, a second experimental group was set up, in which participants played a similar game, missing most numerical aspects. The experiment is described more thoroughly in section 2.

2 Methods

2.1 Participants

In this experiment, 20 students participated (average age was 20.95, 70% male). The participants were personally contacted and voluntarily took part in the experiment. All participants signed a consent form prior to the experiment, indicating they understand the instructions to the experiment and their rights as participants.

2.2 The presented game

The game was built in game developing environment Unity. The scripts encapsulating the logic of the game were written in C#. The background images that were used in the background of the game were downloaded from or inspired by freepik.com and pngtree.com.

The game that has been developed is an infinite runner-type game, which is defined as follows. An infinite runner is a game in which the player traverses an endless path and cannot stop their forward movement. Often the game is expanded with obstacles and/or power-ups. The player receives points for the distance travelled. The goal of the game is to achieve the highest possible score in the shortest amount of time.

Two games were developed to conduct the experiment: the experimental game, which has the aforementioned objective of training participants in large number comprehension, and a control game. The control game is similar to the experimental game, but lacking the unique numerical aspect. Both games share the common goal of obtaining the highest score in the smallest amount of time.

2.2.1 Experimental game

The concept of the infinite runner is potentially ideal to train large number understanding, as the distanced travelled can be used as an analogy for traversing a number line. The intent of the game is for the player to pay attention to the numbers on the number line as they pass over it. This can be achieved if the game revolves around the magnitude of the numbers on the number line, the speed with which the player moves forward, and the location of power-ups, which will be discussed later. By experiencing the numbers actively, the participant will presumably construct an improved mental image of large numbers and the comparison of magnitudes.

Figure 1 shows the game screen as it may appear while playing the game. The background and number line change as the player progresses, while the user interface retains the same format.

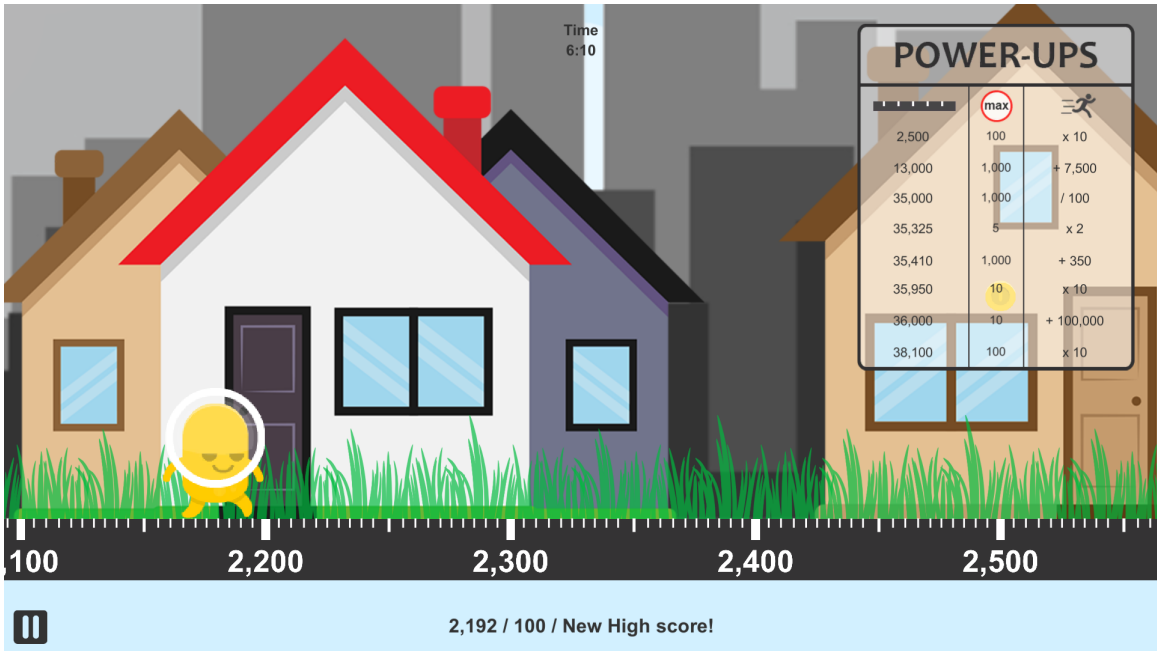


Figure 1: Example game screen with step size 100 and current score of 2192

The platform the player walks on doubles as a number line on which the current score is displayed. In the bottom of the user interface the current score, current step size and high score are shown. The current score corresponds to where the player is walking on the number line. The step size is the interval between the steps the player takes. The magnitude of the step size also determines the interval between numbers on the number line. The high score is displayed if the player has achieved a higher score in a previous game, and "New High score!" otherwise.

To improve the player’s sense of scale the field of view scales with the current step size. When the player is going slower, the environment is viewed from closer up than when the player is moving fast. The number line will also show more intermediate lines as an indicator of more numbers occupying the interval between the two shown numbers. The different number lines are shown in figure 2. To further improve the sense of scale, the background wallpaper changes with the step size. Some example backgrounds are shown in figure 3.

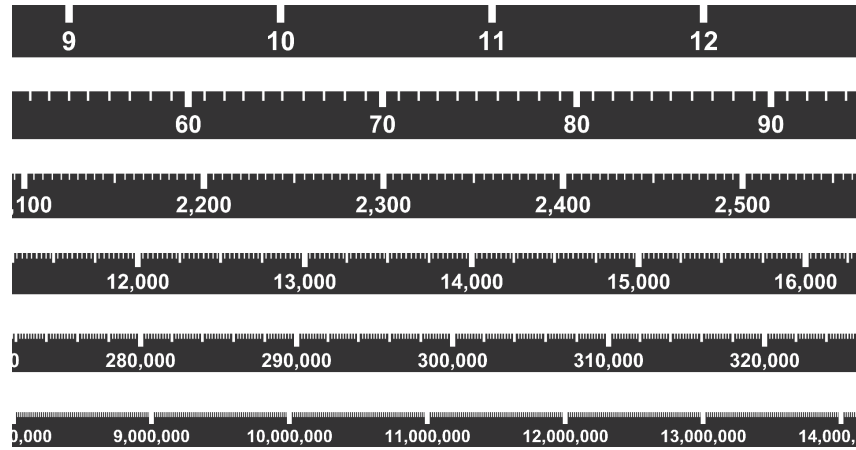


Figure 2: The intermediate lines on the number line for the step size magnitudes of 1; 10; 100; 1000; 10,000; 100,000 and higher, from top to bottom respectively



Figure 3: Example of different background wallpapers

The player will encounter power-ups which they can pick up or avoid. All power-ups have an effect on either the speed with which the player traverses the path, or the total obtained score and thus location. The incoming power-ups are presented on the screen in the top right corner, "Power-ups interface". As a result, the participant can see the next 8 power-ups coming long before they arrive on the screen. An example image of the power-up interface is shown in figure 4.




