Text Mining Childhood Memories

Master thesis by

Lukas Aldag

Submitted in the fulfillment of the requirements for the degree of
Master of Science

Faculty of Behavioural Management and Social Sciences

Human Factors & Engineering Psychology

University of Twente, Enschede, the Netherlands

This Research was supervised by Dr. R. H. J. van der Lubbe and

Prof. Dr. Ir. B. P. Veldkamp

August, 2018

Lukas Aldag

Max-Habermann Straße 1

76189 Karlsruhe

lukas.a94@web.de

Student Number: 1566474

Table of Contents

Abstract

Qualitative research takes a lot of time and resources, but inventions such as text mining programs might be an enrichment for qualitative research, as well as more objective. The primary focus of this study was to analyze described first childhood memories with the text mining program LIWC with the Dutch dictionary versions of 2007 and 2015. The goal was to analyze the validity of the Dutch dictionaries, as well as analyzing their differences. Childhood memories of 513 participants were gathered, which were assessed on their emotional valence by LIWC and five independent participants. The results showed that the assessment of LIWC significantly correlated with the indications of the memory owners, as well as the judgment of the independent participants. That led to the conclusion that LIWC might be indeed a valid tool to analyze emotional valence in memories, and therefore enables analyzes that are much faster and at a larger scale, while also providing a more objective assessment.

Keywords: Text mining, LIWC, Dutch dictionary 2007, 2015, childhood memories

Introduction

Whoever conducted qualitative research at least once in his or her life knows how much time it consumes to go through the data and the coding process. The answer for future qualitative research might be text mining. These programs categorize each used word in specific categories, which enable further analysis. Text mining programs might provide an easy and more objective way to analyze a large number of texts, and much faster than any conventional qualitative research. The current research revolved around the text mining program LIWC, which is the short form of Linguistic Inquiry and Word Count. The data that LIWC analyzed were descriptions of 513 first childhood memories. Everybody knows the feeling of suddenly remembering something that was mistaken to be forgotten. Such a feeling can get triggered by a smell, taste, or a place that seems to be somewhat similar to the location of the memory. Old memories come with a load of sensations and can either be related to a pleasant or bad experience, but who tells us if this memory is what we believe it to be? For decades memories had been studied, and yet it still holds many mysteries which have to be resolved. For instance, Reyna and Lloyd (1997) studied false memories in children and adults by trying to understand the processes behind recalling wrong memories and what makes us believe in them, as several sources can influence memories. Many different factors as beliefs, experience, and mood, can distort memories. Mostly human judges have been used to conduct qualitative research. Therefore it will always be biased by a somewhat subjective judgment. The goal of the current study was to use LIWC as a more objective and stable way of assessing first childhood memories to overcome the problem of subjective assessments. Before LIWC can be used to analyze

childhood memories, it was necessary to examine its validity in doing so. The primary focus of the study was to investigate LIWC's validity in assessing first childhood memories on their emotional valence, in comparison to a subjective assessment done by independent participants. As two different versions of LIWC were used, it was also interesting to analyze the differences between those two versions.

LIWC

LIWC or Linguistic Inquiry and Word Count is a program that is capable of analyzing texts merely by categorizing each word if possible and was initially designed for research within the social, clinical, health field, as well as cognitive psychology. The program provides how many percentages of the words within a text scored within a particular category (Pennebaker, Boyd, Jordan, & Blackburn, 2015). These categories are, for example, emotions, thinking styles, social concerns, language metrics, grammar, cognitive processes, perpetual processes, and many more. The number of categories was relatively stable with each consecutive dictionary, beginning with 68 in 2001, 64 categories in 2007, and 76 in 2015 (Van Wissen, & Boot, 2017). A few categories disappeared in the newer versions. Instead, new categories were created. For instance, the English version of 2015 included summary variables, analytical thinking, clout, authenticity, and emotional tone. Multiple categories compute these summary variables. Unfortunately, these summary variables were kept secret by the owner company and therefore are not available in the Dutch version (Pennebaker et al., 2015). Cohn, Mehl, and Pennebaker (2004) used a comparable summary variable in one of their study. In the study, Cohn et al. calculated a variable called emotional positivity, which was merely the difference between the positive emotion and negative emotion category. This variable indicates how positive/negative a

text was. Higher scores reported higher emotionally positivity, and therefore a lower score indicated less positivity. The emotional positivity was also used during the current research but was called emotional valence instead.

