***"My log has something to tell you"*:* the Value of Log Data in Game-Based Learning**

Mike Klaassen (6218008)

First assessor: Jos Jaspers

Second assessor: Pieter Wouters

June 7, 2021

Word count: 7992

*A reference to a line from the TV series *Twin Peaks* (1990-1991).

Abstract

This study investigated the added value of log data in Game-Based Learning, which is an instructional method to use the entertaining quality of a game for educational purposes such as learning, training and awareness and balances between features of games and instruction. Log data is a form of Game Learning Analytics, which combines the educational goals of learning analytics and the technical capabilities of game analytics. The log data consisted of the player's detailed behavioral actions, which were related to the final performance in an external test after playing the game *Zeldenrust*. Through analysis of the log data files, metrics were established for further exploration and statistical analysis. A significant impact of the metrics on the variance within the post-test score was found, but only game performance and the number of actions in the refrigerator game had a role in this significant result. This corresponds to earlier findings in Game Learning Analytics that the game itself could possibly be an assessment instrument instead of a test. More fine-grained results in exploration showed potentially relevant input for a teacher dashboard in a Game-Based Learning setting, such as strategies after an incorrect answer.

Keywords: Game Learning Analytics, log data, Game-Based Learning, assessment

"My log has something to tell you": the Value of Log Data in Game-Based Learning

Game-Based Learning (GBL) is an instructional method to use the entertaining quality of a game for educational purposes such as learning, training and awareness (Wouters & Van Oostendorp, 2013). An educational game ultimately should balance between features of games and instruction. Learner motivation is supported by using game features and deeper learning is generated through instructional features of the material (Mayer, 2011; Ritterfeld & Weber, 2006). From earlier research, it is shown that GBL can have a positive impact on outcomes such as cognition and knowledge acquisition (Boyle et al., 2016; Kim & Ifenthaler, 2019).

Researchers mostly use summative tests to determine if participants learn from GBL (Alonso-Fernández et al., 2020). These tests could be an opportunity to investigate how much the participant has learned, but not to determine how learning took place (Nguyen et al., 2020). Therefore, more research should be done about possibilities of the game being an assessment instrument instead of using a summative test. This could be possible through *Game Learning Analytics* (GLA), which is "the process of capturing, storing, analyzing, and obtaining information from players' interactions with a serious game" (Alonso-Fernández et al., 2020, p. 2; Freire et al., 2006).

The game could ideally be used as a test through prediction of learning results based on game interactions represented in the form of *log data*, which consists of interactions such as the time spent on a game, the number of completed tasks and social interaction (Alonso-Fernández et al., 2019; Kim & Ifenthaler, 2019). This log data could serve two purposes. Firstly, it could be a method for teachers to look at what elements of the game students struggle with (Westera et al., 2013). Secondly, researchers could get more insight on what kind of behavior is related to the result of a summative test, instead of using a test score to hypothesize what could have possibly happened in the game (Hawlitschek & Köppen, 2014). Therefore, this research will analyze the existing log data of earlier experiments to investigate the value of log data in GBL and if certain behaviors in the log data can be related to performance on a summative test.

Theoretical Framework

Game-Based Learning

According to Wouters and Van Oostendorp's (2013) definition of GBL, a game should be *interactive* (e.g., the game responds to the player's actions), based on agreed *rules* and *constraints* (e.g., being in a simplified world), has a goal that comes with *challenges* (e.g., goals and levels) and provides *feedback* (e.g., scores, adaptivity) (Mayer, 2014; Wouters & Van Oostendorp, 2013). Other features like *competition* and *narrative* could be characteristics of a game, but do not necessarily have to be implemented.

Games have the potential to influence learning in two ways: by directly changing the cognitive process or indirectly by affecting motivation (Wouters & Van Oostendorp, 2017). The cognitive perspective of GBL focuses on the optimization of mental models to learn from a game and the cognitive load that is demanded from the player. Games can help to generate generative processing, which is cognitive processing in order to make sense of the material (Mayer, 2020) and motivate the learner to understand the material (Mayer, 2019), but can also result in extraneous processing, which is cognitive processing that is detrimental to learning (Mayer, 2020). Still, other theoretical perspectives must be considered as well. For example, an interaction with a character within GBL can result in an emotional connection from an affective perspective, which then can increase cognitive engagement (e.g., the willingness to work on a problem in the game) (Plass et al., 2020).

Game Learning Analytics

GLA is a merging of two earlier forms of data-driven analysis of user interaction: *learning analytics* and *game analytics* (Freire et al., 2016). Learning analytics is a technique to collect, analyze and report educational data (e.g., information from digital learning platforms and administrative systems) with the goal to understand and optimize learning (Alonso-Fernández et al., 2020; Loh et al., 2015; Siemens & Long, 2011) and game analytics are techniques within the entertainment gaming industry (e.g., “user metrics” such as return frequency) that are used to investigate how users play a game, which errors occur and how the game can be improved to keep and gain customers (Freire et al., 2016; Loh et al., 2015; Seif El-Nasr et al., 2013). Because even a short GBL session can generate much data (Freire et al., 2016), learning analytics is a fitting method for analysis (Alonso-Fernández et al., 2020). GLA, which is an embedded form of game-based assessment (Ifenthaler et al., 2012) combines the educational goals of learning analytics and the technologies from game analytics (Freire et al., 2016).

Alonso-Fernández et al. (2019) determine several practical applications of GLA. GLA can be performed on-line and off-line, with on-line being focused on users’ behavior during game play and off-line on analysis after data collection to reveal patterns of user interaction (Freire et al., 2016). It can be used to predict the impact (e.g., learning) of the game and is a method to relate performance to player characteristics through off-line GLA (e.g., clustering novice-expert players). Visualizing learning through real-time information on a dashboard (e.g., showing the learner's progress) to inform teachers is another function of GLA. GLA could play a role in the development process of GBL as well, by providing information to support game design and corresponding educational assessment (Alonso-Fernández et al., 2019).

Prediction through Game Learning Analytics

One of the possible functions of GLA is prediction of performance through game interactions (Alonso-Fernández et al., 2019). Before starting to apply GLA in GBL, the to be measured variables within the game must be determined. The Evidence-Centered Design model by Oliveri et al. (2019) can assist in determining the information that needs to be tracked, how this should happen and how data should be interpreted (Perez-Colado et al., 2018).

The most relevant step within Evidence-Centered Design for GLA is the Conceptual Assessment Framework, which consists of several models to determine assessment (Oliveri et al., 2019). The Competence Model contains the skills that must be acquired. The Evidence Model contains which actions are considered as evidence and how these are connected to the skills in the Competence Model. The Task/Action Model determines the environment in which the skills in the Competence Model must be demonstrated. Eventually, the Assembly Model combines data from the Competence, Evidence and Task/Action Models to make assumptions of the skill level of the competencies in the Competence Model and assess learning (Shute & Kim, 2014).

The Conceptual Assessment Framework has been used in *stealth assessment* (Shute & Kim, 2014), which is adaptive formative assessment within GBL (Alonso-Fernández et al., 2019). This formative assessment method is characterized by its nondisruptive measurement of learning, which means that the player should not see or feel that they are being assessed. For example, when a player shows a low level of creativity in problem solving while this is being highly valued by the Competence Model, the game can automatically place more emphasis on the acquisition of creative problem solving by making sure that the player cannot solve problems in an obvious way (Shute & Kim, 2014).

Alonso-Fernández et al. (2020) developed two phases to investigate if game interactions can predict performance on a test. In the Game Validation Phase, the effectivity of the game is tested through pre- and post-tests in combination with game interaction measures. Then, prediction models are determined to look at which game interaction variables can predict performance in combination with a pre- or post-test. The most accurate prediction model could be used automatically as an algorithm for assessment in the Game Deployment Phase as an indicator for teachers to see pupils' levels of acquired knowledge in the game. In the application of the Game Validation Phase, Alonso-Fernández et al. (2020) indicated that the post-test performance possibly could be predicted by solely analyzing game interaction data.

Log Data in Game Learning Analytics

Game interactions could be registered in the form of log data, which is a computer-generated source of information that contains each action that a player performs in a game (Greiff et al., 2014; 2015). The player's behavior is represented in a data set (Hawlitschek & Köppen, 2014). Log data could possibly be a step towards opening the "black box of cognitive processing" (Greiff et al., 2015, p. 2) by giving more insight in how a player has interacted with a problem and if this has resulted in learning (Greiff et al., 2015). It allows for a more fine-grained analysis of player behavior (Hawlitschek & Köppen, 2014).

Westera et al. (2013) performed an exploratory study on existing log data for GBL about environmental policy problems. The log file contained user actions such as opening information sources, watching videos, and the final score on a test, which were connected to a specific user and timestamp. These were considered basic variables but showed that gaming logging data has potential and can bring certain phenomena to the surface, such as the wide variation in the duration of accessing different sources of information.

Liu et al. (2016) investigated log data in the form of the frequency and duration of tool use in GBL and found that high performers on a science knowledge test used the more appropriate tools for advanced stages of problem solving in GBL while low performers used the tools that were deemed “more fun”. By connecting learner performance to learner behavior, scaffolding to assist learning in GBL could be created (Hwang et al., 2012; Liu et al., 2016).

Instead of mostly using frequencies, log data can be explored extensively as well. DiCerbo et al. (2015) targeted which log data to explore by first analyzing descriptive statistics and using statistical visualization methods such as histograms and boxplots. These visualization methods indicate positive and negative outliers in the data, which are explored in detail in the log data to see if they can show something meaningful about the learning (or non-learning) process and which specific actions can indicate difficulties or ease in learning. DiCerbo et al. (2015) indicated outliers through statistical visualization of the activity in an educational SimCity game. Through the log data, it was found that one player chose a strategy that was not corresponding to the learning goals: the player was destroying cities instead of building and sustaining them. Two other outliers were playing the game as asked but struggled with the learning goal. As shown in DiCerbo et al. (2015), log data can assist researchers and teachers to make the distinction between low scoring pupils because they struggle with the material or because they are not seriously performing a task.

Present Study

This research is an exploratory study on the use of log data as a form of GLA. The first research question is: *Which game behaviors do students with different performance levels exhibit in GBL?* Previous studies (Liu et al., 2016; Westera et al., 2013) mostly used frequencies and scores to analyze game behavior. This research will additionally investigate how participants approached a problem by zooming in on their steps in the log data in solving a mathematical problem. Differences between the pre- and post-test will be used to indicate which players progressed or stagnated after playing the game as well as the high and low scoring participants in the post-test score (Liu et al., 2016). These players will be analyzed in detail through log data, which contains many interaction variables (e.g., using a calculator). This part of the research will be called the Exploration Phase.

Indicating meaningful game behaviors based on progress (or stagnation) from the pre- to the post-test score, the pre- and post-test score, *Tempo Test Rekenen* (TTR) score, exploration of the log data and frequencies of selected interaction variables is the first step to determine which variables should be investigated further in the second research question, which is: *To what extent can interaction variables from a log file in GBL predict performance on a post-test?* This research question is related to the Game Validation Phase by Alonso-Fernández et al. (2020) and is the Analysis Phase in this research. After determining which interaction variables contain a potentially meaningful pattern in the first research question, these will be connected to the post-test results to determine if these interaction variables could play a role in explaining performance on a test.

Method

Research Design and Participants

This research uses existing data sets from earlier GBL experiments with the game *Zeldenrust* (Vandercruysse et al., 2015) and an earlier Master Thesis (Van Leeuwen, 2018). The participants in the selected research projects (*N* = 335) were students in prevocational education between the ages of 12 and 16. Appendix 1 gives an overview of the types of research, the number of participants in each research and the content that is available for analysis.

Used Instrumentation of the Existing Research

TTR

The TTR is a national arithmetic test (De Vos, 1992). In this test, students receive a paper with 200 arithmetic problems (e.g., 35-17 = …), and they must solve as many as they can within five minutes. The number of correctly solved problems indicates the level of computational fluency. A score above 100 is deemed "computationally fluent" (Ter Vrugte et al., 2015).

Pre/Post-Test Score

The pre/post-test consisted of sixteen open-ended questions and scores of respectively 310 and 313 participants were available. There were questions for each proportional reasoning problem (missing value, comparison and transformation). These questions were correspondent to the material in the game and its increasing difficulty (Ter Vrugte et al., 2015).

Zeldenrust

Zeldenrust is a mathematics game with a focus on proportional reasoning. The target group were 12-to-16-year-old prevocational students, since they seem to struggle with proportional reasoning and have a low motivational level for mathematics (Vandercruysse et al., 2015). The game consists of three sub games which are connected to a specific proportional reasoning skill with four increasing difficulty levels for each sub game.

The game is situated in a hotel. The main character wants to make some money in the holidays and their uncle and aunt have a hotel where they can work. The three sub games/tasks that the player must complete are filling refrigerators (missing values), mix/blend cocktails (transformation problems) and serve drinks (comparison problems) (Vandercruysse et al., 2015).

Figure 1

Screenshot of the Refrigerator Game in Zeldenrust

Note. http://www2.projects.science.uu.nl/mathgame/zeldenrust/thegame.html

For example, in the refrigerator game (see Figure 1), players must solve missing value problems. These missing value problems are presented in the form of filling a refrigerator. The player first must find out the number of bottles they have to place in the refrigerator by calculating the missing bottles in the proportion. Players can click, drag and drop the bottles in the refrigerator. They confirm their answer by closing the refrigerator door and receive feedback on their answer. If it was incorrect, they receive another attempt (Vandercruysse et al., 2015).

Procedure

The results on the pre-test and post-test will be used as a starting point of the analysis in the Exploration Phase. The score differences between the pre-test and post-test will be computed and a boxplot is generated to visualize the difference score and find outliers (DiCerbo et al., 2015). Another boxplot is generated for the post-test score to see which participants scored exceptionally high or low after playing the game.

Figure 2

Fragment of Log Data

logid	userid	version	game	level	actie	timestamp	datum_tijd	current_assignment	assignment_correct	extra_info
2749660	5053	nosurrand	1	1	3	0	2-11-2015 08:54	1	0	
2749661	5053	nosurrand	1	1	6	0	2-11-2015 08:54	1	0	fanta,1pack
2749662	5053	nosurrand	1	1	6	0	2-11-2015 08:54	1	0	fanta,10pack
2749664	5053	nosurrand	1	1	1	8	2-11-2015 08:54	1	0	cola,5pack
2749665	5053	nosurrand	1	1	6	8	2-11-2015 08:54	1	0	cola,5pack
2749666	5053	nosurrand	1	1	7	8	2-11-2015 08:54	1	0	cola,5pack
2749667	5053	nosurrand	1	1	6	10	2-11-2015 08:54	1	0	cola,5pack
2749668	5053	nosurrand	1	1	7	11	2-11-2015 08:54	1	0	cola,5pack
2749669	5053	nosurrand	1	1	6	11	2-11-2015 08:54	1	0	cola,5pack
2749671	5053	nosurrand	1	1	2	13	2-11-2015 08:55	1	0	cola,5pack
2749672	5053	nosurrand	1	1	1	13	2-11-2015 08:55	1	0	cola,5pack
2749673	5053	nosurrand	1	1	21	13	2-11-2015 08:55	1	0	cola,5pack
2749674	5053	nosurrand	1	1	1	35	2-11-2015 08:55	1	0	cola,5pack
2749675	5053	nosurrand	1	1	6	36	2-11-2015 08:55	1	0	cola,5pack
2749676	5053	nosurrand	1	1	7	37	2-11-2015 08:55	1	0	cola,5pack
2749677	5053	nosurrand	1	1	2	37	2-11-2015 08:55	1	0	cola,5pack
2749678	5053	nosurrand	1	1	6	37	2-11-2015 08:55	1	0	cola,5pack
2749679	5053	nosurrand	1	1	1	39	2-11-2015 08:55	1	0	cola,1pack

Each game interaction was registered in a log file as seen in Figure 2. This log file consists of the user-ID, action-ID, game version, which sub game, which level, which action, timestamp, current assignment, if the assignment is correct and extra information about an action (see Appendix 2 for further details of the different registered actions in the log file). In order to answer the first research question, the goal was to find potentially meaningful events in the log files of the selected players (progression/stagnation from pre- to post-test and exceptionally high and low scoring players) by trying to gain an understanding of their process in the game. This was done by analyzing different individual log data files by reading through the log data. While reading, the goal was to imagine what the player could have possibly thought during the gaming session and find significant actions that could characterize the player.

Another exploration method was to think of how frequencies and actions could indicate potentially relevant information on a dashboard in an off-line (seeing the data after playing) or on-line setting (seeing the data during playing) (Alonso-Fernández et al., 2019): *Which indicators would a teacher need to monitor and possibly intervene in a game situation*? Every time when a potentially meaningful action was encountered, the participant number was written down in combination with the action and time code and the interpretation of the action (Appendix 6). Potential interaction types for further exploration could be the number of steps taken to finish the game and how a participant approached a mathematical problem (e.g., use of tools). This exploration process will result in the creation of *metrics*, which will be used to analyze a larger portion of the log data.

In the Analysis Phase, a multiple regression analysis with an α of 0.05 will be conducted in which the dependent variable (post-test) will be correlated with several independent metrics to determine to which extent they could predict the result on the post-test most accurately.

Results

Data cleaning

The first step in exploring the data was to perform a check on the entire statistical data file and log data file. Thirty-one players were missing, meaning that they do not have a connection with a user-ID in the game and therefore cannot be used for the multiple regression analysis. These players still have value for the Exploration Phase in finding potentially meaningful actions that could be formulated into metrics for analysis. The minimum and maximum score of the pre-, post-test and TTR were generated to see if there were any odd scores (e.g., 99 on the post-test when the highest possible score is 16). There were no scoring mistakes in the data set.

Then, an overview of the number of actions per minute was generated. There were extraordinarily high amounts of actions in the log data of 51 players. The highest number of actions was 46,336 in one minute. The player shows an extreme degree of repetition in the refrigerator game (Appendix 3). This might be the cause of the game over-registering actions, indicated by many of the same actions happening in one second. Another possible explanation from the researcher who provided the existing data was that players found out that you could fill the entire room with bottles. Still, the extremely high numbers that some players generate are not humanly possible, even when "fooling around" in the game. Therefore, an extreme amount of repetition of an action will be reduced to a maximum of two different actions in one second (e.g., taking a bottle out of the refrigerator and grabbing another one) for the Analysis Phase. After this reduction, there mostly were players who had an extremely low number of actions. When looking at their log data files, it seems that their sessions are incomplete, playing for approximately five minutes instead of twenty to thirty minutes. These players (n = 10) were omitted from the data set in the Analysis Phase.

Start of the Exploration Phase: exploration of the Zeldenrust data set

Another aspect that had to be checked were the differences between the various experiments. The Zeldenrust game was customizable in terms of levels, assignments, maximum attempts and assistance tools (Vandercruysse et al., 2015). Most participants have received the same number of assignments and attempts (Appendix 4). The Master Thesis experiment (Van

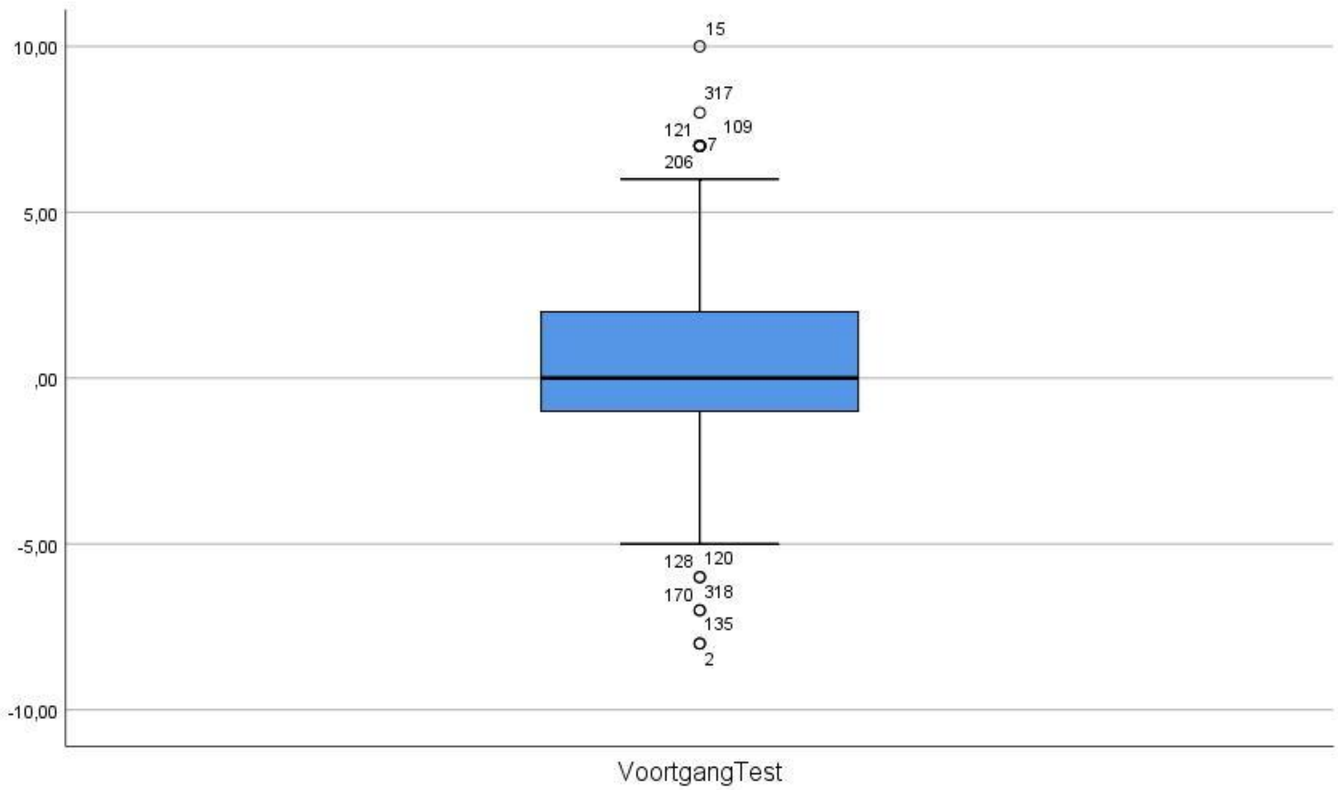
Leeuwen, 2018) has the biggest differences, being a smaller experiment and the Surprise 2015 version of the game contains more assignments. This is due to the *value-added* nature of these studies (Mayer, 2014). According to the researcher who provided the existing data, there were more assignments needed to measure the effect of the surprising intervention. In this case, it meant that due to a surprising event, the player had to change a solving strategy. These different characteristics of the games must be considered in Exploration and Analysis, since these can also influence the frequencies of metrics that are to be determined. When looking at the log data and to which extent these different characteristics are registered, some value-added events are registered as a code in the log data (e.g., part-task practice) while others are not (e.g., a surprising event). Also, the extra information in the log data contains context about the action taken (e.g., the number of bottles that are placed in the refrigerator) but does not show the actual problem that the player is trying to solve. The absence of this context resulted in a more limited and general exploration of the log data files.

DiCerbo et al. (2015) recommend using graphic representations such as histograms and boxplots of the data as a starting point for exploration. A boxplot was generated for the pre-, post- and TTR test score to see which participants scored exceptionally high or low (Appendix 5). This mainly indicated the high scoring outliers on the pre-, post- and TTR test score. When looking at the frequencies per score, there are respectively 33 and 53 players out of the entire participant group who score 10 out of 16 or higher in the pre- or post-test. The Zeldenrust game was targeted towards the mathematics skills that students in prevocational education struggled the most with (Vandercruysse et al., 2015). Therefore, it is plausible for players to score poorly on the pre- and/or post-test with a relatively small group scoring high on both tests.

The TTR scores (De Vos, 1992) were explored as well. For the Zeldenrust experiments, a minimum score of 100 was indicated as computationally fluent (Ter Vrugte et al., 2015). Therefore, a data set was created in which only players with a TTR below 100 (n = 74) were selected. The players with a score between 60 and 100 have differing results in terms of game percentage (the ratio of correct to incorrect answers in the game) and use of assistance tools (e.g., a calculator). Some players score below 100 on the TTR, but have a high score (e.g., 13 points out of 16) on the post-test (Appendix 8). The correlation between the TTR and both tests is moderate (Appendix 9). This could indicate that a TTR score is a smaller factor in the overall performance.

Figure 3

Boxplot of Progression or Stagnation from Pre- to Post-Test Score



Log data analysis: establishment of metrics

The boxplot of the difference score (post-test score minus pre-test score) brought a clearer division and positive and negative outliers were found (Figure 3). There was a group of players who progressed from the pre- to post-test score and players who stagnated. One of the most striking interactions in the Exploration Phase was found within the group of players who stagnated and were negative outliers in the boxplots. They do not seem to be bad players from a mathematics performance perspective, considering their pre-test score. These players start quite good in the game as well but seem to lose steam after a few assignments.

This was the establishment of a potential first metric: finishing tasks without truly performing them. General descriptions of the established metrics in the Exploration Phase and metrics that were deemed meaningful in earlier articles (Alonso-Fernández et al., 2020; Liu et al., 2016) are found in Table 1 and will be described in further detail.

Table 1

Potential Metrics

Metric	Visibility in data
Fed up	Pressing twice or more on the Done button without any other action in between.
Calculator	Frequency of accessing the calculator.
Game manual	Frequency of accessing assistance in how the game functions.
Assistance board	Frequency of accessing assistance for solving arithmetical problem.
Handbook	Frequency of accessing descriptions of how to solve problems in each game.

Game percentage	Percentage (e.g., 25/32 correct = 78%).
Downtime	No action for a minimum of 30 seconds**.**
Solving strategy after incorrect answer	The strategies taken after an incorrect answer.
Effectivity solving strategy after giving an incorrect answer	Combination of solving strategy and if the answer is correct or incorrect**.**
Progression or stagnation after game	The difference between the pre- and post-test score.
Effectivity of using assistance tools (correct answer)	When an assistance tool is applied**,** and this results in a correct answer**.**
Effectivity of using assistance tools (incorrect answer)	When an assistance tool is applied, and this results in an incorrect answer**.**

Metric: Fed Up

This is visible in the log data when a player presses the Done button twice (Action 4 in the log data, see Appendix 2 and 7) or more in a row without any actions in between pressing this button in succession. This could be interpreted as the player wanting to get on with the game and being fed up with it. If the post-test is conducted after the game session, this mood could possibly have been transferred to the post-test and influence performance. This behavior was not found with players who progressed remarkably and were outliers in the boxplots as well.

Metric: Progression after the game

Progression after the game is considered as the progression from the pre- to the post-test and not the progression within the game session in terms of levels. Players who stagnated progressed further in terms of levels because they possibly rushed through the game, while there are also players who got far because they performed well and quickly solved problems. These were also players who were positive outliers in the boxplot, had high scores on the pre-, post-test and TTR and sporadically used assistance tools in the game (Appendix 8). This potentially indicates that metrics that focus on quality are more important for performance than quantity. It is not about the number of steps a player takes to solve a problem, but if the problem is eventually solved correctly. There are players that have a low number of actions, but poor performance in the game and pre- and/or post-test. A subset of these players was mentioned earlier, because they stagnated (Appendix 7).

Exploring specific conditions in the log data

After exploring the variables in the data set from the combined experiments, the eight different conditions (Appendix 10) were split into different data sets. The data sets of the *control* and *part-task/worked example* conditions were used for a more extensive exploration of the log data. The control condition represents a default version of the game and therefore is the easiest to try to understand and the part-task/worked example version has an added value which could influence game performance. The first step of data visualization (DiCerbo et al., 2015) was repeated by generating histograms and boxplots (Appendix 11). Within these conditions, the highest and lowest scoring players were explored further in the log data. This brought some more interesting phenomena to the surface.