POWER-UPS		
		
300,000	100,000	/ 100
302,000	100	+ 200,000
350,000	100,000	/ 1000
350,350	10	+ 300,000
350,500	100	x 100
400,000	100,000	+ 200,000
503,000	1,000	x 10
530,000	10,000	+ 70,000

Figure 4: Power-ups interface. Left: location on number line, middle: maximum pick-up speed, right: effect on score / step size

Each power-up comprises three attributes which are shown in the interface in the top-right corner of the game screen:

- The location of the power-up on the number line;
- The maximum speed with which the player can pick up this power-up;
- The effect of the power-up.

The list of power-ups is shown such that the game has some strategy aspects, as the player must make evaluated choices regarding the power-ups they pick up. The motive behind the strategy aspect is to have the participant pay more attention to the number line and which power-ups to pick up or miss. Moreover, having the player run at varying speeds at different stages in the game causes them to experience the magnitude of their movement speed and score more effectively.

For example, the player may run at a low speed of 10 at a large score of 2,000,000. The player may have to travel from 2,000,000 to 2,002,000 for the next power-up. Such a seemingly negligible distance will prove to take a long time when moving with a step size of only 10.

The player may pick up a seemingly negative power-up in order to pick up a more beneficial power-up later on in the game. For example: the player moves at a speed of 50. A power-up with effect $speed \times 20$ and a max speed of 5 is coming up. In order to pick up this power-up the player must pick up the preceding power-up with effect $speed / 10$. The resulting speed of this sequence is more beneficial than missing the $speed / 10$ negative power-up.

The game comprises 46 power-ups in total, the last of which is the finish at 25,000,000. Some power-ups are not avoidable and forcibly lower the step size to have the player experience the difference in magnitude. For example at unit 2,000,000 the player's step size is set back to 1, which lets them feel what little effect the lower step sizes have on a score of a much larger magnitude.

2.2.2 Control game

The control game was developed and tested in the same setting as the experimental game to rule out any change in large number comprehension not attributed to the experimental

game. An example in-game screen of the control game is shown in appendix 1.

The control game lacks the extended number line on which the player runs and instead includes a blank surface with uniformly spaced indicators to retain the sense of movement. Power-ups are not shown in advance but are randomly distributed along the path. Furthermore, power-ups are solely positive or negative. Positive power-ups speed up the player, leading to a faster increase in total score, while negative pick-ups slow the player down. The participant can not see their current speed explicitly as is the case in the experimental game. Display of current score and high score, along with the changing background wallpaper remain in place.

2.3 Procedure

The conducted experiment comprises three stages: completion of a pretest questionnaire, playing either one of the two games, and completion of a post-test questionnaire. Before the experiment commences the examiner provides the participant with instructions to the questionnaire and the game. The participant signs a consent form indicating they understand the instructions and agree with the terms of the experiment.

The selection of which game was played was determined randomly. Participants picked a number from a collection of numbers under 20 from a bag, which corresponds to their anonymous participant ID. If the number they picked was even, the participant played the experimental game, if the number was odd the participant played the control game.

The complete experiment took 25 minutes on average, including the pre- and post-test questionnaires. The participants had no time limit on completing the questionnaires, but were advised to act intuitively and not calculate the optimal answer wherever they thought it possible. The experimental condition took longer than the control condition because the game is more difficult to explain and to play. Also, the experimental game offered more incentive to pause and restart. Playing the experimental game took 20 minutes on average. The control version of the game took 7 minutes for all participants, which was the time limit on the game. The games were all played on the same laptop and the questionnaires were filled in either on the laptop or on a mobile phone.

During the playing of the experimental game data was automatically gathered and saved to a text file. This data includes the location and effect of the power-ups the player picked up, when the participant paused the game and for how long, when the participant restarted or finished the game and their play time. This information will be incorporated in the analysis to understand the behaviour of a participant and to find a connection to their obtained results. A sample of this data is provided in appendix 2.

2.4 Questionnaires

The questionnaires that were filled in by the participant both consist of 12 multiple choice questions regarding numbers. These questions were divided into three categories of four questions each: relative price-difference, impact ratings and satisfaction ratings. These three question categories were chosen based on previous research which validate them as measures for large number understanding (Azar, 2011; Landy et al., 2013; Guay et al., Under review). As each of these studies relate them to a broad measure of numerical cognition, the results of each measure may vary in this particular study. These three benchmarks were chosen because of their solid relation to large number comprehension and in the interest of distinguishing results of the game in different fields that are affected by large number understanding.

Asking similar questions of each category in the pretest and post-test questionnaire can reveal whether the experimental game affect the participant's response Both complete questionnaires are provided in appendix 3.

2.4.1 Price differences

In the price differences category the participant is given an anecdote and a question regarding the price of a presented product, given the price of a similar product. The participant must then select one of four multiple choice answers which correspond to the best approximation of the price they would be willing to offer for the presented product. For example:

Say you want to buy a laptop. You are interested in a specific model that comes in two screen sizes, 15" and 13". The laptops are otherwise identical. assume you do graphic design and therefore prefer the larger screen. If the 15" laptop costs €830,-, how much must the 13" laptop cost for you to prefer it over the 15"?

- a) €520,- b) €650,- c) €700,- d) €780,-

Questions of this type have also been posed in studies regarding relative thinking and implications to business strategy (Azar, 2011). The ambiguity of each question has been limited as much as possible to get the most objective answer. In the question above for instance, the context of being a graphic designer and therefore preferring the larger screen rules out the option of the participant basing their answer on their own preference.