LIWC's text analysis module works with so-called dictionaries, which also can be created by the user. The Dutch dictionary from 2007 includes 4487 words (Van Wissen & Boot, 2017). The newest dictionary is from 2015 and is composed of nearly 6400 words, including the word stems, and also some emoticons. This feature enables the program to analyze text messages from any modern online messenger that uses any kinds of emoticons, as proven by Settanni and Marengo (2015). Every word within the dictionary links to categories that are most common for the word, which enables LIWC to categorize the text.

“Cried” will be used as an example to explain the way how LIWC works. Cried is represented within five different categories, which are: sadness, negative emotion, the overall effect, verb and past focus. This instance is showing the power of LIWC, with this program it is possible to categorize the texts for emotions on a surficial way with positive and negative emotions, but is also able to determine more dimensions like sadness. Another focus might be the usage of words, for example, what kind of words get mostly used within the text? Due to the usage of whole word stems, LIWC can determine the usage of tense. All these categories enable broad analyses and all different kinds of approaches, showing the extensive application of it. LIWC found its use in this study to provide a more objective assessment of the memories on their emotional valence.

Studies already existed in which LIWC was used to analyze the emotional tone of specific texts. Tausczik and Pennebaker (2010) wrote about LIWC and validation was done in

the first place. It also showed that the program seems to be reasonably valid in measuring emotions, at least the English version. LIWC found application in various contexts, such as short diary entries and well being (Tov, Ng, Lin, & Qui, 2013). Tov et al. showed that LIWC is capable of assessing negative/positive emotions which came over with self-reports. Bantum and Owen (2009) used LIWC to analyze cancer narratives on emotional expressions. They compared LIWC with another text mining program called PCAD (Psychiatric Content Analysis and Diagnosis) and independent raters. The outcome was in favor of LIWC, as it showed a better discriminative validity of the emotions in comparison with PCAD. In other words, LIWC was better in categorizing the words into the right emotion categories. Beside the application in such experiments, it was also used to analyze Facebook posts (Settanni, & Marengo, 2015). It showed a correlation between online emotional expression and self-reported emotional well being, especially in younger age groups. LIWC was also used to analyze memories, and if people with specific brain damage or psychological problems, such as anorexia, could be found with LIWC (St-Laurent, Moscovitch, Jadd, & McAndrews, 2014; Brockmeyer, Holtforth, Bents, Herzog, & Friederich, 2013). Almost all of these studies were done with the dictionary version from 2007, that LIWC uses to analyze the texts. The current study focused on explicitly first memories and their emotional valence. These memories got investigated with LIWC and the two newest dictionaries of 2007 and 2015. The reason to use both versions was the fact that the version from 2015 was still experimental (Van Wissen, and Boot, 2017). The 2015 version got created via an auto-translation program. The dictionary was checked by human judges to ensure validity if every word linked with the right categories. As this way of translating a LIWC dictionary was never done before, and necessary studies were still missing, it became more interesting to test if

the newer version was more precise than its forerunner. The dictionary version of 2007, on the other hand, was entirely translated by hand (Boot, Zijlstra, & Geenen, 2017). The words got translated, which then were approved by two further human judges. The problem with translating is, that it is not merely a process of pure translation, as words can have different meanings in another language. For instance, the word ankle is pretty clear in English, and LIWC would categorize it in the category body part. Looking at the Dutch word, however, "enkel" could also be translated as “just” or “only”, and has to be categorized as an adverb. Such a word is called a homonymy and is the most common problem. The problem is somewhat fixed finding the most used sense and categorize it as such. This approach can cause problems in certain situations, in which the writer meant another sense. Another considerable problem is the bag-of-words problem. LIWC is not looking for a connection between multiple words and therefore is not able to detect a denial. For instance, the sentence “I am happy”, happy gets categorized as a positive emotion. The sentence “I am not happy” however, will also be categorized as emotional positive. LIWC is also not able to detect irony, as it is not able to make any sense of the words, which were categorized. Beside all this problems, the accuracy of LIWC has been proven to be high enough (Boot et al. 2017; Pennebaker et al., 2015; Van Wissen et al., 2017).

Memories

This study focused on first childhood memories. Even when recalling memories, there is no guarantee that these remembered memories represent a stimulus or event that took place the way it was thought to be. Memories store within the long-term memory (LTM). The information of the LTM can be retrieved consciously. The LTM gets classified into the declarative (explicit) part and the nondeclarative (implicit) LTM (Smith, & Kosslyn, 2007).