Figure 4

Two Different Scenarios After an Incorrect Answer

2005	controle	1	1	4	212	05/17/20	2	0
2005	controle	1	1	2	226	05/17/20	2	0
2005	controle	1	1	2	226	05/17/20	2	0
2005	controle	1	1	1	232	05/17/20	2	0
2005	controle	1	1	7	232	05/17/20	2	0
2005	controle	1	1	6	233	05/17/20	2	0
2005	controle	1	1	7	233	05/17/20	2	0
2005	controle	1	1	2	233	05/17/20	2	0
2005	controle	1	1	1	234	05/17/20	2	0
2005	controle	1	1	2	235	05/17/20	2	0
2005	controle	1	1	6	235	05/17/20	2	0
2005	controle	1	1	1	236	05/17/20	2	0
2005	controle	1	1	7	236	05/17/20	2	0
2005	controle	1	1	2	237	05/17/20	2	0
2005	controle	1	1	6	237	05/17/20	2	0
2005	controle	1	1	1	238	05/17/20	2	0
2005	controle	1	1	7	238	05/17/20	2	0
2005	controle	1	1	6	238	05/17/20	2	0
2005	controle	1	1	2	239	05/17/20	2	0
2005	controle	1	1	1	241	05/17/20	2	0
2005	controle	1	1	6	241	05/17/20	2	0
2005	controle	1	1	7	241	05/17/20	2	0
2005	controle	1	1	6	241	05/17/20	2	0
2005	controle	1	1	7	242	05/17/20	2	0
2005	controle	1	1	2	243	05/17/20	2	0
2005	controle	1	1	6	243	05/17/20	2	0
2005	controle	1	1	4	244	05/17/20	2	0
2005	controle	1	1	1	252	05/17/20	2	0
2005	controle	1	1	7	252	05/17/20	2	0
2005	controle	1	1	2	253	05/17/20	2	0
2005	controle	1	1	1	254	05/17/20	2	0
2005	controle	1	1	7	254	05/17/20	2	0
2005	controle	1	1	2	255	05/17/20	2	0
2005	controle	1	1	1	256	05/17/20	2	0
2005	controle	1	1	7	257	05/17/20	2	0
2005	controle	1	1	2	257	05/17/20	2	0
2005	controle	1	1	1	259	05/17/20	2	0
2005	controle	1	1	7	259	05/17/20	2	0
2005	controle	1	1	2	260	05/17/20	2	0
2005	controle	1	1	4	262	05/17/20	2	0

Interpretation of actions

212 -> Answer is incorrect.

226 – 244 -> Player tries to solve problem, does not use assistance tools.

244 -> Answer is incorrect for the second time.

244-252 -> Short break between incorrect answer and next action -> Not thinking thoroughly in how to solve the problem?

252-260 -> Player tries to solve the problem without a real direction.

262 -> Answer is incorrect for the third time -> Player is forced to move on to the next assignment.

1102	pws	1	1	4	169	04/23/2	1	0
1102	pws	1	1	1	177	04/23/2	1	0
1102	pws	1	1	7	177	04/23/2	1	0
1102	pws	1	1	2	178	04/23/2	1	0
1102	pws	1	1	36	180	04/23/2	1	0
1102	pws	1	1	3	182	04/23/2	1	0
1102	pws	1	1	18	182	04/23/2	1	0
1102	pws	1	1	22	208	04/23/2	1	0
1102	pws	1	1	23	208	04/23/2	1	0
1102	pws	1	1	36	247	04/23/2	1	0
1102	pws	1	1	36	247	04/23/2	1	0
1102	pws	1	1	22	248	04/23/2	1	0
1102	pws	1	1	1	254	04/23/2	1	0
1102	pws	1	1	7	254	04/23/2	1	0
1102	pws	1	1	6	254	04/23/2	1	0
1102	pws	1	1	7	255	04/23/2	1	0
1102	pws	1	1	6	255	04/23/2	1	0
1102	pws	1	1	7	255	04/23/2	1	0
1102	pws	1	1	2	255	04/23/2	1	0
1102	pws	1	1	1	257	04/23/2	1	0
1102	pws	1	1	6	258	04/23/2	1	0
1102	pws	1	1	7	258	04/23/2	1	0
1102	pws	1	1	2	259	04/23/2	1	0
1102	pws	1	1	6	259	04/23/2	1	0
1102	pws	1	1	1	260	04/23/2	1	0
1102	pws	1	1	6	260	04/23/2	1	0
1102	pws	1	1	7	260	04/23/2	1	0
1102	pws	1	1	6	261	04/23/2	1	0
1102	pws	1	1	2	261	04/23/2	1	0
1102	pws	1	1	1	262	04/23/2	1	0
1102	pws	1	1	6	263	04/23/2	1	0
1102	pws	1	1	7	265	04/23/2	1	0
1102	pws	1	1	2	265	04/23/2	1	0
1102	pws	1	1	1	266	04/23/2	1	0
1102	pws	1	1	6	266	04/23/2	1	0
1102	pws	1	1	2	266	04/23/2	1	0
1102	pws	1	1	1	267	04/23/2	1	0
1102	pws	1	1	6	267	04/23/2	1	0
1102	pws	1	1	2	268	04/23/2	1	0
1102	pws	1	1	1	269	04/23/2	1	0
1102	pws	1	1	6	269	04/23/2	1	0
1102	pws	1	1	2	270	04/23/2	1	0
1102	pws	1	1	7	270	04/23/2	1	0
1102	pws	1	1	6	270	04/23/2	1	0
1102	pws	1	1	1	270	04/23/2	1	0
1102	pws	1	1	7	270	04/23/2	1	0
1102	pws	1	1	6	270	04/23/2	1	0
1102	pws	1	1	6	271	04/23/2	1	0
1102	pws	1	1	2	272	04/23/2	1	0
1102	pws	1	1	4	274	04/23/2	1	1

Interpretation of actions

169 –> Answer is incorrect.

177-182 –> Player tries to solve it but is struggling.

182-248 -> Using multiple assistance tools (18 – game manual, 22 – assistance board, 23 – handbook).

254-272 -> Player is solving problem after using assistance tools.

274 -> Answer is correct.

Metric: Downtime

The amount of "downtime" that players take could be meaningful for performance. "Downtime" was conceptualized as thirty seconds or more without any actions in the log data file for exploration. While it is not fully clear what the reason of "downtime" could be, possible interpretations could be not paying attention to the game, thinking about an answer or struggling with finishing the assignment (Appendix 12).

Metric: Solving strategies after incorrect answer and the eventual result

Solving strategies after an incorrect answer was another interesting pattern to watch in the log data. This was conceptualized by the strategy that a player takes after an answer is incorrect and eventually giving the correct answer. From just exploring and reading the log data, it seems that some players just go on quickly, while others use assistance tools (see Figure 4), and others seem to think about their answer without any action in between and then fill in the correct answer in a short amount of time (see Appendix 13).

Dividing players into types

Then, an attempt was made to divide players into types. For the division in types, the game percentage was calculated as well, which was the ratio of correct to incorrect answers (e.g., 25 correct answers out of 32 = 78%). The division could have been a convenient step to establish predictor models. Still, this was not successful. Two possible types were indicated: players who were proficient in the pre-, post-test and game percentage and did not use many assistance tools (see Appendix 8), and players who are fed up with the game after a while (see Appendix 7). In between these two types of players, there are no other potential types. There are players who have a great score in the tests and score poorly in the game and vice versa. Still, the attempt to establish types of players brought new information to the surface.

Metric: Effectivity of using assistance tools

When looking at a small sample of players (*n* = 57), there is a difference in assistance tool usage (Appendix 8). This difference is reminiscent of the results in Liu et al.'s (2016) research on scientific problem solving. The different assistance tools have various purposes and could indicate diverse levels of assistance needed. The proficient players (players who score 10 or above on the post-test) only sporadically use a calculator, possibly to check an arithmetical computation. They do not use the other assistance tools: the game manual, the assistance column for arithmetical steps (assistance board) and the manual with guides on how to solve the mathematical problems in the game (handbook). The game manual is used by players who have trouble in understanding the game mechanics, not the mathematical problem solving per se.

The game manual could have had less impact on the external post-test, since it is unique to the gaming environment. The assistance column for computational steps and the "how to" handbook guide could indicate bigger problems in solving a mathematical problem and be of relevance for performance in the post-test. Therefore, the effectivity of using an assistance tool could also be an insightful metric. "Effectivity" is conceptualized in terms of if the answer is correct after using an assistance tool. If a player uses many assistance tools, but has many incorrect answers, this could doubt the effectivity of using the assistance tool in the game session of that player. A note should be made that the earliest version of the game (Curiosity 2013) only contained the calculator as an assistance tool.

Chosen metrics for the Analysis Phase

It seems that the number of actions in a level are not potentially indicative for performance on the post-test. What could be of relevance are the strategies that the player uses to solve problems within the game and if these strategies eventually lead to success. Therefore, the metrics that will be considered in the statistical analysis are a quality indicator in the form of overall performance (game percentage) and use of assistance tools (frequencies of the various assistance tools). The progression or stagnation from the pre- to the post-test will not be considered, since this is not a metric that has been measured in the gaming environment. Quantity in the form of frequencies of actions of sub games in the first level will be computed to test if the prediction that its impact on the post-test is low can be confirmed. These are metrics that could be seen as variables for off-line dashboards after a game session has been performed.

The game manual is not considered, since this is a metric that does not give information about the player's proficiency. The fine-grained metrics that have to do with problem solving in the game (effectivity, downtime, solving strategies) could be seen as metrics for on-line analysis that are measured during a gaming session and therefore are more dynamic and difficult to put in statistical frequencies that could be of meaning. Therefore, these metrics were not considered for further statistical analysis, but their potential for teacher dashboard designs will be further described in the Discussion section.

Analysis Phase

To estimate the proportion of variance in the performance in the post-test that can be accounted for by game percentage, frequency of using assistance tools, and the number of actions in each sub game in the first level, a standard multiple regression analysis was performed.

Prior to interpreting the results of the multiple regression analysis, several assumptions were evaluated. First, boxplots indicated that there were some outliers in the data. Still, these numbers are correspondent to the actions that were registered in the log data and will not be deleted for further analysis, since these are representative of the actions of the respective players (Appendix 14). Second, inspection of the normal probability plot of standardized residuals as well as the scatterplot of standardized residuals against standardized predicted values indicated that the assumptions of normality, linearity and homoscedasticity of residuals were met (Appendix 15).

Third, Mahalanobis Distance did exceed the critical x^2 for $df = 7$ (at $\alpha = .001$) of 24.32, indicating that there were multivariate outliers with a large influence on the overall result (Appendix 15). When deleting the outliers with a Mahalanobis Distance above 15 (Field, 2014), there were differences in outcomes, and the maximum Mahalanobis Distance did not exceed the critical x^2 anymore (see Appendix 16). Therefore, the analysis with the deleted Mahalanobis Distance outliers ($n = 168$) is reported, since this could be considered as a more reliable result. Fourth, relatively high tolerances for both predictors in the regression model indicated that multicollinearity would not interfere with the ability to interpret the outcome of the multiple regression analysis (Appendix 15).

In combination, game percentage, frequencies on the assistance tools and frequencies in the sub games of the first level accounted for a significant 25% of the variability in the post-test, $R^2 = .25$, adjusted $R^2 = .21$, F (7, 161) = 7.55, p = <.001. Unstandardized (B) and standardized (β) regression coefficients and squared semi-partial (or "part") correlations (sr^2) for each predictor in the regression model are reported in Table 2.

Table 2

Unstandardized (B) and Standardized (β) Regression Coefficients, and Squared Semi-Partial Correlations (sr^2) for Each Predictor in a Regression Model Predicting a Score on the Post-Test

Variable	*B* [95%, CI]	β	sr^2
Game percentage	0.06 [0.03, 0.09]**	0.34	.29
Calculator	0.01 [-0.05, 0.08]	0.03	.02
Assistance board	-0.01 [-0.06, 0.04]	-0.04	-.04
Handbook	-0.23 [-0.60, 0.13]	-0.09	-.09
Actions refrigerator game	-0.01 [-0.01, 0.00]**	-0.19	-.18
Actions comparison game	-0.02 [-0.12, 0.08]	-0.03	-.03
Actions blender game	-0.00 [-0.02, 0.00]	-0.08	-.07

Note. *N* = 168. CI = confidence interval

**p* <.05, ***p* <.01

Discussion

This research was an explorative study on the use of log data in GBL. Data from five experiments with the mathematics game Zeldenrust were used for exploration and analysis. Through exploration and prior recommendations from the field of GLA, metrics were established for further statistical analysis. A multiple regression analysis between a set of metrics and the score on the post-test was applied to test the possible connections between gaming behaviors and the result on a mathematics test and potential applications for measuring points for a teacher dashboard.

In the Analysis Phase, a multiple regression analysis with the metrics game percentage, frequencies on the assistance tools and number of actions in the sub games in level 1 found a significant impact of the metrics on the variance within the post-test score with a large effect size of 0.33 (Cohen, 1988). Game percentage and the number of actions in the refrigerator game had a role in this significant result, being the only individual predictors with a significant *t*-score. The refrigerator game plays a small part in the overall variance (-0,3%). The negative percentage even indicates that omitting the variable from the regression model would increase the percentage of explained variance and the negative correlation gives an indication that the post-test score increases if the number of actions in the refrigerator game decreases. This is opposite to the expectations that were expressed in the Exploration Phase, in which quantity was not seen as a potentially defining factor in performance.

Game percentage explains 8,4% unique variance out of the entire predictor model (25%). These results are reminiscent of Alonso Fernández et al.'s (2020) findings that game scores could possibly be used as a learning result instead of a post-test. The power of the results is 1.00 (Appendix 17), which means that these results could be confirmed with confidence. Still, the conclusions must be interpreted with caution.

Limitations and suggestions

The frequencies of the metrics could have differed due to different durations of playing the game and interventions within the various experiments. For example, a part-task practice before commencing a gaming task or a surprising event in the game could have had an impact on eventual performance in the game and use of assistance tools. Therefore, the results of the Analysis Phase could give a general indication of the impact of assistance tool usage, actions in the sub games and game percentage on the eventual test but are not fully externally generalizable because of the different versions and durations of the game.

It could be argued that the Exploration Phase of this research consists of many speculative findings that could not be fully supported. When it is interpreted that a player is possibly fed up due to the appearance of a repetition of the Done button in the log data without other actions in between, there is no "real" proof, because it is not possible to see the player in action and ask if they were truly fed up with the game. Another aspect is the exorbitant number of actions in a minute: when is this a result of over-registration or a result of "fooling around" in the game? To truly confirm the causes of the actions, first-hand accounts of the players themselves would be needed.

This is due to the nature of the research: trying to find out performance indicators in log data and connecting these to earlier statistical findings. Most of the Exploration Phase consisted of reading through log data files and learning how to interpret data and "investigate" the mind of a player. The Exploration Phase was structured as much as possible with the preliminary step in Exploratory Data Analysis (DiCerbo et al., 2015). Through this method, outliers could be found and a division between stagnating players (decrease from pre- to post-test) and progressing players (increase from pre- to post-test) could be seen. While the establishment of metrics besides the ones recommended in earlier literature could be seen as a product of individual interpretation of log files, attempts to counterbalance these personal interpretations were done by trying to find multiple cases of an appearance of a potential metric.

The data of this study consisted of several GBL experiments with the same game, but slightly different versions. This was due to the value-added nature of the experiments. The log data could have possibly brought more insight into the effect of the intervention. Especially the Surprise version of the game had a large change, because a character in the game changed the number of bottles in the refrigerator while the player was working on a solution, which changed the entire solving strategy and forced the player to be flexible in mathematical problem solving.

In Wouters et al.'s (2015) article on the experiment, in which they categorized players into three levels of mathematical skills, it is reported that the more computationally proficient players could respond more properly to these changes. It would have been an addition to further support these findings with log data interactions. This was not possible, because the surprising event does not have a log data code and therefore is not measured and registered. For future research with value-added GBL, it would be helpful for further log data analysis if the value-added event (e.g., a surprise) was included in the log data, despite not being an action of the player.

The same suggestion could be applied for an action code that clarifies which specific arithmetical problem is being solved. The log data file contained a comment section in which extra information of an action was given such as the specific computation that was made on a calculator or the amount of juice that was poured into a blender (Appendix 18). By making the specific arithmetical problem visible in the log data, further opportunities for investigation of different problem-solving methods could be generated. This could support more comparison between varying levels of players in how problems are solved. Log data could be used to support statistical data, and due to this caveat in the log data, this was not fully possible.

Applications of log data

Teacher dashboard

One of the application methods of GLA is a teacher dashboard (Alonso-Fernández et al., 2019). This application method was used for the exploration of the data with a division between hypothetical static (frequencies, results) and dynamic dashboards (behavioral patterns). While frequencies can give much information and could be connected statistically, behavioral patterns could give more context to these frequencies. For example, the interaction pattern of players getting quickly through the game by pressing the Done button in succession could be a measure in a dashboard. If the dashboard recognizes these series of actions, it could trigger a message or notification on a teacher dashboard. It seems that these players do not learn from the game and even stagnate in the post-test, despite a moderate score in the pre-test. If the teacher can identify these players, it could be a reason for intervention in the classroom.

While some metrics (e.g., downtime) that were established were not statistically computable because they still are abstract, these could still be potential indicators for dashboard development. The amount of downtime and the following result (incorrect or correct answer) could be connected in a dashboard. The same goes for the application of an assistance tool and the effect of this tool. This could be helpful for teachers to separate players who are not performing actions because they are thinking about a solution or players who are struggling to find the next step, in combination with the information they already have of players because of teaching experience. This is reminiscent of DiCerbo et al.'s (2015) findings on separating players who are fooling around and players who are struggling. The interpretation of different scenarios in solving a problem after an incorrect answer already gives an indication that this data could be of relevance.

Stealth assessment

A different application of GLA is to establish an adaptive learning system. As shown in stealth assessment (Shute & Kim, 2014), the game can adapt to the player's behavior. In the Exploration Phase, different scenarios after having an incorrect answer were established. Some players go on quickly, while others seem to take some time to think about how to solve the problem by using assistance tools. Other players initially do not take any actions and possibly think about the solution outside the game environment and eventually solve the problem in a short amount of time. If it is desirable to develop a more elaborate thought process in mathematical problem solving, the game could force players who want to get quickly through the game with incorrect answers to think longer about their answers. After having an incorrect answer, they cannot go on with the game immediately, but they are instructed to take some time to think about the solution or use assistance tools. After a while, the player can continue with solving the problem in the game.

Another aspect of stealth assessment (Shute & Kim, 2014) could be to conduct the post-test in the gaming environment as a form of summative assessment. From the multiple regression analysis, it shows that game percentage could be indicative of the score of the post-test. Since the player is already accustomed to the gaming environment, the questions in the post-test could be represented in a set of tasks, but the player cannot use the assistance tools anymore in this phase of the game.

Future directions

As stated by Hawlitschek & Köppen (2014), log data could be a method to investigate the possible reasons why a certain score on a post-test has been achieved, instead of hypothesizing what could have happened based on one score. For example, it is difficult to hypothesize based on scores what happened in the sessions of moderate players in the pre-test who stagnated remarkably. The log data gives more insight in what the causes could be, which do not seem to be related to the skill level of mathematics. The set of stagnating players in the post-test seemingly are not bad at mathematics but seem to get fed up with the experiment or the game and this is potentially a reason why they stagnate in the post-test. Log data also is a method to investigate how learning took place in a game (Nguyen et al., 2020). When looking at players who significantly progressed on the post-test, they seem to use assistance tools more extensively and take their time to think of a new solution when they give an incorrect answer.

The established metrics could be seen as "potential". This means that more support is needed to confirm that these metrics could be relevant indicators in Zeldenrust or GBL in general. A future direction could be to perform an experiment with the control version of Zeldenrust, since this version has no interventions that could have an impact on the course of the game. Then, the metrics could become internally valid due to having the same intervention and duration of the game session. This experiment would not be focused on the learning value of GBL, but the added value of log data for information about learners. Instead of focusing on statistical results, the analysis of the experiment should emphasize more on log data exploration, with statistics possibly confirming the findings in the log data.

Conclusion

The design of log data structures should be an integral part to the general design of educational games. By specifying what must be measured, a more fine-grained picture of a player's game performance can be generated. In the log data of this explorative study, elements such as the description of the specific task and interventions such as a surprise were not fully covered in the log data. Every action must be registered as detailed as possible, since the absence of earlier mentioned elements resulted in a more limited exploration. Log data can be of value in

GBL research, giving more clarity to statistical results and the ability for researchers to be more detailed in their confirmation or refutation of earlier hypotheses. While the exploration of log data can be an intensive process, using a method that is reminiscent of coding interviews, much context behind the statistical findings can be generated and result in a stronger conclusion of a GBL research. This can eventually result in the establishment of metrics, which can potentially be transferred to measuring points in a teacher dashboard. Then, log data can become a valuable formative assessment tool for teachers in the classroom and potentially support further learning through GBL.

References

Allen, P., Bennett, K., & Heritage, B. (2014). *SPSS Statistics Version 22: A Practical Guide.* Cengage.

Alonso-Fernández, C., Martínez-Ortiz, I., Caballero, R., Freire, M., & Fernández-Manjón, B. (2020). Predicting students' knowledge after playing a serious game based on learning analytics data: A case study. *Journal of Computer Assisted Learning*, *36*(3), 350-358. https://doi.org/10.1111/jcal.12405

Alonso-Fernández, C., Calvo-Morata, A., Freire, M., Martínez-Ortiz, I., & Fernández-Manjón, B. (2019). Applications of data science to game learning analytics data: A systematic literature review. *Computers & Education*, *141*, 1-14. https://doi.org/10.1016/j.compedu.2019.103612

Boyle, E. A., Hainey, T., Connolly, T. M., Gray, G., Earp, J., Ott, M., Lim, T., Ninaus, M., Ribeiro, C. & Pereira, J. (2016). An update to the systematic literature review of empirical evidence of the impacts and outcomes of computer games and serious games. *Computers & Education, 94*, 178-192. https://doi.org/10.1016/j.compedu.2015.11.003

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences (2nd edition).* Erlbaum.

De Vos, T. (1992). *Tempo-Test-Rekenen: test voor het vaststellen van het rekenvaardigheidsniveau der elementaire bewerkingen (automatisering) voor het basis- en voortgezet onderwijs: handleiding.* Berkhout.

DiCerbo, K. E., Bertling, M., Stephenson, S., Jia, Y., Mislevy, R. J., Bauer, M., & Jackson, G. T. (2015). An application of exploratory data analysis in the development of game-based assessments. In C. S. Loh, Y. Sheng, D. Ifenthaler (Eds.), *Serious games analytics* (pp. 319-342). Springer.

Field, A. (2018). *Discovering Statistics Using IBM SPSS Statistics (5th Edition).* SAGE.

Freire, M., Serrano-Laguna, Á., Manero, B., Martínez-Ortiz, I., Moreno-Ger, P., & Fernández-Manjón, B. (2016). Game learning analytics: learning analytics for serious games. In M. J. Spector, B. B. Locke, & M. D. Childress (Eds.), *Learning, design, and technology* (pp. 1-29). Springer.

Frost, M. & Lynch, D. (Executive Producers). (1990-1991). *Twin Peaks* [TV series]. ABC.

Greiff, S., Wüstenberg, S., & Avvisati, F. (2015). Computer-generated log-file analyses as a window into students' minds? A showcase study based on the PISA 2012 assessment of problem solving. *Computers & Education*, *91*, 92-105. https://doi.org/10.1016/j.compedu.2015.10.018

Greiff, S., Wüstenberg, S., Csapó, B., Demetriou, A., Hautamäki, J., Graesser, A. C., & Martin, R. (2014). Domain-general problem solving skills and education in the 21st century. *Educational Research Review*, *13*, 74-83. https://doi.org/10.1016/j.edurev.2014.10.002

Hawlitschek, A., & Köppen, V. (2014, October). Analyzing player behavior in digital game-based learning: Advantages and challenges. In *8th European Conference on Games Based Learning: ECGBL2014* (pp. 199-206).

Hwang, W. Y., Shadiev, R., Wang, C. Y., & Huang, Z. H. (2012). A pilot study of cooperative programming learning behavior and its relationship with students' learning performance.

Computers & education, *58*(4), 1267-1281. https://doi.org/10.1016/j.compedu.2011.12.009

Ifenthaler, D., Eseryel, D., & Ge, X. (2012). Assessment for game-based learning. In *Assessment in game-based learning* (pp. 1-8). Springer.

Kim, Y. J., & Ifenthaler, D. (2019). Game-based assessment: The past ten years and moving forward. In D. Ifenthaler, & Y. J. Kim (Eds.), *Game-Based Assessment Revisited* (pp. 3-11). Springer.

Liu, M., Lee, J., Kang, J., & Liu, S. (2016). What we can learn from the data: A multiple-case study examining behavior patterns by students with different characteristics in using a serious game. *Technology, Knowledge and Learning*, *21*(1), 33-57. https://doi.org/10.1007/s10758-015-9263-7

Loh, C. S., Sheng, Y., & Ifenthaler, D. (2015). Serious games analytics: Theoretical framework. In *Serious games analytics* (pp. 3-29). Springer.

Mayer, R. E. (2020). Cognitive Foundations of Game-Based Learning. In J. L. Plass, R. E. Mayer, & B. D. Homer (Eds.), *Handbook of Game-Based Learning* (pp. 83-110). MIT Press.

Mayer, R. E. (2019). Thirty years of research on online learning. *Applied Cognitive Psychology*, *33*(2), 152-159. https://doi.org/10.1002/acp.3482

Mayer, R. E. (2014). *Computer games for learning: An evidence-based approach*. MIT Press.

Mayer, R. E. (2011). Multimedia learning and games. In S. Tobias, & J. D. Fletcher (Eds.), *Computer Games and Instruction* (pp. 281–305). Information Age Publishing.

Nguyen, H. A., Hou, X., Stamper, J., & McLaren, B. M. (2020). Moving beyond Test Scores: Analyzing the Effectiveness of a Digital Learning Game through Learning Analytics. *International Educational Data Mining Society*.

Oliveri, M. E., Lawless, R., & Mislevy, R. J. (2019). Using evidence-centered design to support the development of culturally and linguistically sensitive collaborative problem-solving assessments. *International Journal of Testing*, *19*(3), 270-300. https://doi.org/10.1080/15305058.2018.1543308

Perez-Colado, I., Alonso-Fernandez, C., Freire, M., Martinez-Ortiz, I., & Fernandez-Manjon, B. (2018, April). Game learning analytics is not informagic!. In *2018 IEEE Global Engineering Education Conference (EDUCON)* (pp. 1729-1737). IEEE.

Plass, J. L., Mayer, R. E., Homer, B. D., & Kinzer, C. K. (2020). Theoretical Foundations of Game-Based and Playful Learning. In J. L. Plass, R. E. Mayer, & B. D. Homer (Eds.), *Handbook of Game-Based Learning* (pp. 3-24). MIT Press.

Ritterfeld, U., & Weber, R. (2006). Video games for entertainment and education. In P. Vorderer, & J. Bryant (Eds.), *Playing video games: Motives, responses, and consequences* (pp. 399-413). Taylor & Francis.

Seif El-Nasr, M., Drachen, A., & Canossa, A. (2013). *Game analytics–maximizing the value of player data*. Springer.

Shute, V. J., & Kim, Y. J. (2014). Formative and stealth assessment. In J. M. Spector, M. D. Merrill, J. Elen, & M. J. Bishop (Eds.), *Handbook of research on educational communications and technology* (pp. 311-321). Springer.

Siemens, G., & Long, P. (2011). Penetrating the fog: Analytics in learning and education. *EDUCAUSE review*, *46*(5), 30-40.