2.4.2 Impact rating

In the impact rating category the participant rates the impact of a given political or economic judgment on a 9-point scale. It is shown that people with a higher numerical ability give a lower impact rating to seemingly impressive propositions (Guay et al., Under review). An example question:

Rate the impact of a €100 million increase in the national education budget.

2.4.3 Satisfaction rating

In the satisfaction rating category the participant must rate their satisfaction with the presented anecdote of numerical nature on a 9-point scale. For example:

After days of negotiating with the estate agent on the price of the house you want she lowers the initial price of €275,000 by €10,000.

This type of question has also been asked in a previous study relating the satisfaction rating to number-line estimations. Landy et al. showed that participants with a more linear sense of scale rated the given solutions to political problems less satisfactory than those with a more distorted sense of numeric scale (Landy et al., 2013).

3 Analysis

To analyze the obtained data most effectively, the answers to the questions in each question category were averaged and normalized. Normalization was done per question category over both experimental groups combined. The scores were linearly scaled between 0 and 1. The price difference category had four multiple choice options. Since for this study the difference in responses between the pretest and post-test is of most interest, the answers were transformed to a score of 1 through 4 for the respective multiple choice options. The following data sets originated from the preprocessing of the data. For each participant an average response to each of the question categories in the pretest and post-test questionnaires, and for both the pretest and post-test questionnaires a normalized average response to the complete questionnaire. The data set used for the analysis can be found in appendix 4.

To analyze the effectiveness of the game a "split-plot" repeated-measures-MANOVA test was conducted (Ellis, 2013). The between-factor in the analysis is Group (experimental /

control). The within-factor is Time (before / after playing the game). The dependent variables that were tested for are mean scores of the three question categories price-difference, impact rating and satisfaction rating, and the normalized mean score on the complete questionnaire.

Design:

Dependent variable 1	= Price-difference
Dependent variable 2	= Impact rating
Dependent variable 3	= Satisfaction rating
Dependent variable 4	= Total questionnaire score
Within-subject factor	= Time (before, after)
Between-subject factor	= Group (experimental, control)

4 Hypotheses

The hypotheses for this study are shown in Table 1.

Table 1: Null hypotheses

Dependent Variable	Con trast	Group	Time	Time X Group
CQ	Pre + Post Pre vs. Post	$\mu_{exp\bullet} \geq \mu_{cont\bullet}$	$\mu_{\bullet Pre} \geq \mu_{\bullet Post}$	$\mu_{ExpPost} - \mu_{ExpPre} \geq \mu_{ContPost} - \mu_{ContPre}$
PD	Pre + Post Pre vs. Post	$\mu_{exp\bullet} \geq \mu_{cont\bullet}$	$\mu_{\bullet Pre} \geq \mu_{\bullet Post}$	$\mu_{ExpPost} - \mu_{ExpPre} \geq \mu_{ContPost} - \mu_{ContPre}$
IR	Pre + Post Pre vs. Post	$\mu_{exp\bullet} \geq \mu_{cont\bullet}$	$\mu_{\bullet Pre} \geq \mu_{\bullet Post}$	$\mu_{ExpPost} - \mu_{ExpPre} \geq \mu_{ContPost} - \mu_{ContPre}$
SR	Pre + Post Pre vs. Post	$\mu_{exp\bullet} \geq \mu_{cont\bullet}$	$\mu_{\bullet Pre} \geq \mu_{\bullet Post}$	$\mu_{ExpPost} - \mu_{ExpPre} \geq \mu_{ContPost} - \mu_{ContPre}$

Table 1: CQ = the average response to the complete questionnaire, PD = the average response to the price-difference questions, IR = the average response to the impact rating questions, and SR = the average response to the satisfaction rating questions.

In words: it is hypothesized that for the experimental group the mean score of each of the categories separate, as well as the complete questionnaire, will significantly decrease, while for the control group it will not.

The price difference score is expected to decrease as, according to Azar, people with higher numeracy are less prone to "full relative thinking", which implies they only consider relative price differences and disregard absolute price differences (Azar, 2011). Therefore, they are more inclined to choose a higher price for the proposed alternative product, than a more numerate participant. A higher price selection corresponds to a higher score on the questionnaire, thus it is expected the experimental group will decrease in score after playing the game.

Impact ratings for political scenarios have been shown to correlate with large number numeracy (Guay et al., Under review). Participants with a more linear model of large numbers rate the impact of seemingly large numeric schemes lower than participants with a poorer large number comprehension. Consequently, the experimental group is expected to decrease their impact ratings after playing the game while the control group should remain the same.

Satisfaction ratings given to numerical propositions are also expected to decrease with numeracy (Landy et al., 2013). This was tested in a similar setup to the impact ratings study.

As a result of all categories being expected to decrease in score, the complete averaged questionnaire score should also decrease significantly for the experimental group.

5 Results

5.1 Multivariate test

A 2 (Group: experimental, control) \times 2 (Time: before, after) repeated-measures-MANOVA was conducted with each of the variables mentioned in the previous section as dependent variables. Time, before and after playing the game, was used as within-subject factor and Group, experimental vs. control, as between-subject factor.

Only the Time \times Group-interaction effect is relevant for the question whether the experimental game is effective. This effect was significant: $F(3, 16) = 4.059, p < .05, \eta^2 = .432$.

The profile plot of the estimated marginal means of the total questionnaire is shown in Figure 5.

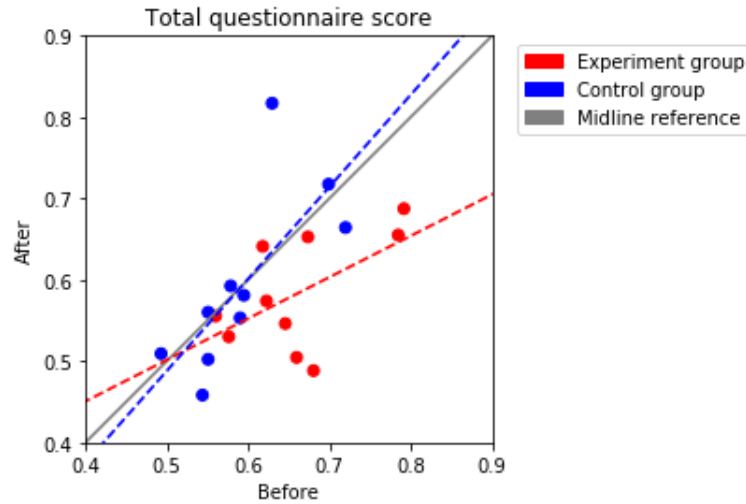


Figure 5: Interaction effect for the total questionnaire score.