Explicit memories classify into two categories, episodic and semantic memories (Tulving, 1972). An episodic memory contains specific and personal events of the past, whereas the semantic memory represents the general knowledge one has about the world and their meaning. The significant difference between those two memories is lack of context in the semantic memory. Episodic memories contain the information itself but also more specific context about the information. Semantic memories miss the specific context, and these memories are an accumulation of information that builds one greater unified memory. The retrieval and encoding of new episodic and semantic memories depend on a specific set of brain structures in the temporal lobes (Corkin, 1984; Scoville and Milner, 1957). Anyhow, the dependency is not always existing, as with time, the episodic memories get stored outside of the medial temporal lobes (Squire, Stark, & Clark, 2004).

Episodic memories get created in three stages, encoding, consolidation, and retrieval. Encoding is a term for various processes that take place to transform information into a memory representation. The encoding process gets influenced by different factors, attention, elaboration, and conscious retrieval of the information (Smith, & Kosslyn, 2007). In other words, it is necessary to attend towards the stored stimulus or information (Craik et al., 1996). The second-factor elaboration is a process during which the interpretation of information takes place and gets connected to already existing information. This process makes information meaningful, and conscious elaboration consolidates the information. The third factor is retrieval and contains the conscious retrieval of information at a later stage, which further consolidates the information and facilitates its retrieval. Via pattern completion is the retrieval achieved, one or multiple cues trigger the cognitive pattern that was available during the encoding (McClelland, 1995;

Nakazawa et al., 2002). These processes are all partly connected to the frontal lobes, which acts as a convergence zone for all the incoming information (Lavenex, & Amaral, 2000; Suzuki, & Amaral, 1994). In other words, an episodic memory consists of a conjunction of linked features, that in their entirety build the memory.

The retrieved representations are not identical copies of the information that was present during the encoding. The retrieval of memories is an interaction between the medial temporal lobes and other cortical regions. Memories are recallable by using a retrieval plan. Such a plan is developed in the frontal lobes and contains a selection and representation of the cues that will be used to probe the specific memory. The cues that help the pattern activation can be external as well as internal cues. Many events that seem forgotten are not per se forgotten, as they are still stored somewhere in the brain, but they are not recallable due to missing cues to probe the memory effectively. If a memory is not complete, it is possible to trigger further cues by matching the context or state of the memory. A context/state-dependent effect exist, which refers to better retrieval when the physical environment/internal state at retrieval matches that at encoding (Eich, Weingartner, Stillman, & Gillin, 1975). This effect also explains the experience of memories that seem long forgotten, but suddenly come out of nowhere, because the environment and/or, an emotional state triggers a specific pattern completion.

As pleasant as it is to reminisce, the memories are probably flawed. Schacter (1999) categorized multiple flaws of our memories, bias, misattribution, and suggestion, consistency bias (Wilson and Ross, 1989), and the misinformation effect (Loftus, 2005). Every human being is biased, the cultural experiences and background knowledge influence not only the present actions one takes but also the memories of specific actions. One of the most critical flaws in

childhood memories is created through suggestions, the recall of false information during the recall implanted information by others. This effect is called the misinformation effect, misremembering of an event in line with false information (Loftus, 2005). If memories cannot be recalled correctly or with certainty, people tend to accept information, even if they are wrong, given to them to fully recreate the memory. According to this, childhood memory in specific might be extremely flawed by suggestions, as most people asked their parents or sisters once about an event that might have happened to them. Therefore, it is reasonable to say that they probably do not reflect the real past. Thus it is necessary to analyze these memories in the most objective way possible to prevent any further distortion. A text mining program may achieve this.

Early childhood memories

Back in 1895, Miles was one of the first researchers to study early childhood memories. She asked one hundred students in a questionnaire to write down their earliest memories that they were able to recall. The overarching conclusion of her study was that emotions play an important role in childhood memories. Many other researchers started to analyze first/early childhood memories and also concluded that emotions were an essential characteristic of these memories. Nevertheless, the researchers could not conclude which kind of emotions were more important and found different results. Dudycha and Dudycha (1933), Cowan and Davidson (1984), Howes, Siegel, and Brown (1993), and Mullen (1994) found negative emotions to be more prominent within early memories and therefore claimed them to be more critical. In contrast to the studies mentioned above, Waldfogel (1948), Kihlstrom and Harackiewicz (1982), and Saunder and Norcross (1988) found more memories with positive emotions and therefore

concluded that they were more critical. All these studies have a comparable approach and in either the participants or the researchers assessed the memories. This approach might be a reason for the different findings. Therefore a more objective way could lead to a more precise outcome, which could approve to be more stable over multiple studies. The necessary objectivity might be achievable by using a text mining program. Many studies also came up with a different age of onset, but overall the conclusion was, first memories were made during the third and fourth year (Draaisma, 2005, Howes et al., 1993, Jack & Hayne, 2007, Mullen, 1994, Peterson, Grant & Boland, 2005, Tustin & Hayne, 2010).