Ter Vrugte, J., de Jong, T., Wouters, P., Vandercruysse, S., Elen, J., & van Oostendorp, H. (2015). When a game supports prevocational math education but integrated reflection does not. *Journal of Computer Assisted Learning*, *31*(5), 462-480. https://doi.org/10.1111/jcal.12104

Vandercruysse, S., ter Vrugte, J., de Jong, T., Wouters, P., van Oostendorp, H., Verschaffel, L., Van Dooren, W., & Elen, J. (2015). "Zeldenrust": A mathematical game-based learning environment for prevocational students. In J. Torbeyns, E. Lehtinen, & J. Elen (Eds.), *Describing and studying domain-specific serious games* (pp. 63-81). Springer.

Van Leeuwen, M. (2018). *Innovatie in het onderwijs: Een vergelijkend onderzoek naar game-based, traditioneel en begeleid zelfontdekkend wiskunde leren* [Master's thesis, Utrecht University].

Westera, W., Nadolski, R., & Hummel, H. (2013, October). Learning analytics in serious gaming: uncovering the hidden treasury of game log files. In A. De Gloria (Eds.), *International Conference on Games and Learning Alliance* (pp. 41-52). Springer.

Wouters, P., & Van Oostendorp, H. (2017). Overview of instructional techniques to facilitate learning and motivation of serious games. In *Instructional techniques to facilitate learning and motivation of serious games* (pp. 1-16). Springer.

Wouters, P., van Oostendorp, H., Ter Vrugte, J., de Jong, T., & Elen, J. (2015). The role of surprise in game-based learning for mathematics. In *International Conference on Games and Learning Alliance* (pp. 401-410). Springer.

Wouters, P., & Van Oostendorp, H. (2013). A meta-analytic review of the role of instructional support in game-based learning. *Computers & Education*, *60*(1), 412-425. https://doi.org/10.1016/j.compedu.2012.07.018

Appendices

Appendix 1: Description of existing data sets

All data is anonymized, and participants were between 12 and 16 years old.

Existing master thesis research:

Name	***Participants***	***Version of game***	***Content of existing data***
Merel van Leeuwen (2018)	39 (game condition)	Part-task practice/worked example	Log data, basic participant information (participant number, condition, group, gender, age, level of education), Family Affluence Scale III-questionnaire, Tempo Test, pre-test, post-test, motivation rate, retention rate

Earlier research by Pieter Wouters et al. with the mathematics game Zeldenrust:

Study	***Version of game***	***Participants***	***Content of existing data***
Curiosity in a game (2014)	***Curiosity*** *It is not completely clear from the get-go how the problem can be solved.*	*25*	*Log data, basic participant information (participant number, age, game code, condition, level of education, gender), Tempo Test score, log data, score on pre-test and post-test*
Surprise in a game (2015)	***Surprise*** *An unexpected event occurs. A character in the game changes*	*113*	*Log data, basic participant information (participant number, game code, condition, level of education, gender), Tempo Test*

	the characteristics of the problem.		*score, score on pre-test and post-test*
Part-task/worked example in a game (2015)	*Part-task/worked example* ***Part-task:*** *stand-alone exercises to practice skills that are used in the game* ***Worked example:*** *expert explanation of how a mathematical problem must be solved*	*71*	*Log data, basic participant information (participant number, game code, condition, level of education, gender), Tempo Test score, score on pre-test and post-test*
Curiosity in a game – pilot (2013)	***Curiosity*** *See description of Curiosity in a Game (2014)*	*83*	*Log data, basic participant information (participant number, level of education, condition, gender), log data, score on pre-test and post-test*

Appendix 2: Detailed information on the different measured variables in the log data of the Zeldenrust-game

The log data contains the user-id, timestamp, log-id (code for each action, chronologically ordered).

- Game: 1 = refrigerator game, 2 = comparison game, 3 = blender game
- Actie: see table below
- Correct: 0 = incorrect, 1 = correct
- Level and Assignment are levels (1 t/m 4)

Action code	Description
1	Picking up bottle/box/pitcher
2	Letting go of bottle/box/pitcher
3	Beginning of assignment
4	Player presses Done button/ closes refrigerator door/ Blender button/ cup on serving plate

5	Tutorial
6	Bottle in refrigerator
7	Bottle out of refrigerator
10	Bottle/box is poured into blender
11	Calculator
18	Game manual
20	Clicking on unknown box in curiosity version
21	Clicking on a bottle (refrigerator)
22	Assistance board
23	Handbook
37	Results part-task practice

Appendix 3: Player with an extreme number of actions and screen shot of seconds with many actions.

GV	GW	GX
5010	5011	5012
74	30	47
791	81	24
246		2
222	34	1
112	38	36
48	48	96
49	27	78
66	6	74
71	3	45
44	24	70
46	20	18
29	44	24
38	91	17
41	70	55
54	16	21
	50	
8	44	
	409	1
	2172	8
	46336	37
	26366	4
	20	1
	27	13
	13	9
	21	17

B	C	D	E	F	G
5011	surseq	1	2	1	1126
5011	surseq	1	2	2	1126
5011	surseq	1	2	1	1126
5011	surseq	1	2	2	1126
5011	surseq	1	2	1	1126
5011	surseq	1	2	2	1126
5011	surseq	1	2	1	1126
5011	surseq	1	2	2	1126
5011	surseq	1	2	1	1127
5011	surseq	1	2	2	1127
5011	surseq	1	2	1	1127
5011	surseq	1	2	1	1127
5011	surseq	1	2	2	1127
5011	surseq	1	2	1	1127
5011	surseq	1	2	1	1127
5011	surseq	1	2	2	1127
5011	surseq	1	2	1	1127
5011	surseq	1	2	1	1127
5011	surseq	1	2	2	1127
5011	surseq	1	2	1	1127
5011	surseq	1	2	1	1127
5011	surseq	1	2	2	1127
5011	surseq	1	2	1	1127
5011	surseq	1	2	2	1128
5011	surseq	1	2	1	1128
5011	surseq	1	2	2	1128
5011	surseq	1	2	1	1128
5011	surseq	1	2	6	1128
5011	surseq	1	2	7	1128
5011	surseq	1	2	2	1128
5011	surseq	1	2	1	1128
5011	surseq	1	2	2	1128
5011	surseq	1	2	1	1128
5011	surseq	1	2	6	1128
5011	surseq	1	2	7	1128
5011	surseq	1	2	2	1128
5011	surseq	1	2	1	1128
5011	surseq	1	2	2	1128
5011	surseq	1	2	1	1128
5011	surseq	1	2	6	1128
5011	surseq	1	2	7	1128
5011	surseq	1	2	2	1128

Appendix 4: Characteristics of the different types of games (games, levels, assignments, attempts)

L = levels

G = game

G1 = refrigerator game

G2 = comparison game

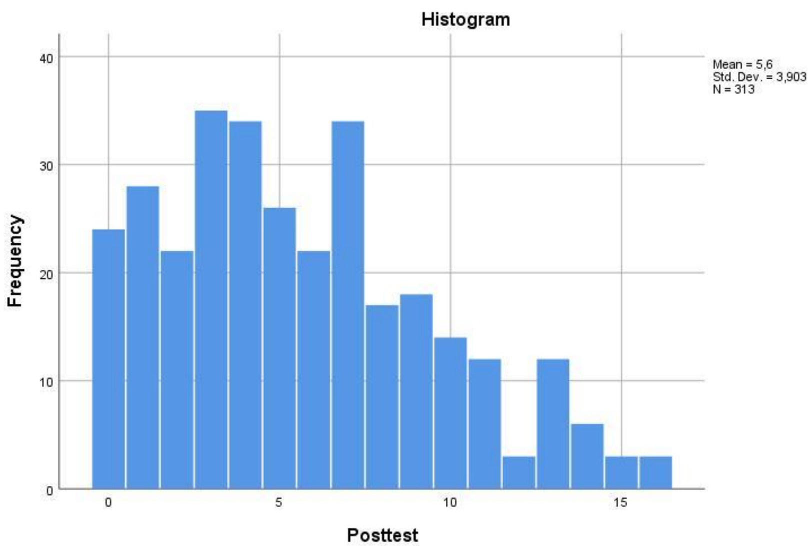
G3 = blender game

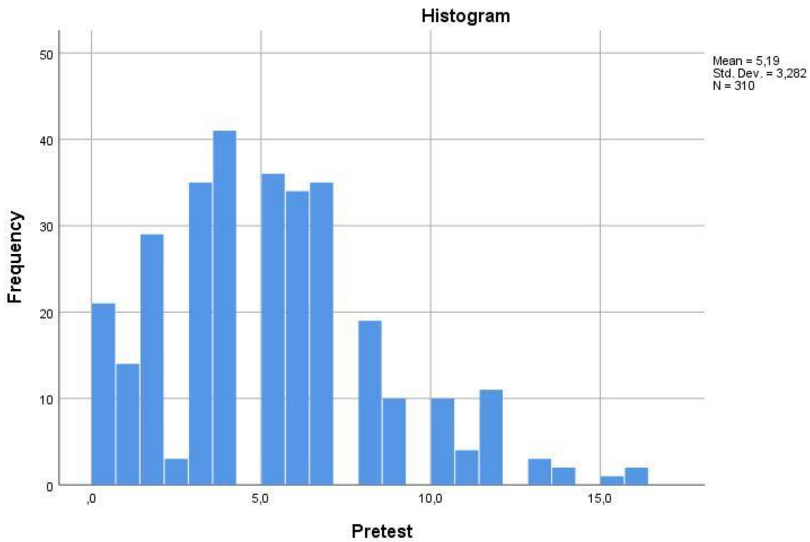
A = assignments

Att = attempts

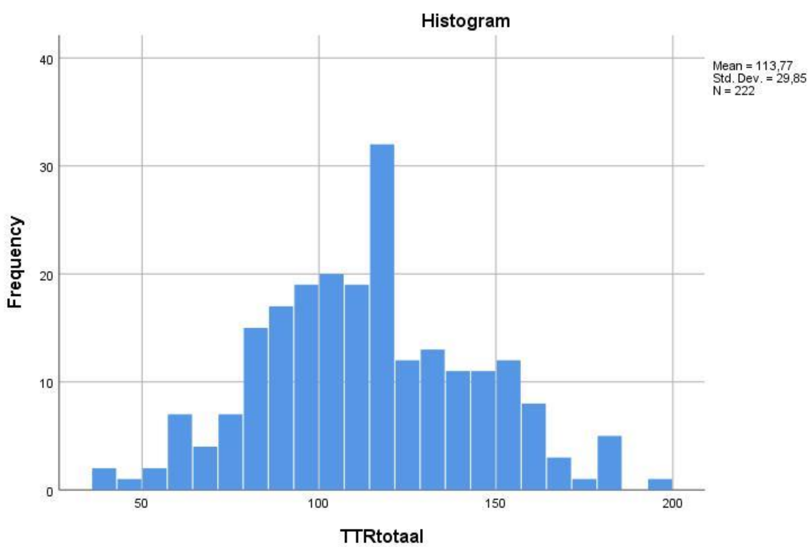
Version	Number	Games	LG1	AG1	LG2	AG2	LG3	AG3	Att
Master Thesis	1000	3	4	2	4	4	4	2	2
Curiosity 2013	2000	3	4	4	4	4	4	4	3
Curiosity 2014	3000	3	4	4	4	4	4	4	3
Part-task/Wor-ked Example 2015	4000	3	4	4	4	4	4	4	3
Surprise 2015	5000	3	3	6	3	4	3	6	3

Appendix 5: Histograms and boxplots data set (pre-test, post-test, TTR, difference score/Voortgang Totaal)

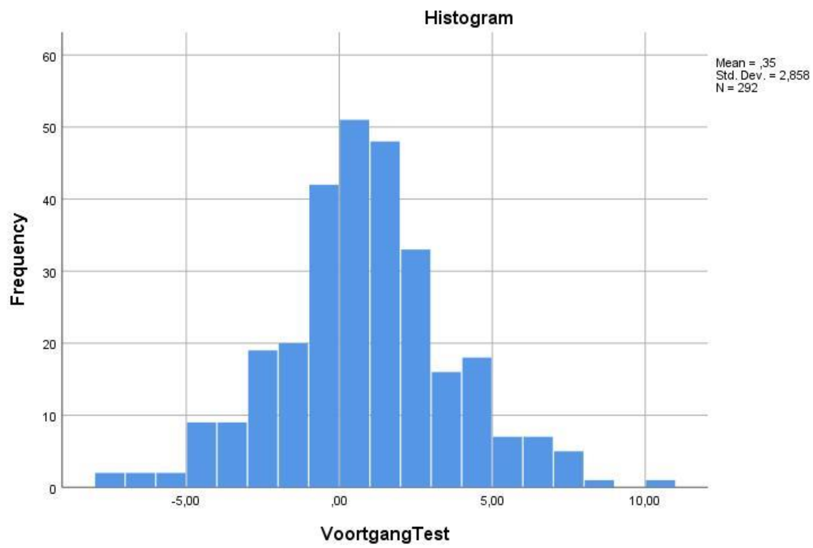


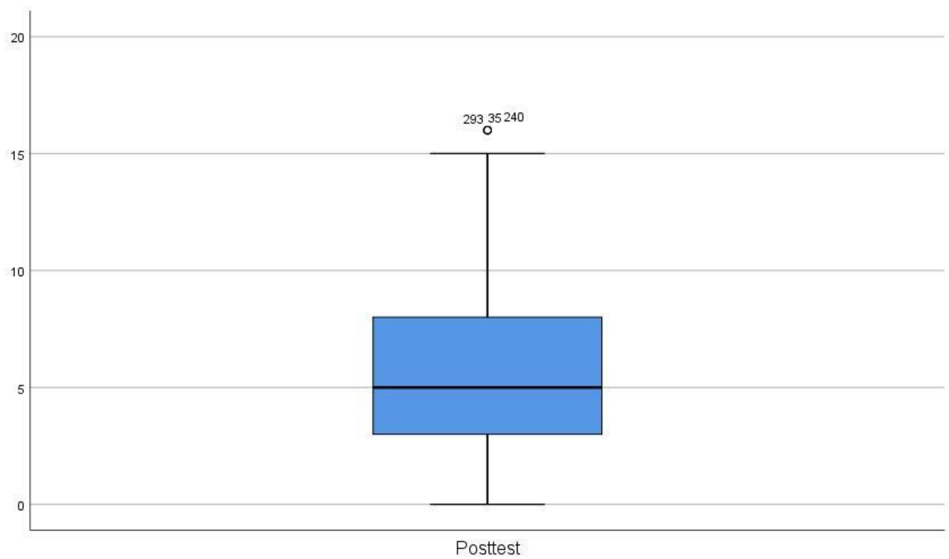


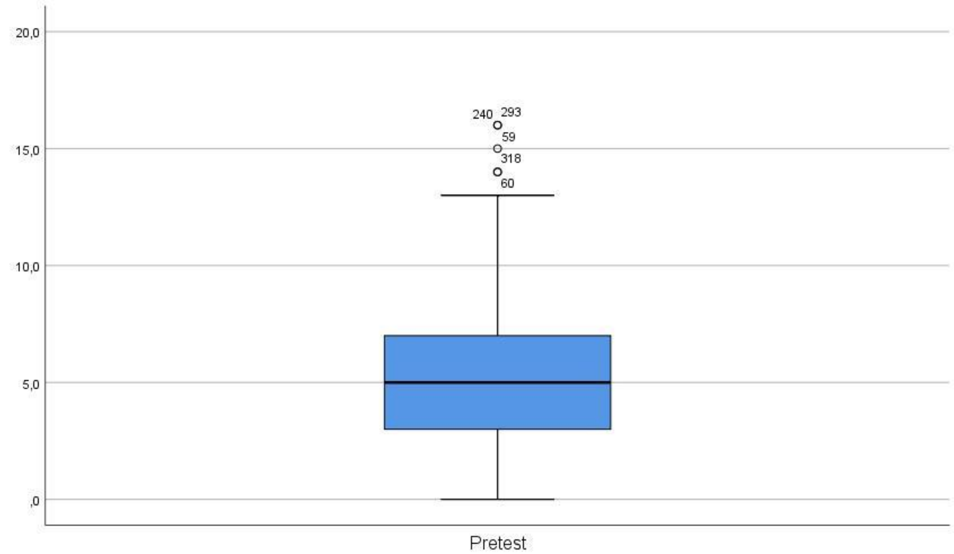
Histogram
Mean = 5,19
Std. Dev. = 3,282
N = 310
Frequency
Pretest

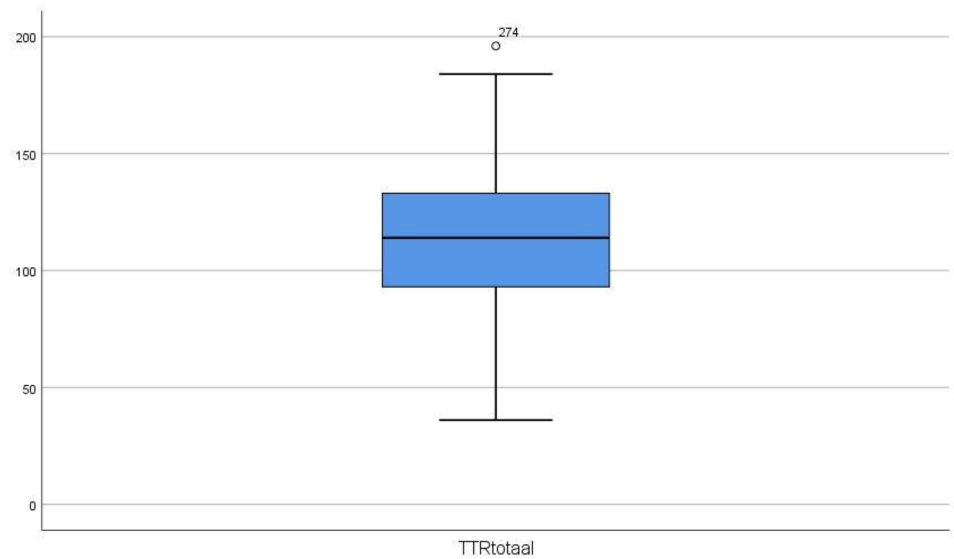


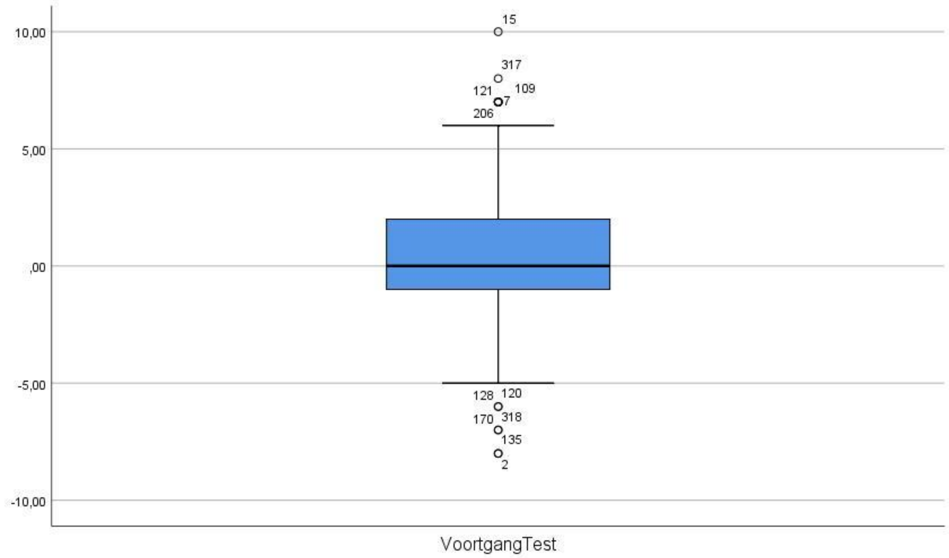
Histogram
Mean = 113,77
Std. Dev. = 29,85
N = 222
Frequency
TTRtotaal


Histogram
Mean = ,35
Std. Dev. = 2,858
N = 292
Frequency
60
50
40
30
20
10
0
-5,00
,00
5,00
10,00
VoortgangTest


20
15
10
5
0
293 35 240
Posttest


20,0
15,0
10,0
5,0
,0
240
293
59
318
60
Pretest


200
150
100
50
0
274
TTRtotaal



Information of stagnating and progressing players:

Stagnation:

Number boxplot	Number participant	Pre-test	Post-test	TTR	Stagnation
2	1002	12	4	162	-8
120	2077	6	0		-6
128	3002	8	2	126	-6
135	3009	8	0	49	-8
170	4019	8	1	138	-7
318	5097	14	7	131	-7

Progression:

Number boxplot	Number participant	Pre-test	Post-test	TTR	Progression
7	1007	4	11	64	+7
15	1015	3	13	81	+10
109	2066	6	13		+7
121	2078	7	14		+7
206	4055	3	10	90	+7
317	5096	7	15	81	+8

Appendix 6: example of log data exploration

Log data notes were written in Dutch on paper. Below, you will see a fragment of these notes.

1102 296-319 → rekenmachine
antwoord is juist

Game 3: level 1 → struggles 509
veel hulpkolommen
rekenmachine
weer game manual
snapt spelregels wel?
scoort redelijk op Pre en Post, goede TTR-score

- vaak is korte raadpleging van rekenmachine om som op te lossen
- langdurigere kan problemen betekenen

LANGE downtime → 1964-2372
hulpmiddelen hebben grotendeels geen nut

2663 10-10 84% op game

90-94 rekenmachine → antwoord is juist
363 - rekenmachine → antwoord is juist

2x rekenmachine → uiteindelijk antwoord juist

822 - rekenmachine → antwoord is juist

Dezelfde hypothese als hier
- Rekenmachines gebruikt voor oplossingen

andere hulpmiddelen kunnen indicator voor groter probleem zijn

Log data of player 1102 (game 3, level 1):

Participant	***Pre-test***	***Post-test***	***TTR***	***Progression***
1102	***8***	***7***	***129***	***-1***

Interpretation of log data:

577 – player starts using assistance board

610 – after repeated action with an assistance board, the player opens the game manual -> does the player understand the game?

632 – player starts solving the problem

675-735 – player performs multiple actions on a calculator -> struggling with calculation to solve the problem?

782-785 – two incorrect answers in succession

1102	pws	3	1	22	577	04/23/201	1	0	
1102	pws	3	1	22	578	04/23/201	1	0	
1102	pws	3	1	22	579	04/23/201	1	0	
1102	pws	3	1	34	580	04/23/201	1	0	
1102	pws	3	1	22	581	04/23/201	1	0	
1102	pws	3	1	22	583	04/23/201	1	0	
1102	pws	3	1	34	585	04/23/201	1	0	
1102	pws	3	1	22	589	04/23/201	1	0	
1102	pws	3	1	18	610	04/23/201	1	0	
1102	pws	3	1	1	632	04/23/201	1	0	juice,1unitsperpour
1102	pws	3	1	2	632	04/23/201	1	0	juice,1unitsperpour
1102	pws	3	1	1	632	04/23/201	1	0	juice,1unitsperpour
1102	pws	3	1	2	635	04/23/201	1	0	juice,1unitsperpour
1102	pws	3	1	1	636	04/23/201	1	0	juice,1unitsperpour
1102	pws	3	1	2	637	04/23/201	1	0	juice,1unitsperpour
1102	pws	3	1	1	658	04/23/201	1	0	juice,1unitsperpour
1102	pws	3	1	10	658	04/23/201	1	0	juice,1unitsperpour
1102	pws	3	1	2	659	04/23/201	1	0	juice,1unitsperpour
1102	pws	3	1	1	664	04/23/201	1	0	juice,1unitsperpour
1102	pws	3	1	2	666	04/23/201	1	0	juice,1unitsperpour
1102	pws	3	1	11	675	04/23/201	1	0	60/40 = 1.5
1102	pws	3	1	11	682	04/23/201	1	0	60* = 0
1102	pws	3	1	11	685	04/23/201	1	0	0*1.5 = 0
1102	pws	3	1	11	693	04/23/201	1	0	60* = 0
1102	pws	3	1	11	697	04/23/201	1	0	0*1.5 = 0
1102	pws	3	1	11	702	04/23/201	1	0	60*1.5 = 90
1102	pws	3	1	22	719	04/23/201	1	0	
1102	pws	3	1	18	726	04/23/201	1	0	
1102	pws	3	1	23	730	04/23/201	1	0	
1102	pws	3	1	22	735	04/23/201	1	0	
1102	pws	3	1	23	735	04/23/201	1	0	
1102	pws	3	1	1	780	04/23/201	1	0	juice,1unitsperpour
1102	pws	3	1	2	781	04/23/201	1	0	juice,1unitsperpour
1102	pws	3	1	4	782	04/23/201	1	0	
1102	pws	3	1	4	785	04/23/201	1	0	

Assistance tool did not seem to have an effect.

Log data of player 2063 (Game 3, Level 1)

Participant	***Pre-test***	***Post-test***	***TTR***	***Progression***
2063	***10***	***10***		***0***

Interpretation of the log data:

517-530: player performs two actions on calculator

530-556: thinking about how to solve the problem?

556-558: tries to solve the problem

567: incorrect answer

567-612: downtime -> thinking about a new strategy to solve the problem?

612-623: two actions on the calculator

625 -> players starts solving problem

642 -> correct answer

2063	curiosity	3	1	11	517	06/24/201	4	0	13/6 = 2.1666666666666665
2063	curiosity	3	1	11	530	06/24/201	4	0	2.1666666666666665*12 = 26
2063	curiosity	3	1	1	556	06/24/201	4	0	yoghurt,1unitsperpour
2063	curiosity	3	1	10	557	06/24/201	4	0	yoghurt,1unitsperpour
2063	curiosity	3	1	10	557	06/24/201	4	0	yoghurt,1unitsperpour
2063	curiosity	3	1	2	558	06/24/201	4	0	yoghurt,1unitsperpour
2063	curiosity	3	1	4	567	06/24/201	4	0	
2063	curiosity	3	1	11	612	06/24/201	4	0	2*6 = 12
2063	curiosity	3	1	11	623	06/24/201	4	0	2* = 0
2063	curiosity	3	1	20	625	06/24/201	4	0	juice,1,2,3pack
2063	curiosity	3	1	20	626	06/24/201	4	0	yoghurt,1,2,3pack
2063	curiosity	3	1	1	629	06/24/201	4	0	yoghurt,1unitsperpour
2063	curiosity	3	1	10	630	06/24/201	4	0	yoghurt,1unitsperpour
2063	curiosity	3	1	2	630	06/24/201	4	0	yoghurt,1unitsperpour
2063	curiosity	3	1	1	632	06/24/201	4	0	juice,1unitsperpour
2063	curiosity	3	1	10	632	06/24/201	4	0	juice,1unitsperpour
2063	curiosity	3	1	2	632	06/24/201	4	0	juice,1unitsperpour
2063	curiosity	3	1	4	642	06/24/201	4	1	

Appendix 7: example of stagnating player

Participant	**Pre-test**	**Post-test**	**TTR**	**Stagnation**
4065	**5**	**0**	**102**	**-5**

Pressing thrice on the Done button (Action number 4) without any action in between. This example consists of an entire level in which the player performs these series of actions.