5.2 Univariate tests

The interaction effect of Time \times Group on the price-difference questions category was significant ($F(1,18) = 13.322, p < .05, \eta^2 = .425$).

One outlier was observed in the pretest data of this category which may have contributed to the significance, though the effect is strong and can therefore not solely be contributed to this participant. Removing the outlier from the analysis does not affect the outcome.

The marginal means reveal the experimental group in the pretest questionnaire scored higher on the price-difference questions ($M = .661$) than the control group ($M = .595$), though this effect was abolished in the post-test questionnaire ($M = 5.84$ and 5.96). This affirms the experimental game reduces the participant's response to price-difference questions, as hypothesized.

The interaction effect of Time \times Group on only the impact rating question category was not significant ($p > .05$). The interaction effect of Time \times Group on only the satisfaction rating was also not significant ($p \gg .05$). The possible reasons for the non-significant results of the latter two question categories are given in the discussion section.

The complete report of the analysis can be found in appendix 5.

5.3 In-game data

The in-game data did not conclusively say anything about the obtained results. The total number of times the participant paused the game and the total time the game was paused did not significantly correlate with the difference between the pretest and post-test questionnaire scores.

6 Conclusion

This research proposed an educational game to improve large number comprehension. The research question was *How can an infinite runner game on an extended number line improve large number comprehension?*

The analysis tested whether playing the proposed game has a positive effect on the response to numerical questions. These questions function as a representation of large number comprehension, backed up by previous research (Landy et al., 2013; Azar, 2011; Guay et al., Under review).

The multivariate repeated-measures-MANOVA revealed the experimental game to be effective in improving large number comprehension. Especially price-difference questions were affected. Accordingly, it can be concluded that the proposed game improves responses to price-different questions.

The hypotheses for the complete questionnaire and the price-difference questions hold true, while the hypotheses for the impact ratings and satisfaction ratings are rejected.

Hence, the research question may be answered positively. An infinite runner game with strategic aspects as described in section 2.2.1 can improve large number comprehension, if large number comprehension is measured with the tools described in section 2.4.

7 Discussion

7.1 Interpretation of results

These results may provide an extension of the research base regarding numerical cognition, as well as a novel training method to improve comprehension of large numbers.

The statistical significance of the results of the total questionnaire indicate the experimental game may indeed be an effective training method for understanding of large numbers. Appropriately, as numeracy improved, participants responded with smaller prices to price-difference questions, in accordance with previous results (Azar, 2011). Impact ratings alone did not decrease after playing the game, in contrast to previous research (Guay et al., Under review), which may indicate this measure was inadequately assessed, or that the game did not affect this particular aspect in numerical cognition. The same verdict holds for satisfaction ratings, which insufficiently corresponded with previous results (Landy et al., 2013).

7.2 Research design

The design of the research was appropriate for this research question as it was a controlled experimental setting. The within-subject factor embraced two moments of measurements, which was the most adequate division to test the effectiveness of the game in a short period of time. Since this research employs an experimental design, the results can confidently be interpreted to be caused by the proposed game.

Admittedly, the sample size in this study is small (20 students), and is therefore more prone to random effects than a larger sample. However, the sample was adequate to show an effect of the game on the participant's responses to the questionnaires. Furthermore, although an educational game such as the one proposed in this study is new, one may

argue the finish line at unit 25,000,000 is too little to be considered a large number. The intent of the game is to notice the difference in magnitude, which is indeed demonstrated by unavoidably changing the movement speed at various magnitudes along the way. The result may have been more prominent when increasing the magnitudes covered by the game. A stronger effect may have also occurred when playing the game for a longer amount of time, though this was infeasible in this research as the participants were recruited based on voluntary contribution.

The experiment was not double blind, as the experimenter knew which of the two games to present to the participant. The experimenter did not interact with the participants for reasons other than answering questions regarding the instructions.

The results may have been influenced by the timing difference between the two conditions. As the experimental group completed the post-test questionnaire later relative to the pretest questionnaire than the control group, the gathered information during this period may have matured more in the experimental group.

The results may also have been influenced by the fact that the experimental game is more demanding in terms of concentration and working memory than the control game. As active rehearsal provides better recall of skills and knowledge than passive revision (Norbert, James, & Otmar, 2009), so too may the experimental game provide a better integration of information than the control game.

7.3 Limitations of questionnaires

Since large number comprehension is not an often evaluated quality, finding the right measure to do so is a difficult task on its own. The measures used in this research were derived from previous research and built on the premise that a clear difference was previously observed before and after training.

The chosen question categories may be intrinsically limiting however. Although the questionnaire was carefully constructed to be as objective as possible, it is inevitable that some participants inherently respond differently than others. For example if a participant is a more conservative spender, they may choose lower prices in the price-difference questions overall than someone who is not concerned about money. In some cases in the price-difference questions, the proposed "worse" product was actually preferred by some participants. A few participants would prefer to visit an amusement park on a cloudy and rainy day because it would be less busy, for example. These limitations restrict the main goal of the questionnaire. Despite these limitations, the price-difference category still showed a significant effect.

The results of the impact rating questions decreased in both groups, which implies the questions in the pretest questionnaire were not tied closely enough to those in the post-test questionnaire. The experimental group did decrease more than the control group, though the effect was not significant. If these questions were constructed more carefully the effect may have proven significant.

As for the satisfaction rating questions, the results showed no significant effect at all. Both groups stayed on the same level before and after playing their respective game. This could be the result of the questions not coming across as numerically, but more as a matter of principle. The fact that both groups stayed on same level before and after playing the game, while observing a difference between the two groups affirms this notion. In retrospect, more numerically heavy questions such as those in the research by (Guay et al., Under review) might have provided a more accurate measure.

7.4 Future research

Future research may provide a more solid backing to this project. Firstly, a well-established and all-encompassing measurement of large number comprehension is needed to accurately conduct research such as the current. Furthermore, related to this research, it may be interesting to see whether the results get stronger the more the game is played. Also, re-

production with a larger sample and a more objective measure would provide a more solid justification of the found results.