A possible explanation for the different outcomes might be the way they gathered their data. All of them asked for a written description of early memories, but a few were asking for 'the' earliest memory, while others asked for many earlier memories. Most of the studies scored the emotions similarly. They were asking the participants to describe an emotion to their memory, by choosing between positive/pleasant, negative/unpleasant, and neutral/no emotion (Howes et al., 1993, Kihlstrom et al., 1982, Mullen, 1994, Saunders et al., 1988, Waldfogel, 1948). Another difference was the way they asked the participants to state their earliest memories. Dudycha et al. (1933) for instance, asked for memories the participants were entirely sure about, that they remember them correctly, instead of an event that they got told. This approach restricted the repertoire of first memories because participant will less tend to pick a memory that they do not remember that well, even though it might be their potential first memory. The memory gathering was also done with certain restrictions, as it has to be sure that the memory was real and not a told event.

The overall goal of the current study was the attempt of using LIWC as an analysis tool for first memories on their emotional valence. LIWC might provide a more efficient and objective analysis in comparison to conventional qualitative analyses. Therefore it was necessary to explore its validity in assessing descriptions of first childhood memories, which also was the central question of the current study. As the collected memories were all written in Dutch, the question arose, if a translated LIWC dictionary can assess the emotional valence in childhood memories. This study not only focused on LIWC and how valid it is in evaluating the emotional valence in descriptive first childhood memories but also on how valid the the two different Dutch dictionaries were. Too few studies have been conducted so far, that focus not only on translated dictionaries but also on the comparison of two different versions. An older version of LIWC might prove to be more useful in specific scenarios, as with more categories and the change of them, might also have changed the sensitivity for particular categories. This study might provide essential outcomes that could help to establish the usage of LIWC Dutch dictionaries to study memories, while it also delivered information about the differences between the dictionary version of 2007 and 2015.

The versions of LIWC's assessment of the memories descriptions got validated by comparing them with the evaluation of independent participants and the actual owner of the memories. The data gathering took place at a Dutch university, and all participants were psychology students. Each participant wrote down the earliest memory he or she could remember, as well as assessing the emotional valence of the memory. Each memory description also got evaluated by a participant, which served as a subjective evaluation. A small subgoal of

this research was to provide further data, for the question if people instead tend to recall positive or negative childhood memories, and towards the onset of the first memory.

Method

Participants

513 participants provided their childhood memories in written form. All of the participants were first-year psychology students from the year 2011 and 2012. The average age was 21.6 years (*SD* = 3.87), ranging from 17 to 54 years. 120 participants were male (23.4%), and 393 were female (76.6%). Five participants enlisted for the second part of the research, three female, and two male. The average age was 23 years (*SD* = 3.25), ranging from 19 to 29 years. All participants of the second part were also students of the University of Twente but were recruited via the test subject pool SONA or approached by the researcher. Each participant gave the consent that their data may be used for research purpose.

Materials

The data of the memories were first written down by the students and later digitalized and merged in one data set. The memory dataset was analyzed with the text mining program LIWC version 1.4.0 from 05.04.2017. The LIWC dictionaries used to examine the data were the Dutch dictionary version of 2007, and 2015. RStudio 1.1.383 served as a tool to investigate the outcomes.

Design

The data consisted of eight different variables that got compared with each other. The first variable consisted of the assessment of the actual owner of the written memory. The second

variable consisted of the evaluations of independent participants. The next three variables, consist of the positive emotion, negative emotion, and emotional valence value, obtained with the Dutch LIWC dictionary from 2007. The last three variables were the same as the ones before with the difference that they were the product of the Dutch dictionary 2015. The outcome from LIWC was correlated with the assessment of the participants as well as the owners, and between the two version of 2007 and 2015.

Data Analysis Plan

The analysis started off with an exploration of the dataset. The important variables for the current research were Posemo 07/15, Negemo 07/15, Emovalence 07/15, and assessment owner/participant. Posemo 07 and 15 included the percentages of words categorized as emotional positive, either for version 2007 and 2015. Therefore the values in this set could reach from 0 to possible 100. The same applied to Negemo 07 and 15, which reflect the number of words categorized as emotional negative. The variable emotional valence 07 and 15 were computed variables in the data set, achieved by subtracting the Negemo variable from the Posemo variable. The Emovalence variable reflected the overall emotional valence of the written memories, therefore indicating if more words were used that got categorized as emotional positive or negative. A value below zero means more negative words in the text, a value above zero, indicates the usage of more positive words. If the value was close to zero or zero, the amount of negative and positive words were balanced, which indicated the usage of uncategorizable words or a comparable amount of negative/positive words. The last two variables were assessment owner and participants. These variables included the assessment of

the emotional valence of the childhood memory. The assessment was done in one of three possible categories, 0 = negative, 1 = neutral, 2 = positive.