Assignment 1: 2301-2303-2305

Assignment 2: 2308-2310-2312

Assignment 3: 2315-2317-2319

Assignment 4: 2321-2324-2326

4065	controle	3	4	3	2298	03/19/201	1	0
4065	controle	3	4	4	2301	03/19/201	1	0
4065	controle	3	4	4	2303	03/19/201	1	0
4065	controle	3	4	4	2305	03/19/201	1	0
4065	controle	3	4	3	2307	03/19/201	2	0
4065	controle	3	4	4	2308	03/19/201	2	0
4065	controle	3	4	4	2310	03/19/201	2	0
4065	controle	3	4	4	2312	03/19/201	2	0
4065	controle	3	4	3	2314	03/19/201	3	0
4065	controle	3	4	4	2315	03/19/201	3	0
4065	controle	3	4	4	2317	03/19/201	3	0
4065	controle	3	4	4	2319	03/19/201	3	0
4065	controle	3	4	3	2320	03/19/201	4	0
4065	controle	3	4	4	2321	03/19/201	4	0
4065	controle	3	4	4	2324	03/19/201	4	0
4065	controle	3	4	4	2326	03/19/201	4	0

Appendix 8: Table of selection of players from entire data set

Number	**Pre**	**Post (and TTR)**	**Correct**	**Incorrect**	**Percentage**	**Assistance tools Frequency**
1009	**6**	**8 (TTR 94)**	**15**	**14**	**51,72**	**Calculator: 13** **Assistance board: 43** **Handbook: 2**
1010	**10**	**8 (TTR 123)**	**17**	**15**	**53,13**	**Calculator: 5** **Game manual: 1** **Assistance board: 4** **Handbook: 1**
1012	**7**	**7 (TTR 114)**	**4**	**8**	**33,33**	**Calculator: 5** **Assistance board: 69** **Handbook: 2**
1013	**6**	**7 (TTR 109)**	**4**	**21**	**16**	**Calculator: 2**

						Game manual: 1 Assistance board: 4 Handbook: 4
1019	7	7 (TTR 111)	11	20	35,48	Calculator: 3 Assistance board: 11 Handbook: 1
1021	6	9 (TTR 120)	7	15	31,82	Calculator: 8 Assistance board: 2 Handbook: 1
1024	7	8 (TTR 151)	9	28	24,32	Game manual: 1 Assistance board: 32 Handbook: 3
1093	9	9 (TTR 157)	14	9	60,87	Assistance board: 21 Handbook: 2
1099	7	7 (TTR 133)	16	13	55,17	Assistance board: 62 Handbook: 4
1102	8	7 (TTR 129)	9	27	25	Calculator: 39 Game manual: 6 Assistance board: 29 Handbook: 5
1104	9	7 (TTR 119)	19	18	51,35	Calculator: 5 Assistance board: 1 Handbook: 2
5053	12	13 (TTR 196)	23	4	85,19	None
2016	15	14	29	8	78,38	None
2017	14	15	23	8	74,19	Calculator: 7
5018	16	16 (TTR 117)	17	9	65,38	None
1100	10	16 (TTR 77)	20	4	83,33	Game manual: 1 Assistance board: 3 Handbook: 1
2021	12	14	23	7	76,67	None
2025	11	15	32	12	77,73	Calculator: 3
5038	13	13 (TTR 180)	20	3	86,97	None
5072	16	16 (TTR 140)	25	4	86,21	Calculator: 3 Assistance board: 9
1094	10	11 (TTR 114)	14	6	70	Calculator: 2 Assistance board: 11 Handbook: 9
2063	10	10	21	4	84	Calculator: 20
2069	11	14	28	5	84,8	Calculator: 25
1106	10	11 (TTR 127)	12	18	40	Calculator: 3 Assistance board: 6 Handbook: 5
2019	5	7	19	20	48,72	None
2074	10	13	37	11	77,08	Calculator: 10
5013	11	14 (TTR 160)	28	13	68,29	Assistance board: 26 Handbook: 2
5031	10	14 (TTR 167)	28	18	60,87	None

5039	12	13 (TTR 148)				
5040	12	12 (TTR 140)	16	10	61,54	Calculator: 1 Assistance board: 2
5056	13	11 (TTR 169)	28	13	68,29	Calculator: 9 Assistance board: 2
5060	12	11 (TTR 165)				
1002	12	4 (TTR 162)	22	32	40,74	Calculator: 8 Game manual: 1 Assistance board: 1 Handbook: 1
1008	6	1 (TTR 141)	8	22	26,67	Calculator: 13 Game manual: 1 Assistance board: 7 Handbook: 6
3009	8	0 (TTR 49)	9	14	39,13	Calculator: 6 Assistance board: 9
4024	5	0 (TTR 59)	13	21	38,24	Calculator: 11 Assistance board: 22 Handbook: 2
4065	5	0 (TTR 102)	26	56	31,71	Calculator: 64 Assistance board: 3 Handbook: 3
4009	7	2 (TTR 90)	9	60	13,04	Calculator: 6 Assistance board: 2
3002	8	2 (TTR 126)	11	15	42,31	None
2023	8	3	13	39	25	Calculator: 3
1006	0	7 (TTR 115)	7	12	36,84	Calculator: 7 Assistance board: 9 Handbook: 4
1007	4	11 (TTR 64)	11	2	84,62	Calculator: 24 Assistance board: 4 Handbook: 1
1015	3	13 (TTR 81)	18	11	62,07	Calculator: 12 Assistance board: 4 Handbook: 2
1100	10	16 (TTR 77)	20	4	83,33	Game manual: 1 Assistance board: 3 Handbook: 2
2055	0	6	4	3	57,14	Calculator: 1
2066	6	13	32	5	86,49	Calculator: 4
2067	5	11	24	13	64,86	Calculator: 9
2071	4	10	20	17	54,05	Calculator: 32
2076	7	13	30	7	81,08	Calculator: 5
5023	7	13 (TTR 82)	11	2	84,62	Calculator: 5 Assistance board: 7
5028	4	10 (TTR 135)	26	3	89.66	Calculator: 4 Assistance board: 3
5096	7	15 (TTR 81)	12	21	36,37	Assistance board: 1 Handbook: 1

4048	**3**	**3 (TTR 127)**				
4040	**1**	**0 (TTR 61)**	**9**	**21**	**30**	**Calculator: 10** **Assistance board: 9** **Handbook: 2**
3017	**4**	**2 (TTR 95)**	**9**	**18**	**33,33**	**Calculator: 8** **Assistance board: 3**
5034	**5**	**2 (TTR 116)**				
1011	**5**	**3 (TTR 132)**	**12**	**11**	**52,17**	**Calculator: 4** **Game manual: 1** **Assistance board: 46** **Handbook: 1**
5002	**2**	**2 (TTR 87)**	**9**	**21**	**30**	**Calculator: 1** **Assistance board: 8**
3019	**3**	**5 (TTR 109)**	**4**	**19**	**17,39**	**Calculator: 23** **Assistance board: 2**
4046	**1**	**3 (TTR 117)**	**25**	**21**	**54,35**	**Assistance board: 19** **Handbook: 2**
1022	**4**	**4 (TTR 126)**	**8**	**5**	**61,54**	**Assistance board: 20** **Handbook: 2**

Example of a proficient player: 5053

Participant	**Pre-test**	**Post-test**	**TTR**	**Progression**
5053	**12**	**13**	**196**	**+1**

Interpretation of log data:

Players commences level (Game 3, Level 2).

Interpretation of actions per assignment:

- *Player sees the assignment (example: 1206 – assignment 2 of Game 3, Level 2).*
- *Player takes a while to think about the steps towards the solution (see time stamps 1206-1232, 1248-1322, 1329-1338), solves the problem and gives the correct answer. This version contained all assistance tools, but there were no assistance tools needed.*

5053	nosurran	3	2	1	1188	2-11-2015 09:14	1	0	yoghurt,1unitsperpour
5053	nosurran	3	2	10	1188	2-11-2015 09:14	1	0	yoghurt,1unitsperpour
5053	nosurran	3	2	10	1189	2-11-2015 09:14	1	0	yoghurt,1unitsperpour
5053	nosurran	3	2	10	1191	2-11-2015 09:14	1	0	yoghurt,1unitsperpour
5053	nosurran	3	2	10	1192	2-11-2015 09:14	1	0	yoghurt,1unitsperpour
5053	nosurran	3	2	2	1195	2-11-2015 09:14	1	0	yoghurt,1unitsperpour
5053	nosurran	3	2	1	1197	2-11-2015 09:14	1	0	juice,1unitsperpour
5053	nosurran	3	2	10	1198	2-11-2015 09:14	1	0	juice,1unitsperpour
5053	nosurran	3	2	10	1198	2-11-2015 09:14	1	0	juice,1unitsperpour
5053	nosurran	3	2	10	1200	2-11-2015 09:14	1	0	juice,1unitsperpour
5053	nosurran	3	2	10	1202	2-11-2015 09:14	1	0	juice,1unitsperpour
5053	nosurran	3	2	2	1203	2-11-2015 09:14	1	0	juice,1unitsperpour
5053	nosurran	3	2	4	1204	2-11-2015 09:14	1	1	
5053	nosurran	3	2	3	1206	2-11-2015 09:14	2	0	
5053	nosurran	3	2	1	1232	2-11-2015 09:15	2	0	yoghurt,1unitsperpour
5053	nosurran	3	2	10	1233	2-11-2015 09:15	2	0	yoghurt,1unitsperpour
5053	nosurran	3	2	10	1233	2-11-2015 09:15	2	0	yoghurt,1unitsperpour
5053	nosurran	3	2	10	1234	2-11-2015 09:15	2	0	yoghurt,1unitsperpour
5053	nosurran	3	2	2	1235	2-11-2015 09:15	2	0	yoghurt,1unitsperpour
5053	nosurran	3	2	1	1235	2-11-2015 09:15	2	0	yoghurt,1unitsperpour
5053	nosurran	3	2	2	1235	2-11-2015 09:15	2	0	yoghurt,1unitsperpour
5053	nosurran	3	2	4	1240	2-11-2015 09:15	2	1	
5053	nosurran	3	2	3	1248	2-11-2015 09:15	3	0	
5053	nosurran	3	2	1	1322	2-11-2015 09:16	3	0	yoghurt,5unitsperpour
5053	nosurran	3	2	10	1323	2-11-2015 09:16	3	0	yoghurt,5unitsperpour
5053	nosurran	3	2	10	1324	2-11-2015 09:16	3	0	yoghurt,5unitsperpour
5053	nosurran	3	2	10	1324	2-11-2015 09:16	3	0	yoghurt,5unitsperpour
5053	nosurran	3	2	2	1325	2-11-2015 09:16	3	0	yoghurt,5unitsperpour
5053	nosurran	3	2	1	1326	2-11-2015 09:16	3	0	yoghurt,1unitsperpour
5053	nosurran	3	2	10	1326	2-11-2015 09:16	3	0	yoghurt,1unitsperpour
5053	nosurran	3	2	2	1327	2-11-2015 09:16	3	0	yoghurt,1unitsperpour
5053	nosurran	3	2	4	1328	2-11-2015 09:16	3	1	
5053	nosurran	3	2	3	1329	2-11-2015 09:16	4	0	
5053	nosurran	3	2	1	1338	2-11-2015 09:17	4	0	yoghurt,1unitsperpour
5053	nosurran	3	2	10	1338	2-11-2015 09:17	4	0	yoghurt,1unitsperpour
5053	nosurran	3	2	2	1339	2-11-2015 09:17	4	0	yoghurt,1unitsperpour
5053	nosurran	3	2	4	1340	2-11-2015 09:17	4	1	

Appendix 9: correlation between pre-test, post-test and TTR score

Correlations

		Pretest	Posttest	TTRtotaal
Pretest	Pearson Correlation	1	,705**	,416**
	Sig. (2-tailed)		,000	,000
	N	310	292	221
Posttest	Pearson Correlation	,705**	1	,404**
	Sig. (2-tailed)	,000		,000
	N	292	313	203
TTRtotaal	Pearson Correlation	,416**	,404**	1
	Sig. (2-tailed)	,000	,000	
	N	221	203	222

**. Correlation is significant at the 0.01 level (2-tailed).

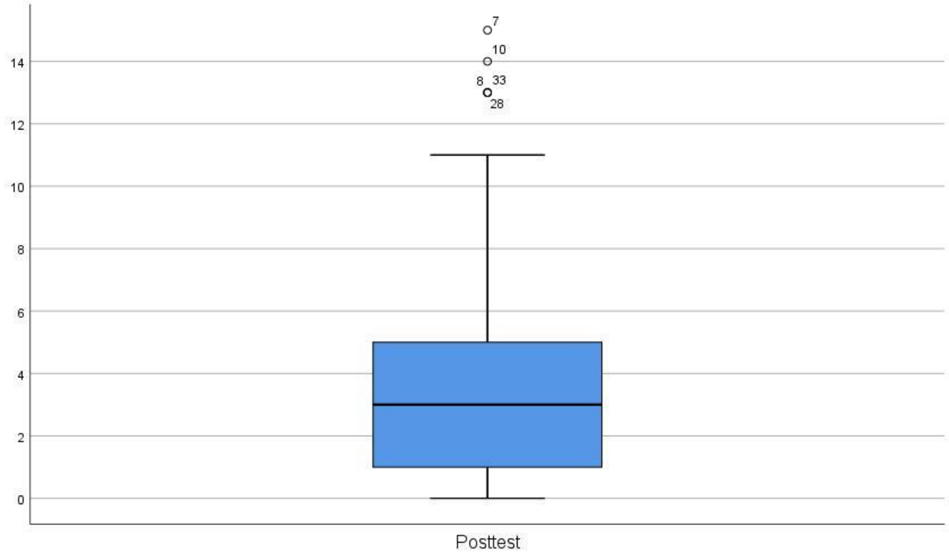
Appendix 10: different conditions in data set

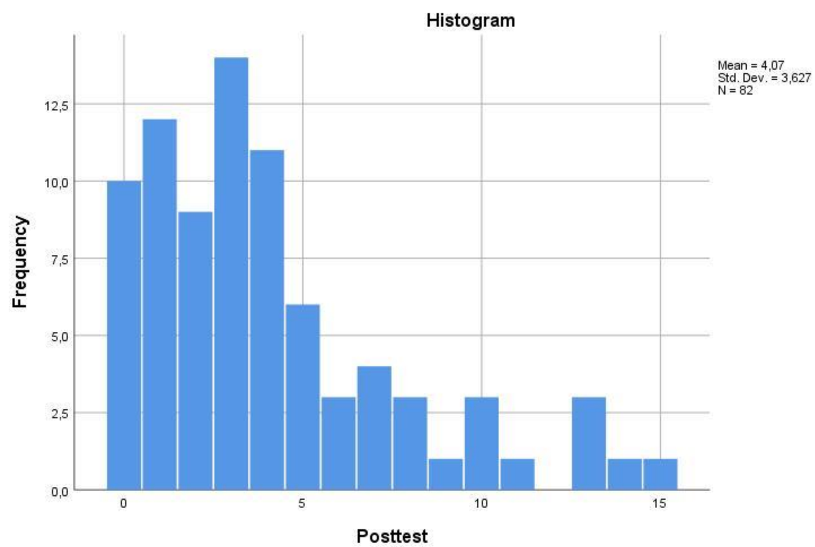
Versions	Description	Participants	Unique code	Meaning/notes
PWS/Master	Part-task practice and worked example	80	37	Results of part-task practice
Control	Default game	83		The 2013 version only has a calculator as an assistance tool.
Gameplay	Not clear -> no log data or statistics available	11		
Curiosity	Player must activate unknown information by clicking on it.	48	20	Clicking on a question mark The 2013 version only has a calculator as an assistance tool.
Surprise Sequenced	Surprising events in sequences	25		Surprising events changes steps to solve problem.

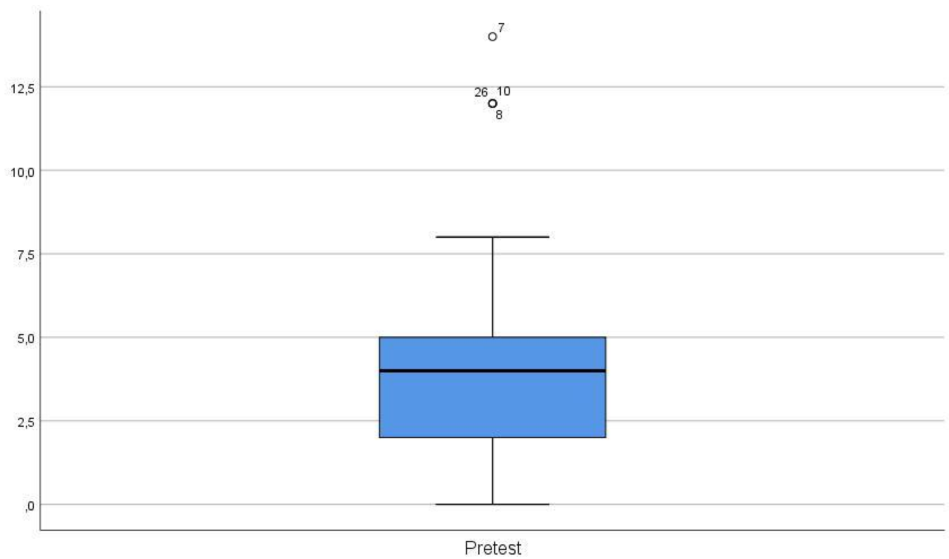
				Character throws bottles out of the refrigerator.
No Surprise Sequenced	No surprising events in sequences	30		
Surprise Random	Surprising events in random sequence	28		
No Surprise Random	No surprising events in random sequence	30		

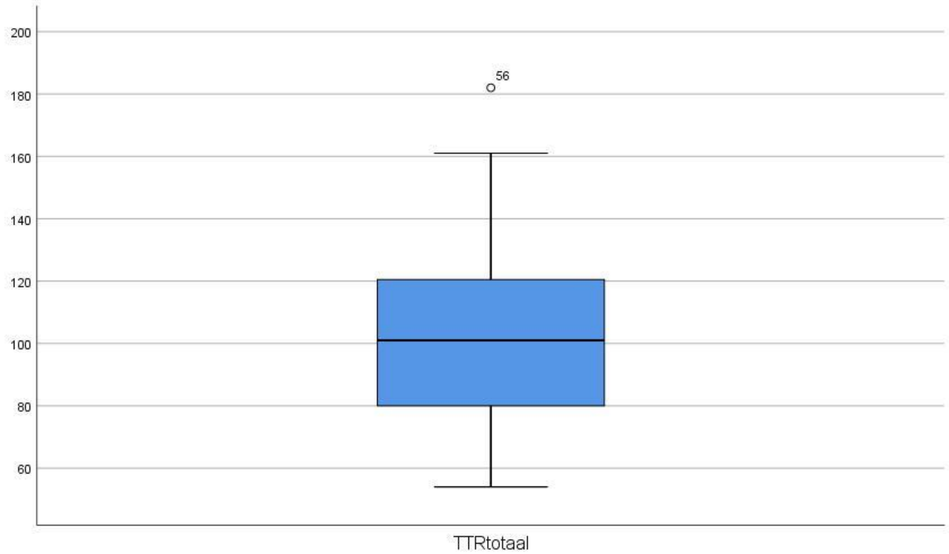
Appendix 11: histograms and boxplots of Control and Part-Task/Worked Example Conditions (pre-test, post-test, TTR and progression score)

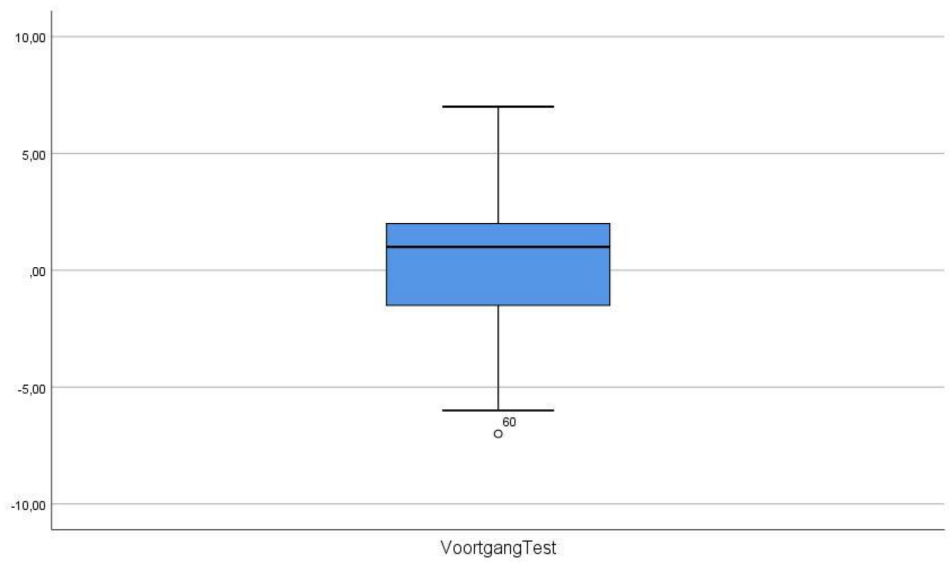
Control:

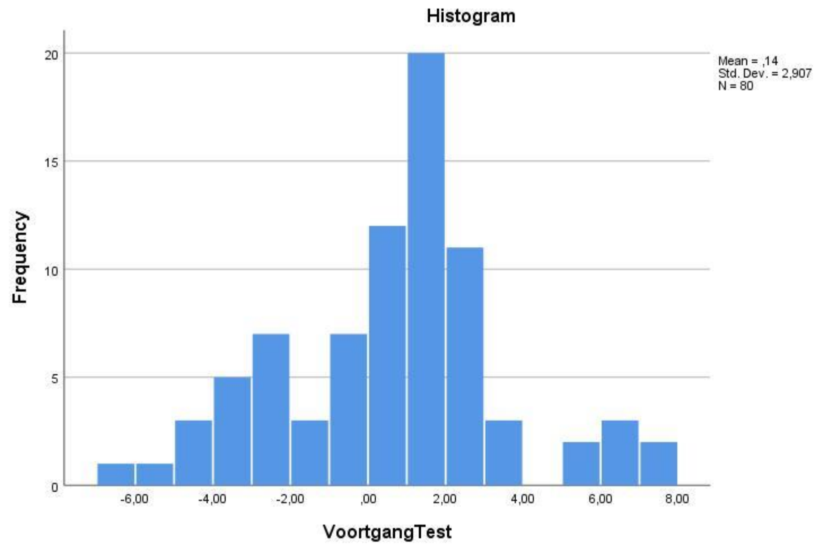



Histogram
Mean = 4,07
Std. Dev. = 3,627
N = 82
Frequency
0,0
2,5
5,0
7,5
10,0
12,5
0
5
10
15
Posttest

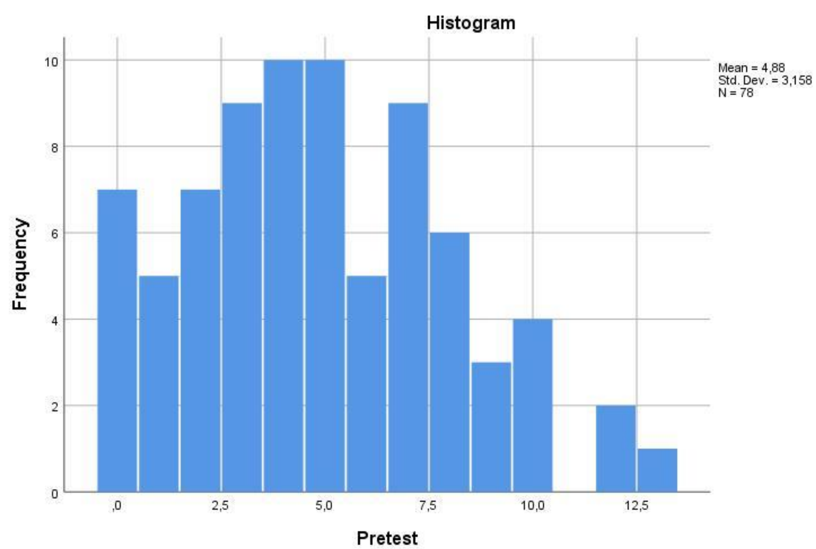

,0
2,5
5,0
7,5
10,0
12,5
7
26
10
8
Pretest

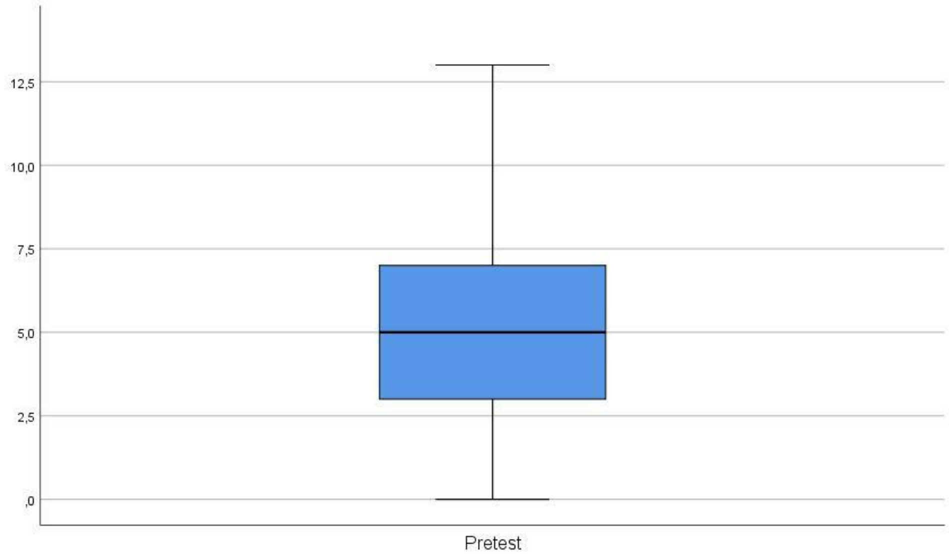

200
180
160
140
120
100
80
60
56
TTRtotaal


10,00
5,00
,00
-5,00
-10,00
60
VoortgangTest

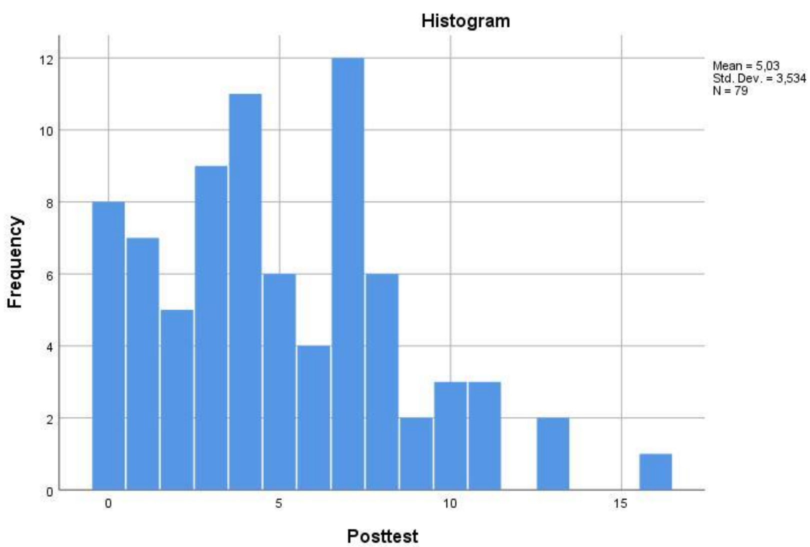

Histogram
Mean = ,14
Std. Dev. = 2,907
N = 80
Frequency
0
5
10
15
20
-6,00
-4,00
-2,00
,00
2,00
4,00
6,00
8,00
VoortgangTest

PWS:

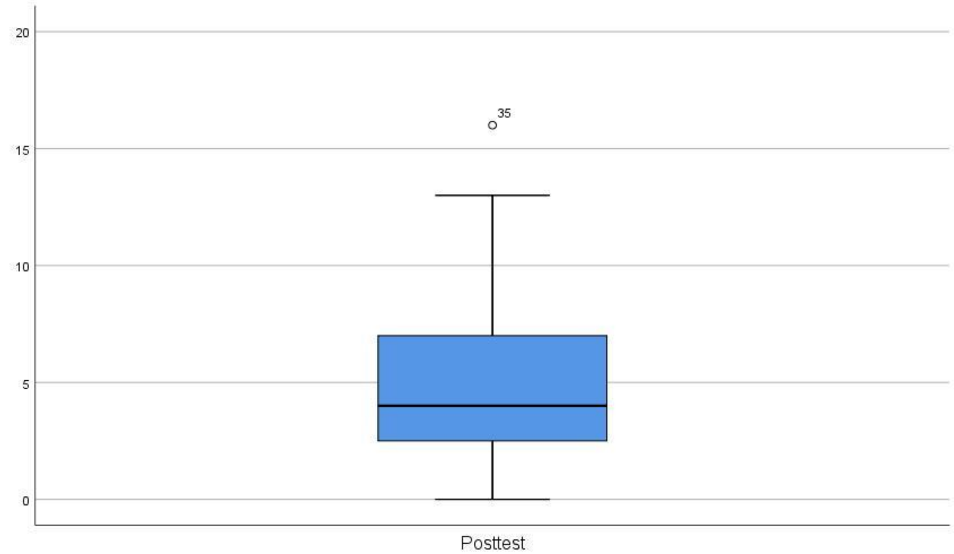

Histogram
Mean = 4,88
Std. Dev. = 3,158
N = 78
Frequency
0
2
4
6
8
10
,0
2,5
5,0
7,5
10,0
12,5
Pretest

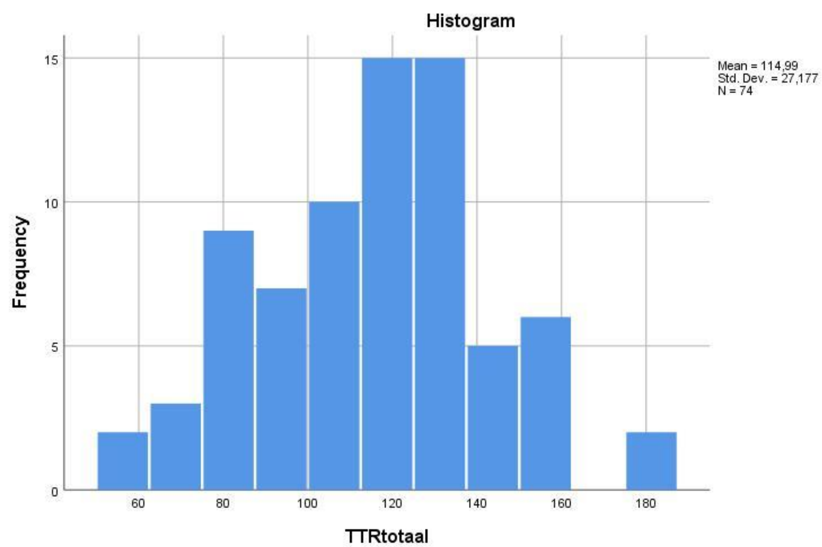


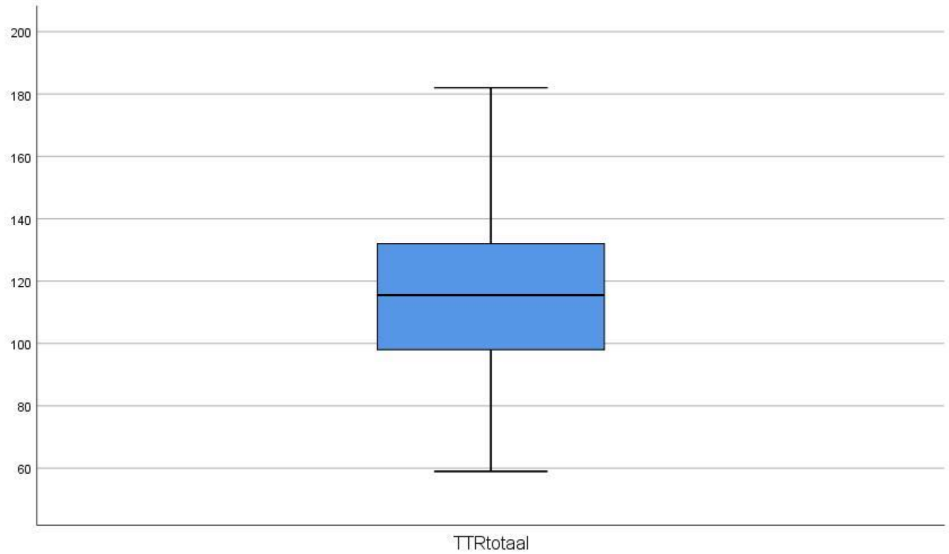
12,5
10,0
7,5
5,0
2,5
,0
Pretest

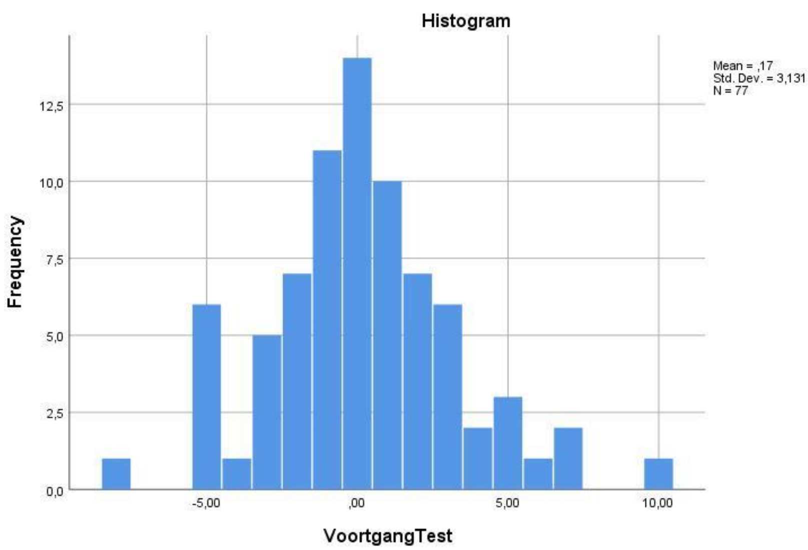


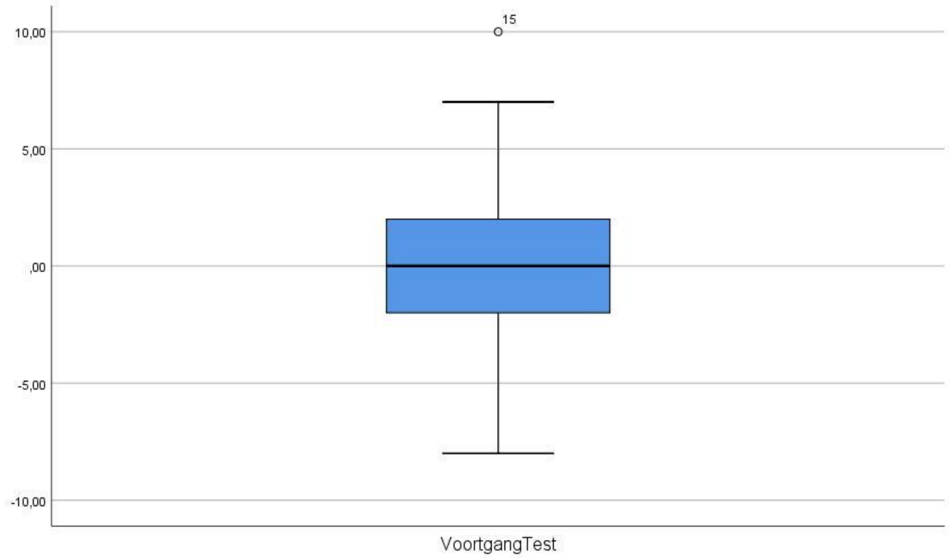
Histogram
Mean = 5,03
Std. Dev. = 3,534
N = 79
Frequency
Posttest


20
15
10
5
0
35
Posttest


Histogram
Mean = 114,99
Std. Dev. = 27,177
N = 74
Frequency
15
10
5
0
60
80
100
120
140
160
180
TTRtotaal


200
180
160
140
120
100
80
60
TTRtotaal


Histogram
Mean = ,17
Std. Dev. = 3,131
N = 77
Frequency
12,5
10,0
7,5
5,0
2,5
0,0
-5,00
,00
5,00
10,00
VoortgangTest



Appendix 12: example of downtime in a game session

Participant	Pre-test	Post-test	TTR	Progression
2005	4	1		-3

Interpretation of log data:

1109: answer is incorrect

1109-1140: player thinks about new strategy to solve the problem

1140-1222: long series of actions with calculator (player struggles in finding the arithmetical steps to solve the problem), but eventually finds it

1230-1234: quickly solving the problem

1235: correct answer

2005	controle	3	2	4	1109	05/17/201	2	0	
2005	controle	3	2	11	1140	05/17/201	2	0	9* = 0
2005	controle	3	2	11	1140	05/17/201	2	0	0* = 0
2005	controle	3	2	11	1140	05/17/201	2	0	0* = 0
2005	controle	3	2	11	1140	05/17/201	2	0	0* = 0
2005	controle	3	2	11	1141	05/17/201	2	0	0* = 0
2005	controle	3	2	11	1141	05/17/201	2	0	0* = 0
2005	controle	3	2	11	1146	05/17/201	2	0	33 = 33
2005	controle	3	2	11	1149	05/17/201	2	0	9 = 9
2005	controle	3	2	11	1152	05/17/201	2	0	9*33 = 297
2005	controle	3	2	11	1180	05/17/201	2	0	24/ = 0
2005	controle	3	2	11	1183	05/17/201	2	0	0/24 = 0
2005	controle	3	2	11	1188	05/17/201	2	0	24 = 24
2005	controle	3	2	11	1193	05/17/201	2	0	24/6 = 4
2005	controle	3	2	11	1222	05/17/201	2	0	9*4 = 36
2005	controle	3	2	1	1230	05/17/201	2	0	yoghurt,1unitsperpour
2005	controle	3	2	10	1231	05/17/201	2	0	yoghurt,1unitsperpour
2005	controle	3	2	10	1232	05/17/201	2	0	yoghurt,1unitsperpour
2005	controle	3	2	10	1233	05/17/201	2	0	yoghurt,1unitsperpour
2005	controle	3	2	2	1234	05/17/201	2	0	yoghurt,1unitsperpour
2005	controle	3	2	4	1235	05/17/201	2	1	

Appendix 13: different scenarios after having an incorrect answer

Participant	Pre-test	Post-test	TTR	Progression
2005	4	1		-3

Interpretation of action:

212 -> answer is incorrect

226 – 244 -> player tries to solve problem, does not use assistance tools

244 -> answer is incorrect for the second time

244-252 -> short downtime between incorrect answer and next action -> not thinking thoroughly in how to solve the problem?

252-260 -> player tries to solve the problem without a real direction.

262 -> answer is incorrect for the third time -> player is forced to move on to the next assignment

82593	2005	controle	1	1	4	212	05/17/20	2	0	
82750	2005	controle	1	1	2	226	05/17/20	2	0	fanta,5pack
82751	2005	controle	1	1	2	226	05/17/20	2	0	fanta,5pack
82752	2005	controle	1	1	1	232	05/17/20	2	0	cola,10pack
82753	2005	controle	1	1	7	232	05/17/20	2	0	cola,10pack
82754	2005	controle	1	1	6	233	05/17/20	2	0	cola,10pack
82755	2005	controle	1	1	7	233	05/17/20	2	0	cola,10pack
82756	2005	controle	1	1	2	233	05/17/20	2	0	cola,10pack
82757	2005	controle	1	1	1	234	05/17/20	2	0	cola,5pack
82758	2005	controle	1	1	2	235	05/17/20	2	0	cola,5pack
82759	2005	controle	1	1	6	235	05/17/20	2	0	cola,5pack
82760	2005	controle	1	1	1	236	05/17/20	2	0	cola,5pack
82761	2005	controle	1	1	7	236	05/17/20	2	0	cola,5pack
82762	2005	controle	1	1	2	237	05/17/20	2	0	cola,5pack
82763	2005	controle	1	1	6	237	05/17/20	2	0	cola,5pack
82764	2005	controle	1	1	1	238	05/17/20	2	0	cola,1pack
82765	2005	controle	1	1	7	238	05/17/20	2	0	cola,1pack
82766	2005	controle	1	1	6	238	05/17/20	2	0	cola,1pack
82767	2005	controle	1	1	2	239	05/17/20	2	0	cola,1pack
82768	2005	controle	1	1	1	241	05/17/20	2	0	cola,5pack
82769	2005	controle	1	1	6	241	05/17/20	2	0	cola,5pack
82770	2005	controle	1	1	7	241	05/17/20	2	0	cola,5pack
82771	2005	controle	1	1	6	241	05/17/20	2	0	cola,5pack
82772	2005	controle	1	1	7	242	05/17/20	2	0	cola,5pack
82773	2005	controle	1	1	2	243	05/17/20	2	0	cola,5pack
82774	2005	controle	1	1	6	243	05/17/20	2	0	cola,5pack
82775	2005	controle	1	1	4	244	05/17/20	2	0	
82813	2005	controle	1	1	1	252	05/17/20	2	0	cola,5pack
82814	2005	controle	1	1	7	252	05/17/20	2	0	cola,5pack
82815	2005	controle	1	1	2	253	05/17/20	2	0	cola,5pack
82816	2005	controle	1	1	1	254	05/17/20	2	0	cola,5pack
82817	2005	controle	1	1	7	254	05/17/20	2	0	cola,5pack
82818	2005	controle	1	1	2	255	05/17/20	2	0	cola,5pack
82819	2005	controle	1	1	1	256	05/17/20	2	0	cola,1pack
82820	2005	controle	1	1	7	257	05/17/20	2	0	cola,1pack
82821	2005	controle	1	1	2	257	05/17/20	2	0	cola,1pack
82822	2005	controle	1	1	1	259	05/17/20	2	0	cola,5pack
82823	2005	controle	1	1	7	259	05/17/20	2	0	cola,5pack
82824	2005	controle	1	1	2	260	05/17/20	2	0	cola,5pack
82825	2005	controle	1	1	4	262	05/17/20	2	0	

Participant	Pre-test	Post-test	TTR	Progression
1102	8	7	129	-1

Interpretation of action:

169 – answer is incorrect (Action 4 – Score 0)

177-182 – player tries to solve it but is struggling.

182-248 -> using multiple assistance tools (18 – game manual, 22 – assistance board, 23 – handbook)

254-272 -> player is solving problem after using assistance tools

274 -> answer is correct

6E+06	1102	pws	1	1	4	168	04/23/2	1	0	
6E+06	1102	pws	1	1	1	177	04/23/2	1	0	cola,10pack
6E+06	1102	pws	1	1	7	177	04/23/2	1	0	cola,10pack
6E+06	1102	pws	1	1	2	178	04/23/2	1	0	cola,10pack
6E+06	1102	pws	1	1	36	180	04/23/2	1	0	
6E+06	1102	pws	1	1	3	182	04/23/2	1	0	
6E+06	1102	pws	1	1	16	182	04/23/2	1	0	
6E+06	1102	pws	1	1	22	208	04/23/2	1	0	
6E+06	1102	pws	1	1	23	208	04/23/2	1	0	
6E+06	1102	pws	1	1	36	247	04/23/2	1	0	
6E+06	1102	pws	1	1	36	247	04/23/2	1	0	
6E+06	1102	pws	1	1	22	248	04/23/2	1	0	
6E+06	1102	pws	1	1	1	254	04/23/2	1	0	cola,10pack
6E+06	1102	pws	1	1	7	254	04/23/2	1	0	cola,10pack
6E+06	1102	pws	1	1	6	254	04/23/2	1	0	cola,10pack
6E+06	1102	pws	1	1	7	255	04/23/2	1	0	cola,10pack
6E+06	1102	pws	1	1	6	255	04/23/2	1	0	cola,10pack
6E+06	1102	pws	1	1	7	255	04/23/2	1	0	cola,10pack
6E+06	1102	pws	1	1	2	255	04/23/2	1	0	cola,10pack
6E+06	1102	pws	1	1	1	257	04/23/2	1	0	cola,10pack
6E+06	1102	pws	1	1	6	258	04/23/2	1	0	cola,10pack
6E+06	1102	pws	1	1	7	258	04/23/2	1	0	cola,10pack
6E+06	1102	pws	1	1	2	259	04/23/2	1	0	cola,10pack
6E+06	1102	pws	1	1	6	259	04/23/2	1	0	cola,10pack
6E+06	1102	pws	1	1	1	260	04/23/2	1	0	cola,10pack
6E+06	1102	pws	1	1	6	260	04/23/2	1	0	cola,10pack
6E+06	1102	pws	1	1	7	260	04/23/2	1	0	cola,10pack
6E+06	1102	pws	1	1	6	261	04/23/2	1	0	cola,10pack
6E+06	1102	pws	1	1	2	261	04/23/2	1	0	cola,10pack
6E+06	1102	pws	1	1	1	262	04/23/2	1	0	cola,10pack
6E+06	1102	pws	1	1	6	263	04/23/2	1	0	cola,10pack
6E+06	1102	pws	1	1	7	265	04/23/2	1	0	cola,10pack
6E+06	1102	pws	1	1	2	265	04/23/2	1	0	cola,10pack
6E+06	1102	pws	1	1	1	266	04/23/2	1	0	cola,10pack
6E+06	1102	pws	1	1	6	266	04/23/2	1	0	cola,10pack
6E+06	1102	pws	1	1	2	266	04/23/2	1	0	cola,10pack
6E+06	1102	pws	1	1	1	267	04/23/2	1	0	cola,10pack
6E+06	1102	pws	1	1	6	267	04/23/2	1	0	cola,10pack
6E+06	1102	pws	1	1	2	268	04/23/2	1	0	cola,10pack
6E+06	1102	pws	1	1	1	269	04/23/2	1	0	cola,10pack
6E+06	1102	pws	1	1	6	269	04/23/2	1	0	cola,10pack
6E+06	1102	pws	1	1	2	270	04/23/2	1	0	cola,10pack
6E+06	1102	pws	1	1	7	270	04/23/2	1	0	cola,10pack
6E+06	1102	pws	1	1	6	270	04/23/2	1	0	cola,10pack
6E+06	1102	pws	1	1	1	270	04/23/2	1	0	cola,10pack
6E+06	1102	pws	1	1	7	270	04/23/2	1	0	cola,10pack
6E+06	1102	pws	1	1	6	270	04/23/2	1	0	cola,10pack
6E+06	1102	pws	1	1	6	271	04/23/2	1	0	cola,10pack
6E+06	1102	pws	1	1	2	272	04/23/2	1	0	cola,10pack
6E+06	1102	pws	1	1	4	274	04/23/2	1	1	

Participant	Pre-test	Post-test	TTR	Progression
2021	12	14		+2

Interpretation of action:

156 -> participant has an incorrect answer

156-176 -> thinking about how to solve the problem (no assistance tools)

177 -> knowing how to solve the problem – filling in the steps to solve the problem immediately

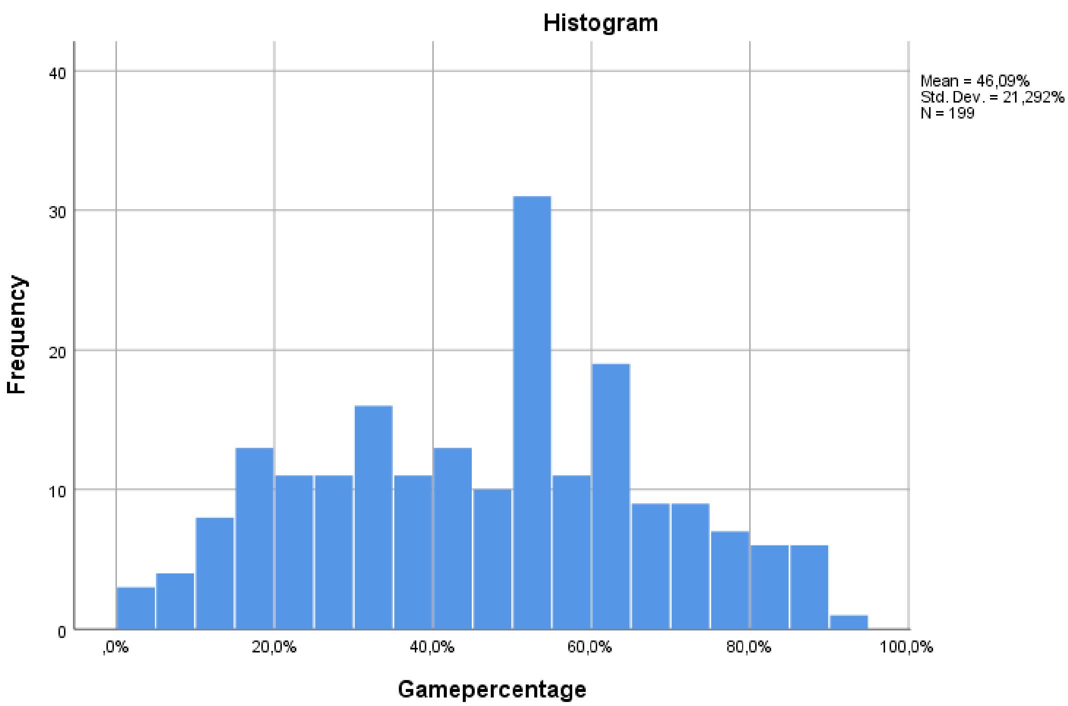
185 -> answer is correct

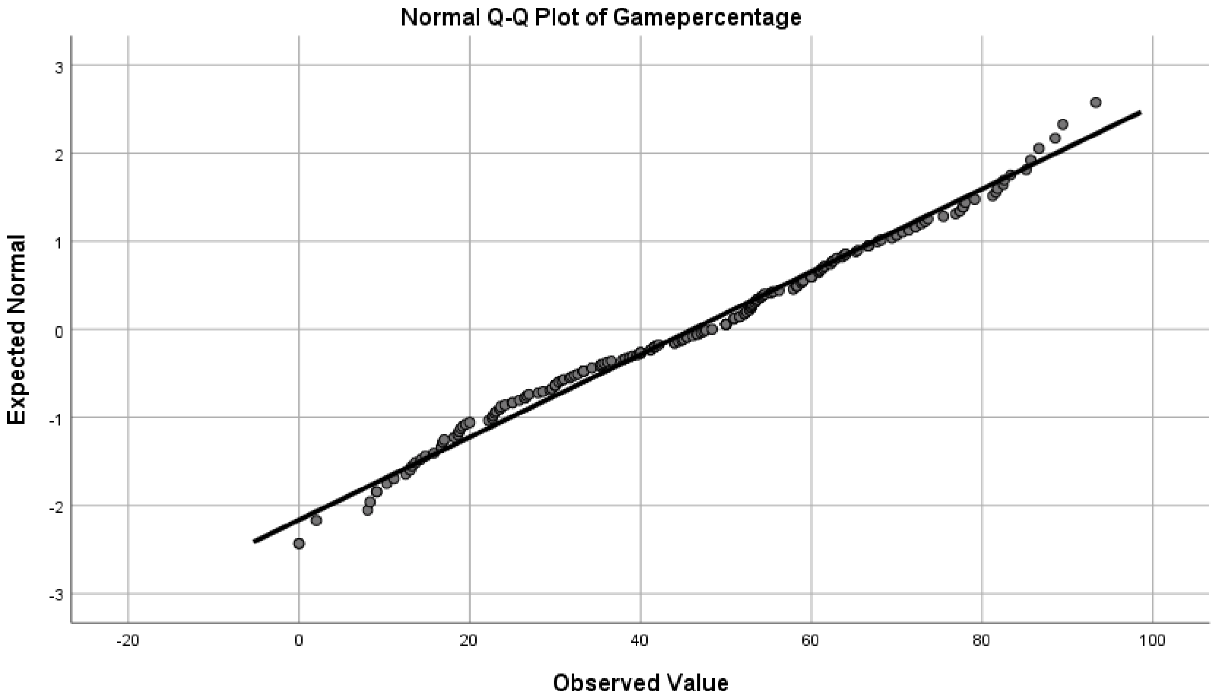
45	156147	2021	controle	3	1	4	156	06/14/20:	1	0	
46	156169	2021	controle	3	1	1	176	06/14/20:	1	0	juice,1unitsperpour
47	156170	2021	controle	3	1	10	177	06/14/20:	1	0	juice,1unitsperpour
48	156171	2021	controle	3	1	2	177	06/14/20:	1	0	juice,1unitsperpour
49	156172	2021	controle	3	1	1	179	06/14/20:	1	0	juice,1unitsperpour
50	156173	2021	controle	3	1	10	179	06/14/20:	1	0	juice,1unitsperpour
51	156174	2021	controle	3	1	2	180	06/14/20:	1	0	juice,1unitsperpour
52	156175	2021	controle	3	1	1	182	06/14/20:	1	0	juice,1unitsperpour
53	156176	2021	controle	3	1	2	182	06/14/20:	1	0	juice,1unitsperpour
54	156177	2021	controle	3	1	4	185	06/14/20:	1	1	

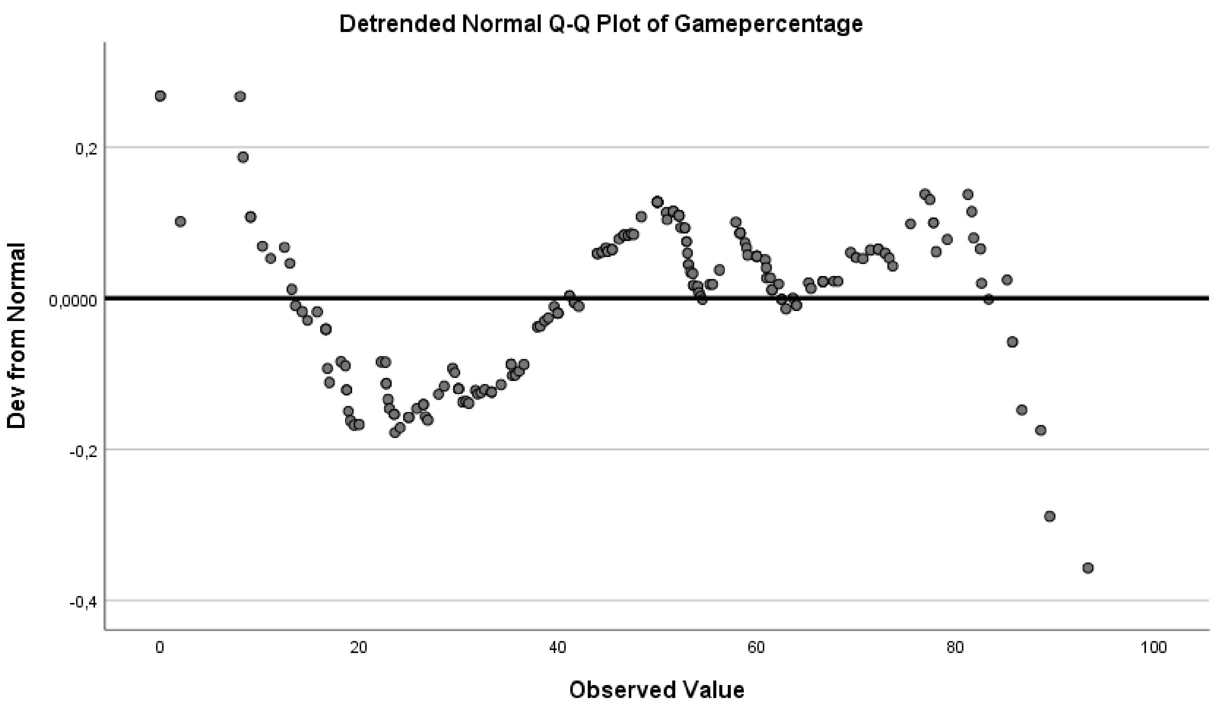
ANALYSIS PHASE

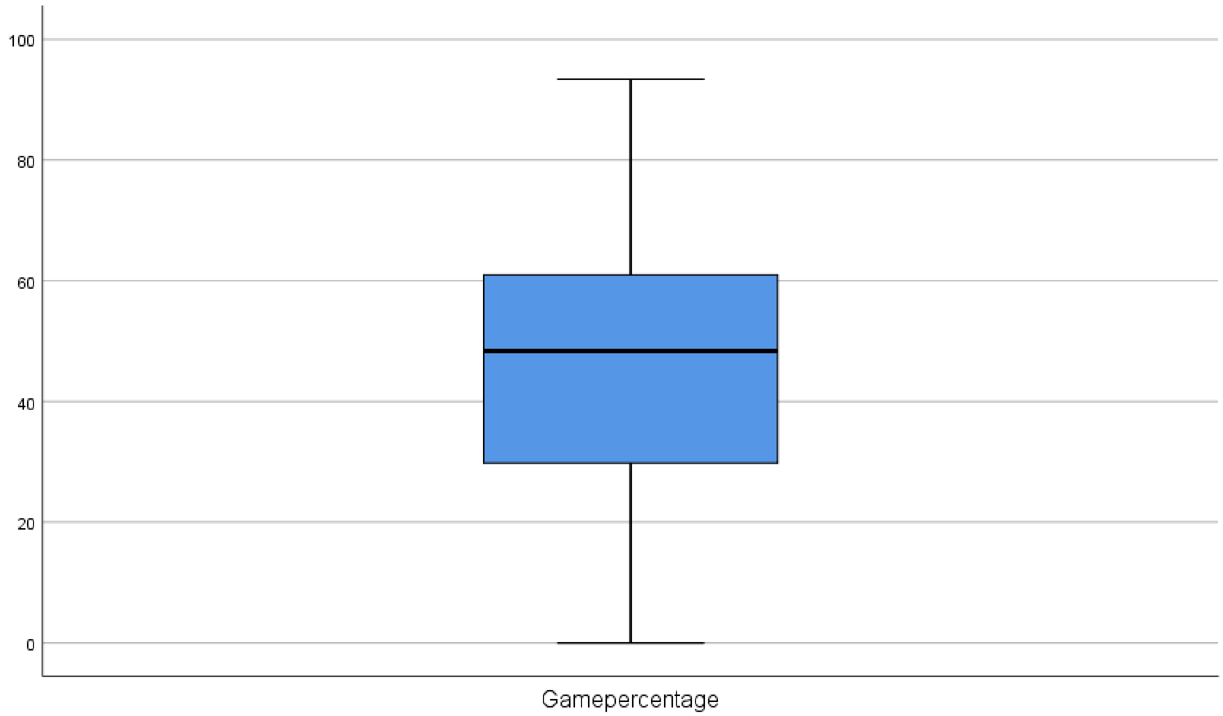
Appendix 14: SPSS Explore – Checking assumptions of independent variables for multiple regression analysis