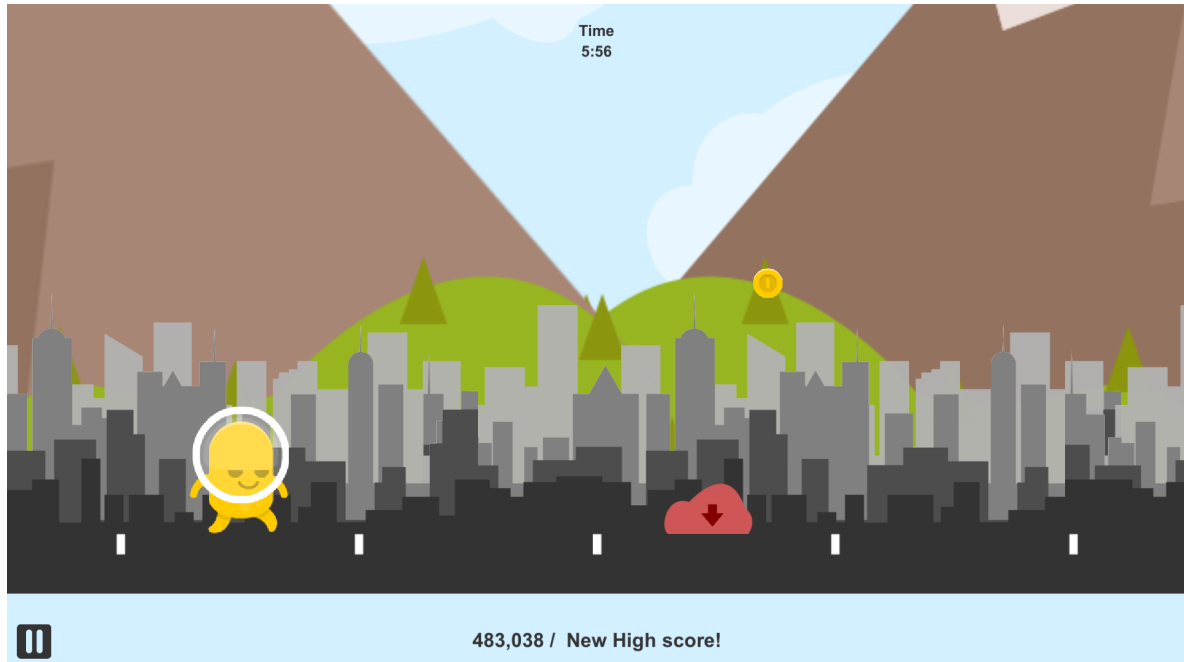
If any future research strengthens the proof of effectiveness of this game, people can effectively be trained on large number comprehension in the form of a game. This should help the public to be more aware of large numbers and what they truly mean.

References

- Azar, O. H. (2011). Relative thinking in consumer choice between differentiated goods and services and its implications for business strategy. *Judgment and Decision Making*, 6(2), 176.
- Barrio, P. J., Goldstein, D. G., & Hofman, J. M. (2016). Improving comprehension of numbers in the news. In *Proceedings of the 2016 chi conference on human factors in computing systems* (pp. 2729–2739).
- Barth, H. C., & Paladino, A. M. (2011). The development of numerical estimation: Evidence against a representational shift. *Developmental science*, 14(1), 125–135.
- Caryn, C. (1989). The psychophysics of spending. *Journal of Behavioral Decision Making*, 2(2), 69-80.
- David, L., Arthur, C., & Erin, O. (2017). Categories of large numbers in line estimation.
- Ellis, J. L. (2013). *Statistiek voor de psychologie*.
- Fechner, G. T. (1860). Gustav theodor fechner elemente der psychophysik 1860. *British Journal of Statistical Psychology*, 13(1), 1-10.
- Grewal, D., & Marmorstein, H. (1994). Market price variation, perceived price variation, and consumers' price search decisions for durable goods. *Journal of Consumer Research*, 21(3), 453–460.
- Guay, B., Davis, Z., DeLaunay, M., Charlesworth, A., & Landy, D. (Under review). Number comprehension impacts political judgments.
- Landy, D., Charlesworth, A., & Ottmar, E. (2014). Cutting in line: discontinuities in the use of large numbers by adults. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 36).
- Landy, D., Silbert, N., & Goldin, A. (2013). Estimating large numbers. *Cognitive science*, 37(5), 775–799.
- Laurillard, D. (2016). Learning number sense through digital games with intrinsic feedback. *Australasian Journal of Educational Technology*, 32(6).
- Moeller, K., Fischer, U., Link, T., Wasner, M., Huber, S., Cress, U., & Nuerk, H.-C. (2012). Learning and development of embodied numerosity. *Cognitive processing*, 13(1), 271–274.
- Moyer, R. S., & Landauer, T. K. (1967). Time required for judgements of numerical inequality. *Nature*, 215(5109), 1519.
- Noemí, P.-M., & Máximo, S. H. (2014). Educational games for learning. *Universal Journal of Educational Research*, 2(3), 230–238.
- Norbert, M., James, C. J., & Otmar, V. (2009). Active versus passive teaching styles: An empirical study of student learning outcomes. *Human Resource Development Quarterly*, 20(4), 397-418.
- Petrova, D. G., Pligt, J., & Garcia-Retamero, R. (2014). Feeling the numbers: On the interplay between risk, affect, and numeracy. *Journal of Behavioral Decision Making*, 27(3), 191–199.
- Siegler, R. S., & Opfer, J. E. (2003). The development of numerical estimation: Evidence for multiple representations of numerical quantity. *Psychological science*, 14(3), 237–250.