The first set of analyses regarded the correlation between the Posemo 07, Negemo 07, with the assessment of the owner/participants. Four Spearman correlations were conducted, as the data showed no normal distribution. Posemo/Negemo 07 were correlated with the evaluation of the owners and the participants. It was expected to have a positive correlation between Posemo 07 and the assessments of the owners/participants, as with each category the amount of positive categorized words should rise, with the highest amount in the positive category as assessed by the owners and participants. The opposite is hypothesized for Negemo 07, as the highest values were expected in the negative category, which is represented by 0. A higher correlation was hypothesized for Neg/posemo 07 towards the owner assessment, due to the more objective way of assessing. The same procedure as for Posemo/Negemo 07 was done for the Posemo/Negemo 15 variables. The same outcomes were expected as for the variables with the 2007 version, with the slight difference that the correlations might be slightly higher. The higher correlation is hypothesized as the newer version is expected to be more precise, and therefore come closer to the assessment of the owner and maybe the participants. The next set of correlations were regarding the differences between the two dictionary versions. Therefore two Spearman correlations were conducted to compare posemo 07 with 15, and Negemo 07 with 15. A relatively high correlation is expected, as the official LIWC website advertises a correlation of 0.96 between these variables and the two different versions. Nevertheless, it is necessary to keep in mind that they tested the English dictionary, and the current research was examining the Dutch

dictionary. Therefore it was possible to expect not the same correlation values, due to the translation, which in the best case should be near the .96.

The next set of correlations included the Emovalence variables. Five Pearson correlations were used to analyze the Emovalence variables, as they showed a normal distribution. Both Emovalence variables 07 and 15 were correlated with the assessment of the owners and participants. For all of these correlations, it was hypothesized that they correlated significantly, but the correlation between Emovalence 15 and owners/participants should be higher in comparison to the 2007 version. As Emovalence 15 got created with the newer dictionary version, it might be expectable that it would score more similar to a human judge, due to refined categories and the ability to recognize more words.

The last analysis was a Chi-squared test, and Cramer's V with the assessment of the owners and the assessment of the participants. The test should show the correlation between these two variables, as well as how close the participants came towards the evaluation of the actual owners of the memories. A significant relationship was expected for these two variables.

Procedure

Memory and emotion recognition gathering

The gathered dataset of memories was a part of the academic training of the psychology students during their first year. Part of the class were multiple workshops in which the students had to work on different ZAPs. ZAPs were little experiments in which the students should experience psychological theories right at hand. The students were also asked to write down their oldest childhood memory. Specific constraints should ensure a real memory. The memory should

not be available in a photo or film, and the student had to be sure that the memory was not a told one. The memory should not exceed a word count of 400. The students had to give further information about the onset of their memory, the current age, and if the memory is related to any sensory modality (e.g., scent, taste, sound), or if the memory if more an episodic-like story (with a beginning and end). The students also had to make a statement about the emotional state of the memory. No predefined answer possibilities were given. Overall the emotional state was defined with one or more words. For instance, the emotional state could have been happy, shock, joy, surprised, and so on. A difference existed between the first and second memory collection. In the second collection, the students only indicated one emotional state. It served as an overall emotional valence of the memory. At last each student decided if their memory could be used for future research.

Data check

Firstly, removing every memory from the data set if the student did not want it to be in any further research. Secondly, as mentioned above, the gathering of the memories differed slightly between the years. While in the first year the students wrote down one or more emotional states, the students of the second year only wrote down one emotional state. The data got prepared for statistical analysis by filtering several contrary emotions. For instance, a student wrote happy and sad. It was not possible to categorize the memory clearly into one of three categories. The current study focused on overall emotional valence which could either be negative, neutral, or positive. Thus every memory with multiple/developing emotional states had to be filtered out. If a student wrote about various positive emotional states, it got a positive rating. The same rule applied to multiple negative words. If the student chose to use the term

surprised in combination with a negative or a positive word, the classification had to be selected depending on the context. If the context was not revealing the overall emotional valence, the memory was not taking into account, as the assessment of the owner should be as untouched as possible. Just one case had to be judged by two individual persons, which led to the same outcome. Every other memory with such a term had to be taken out, due to ambiguity. This filtering process was applied since the two datasets showed slight differences in the assessment of the emotional valence. In order to get one great dataset, it was necessary to adjust one of them. The only possibility was the selection of memories with the statement of one emotional direction.