Game percentage

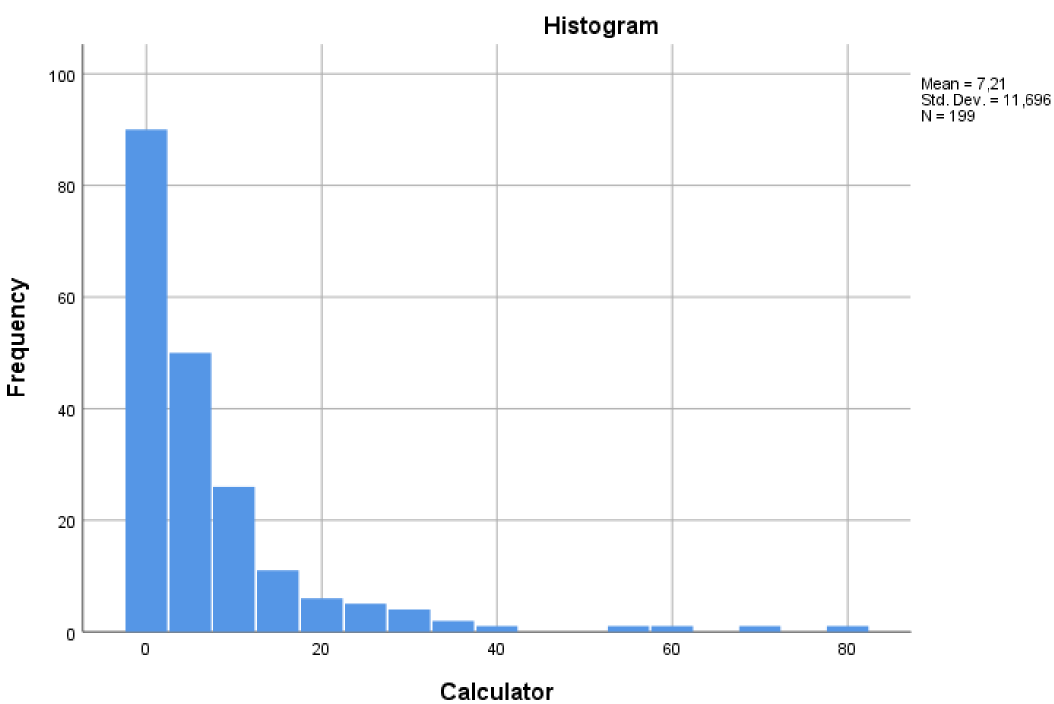


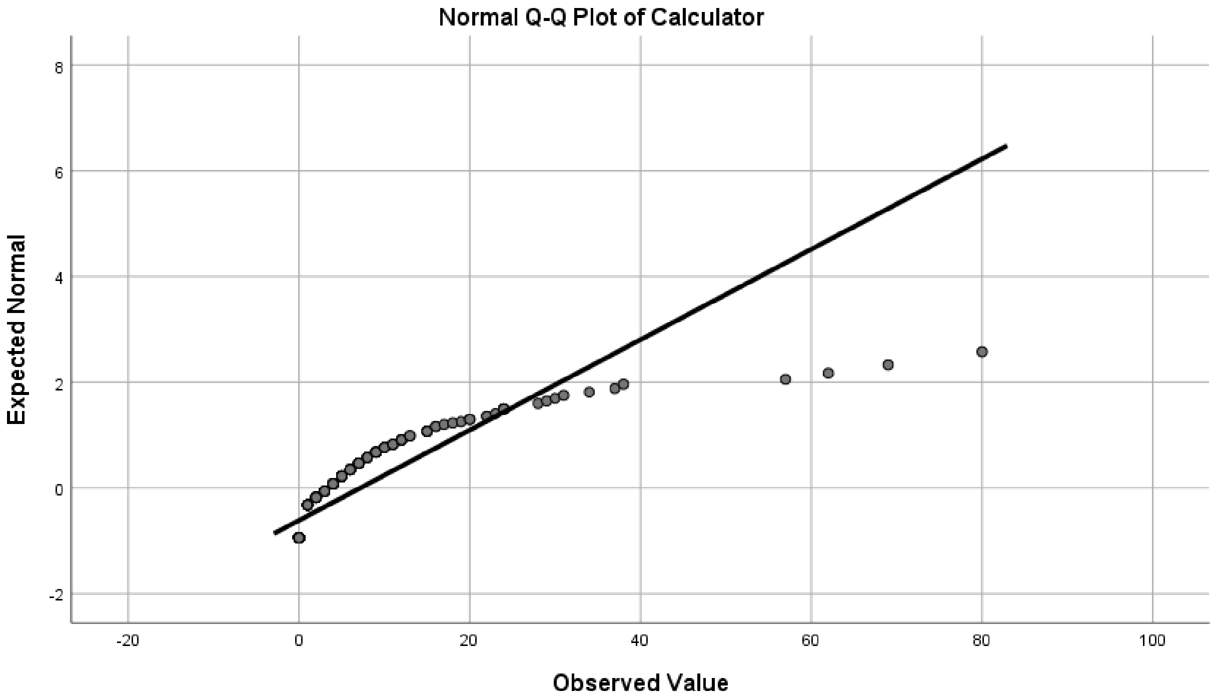

Normal Q-Q Plot of Gamepercentage
Expected Normal
Observed Value
3
2
1
0
-1
-2
-3
-20
0
20
40
60
80
100


Detrended Normal Q-Q Plot of Gamepercentage
Dev from Normal
Observed Value
0,2
0,0000
-0,2
-0,4
0
20
40
60
80
100

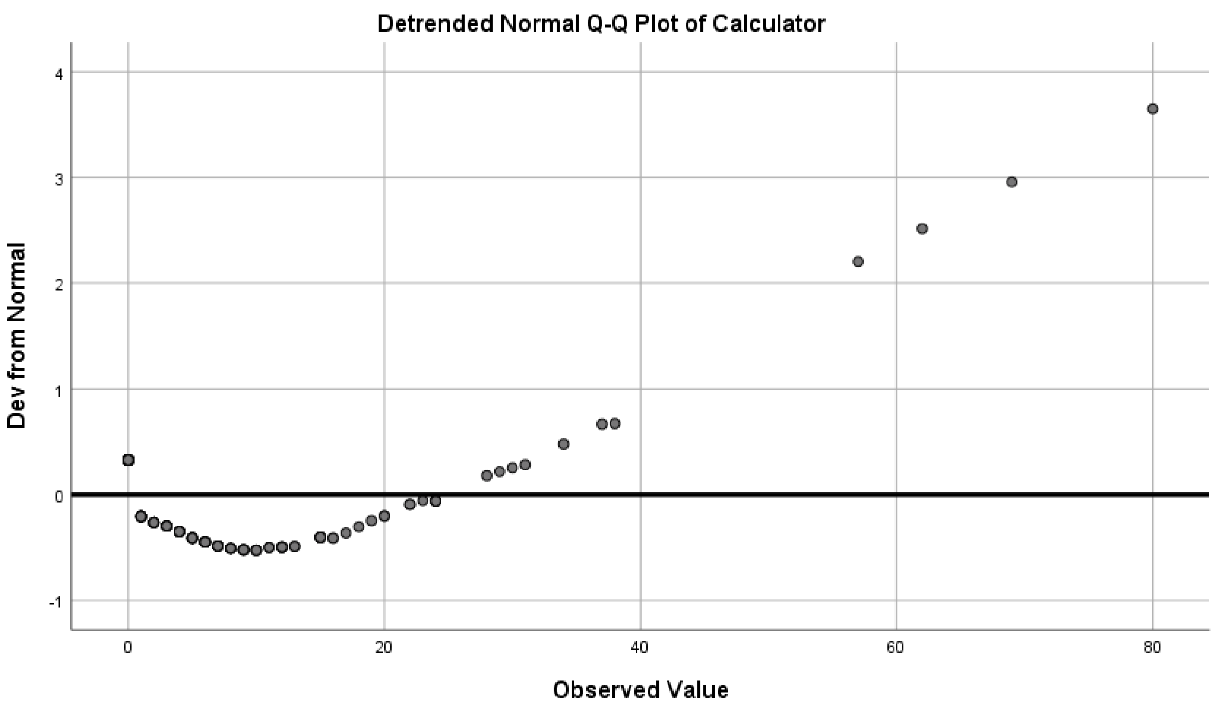

100
80
60
40
20
0
Gamepercentage

Calculator

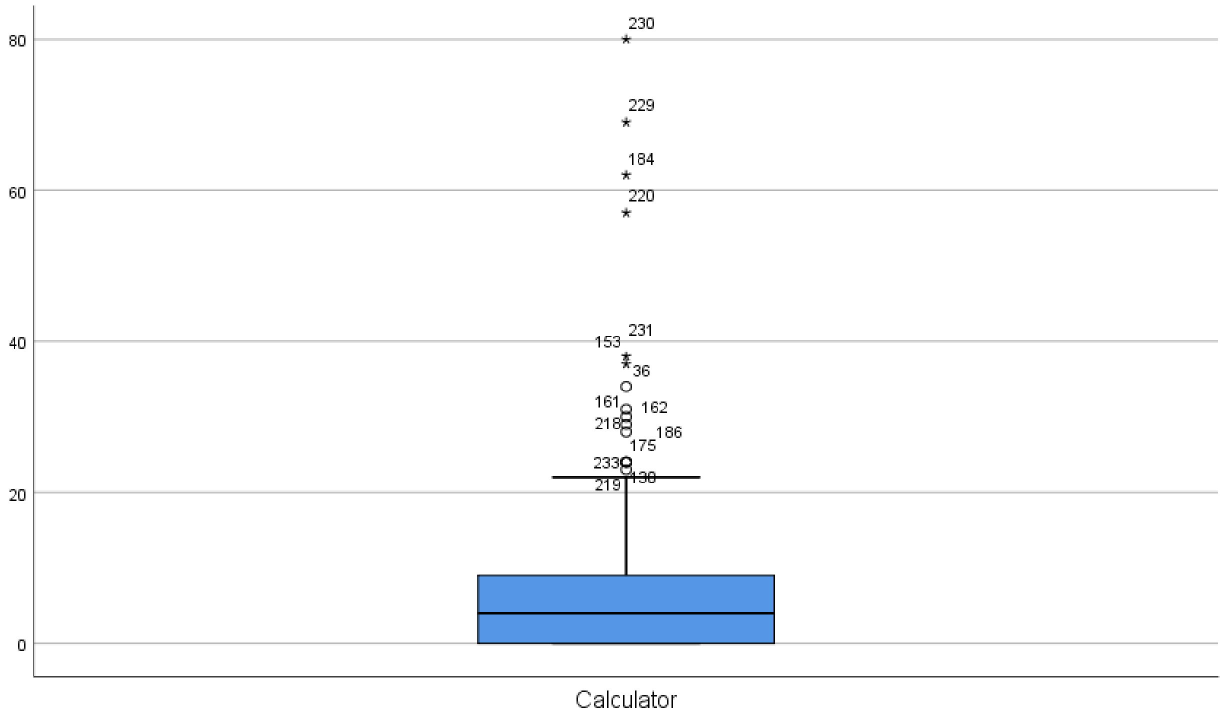

Histogram
Mean = 7,21
Std. Dev. = 11,696
N = 199
Frequency
100
80
60
40
20
0
0
20
40
60
80
Calculator



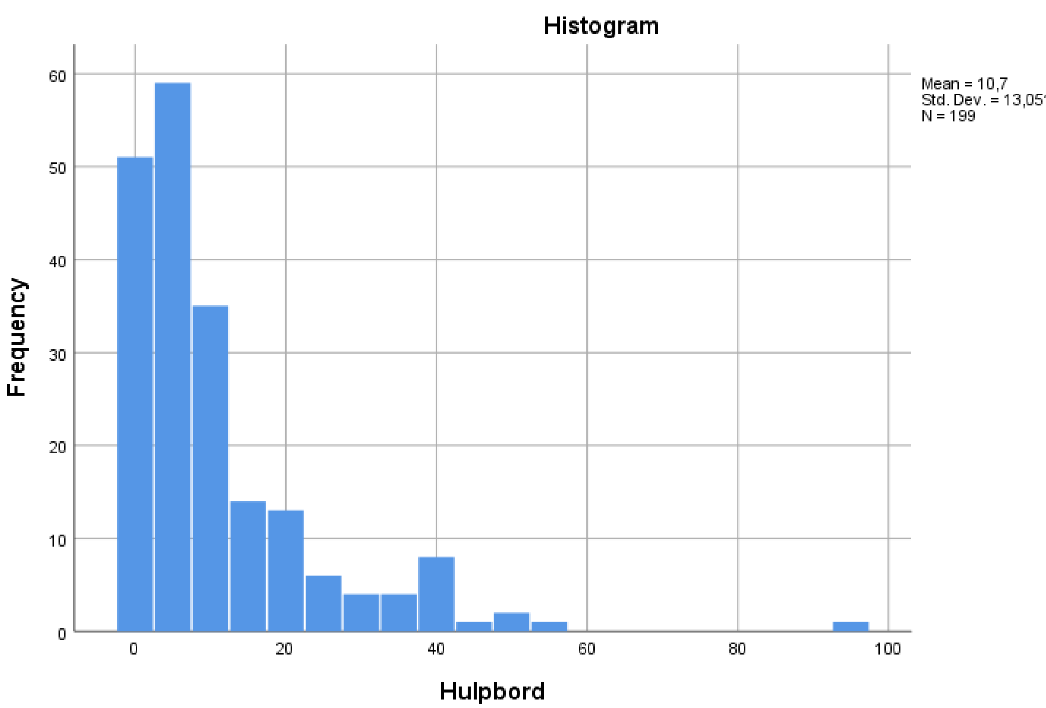
Normal Q-Q Plot of Calculator
Expected Normal
Observed Value
8
6
4
2
0
-2
-20
0
20
40
60
80
100

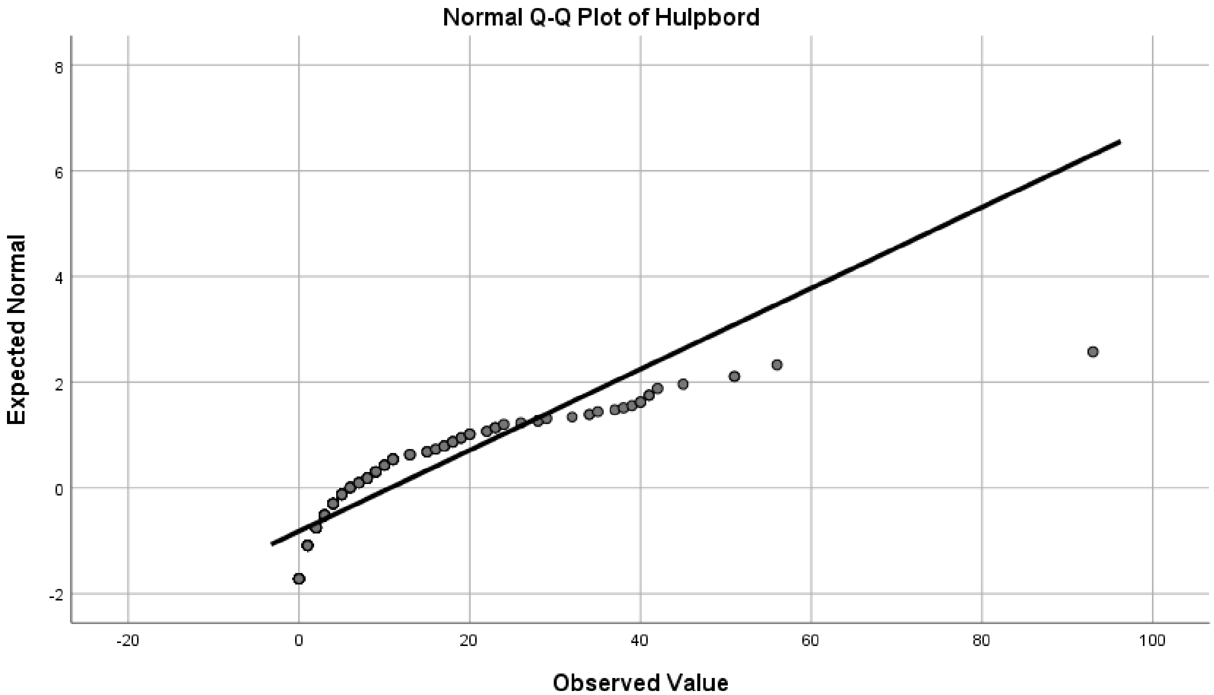


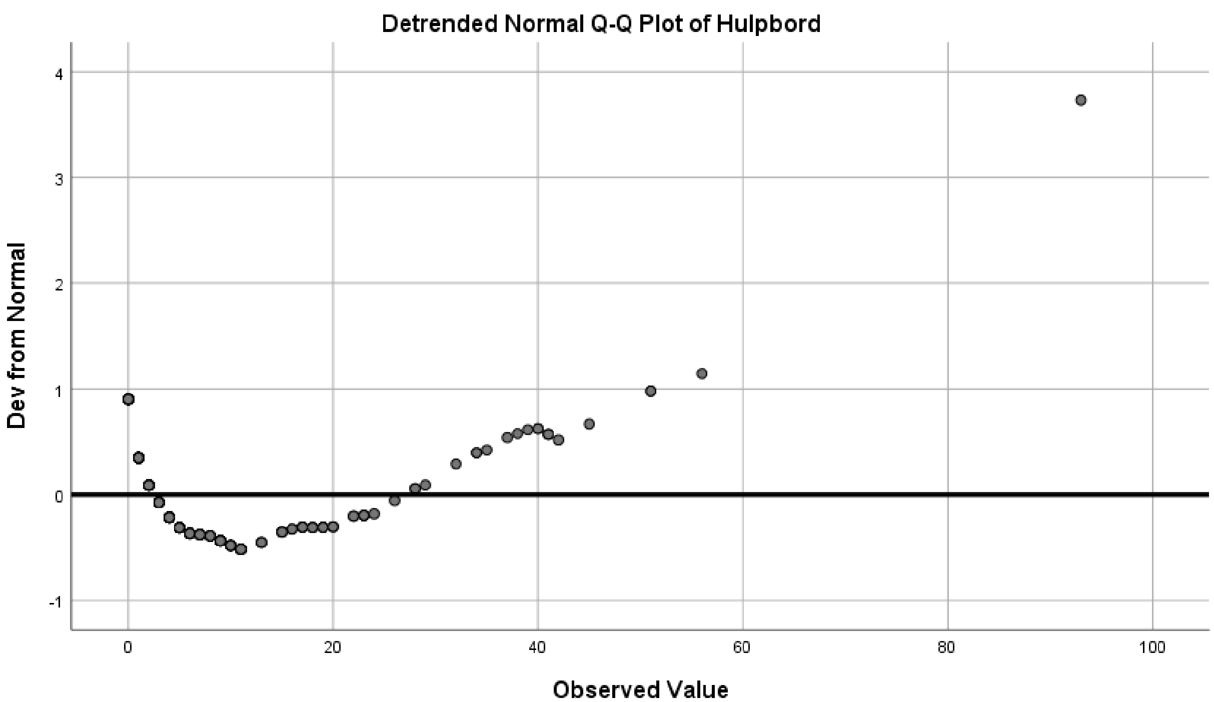
Detrended Normal Q-Q Plot of Calculator
Dev from Normal
Observed Value
4
3
2
1
0
-1
0
20
40
60
80

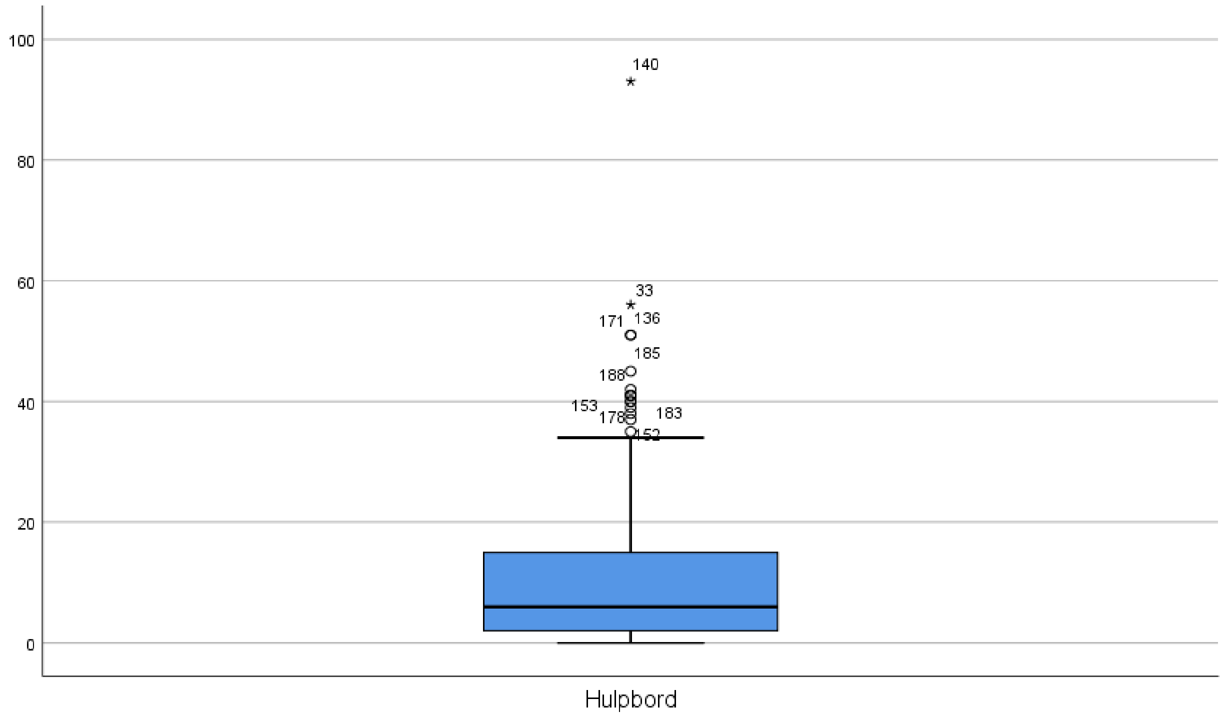


Assistance board:

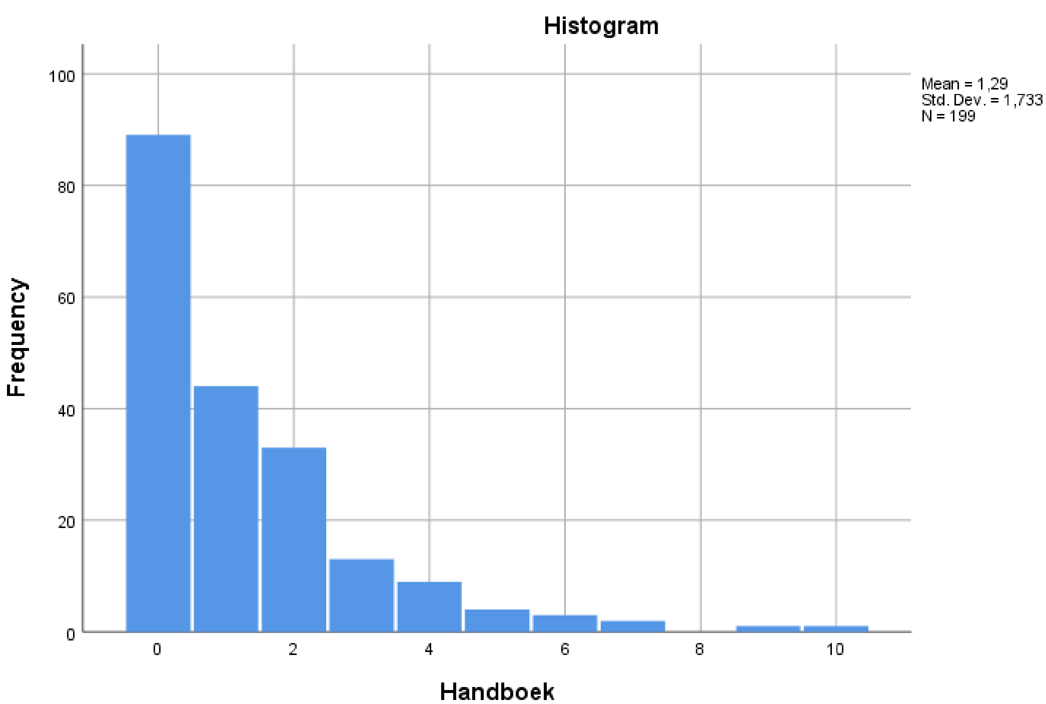


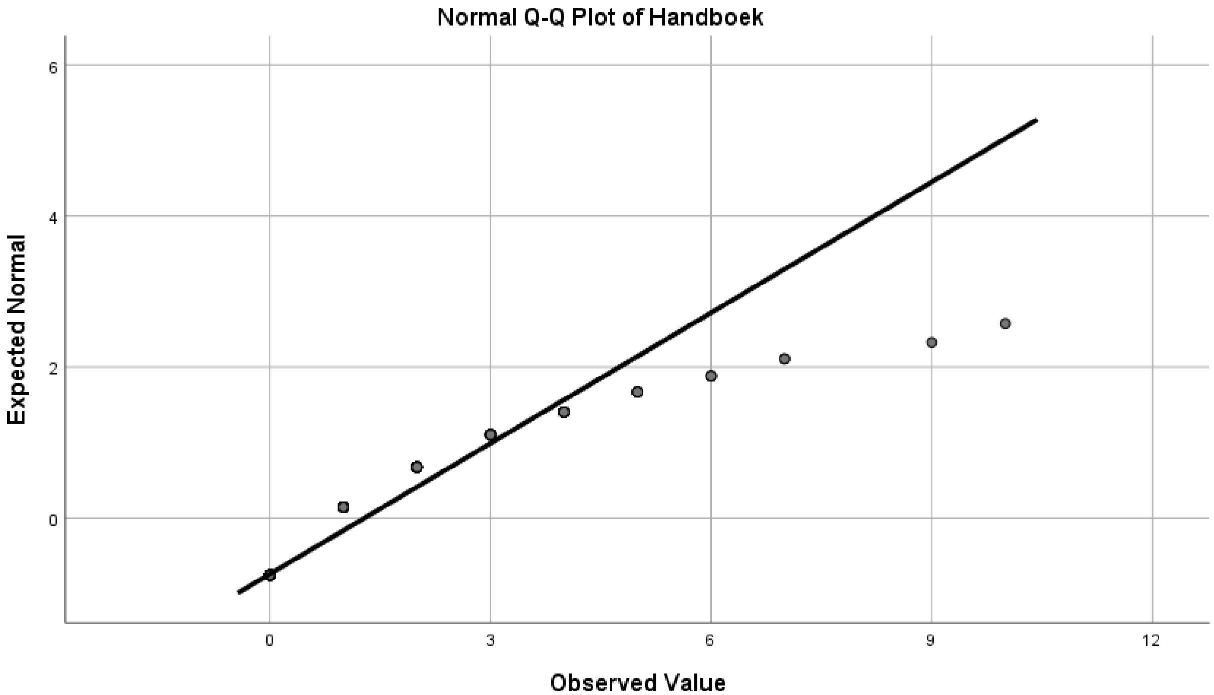

Normal Q-Q Plot of Hulpbord
Expected Normal
Observed Value
8
6
4
2
0
-2
-20
0
20
40
60
80
100