8 Appendices

8.1 Appendix 1. Control Game screen



8.2 Appendix 2: Example of data gathered during the game

Session: 5/30/2018 3:18:16 PM (UID 18)
Location: 20, effect: x 10
Location: 120, effect: + 40
Pause
Resumed (7:221 seconds in pause menu)
Pause
Resumed (15:58 seconds in pause menu)
Location: 440, effect: x 5
Pause
Resumed (7:488 seconds in pause menu)
Location: 1000, effect: / 10
Location: 1100, effect: x 20
Location: 2500, effect: x 10
Location: 13000, effect: + 7,500
Pause
Resumed (15:491 seconds in pause menu)
Pause
Resumed (50:901 seconds in pause menu)
Location: 35410, effect: + 350
Pause
Resumed (41:599 seconds in pause menu)
Location: 150000, effect: x 10
Pause
Resumed (4:209 seconds in pause menu)
Location: 300000, effect: / 100
Pause
Resumed (5:158 seconds in pause menu)
Pause

Resumed (8:137 seconds in pause menu)
 Location: 302000, effect: + 200,000
 Pause
 Resumed (6:906 seconds in pause menu)
 Location: 503000, effect: x 10
 Location: 530000, effect: + 70,000
 Location: 670000, effect: x 10
 Location: 800000, effect: / 100
 Location: 802800, effect: + 1,000,000
 Pause
 Resumed (23:927 seconds in pause menu)
 Location: 1806000, effect: x 100
 Location: 2000000, effect: Speed = 1
 Location: 2000024, effect: + 300
 Location: 2000340, effect: x 10
 Location: 2000500, effect: x 10
 Pause
 Resumed (5:723 seconds in pause menu)
 Location: 2002000, effect: + 8000
 Pause
 Resumed (1:731 seconds in pause menu)
 Location: 2011500, effect: x 10
 Pause
 Resumed (3:877 seconds in pause menu)
 Location: 2030000, effect: + 50,000
 Pause
 Resumed (3:577 seconds in pause menu)
 Pause
 Resumed (3:727 seconds in pause menu)
 Pause
 Resumed (8:303 seconds in pause menu)
 Location: 2125000, effect: x 10
 Location: 2200000, effect: x 10
 Location: 3500000, effect: x 10
 Location: 25000000, effect: Finish!
 Finished in 6.284626:17.07758

8.3 Appendix 3: Questionnaires

8.3.1 Pretest questionnaire

Price-difference questions

Q1: Say you want to buy a laptop. You are interested in a specific model that comes in two screen sizes, 15" and 13". The laptops are otherwise identical. assume you do graphic design and therefore prefer the larger screen. If the 15" laptop costs €830.-, how much must the 13" laptop cost for you to prefer it over the 15"?

- (i) €520.- (ii) €650.- (iii) €700.- (iv) €780.-

Q2: Assume you do the groceries in a supermarket around the corner. You spend €50 on average in this store. There is a cheaper supermarket that sells identical items but it's a 16 minute bike ride to get there. How much should the items cost in this store for you to prefer to go to this store instead?

- (i) €30.- (ii) €35.- (iii) €40.- (iv) €45.-

Q3: Assume you want to buy a house which you expect to be in good condition and are willing to pay €250,000 for. In a later visit you find the bathroom to be leaking and some floorboards in the living room to be rotten. With this new information (taking into consideration repair costs and effort), which value approximates best the new price you are willing to pay?

- (i) €190,000.- (ii) €210,000.- (iii) €225,000.- (iv) €240,000.-

Q4: Assume you want to buy a new car. You have a specific model in mind which costs €32,000 in the dealership (brand new). You also find an advertisement for this same car, second hand with 130,000 kilometers driven and in a colour you prefer less than the colour on the new car. The car has some slight signs of usage but is completely driveable. What amount approximates best the price you are willing to pay for the second hand car?

- (i) €12,000.- (ii) €19,000.- (iii) €22,000.- (iv) €25,000.-

Impact rating questions

Note: these questions are answered by giving a rating on a 9-point scale

Q1: Rate the impact of a €100 million increase in the national education budget.

Q2: Rate the impact of a 15% additional import tax in China on 128 products from the United States.

Q3: Rate the impact of reducing the natural gas extraction from 21 billion cubic metres to 12 billion cubic metres per year in a period of 4 years.

Q4: Rate the impact of a €12.5 million increase in the Marlboro cigarettes revenue.

Satisfaction rating questions

Note: these questions are answered by giving a rating on a 9-point scale

Q1: After days of negotiating with the estate agent on the price of the house you want she lowers the initial price of €275,000 by €10,000.

Q2: Two weeks after buying a new €650.- TV it breaks through no fault of your own. The manufacturer assumes fixing the TV will take three weeks due to a hardware issue with this model and compensates for the wait with €100 cash-back.

Q3: Upon arrival at a hotel your room, for which you paid €120 per night, appears to be overbooked. You are relocated to a smaller room with no windows and receive half of your payment back, you also receive a complimentary breakfast for the rest of your 5 day stay.

Q4: You invested €50,000 in the stock market and after a few nerve wracking weeks of heavy fluctuation your total investments have increased by 2%.

8.3.2 Post-test questionnaire

Price-difference questions

Q1: Assume you are going on holiday to New York. You can get a direct flight to New York for €720.-, which will take 8 hours. You can also take a flight with a layover, taking 12 hours in total. You do not need to worry about your luggage during the transfer. How much should this flight cost for you to prefer it over the direct flight?

- (i) €450.- (ii) €550.- (iii) €620.- (iv) €690.-

Q2: You want to go to an amusement park next week, you can buy a ticket for the sunday, for which a sunny day is forecasted, for €40.-. You can also go on friday for a reduced price, though it will be cloudy with a chance of rain. Assume you don't have a predefined preference for either day. How expensive should the ticket for friday be for you to prefer it over the sunday?

- (i) €20.- (ii) €25.- (iii) €30.- (iv) €35.-

Q3: Assume you have been saving up a couple of years for a trip around the world. On this trip you would visit all the locations you wanted to experience. The trip would cost you €21,000 and would last two months. Now imagine another worldly trip in which you would miss out some of your planned visits and stay in slightly worse hotels. This trip would take 5½ weeks. The rest of the details are equal to the first trip. What amount would approximate best the price you would be willing to pay for this shorter trip?

- (i) €11,000.- (ii) €13,000.- (iii) €16,000.- (iv) €18,500.-

Q4: Imagine you are buying a new house which you expect to be in good condition and are willing to pay €190,000 for. In a later visit you find out the stairs have termites and need to be reconstructed completely the bathroom tiles are also broken and in need of replacement. With this new information, which value approximates best the new price you are willing to pay?