Memory assessment gathering

The second part of the research revolved around the assessment of the memories from independent people. The evaluation was achieved via a survey created with Qualtrics. The participants were able to register for the study via SONA. After the registration, each participant got an email with a link towards the survey. The first page of the questionnaire contained information about the research and how long it would approximately take to answer it, as seen in Appendix A. Furthermore, it included the info that the participants would always have the possibility to retreat from the research, without experiencing any disadvantages. The first site also contained information that all their answers will be used for analysis, as long as they do not retreat or change their mind after finishing the survey. The second page contained information about the actual task they had to do, as well as more details about the research, as shown in Appendix B. The participants received the information, that they had to read memories and assess each of them for their emotional valence, and that this task would take around 120

minutes. The participants were also asked to make a five-minute break after each 20 consecutive memory they assessed, to ensure a focused assessment. They also got informed that they may close the survey and finish it at a later stage if they feel too exhausted, as Qualtrics offers the possibility to save already given answers. An example of how to assess the memories was provided, with an explanation afterward. On this side of the survey, the participants were also asked to enter their age and gender. Each of the next pages of the survey contained around 20 memories so that the participants knew when they should take breaks. The last page of the survey thanked the participant for taking part in the research, as well as making the participant aware, that if any questions were left he or she could contact the researcher via email. Therefore it also contained the email address of the researcher. The last page of the Qualtrics survey is represented in Appendix C.

LIWC Procedure

The gathered memory data was analyzed with LIWC, to do so, each memory got saved in an individual text file. All text files were then examined with LIWC by the analyze text option. Not all categories were essential for the analysis. The current research focused on the categories; dictionary, positive emotions, negative emotions, and word count. The dictionary category indicated how many words, in percent, of the text could be recognized and categorized by LIWC. Positive/negative emotions category contained how many words, in percent, were categorized as either positive or negative. The word count category included the total number of words used for each memory.

Results

The first step of the analysis was the exploration of each variable. The owner assessed 320 written first childhood memories (62.38%) as negative, 30 (5.85%) as neutral, and 163 (31.77%) as positive, which makes a total of 513 memories. The independent participants assessed 238 (43.39%) as negative, 137 (26.71) as neutral, and 138 (26.9%) as positive. Posemo 07 had an average of 1.189 (*SD* = 0.95), with a maximum of 4.85, and a minimum of 0. Negemo 07 averaged at 1.674 (*SD* = 1.32), with a maximum of 7.09 and a minimum of 0. Posemo 15 had an average of 1.734 (*SD* = 1.24), with a maximum of 7.5 and a minimum of 0. Negemo 15 had an average of 1.608 (*SD* = 1.27), with a maximum of 7.8 and a minimum of 0. The average for Emovalence 07 was -.4849 (*SD* = 1.73), with a maximum of 4.21, and a minimum of -6.38. The dictionary value showed an average of recognized words 77.72% (*SD* = 4.63) with the version of 2007, with a maximum of 90.77%, and a minimum of 64.14%. Emovalence 07 ascribed 270 memories a negative value, 80 a neutral value (0), and 163 a positive value. The mean value for Emovalence 15 was .1254 (*SD* = 1.94), with a minimum of -7.09, and a maximum of 5.37. The dictionary value of the 2015 version showed an average of 85.19% (*SD* = 3.84), with a minimum of 72.87%, and a maximum of 95.45%. The Emovalence variable of the newer version assessed 222 memories with a negative value, 58 a neutral value (0), and 233 a positive value.

At first, a correlation of owner assessment and participant assessment was computed with a Chi-Square test, and Cramer's V. The Chi-Squared test showed a strong correlation, $x^2(4, N = 513) = 252.24$, $p < .001$, with $r = .70$, which refers to a $R^2 = .49$. The next analyses were done to examine the correlation of the Pos/Negemo variables with the assessments of the owners and