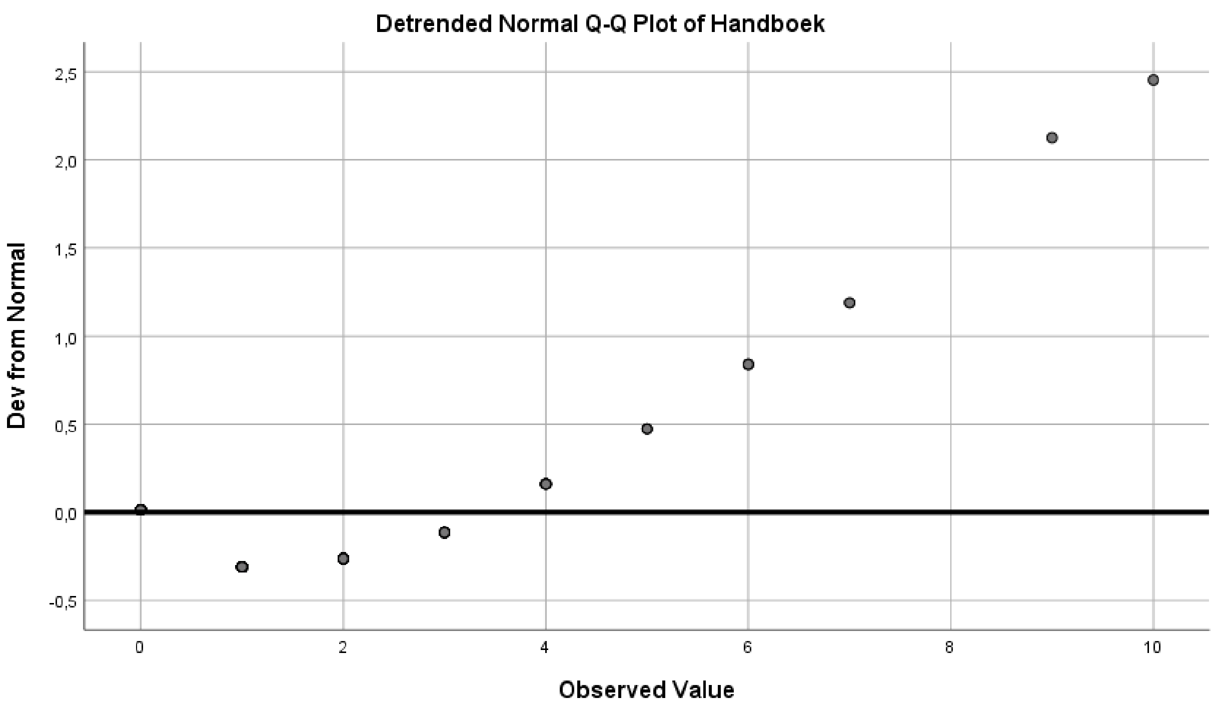

Detrended Normal Q-Q Plot of Hulpbord
Dev from Normal
Observed Value
4
3
2
1
0
-1
0
20
40
60
80
100

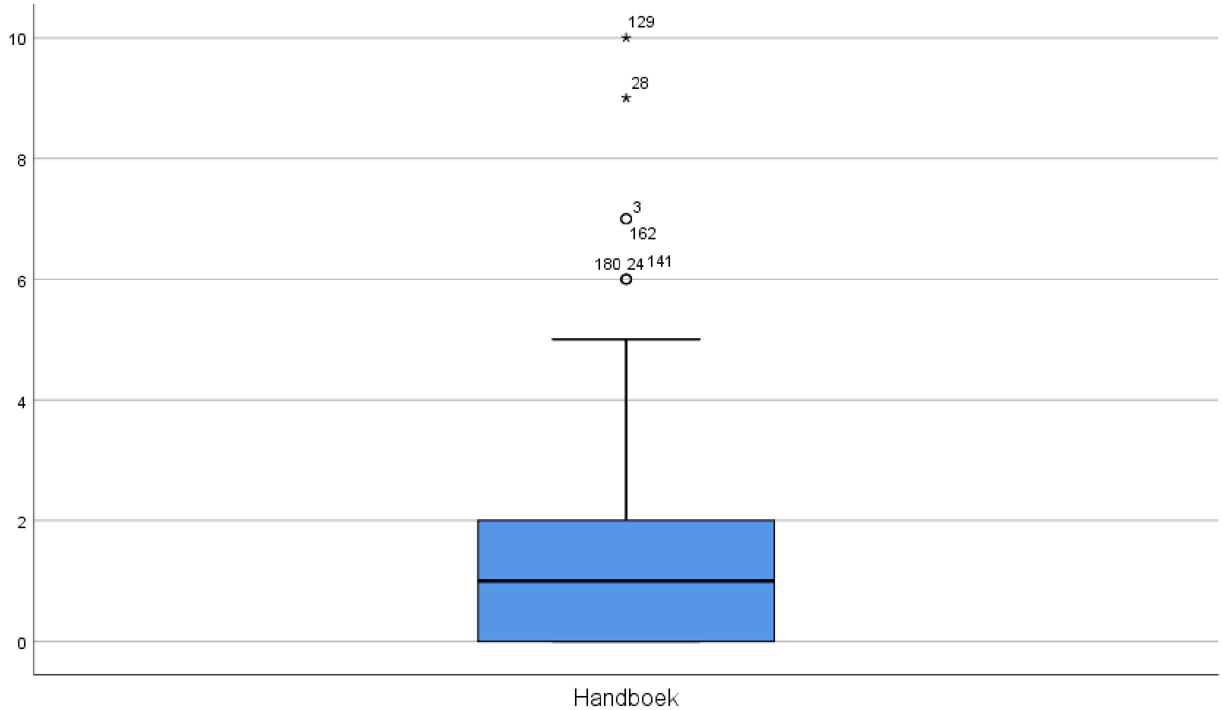

100
80
60
40
20
0
140
33
171
136
185
188
153
178
183
152
Hulpbord

Handbook:

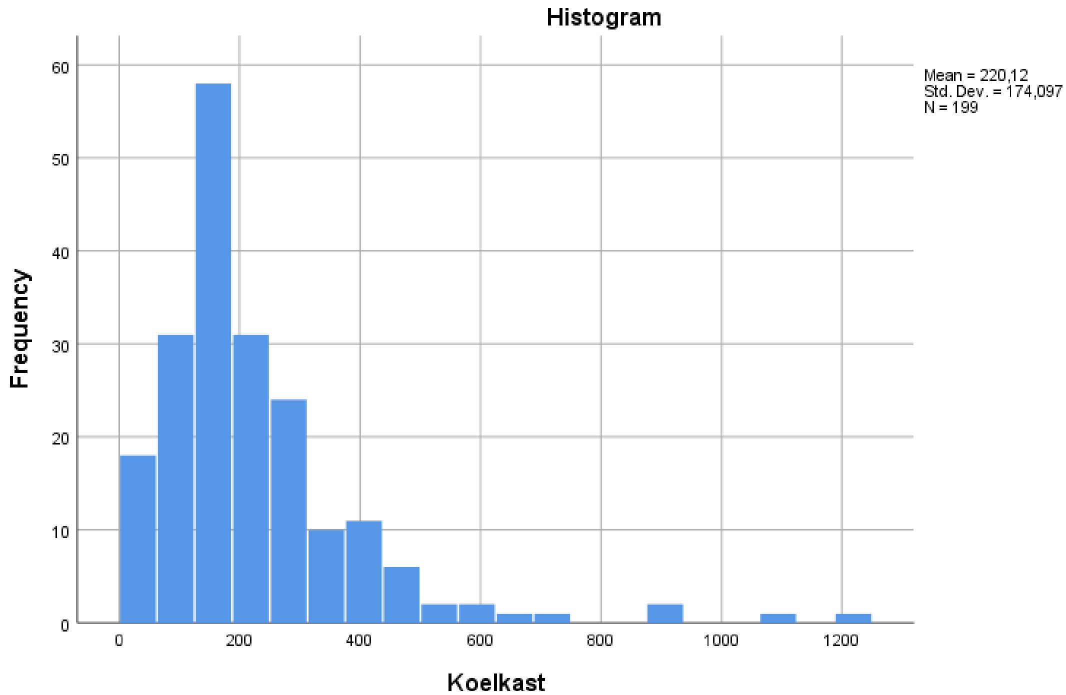

Histogram
Mean = 1,29
Std. Dev. = 1,733
N = 199
Frequency
100
80
60
40
20
0
0
2
4
6
8
10
Handboek

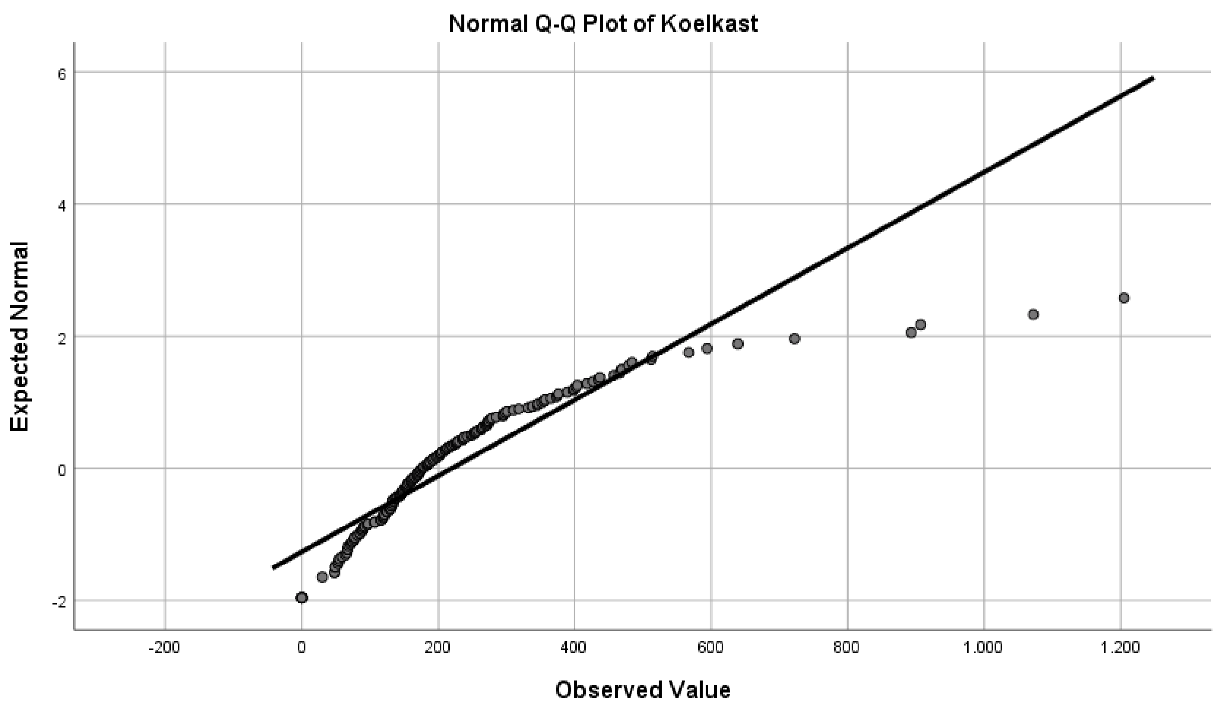

Normal Q-Q Plot of Handboek
Expected Normal
Observed Value
6
4
2
0
0
3
6
9
12

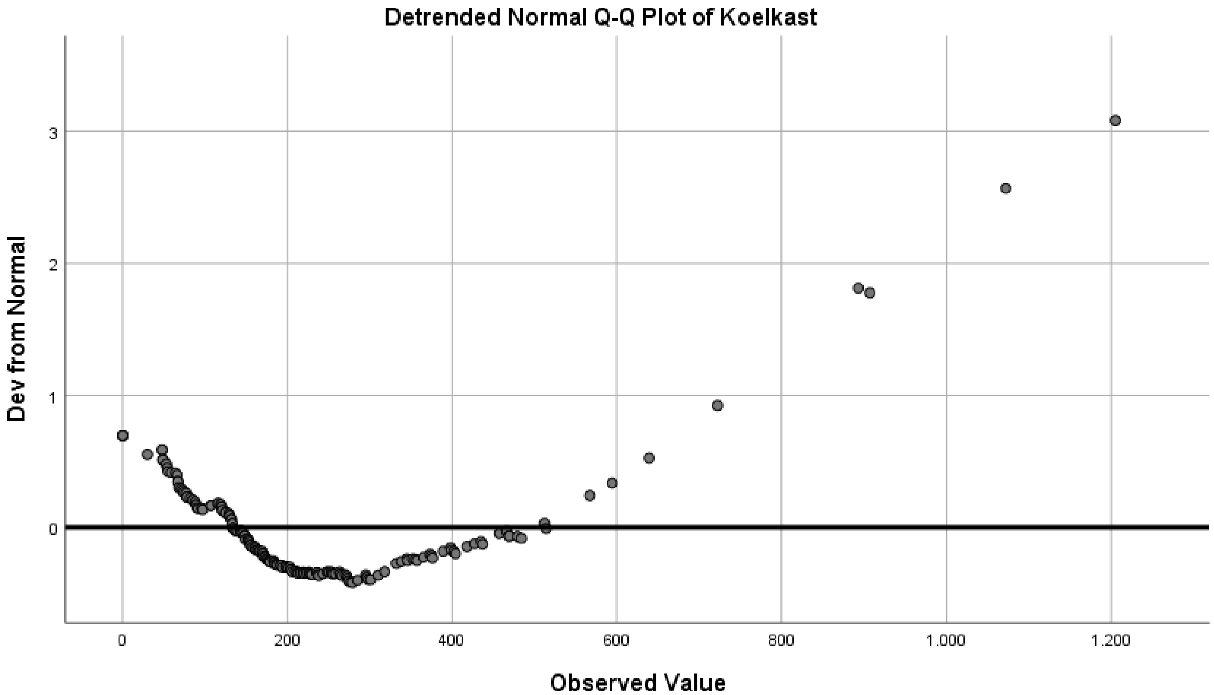

Detrended Normal Q-Q Plot of Handboek
Dev from Normal
Observed Value
2,5
2,0
1,5
1,0
0,5
0,0
-0,5
0
2
4
6
8
10

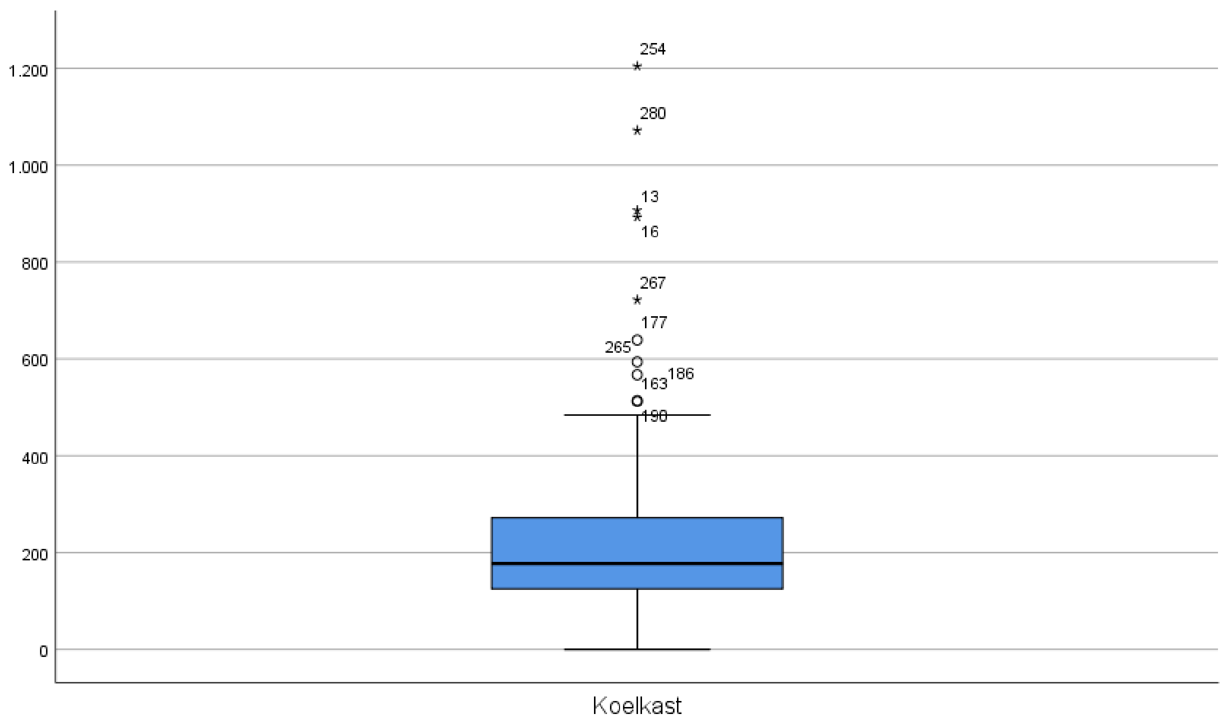


Refrigerator game – level 1:

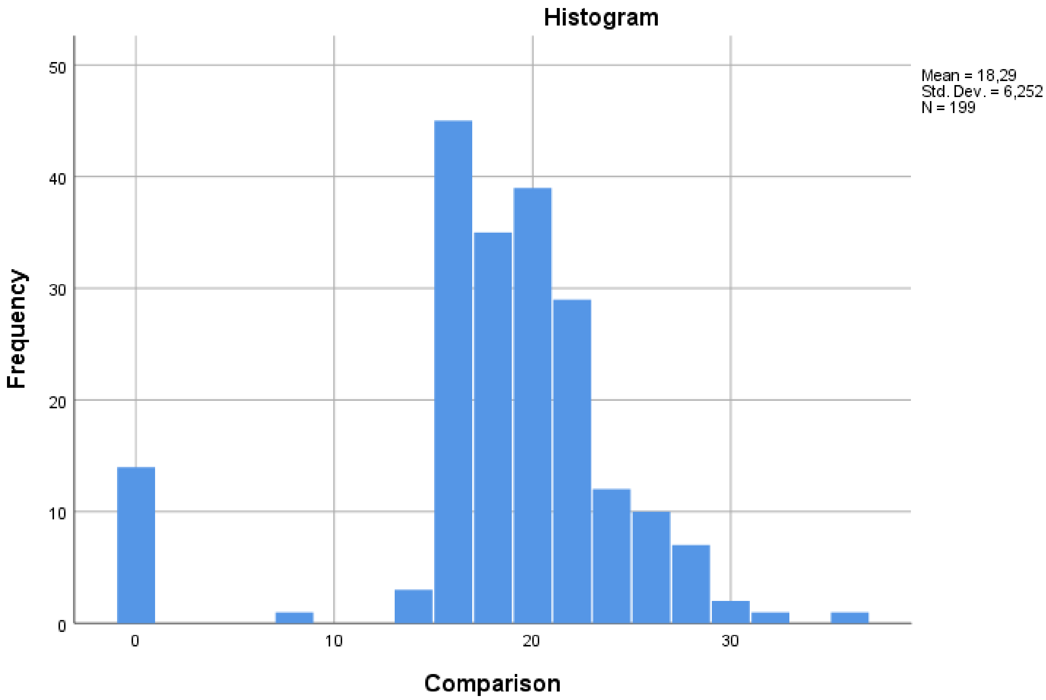

Histogram
Mean = 220,12
Std. Dev. = 174,097
N = 199
Frequency
60
50
40
30
20
10
0
0
200
400
600
800
1000
1200
Koelkast

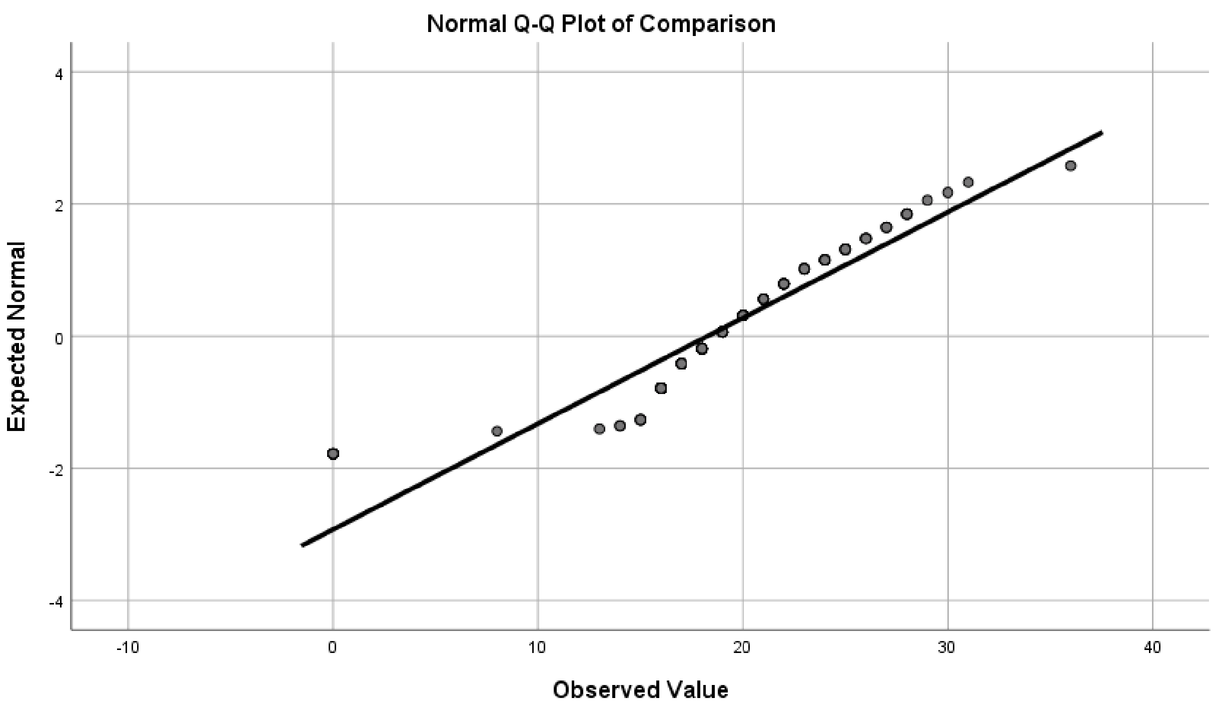

Normal Q-Q Plot of Koelkast
Expected Normal
6
4
2
0
-2
-200
0
200
400
600
800
1.000
1.200
Observed Value