- (i) €120,000.- (ii) €150,000.- (iii) €168,000.- (iv) €177,000.-

Impact rating questions

Note: these questions are answered by giving a rating on a 9-point scale

Q1: Rate the impact of ignoring a €150 billion government debt by the italian government (the total government debt is €2263 billion).

Q2: Rate the impact of a €200 million increase in the national defense budget.

Q3: Rate the impact of a 3% growth in the average consuming behaviour of the Dutch citizen.

Q4: Rate the impact of a €1 million decrease in the annual budget for the Amsterdam municipality.

Satisfaction rating questions

Note: these questions are answered by giving a rating on a 9-point scale

Q1: After days of negotiating with the estate agent on the price of the house you want she lowers the initial price of €220,000 by €12,000.

Q2: You bought a new iPhone for €750 and after 10 days it suffered water damage from leaving it on the porch in the rain for a couple of minutes. Your insurance covers water damage but is suspicious of the nature of the damage and tries to find a middle ground: your phone is repaired for a reduced cost of €50.

Q3: You want to buy a second hand bike. the seller asks €230 for the bike while you budget is €170. After bargaining for a while you manage to settle on €180 and decide to buy it.

Q4: You deposited €50,000 in your retirement savings account, which, after 10 years of interest resulted in €55,000.

8.4 Appendix 4: Data set used for analysis

	GROUP	PRICE_ BEFORE	IMPACT_ BEFORE	SATISFACTIO N_BEFORE	PRICE_ AFTER	IMPACT_ AFTER	SATISF ACTION _AFTER	TOTAL_NO RMALIZED _BEFORE	TOTAL_NO RMALIZED _AFTER
1	1	.6875	.5278	.7222	.5000	.4444	.6944	.6458	.5463
2	1	.6875	.6944	.6389	.6250	.6111	.7222	.6736	.6528
3	1	.6250	.7500	.6667	.4375	.6111	.4167	.6806	.4884
4	1	.5625	.5556	.6111	.5625	.5556	.4722	.5764	.5301
5	1	.8750	.7222	.7778	.5625	.7222	.7778	.7917	.6875
6	1	.5625	.7222	.6944	.3750	.5833	.5556	.6597	.5046
7	1	.6875	.7500	.4167	.5625	.5556	.8056	.6181	.6412
8	1	.5625	.6944	.6111	.5000	.5556	.6667	.6227	.5741
9	1	.6875	.8611	.8056	.4375	.8056	.7222	.7847	.6551
10	1	.6250	.5000	.5556	.5000	.5278	.6389	.5602	.5556
11	0	.6250	.6389	.4722	.5000	.6389	.6389	.5787	.5926
12	0	.5625	.5556	.6667	.6875	.5000	.5556	.5949	.5810
13	0	.5625	.5833	.3333	.7500	.4444	.3333	.4931	.5093
14	0	.3750	.6111	.6667	.3750	.6944	.6111	.5509	.5602
15	0	.6875	.7500	.7222	.6875	.6389	.6667	.7199	.6644
16	0	.6875	.5278	.5556	.6875	.5833	.3889	.5903	.5532
17	0	.4375	.6667	.5278	.3750	.5000	.5000	.5440	.4583
18	0	.6250	.5556	.4722	.5625	.4722	.4722	.5509	.5023
19	0	.6250	.7500	.7222	.6250	.6944	.8333	.6991	.7176
20	0	.5000	.6667	.7222	.8125	.8056	.8333	.6296	.8171

8.5 Appendix 5: Full analysis results

Measure	Time	Dependent Variable
total	1	TOTAL_NORM MALIZED_BE FORE
	2	TOTAL_NOR MALIZED_AF TER
price	1	PRICE_BEFO RE
	2	PRICE_AFT ER
impact	1	IMPACT_BEF ORE
	2	IMPACT_AFT ER
satisfaction	1	SATISFACTIO N_BEFORE
	2	SATISFACTIO N_AFTER

Figure 6: Within-Subject Factors

N		
GROUP	0	10
	1	10

Figure 7: Between-Subject Factors

Multivariate ^{a,b}									
Within Subjects Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power ^d
Time	Pillai's Trace	.384	3.322 ^c	3.000	16.000	.047	.384	9.965	.643
	Wilks' Lambda	.616	3.322 ^c	3.000	16.000	.047	.384	9.965	.643
	Hotelling's Trace	.623	3.322 ^c	3.000	16.000	.047	.384	9.965	.643
	Roy's Largest Root	.623	3.322 ^c	3.000	16.000	.047	.384	9.965	.643
Time * GROUP	Pillai's Trace	.432	4.059 ^c	3.000	16.000	.025	.432	12.178	.738
	Wilks' Lambda	.568	4.059 ^c	3.000	16.000	.025	.432	12.178	.738
	Hotelling's Trace	.761	4.059 ^c	3.000	16.000	.025	.432	12.178	.738
	Roy's Largest Root	.761	4.059 ^c	3.000	16.000	.025	.432	12.178	.738

a. Design: Intercept + GROUP
Within Subjects Design: Time

b. Tests are based on averaged variables.