participants. Posemo 07 had a positive correlation with the owner assessment, r_s (511) = 0.256, p < .001. Posemo 07 correlated strongly with the participant assessment, r_s (511) = .228, p < .001. Negemo 07 correlated negatively with the owner assessment, r_s (511) = -.528, p < .001. Negemo 07 also correlated negatively with the participant assessment, r_s (511) = -.498, p < .001. Posemo 15 correlated positively with the owner, r_s (511) = .326, p < .001, and participant assessment, r_s (511) = .285, p < .001. Negemo 15 correlated negatively with owner, r_s (511) = -.537, p < .001, and participant assessment, r_s (511) = -.497, p < .001. The graphs of the Pos/Negemo 07/15 and owner/participant assessment correlations can be seen in appendix D. The next two correlations were conducted to examine the connection between the two different dictionaries. The analysis of Posemo 07 and Posemo 15, showed a strong correlation, r_s (511) = .712, p < .001 (Figure 1). Negemo 07 and Negemo 15 correlated strongly, r_s = .756, p < .001 (Figure 2).

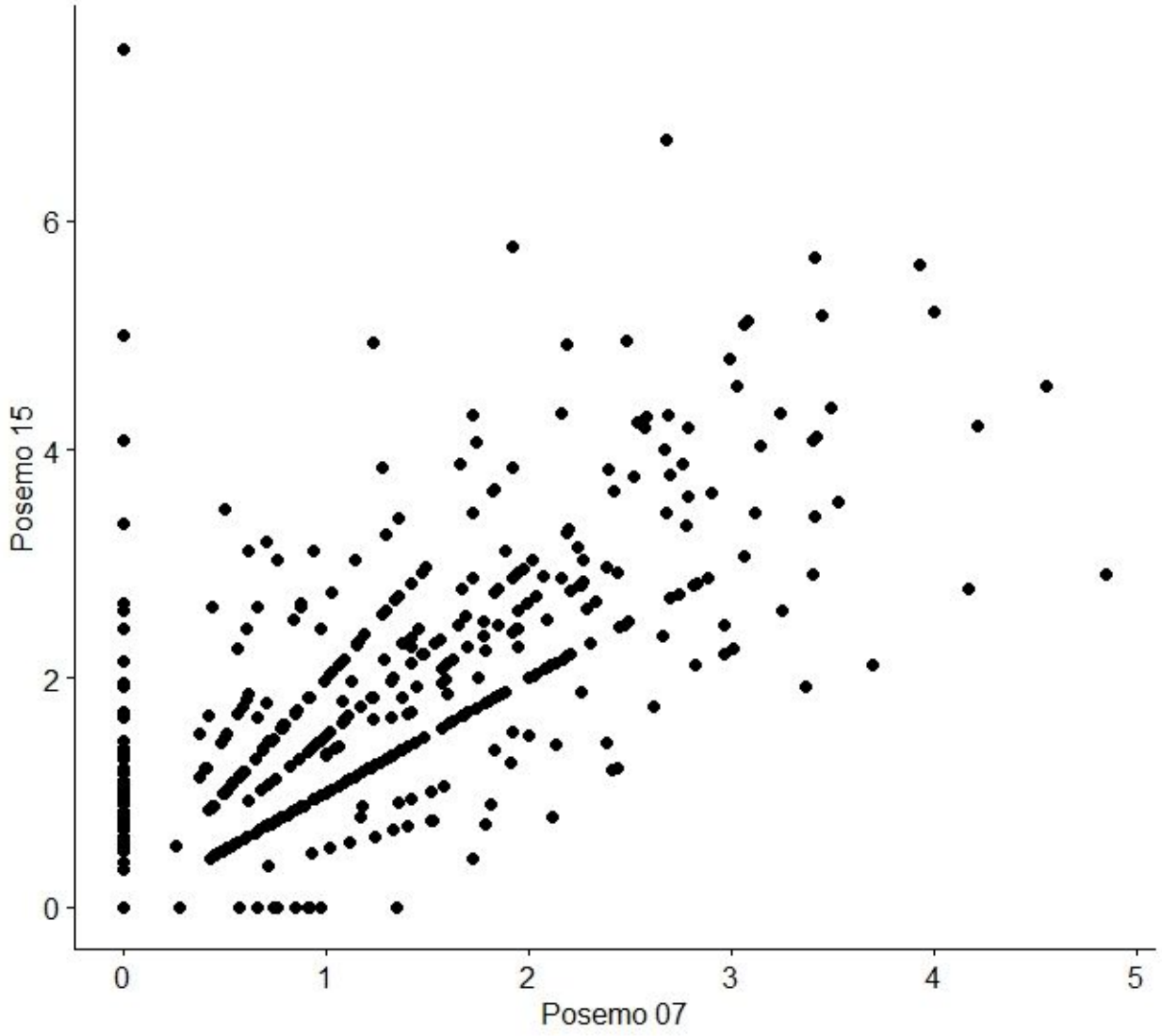


Figure 1. Spearman correlation with Posemo 07 and Posemo 15 resulted in a significant correlation.