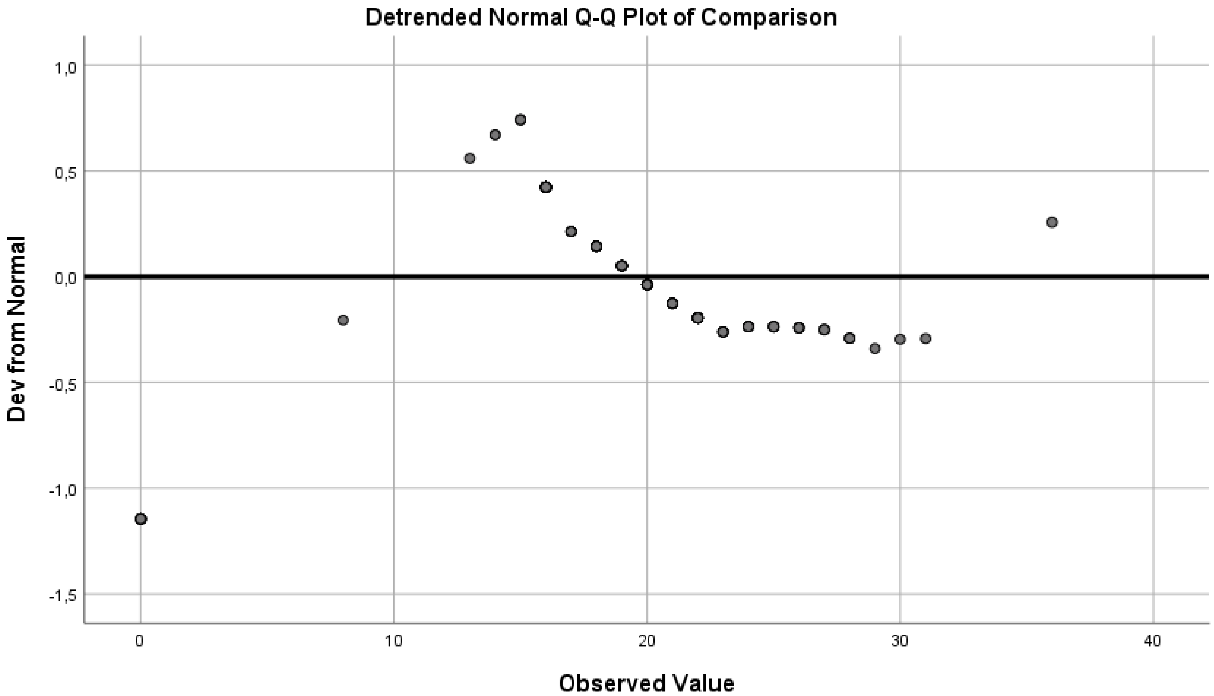


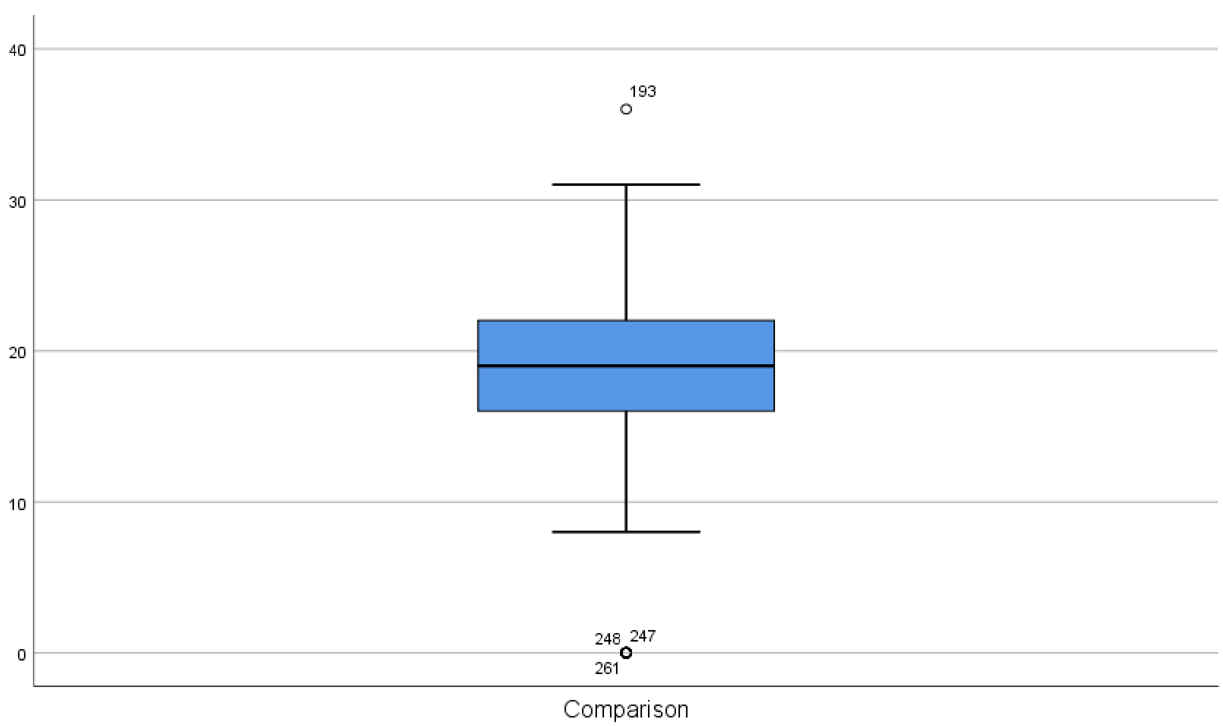


Comparison game – Level 1

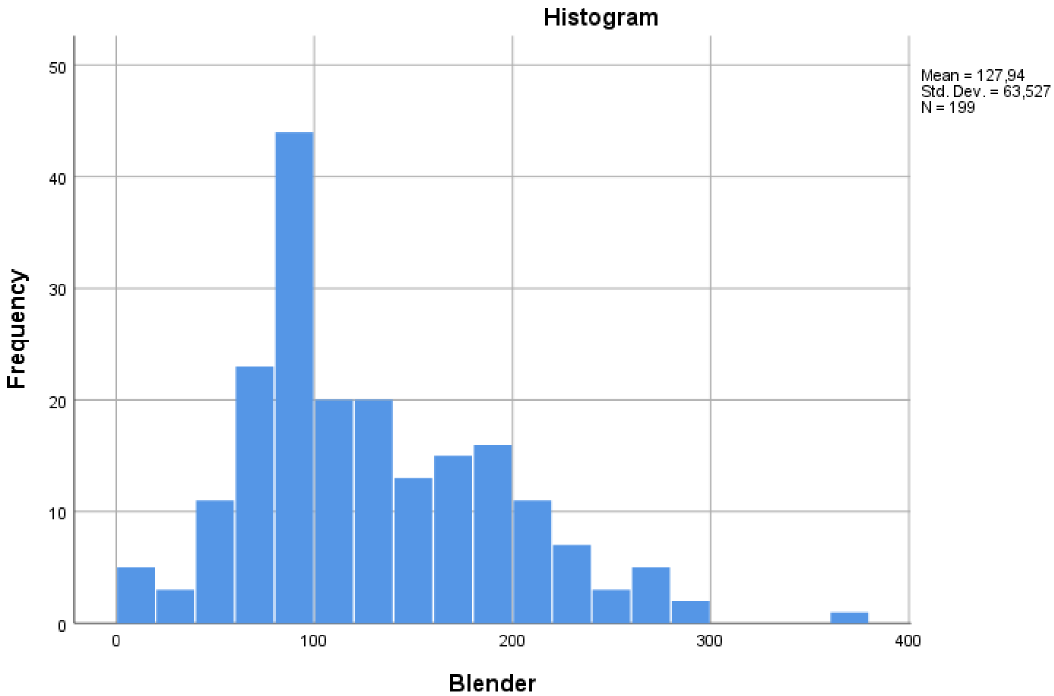

Histogram
Mean = 18,29
Std. Dev. = 6,252
N = 199
Frequency
0
10
20
30
40
50
0
10
20
30
Comparison


Normal Q-Q Plot of Comparison
Expected Normal
4
2
0
-2
-4
-10
0
10
20
30
40
Observed Value

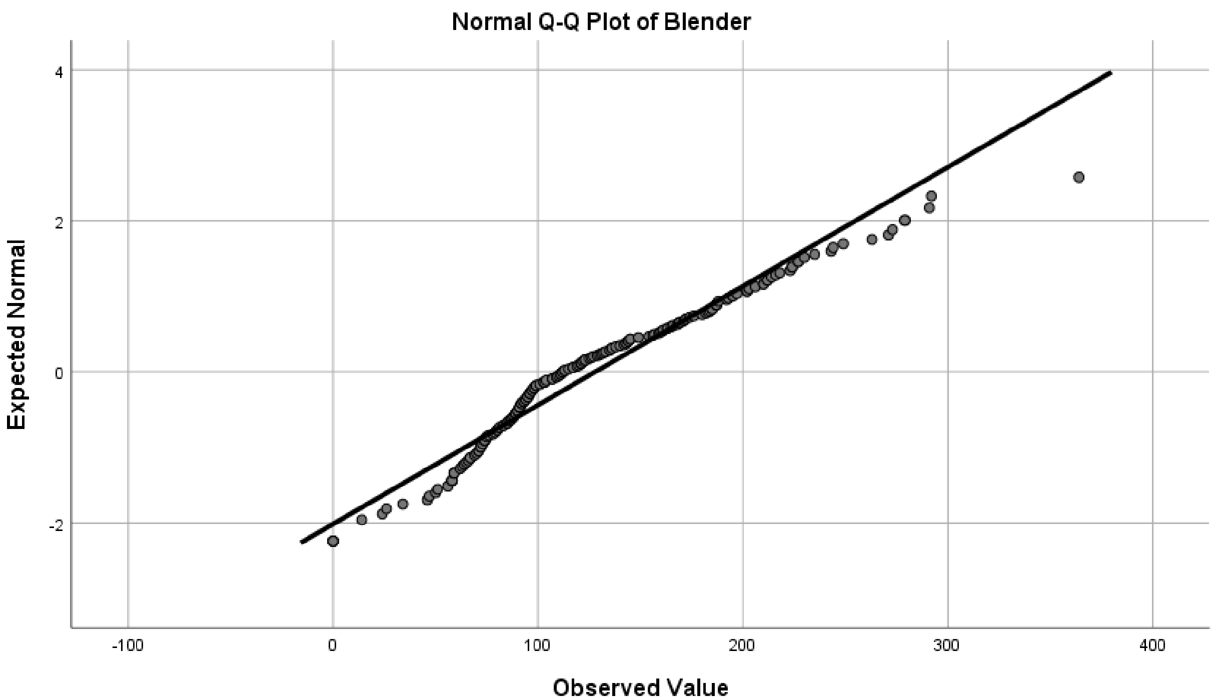




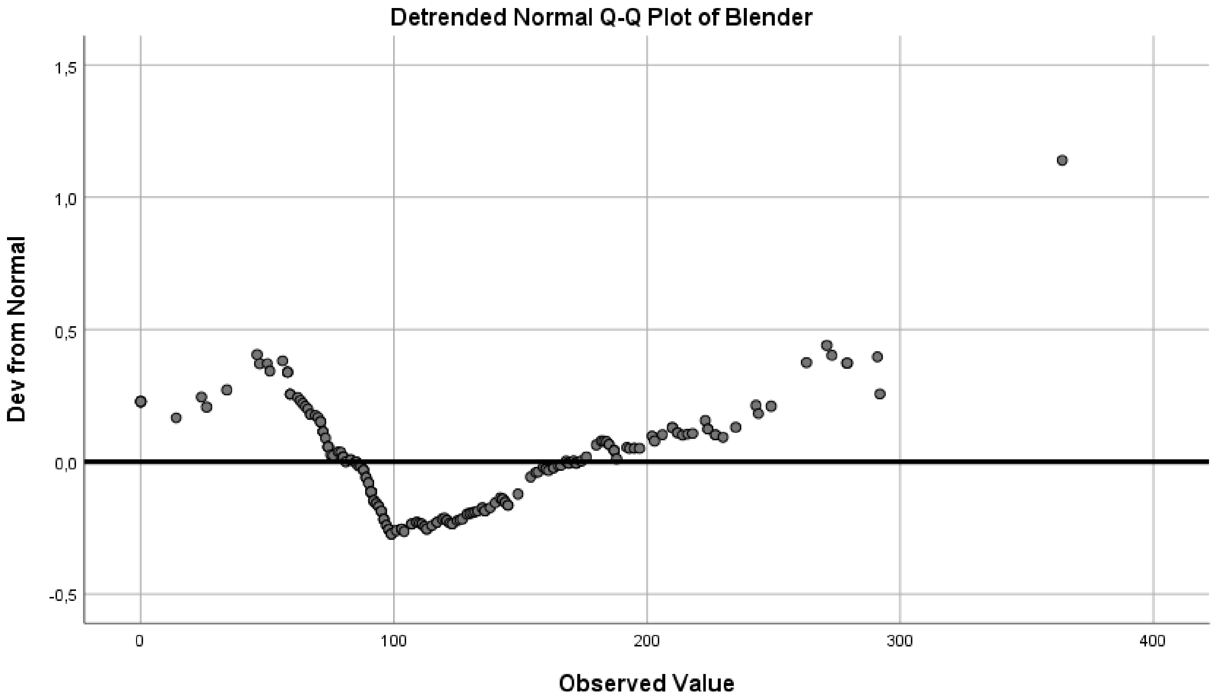
Blender game – Level 1

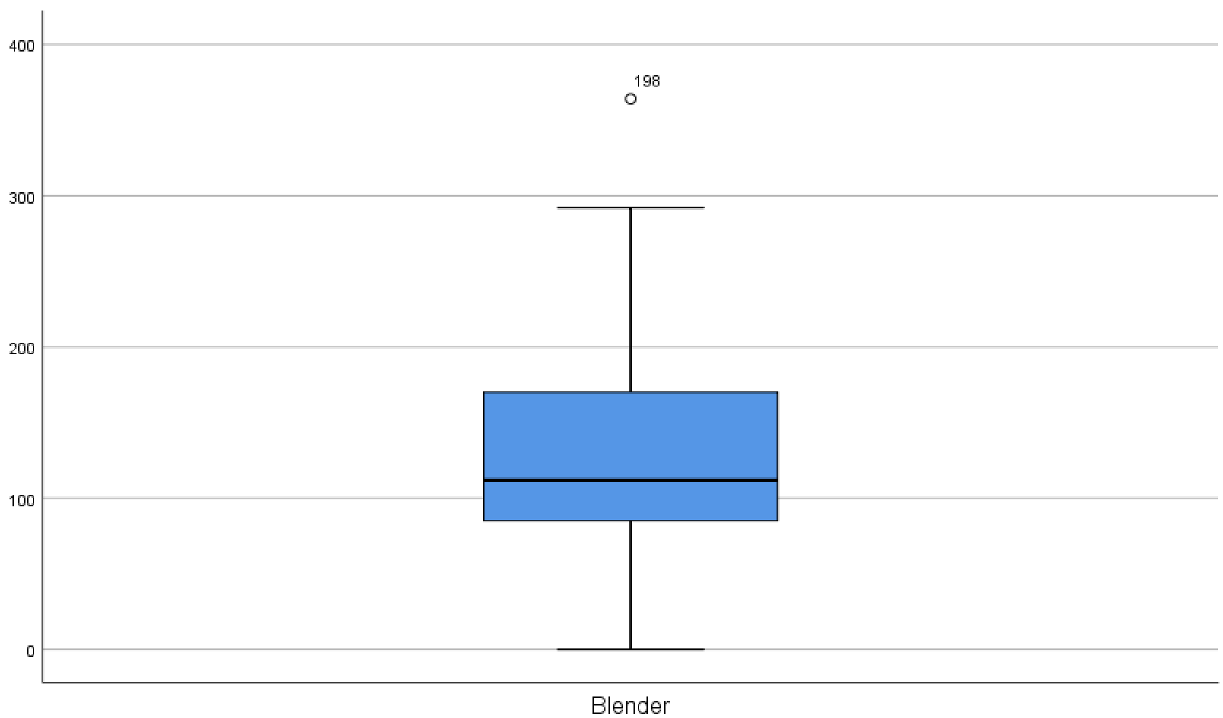


Histogram
Mean = 127,94
Std. Dev. = 63,527
N = 199
Frequency
Blender
0
10
20
30
40
50
0
100
200
300
400



Normal Q-Q Plot of Blender
Expected Normal
Observed Value
4
2
0
-2
-100
0
100
200
300
400





Appendix 15: Multiple Regression Analysis with Mahalanobis outliers

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	Blender, Calculator, Koelkast, Handboek, Comparison, Hulpbord, Gamepercentage[b]	.	Enter

a. Dependent Variable: Posttest

b. All requested variables entered.

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,485[a]	,235	,205	3,369

a. Predictors: (Constant), Blender, Calculator, Koelkast, Handboek, Comparison, Hulpbord, Gamepercentage

b. Dependent Variable: Posttest

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	616,931	7	88,133	7,767	,000[b]
	Residual	2008,528	177	11,348		
	Total	2625,459	184			

a. Dependent Variable: Posttest

b. Predictors: (Constant), Blender, Calculator, Koelkast, Handboek, Comparison, Hulpbord, Gamepercentage

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients			95,0% Confidence Interval for B		Corr	
		B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound	Zero-order	F
1	(Constant)	4,311	1,406		3,067	,002	1,538	7,085		
	Gamepercentage	,069	,014	,380	4,921	,000	,041	,096	,442	
	Calculator	-,024	,021	-,076	-1,148	,253	-,065	,017	-,114	
	Hulpbord	-,022	,021	-,079	-1,062	,289	-,064	,019	-,123	
	Handboek	-,214	,146	-,100	-1,459	,146	-,502	,075	-,161	
	Koelkast	-,002	,002	-,090	-1,246	,214	-,005	,001	-,234	
	Comparison	,000	,044	,000	,003	,997	-,086	,086	-,073	
	Blender	-,003	,005	-,052	-,672	,502	-,012	,006	-,220	

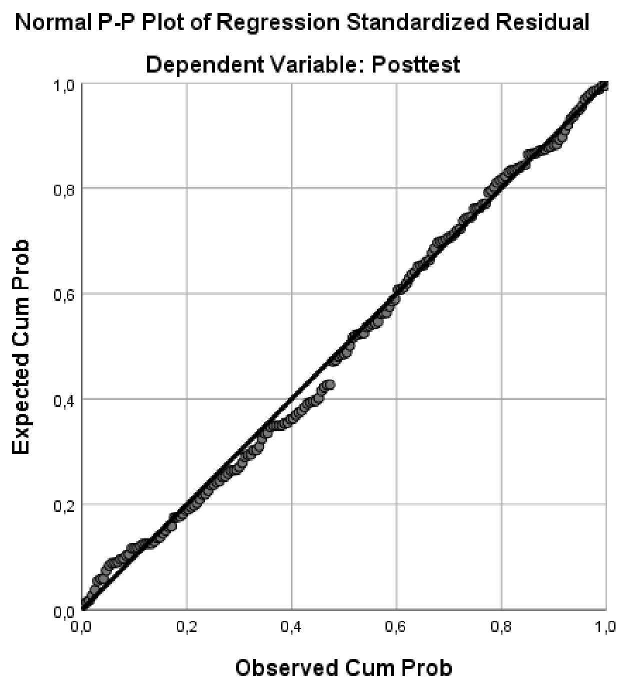
a. Dependent Variable: Posttest

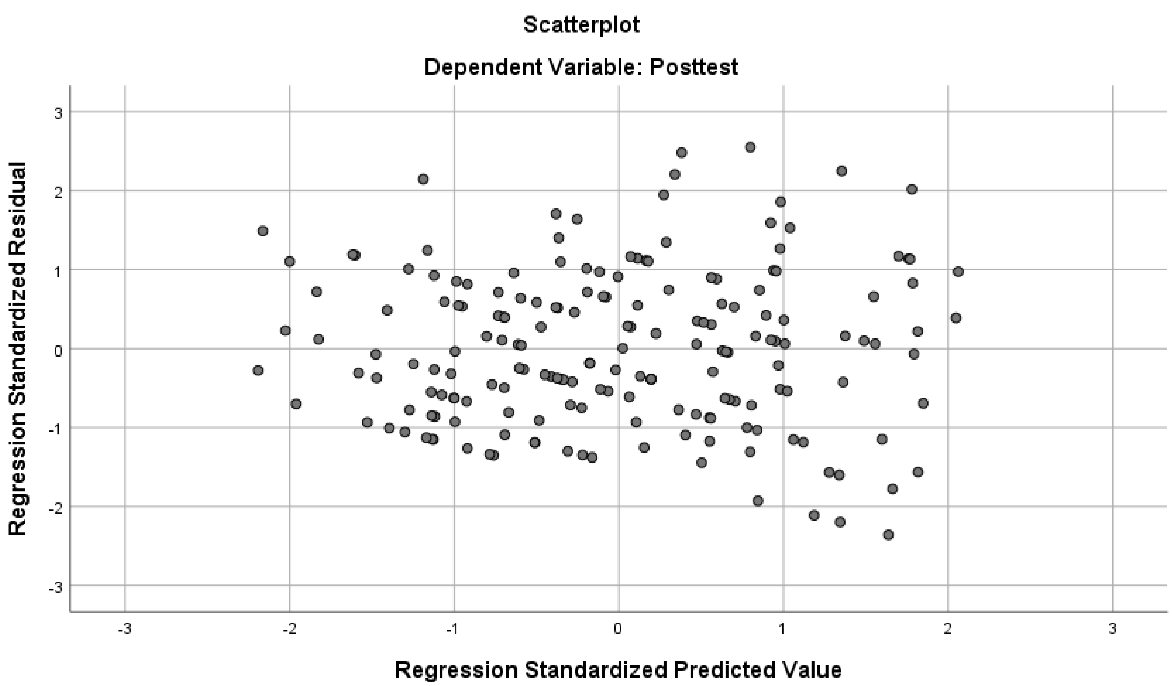
Residuals Statistics[a]

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	1,93	9,72	5,95	1,831	185
Std. Predicted Value	-2,191	2,062	,000	1,000	185
Standard Error of Predicted Value	,343	1,639	,653	,254	185
Adjusted Predicted Value	,89	9,63	5,94	1,862	185
Residual	-7,945	8,592	,000	3,304	185
Std. Residual	-2,358	2,551	,000	,981	185
Stud. Residual	-2,396	2,578	,001	1,001	185
Deleted Residual	-8,203	8,776	,008	3,448	185
Stud. Deleted Residual	-2,429	2,620	,002	1,006	185
Mahal. Distance	,914	42,568	6,962	7,215	185
Cook's Distance	,000	,074	,006	,010	185
Centered Leverage Value	,005	,231	,038	,039	185

a. Dependent Variable: Posttest

Charts





Appendix 16: Multiple Regression Analysis without Mahalanobis outliers

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	Blender, Calculator, Koelkast, Handboek, Comparison, Gamepercentage, Hulpbord[b]	.	Enter

a. Dependent Variable: Posttest

b. All requested variables entered.

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,497[a]	,247	,214	3,368

a. Predictors: (Constant), Blender, Calculator, Koelkast, Handboek, Comparison, Gamepercentage, Hulpbord

b. Dependent Variable: Posttest

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	599,311	7	85,616	7,550	,000[b]
	Residual	1825,826	161	11,341		
	Total	2425,136	168			

a. Dependent Variable: Posttest

b. Predictors: (Constant), Blender, Calculator, Koelkast, Handboek, Comparison, Gamepercentage, Hulpbord

Coefficients[a]

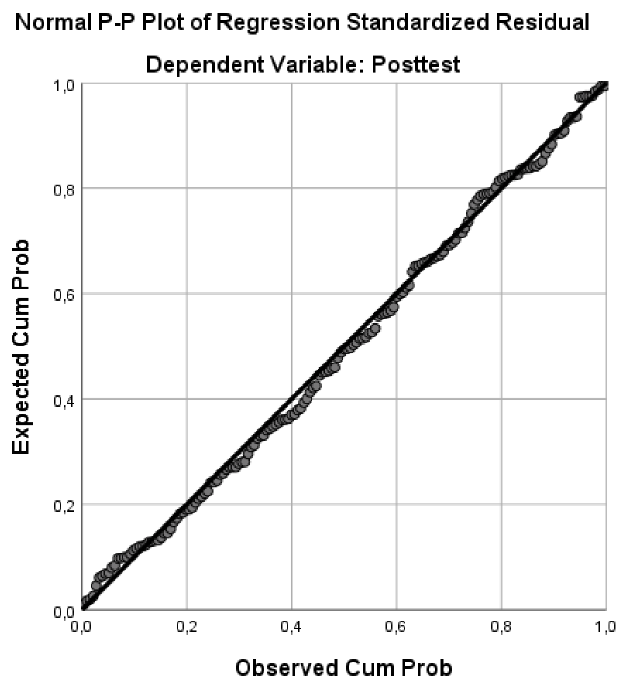
Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	95,0% Confidence Interval for B Lower Bound	95,0% Confidence Interval for B Upper Bound	Correlatio Zero-order	Correlatio Partial
1	(Constant)	5,865	1,583		3,705	,000	2,739	8,992		
	Gamepercentage	,062	,015	,340	4,224	,000	,033	,090	,440	,31
	Calculator	,012	,033	,025	,354	,724	-,054	,077	-,043	,02
	Hulpbord	-,013	,025	-,043	-,535	,594	-,063	,036	-,099	-,04
	Handboek	-,232	,185	-,093	-1,254	,212	-,598	,133	-,170	-,09
	Koelkast	-,006	,002	-,192	-2,626	,009	-,011	-,002	-,284	-,20
	Comparison	-,021	,051	-,030	-,402	,688	-,121	,080	-,140	-,03
	Blender	-,005	,005	-,082	-,987	,325	-,016	,005	-,240	-,07

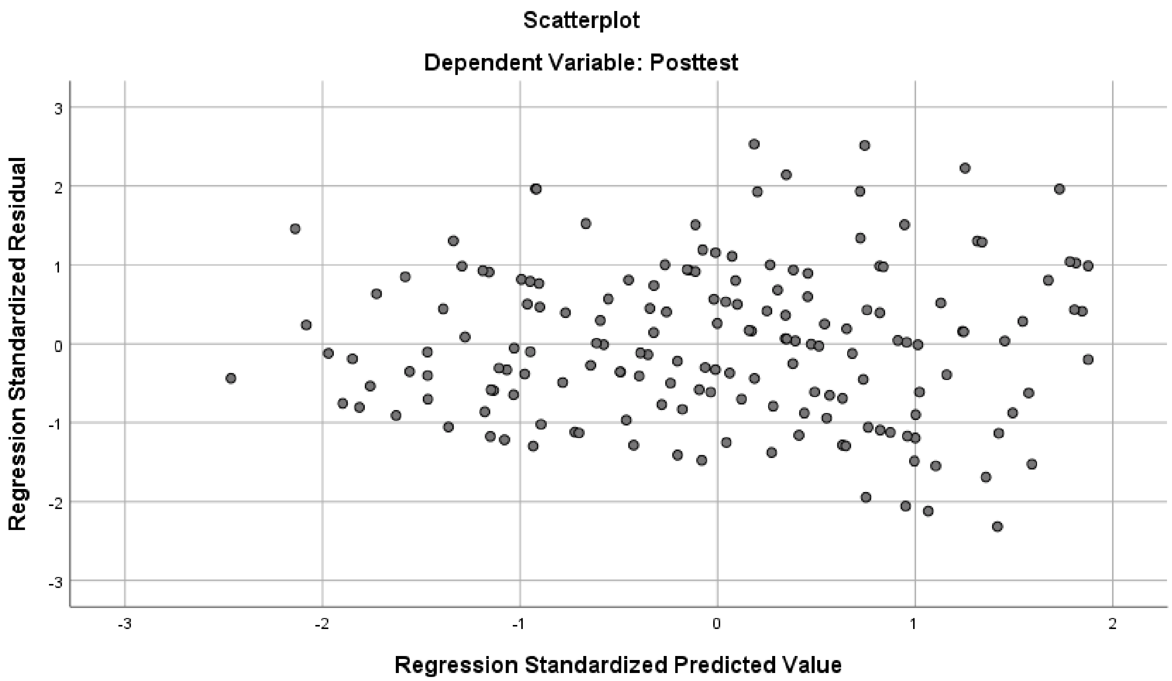
a. Dependent Variable: Posttest

Residuals Statistics[a]

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	1,48	9,67	6,13	1,889	169
Std. Predicted Value	-2,464	1,876	,000	1,000	169
Standard Error of Predicted Value	,374	1,257	,702	,210	169
Adjusted Predicted Value	1,58	9,74	6,13	1,902	169
Residual	-7,805	8,521	,000	3,297	169
Std. Residual	-2,318	2,530	,000	,979	169
Stud. Residual	-2,365	2,549	,000	1,001	169
Deleted Residual	-8,129	8,678	,000	3,449	169
Stud. Deleted Residual	-2,400	2,594	,001	1,007	169
Mahal. Distance	1,083	22,398	6,959	4,806	169
Cook's Distance	,000	,070	,006	,009	169
Centered Leverage Value	,006	,133	,041	,029	169

a. Dependent Variable: Posttest





Appendix 17: Power Analysis Multiple Regression Analysis without Mahalanobis outliers

F tests - Linear multiple regression: Fixed model, R^2 deviation from zero

Analysis: Post hoc: Compute achieved power

Input: Effect size f² = 0.33

α err prob = 0.05

Total sample size = 168

Number of predictors = 7

Output: Noncentrality parameter λ = 55.4400000

Critical F = 2.0672372

Numerator df = 7

Denominator df = 160

Power (1-β err prob) = 0.9999794

Appendix 18: Comments detailing actions in the game

1001	pws	3	1	2	676	03/26/201	1	0	juice,5unitsperpour	
1001	pws	3	1	1	677	03/26/201	1	0	juice,5unitsperpour	
1001	pws	3	1	10	679	03/26/201	1	0	juice,5unitsperpour	
1001	pws	3	1	2	680	03/26/201	1	0	juice,5unitsperpour	
1001	pws	3	1	1	680	03/26/201	1	0	juice,5unitsperpour	
1001	pws	3	1	2	681	03/26/201	1	0	juice,5unitsperpour	
1001	pws	3	1	1	684	03/26/201	1	0	juice,1unitsperpour	
1001	pws	3	1	10	685	03/26/201	1	0	juice,1unitsperpour	
1001	pws	3	1	2	686	03/26/201	1	0	juice,1unitsperpour	
1001	pws	3	1	4	688	03/26/201	1	0		
1001	pws	3	1	11	762	03/26/201	2	0	18+ = 18	
1001	pws	3	1	11	767	03/26/201	2	0	18+15 = 33	