c. Exact statistic

d. Computed using alpha = .05

Figure 8: Tests of Within-Subject Effects

Univariate Tests										
Source	Measure		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power ^a
Time	total	Sphericity Assumed	.015	1	.015	5.728	.028	.241	5.728	.620
		Greenhouse-Geisser	.015	1.000	.015	5.728	.028	.241	5.728	.620
		Huynh-Feldt	.015	1.000	.015	5.728	.028	.241	5.728	.620
		Lower-bound	.015	1.000	.015	5.728	.028	.241	5.728	.620
	price	Sphericity Assumed	.032	1	.032	4.796	.042	.210	4.796	.545
		Greenhouse-Geisser	.032	1.000	.032	4.796	.042	.210	4.796	.545
		Huynh-Feldt	.032	1.000	.032	4.796	.042	.210	4.796	.545
		Lower-bound	.032	1.000	.032	4.796	.042	.210	4.796	.545
	impact	Sphericity Assumed	.032	1	.032	8.382	.010	.318	8.382	.782
		Greenhouse-Geisser	.032	1.000	.032	8.382	.010	.318	8.382	.782
		Huynh-Feldt	.032	1.000	.032	8.382	.010	.318	8.382	.782
		Lower-bound	.032	1.000	.032	8.382	.010	.318	8.382	.782
	satisfaction	Sphericity Assumed	7.716E-5	1	7.716E-5	.007	.932	.000	.007	.051
		Greenhouse-Geisser	7.716E-5	1.000	7.716E-5	.007	.932	.000	.007	.051
		Huynh-Feldt	7.716E-5	1.000	7.716E-5	.007	.932	.000	.007	.051
		Lower-bound	7.716E-5	1.000	7.716E-5	.007	.932	.000	.007	.051
Time * GROUP	total	Sphericity Assumed	.015	1	.015	5.866	.026	.246	5.866	.630
		Greenhouse-Geisser	.015	1.000	.015	5.866	.026	.246	5.866	.630
		Huynh-Feldt	.015	1.000	.015	5.866	.026	.246	5.866	.630
		Lower-bound	.015	1.000	.015	5.866	.026	.246	5.866	.630
	price	Sphericity Assumed	.088	1	.088	13.322	.002	.425	13.322	.932
		Greenhouse-Geisser	.088	1.000	.088	13.322	.002	.425	13.322	.932
		Huynh-Feldt	.088	1.000	.088	13.322	.002	.425	13.322	.932
		Lower-bound	.088	1.000	.088	13.322	.002	.425	13.322	.932
	impact	Sphericity Assumed	.006	1	.006	1.441	.246	.074	1.441	.206
		Greenhouse-Geisser	.006	1.000	.006	1.441	.246	.074	1.441	.206
		Huynh-Feldt	.006	1.000	.006	1.441	.246	.074	1.441	.206
		Lower-bound	.006	1.000	.006	1.441	.246	.074	1.441	.206
	satisfaction	Sphericity Assumed	.000	1	.000	.000	1.000	.000	.000	.050
		Greenhouse-Geisser	.000	1.000	.000	.000	1.000	.000	.000	.050
		Huynh-Feldt	.000	1.000	.000	.000	1.000	.000	.000	.050
		Lower-bound	.000	1.000	.000	.000	1.000	.000	.000	.050

Figure 9: Univariate Tests

GROUP * Time						
Measure	GROUP	Time	Mean	Std. Error	95% Confidence Interval	
					Lower Bound	Upper Bound
total	0	1	.595	.023	.546	.644
		2	.596	.029	.535	.657
	1	1	.661	.023	.612	.711
		2	.584	.029	.523	.644
price	0	1	.569	.031	.503	.635
		2	.606	.038	.527	.685
	1	1	.656	.031	.590	.722
		2	.506	.038	.427	.585
impact	0	1	.631	.031	.565	.696
		2	.597	.035	.524	.670
	1	1	.678	.031	.613	.743
		2	.597	.035	.524	.670
satisfaction	0	1	.586	.039	.504	.669
		2	.583	.047	.484	.683
	1	1	.650	.039	.568	.732
		2	.647	.047	.548	.747

Figure 10: Estimated Marginal Means

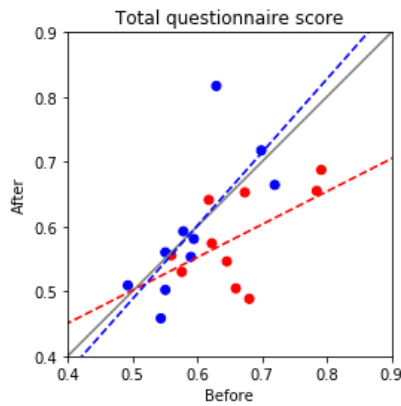


Figure 11: Profile plot total questionnaire

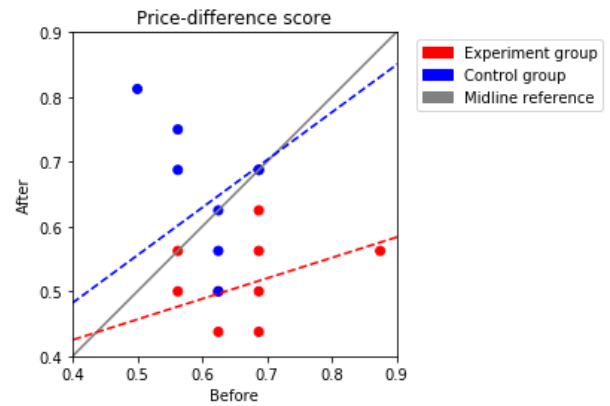


Figure 12: Profile Plot Price Questions

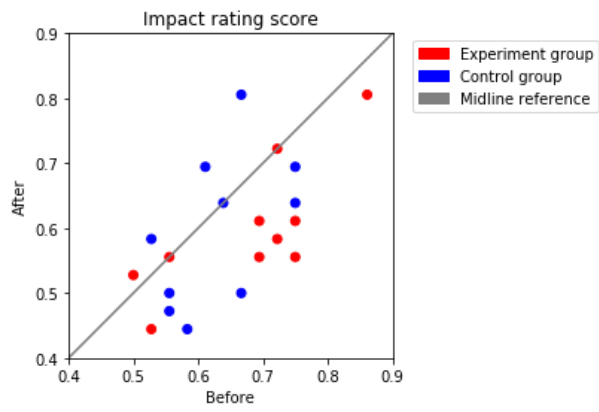


Figure 13: Profile plot Impact Questions (not significant)

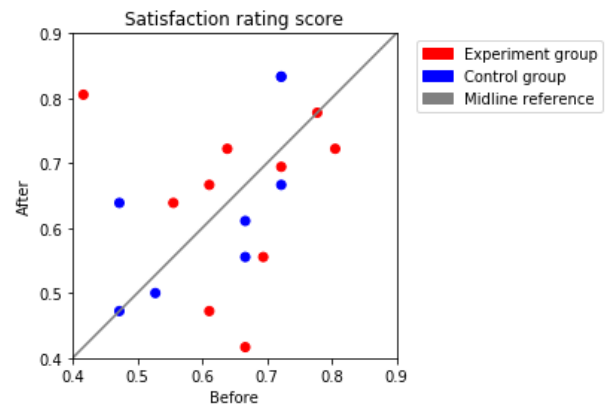


Figure 14: Profile Plot Satisfaction Questions (not significant)