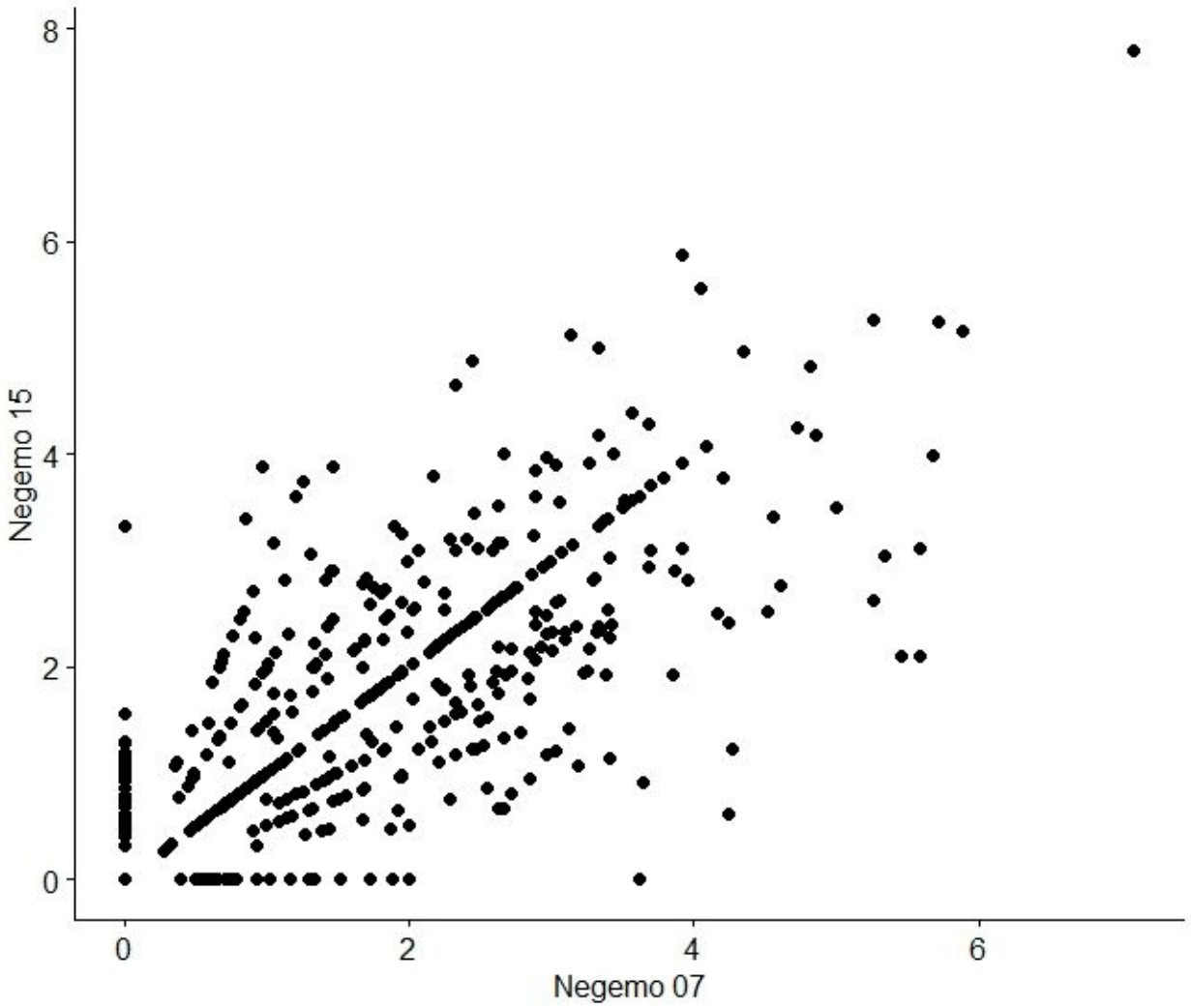


Figure 2. Spearman correlation with Negemo 07 and Negemo 15 resulted in a significant correlation.

The next couple of correlations were conducted to show the relationship between the emovalence variables. The first two correlations were conducted between the emovalence 07/15 and the owner assessment. Emovalence 07 showed a strong correlation with owner assessment, $r(511) = .517$, $p < .001$ (Figure 3). Emovalence 15 resulted in a strong correlation with owner assessment, $r(511) = .533$, $p < .001$ (shown in Figure 4).

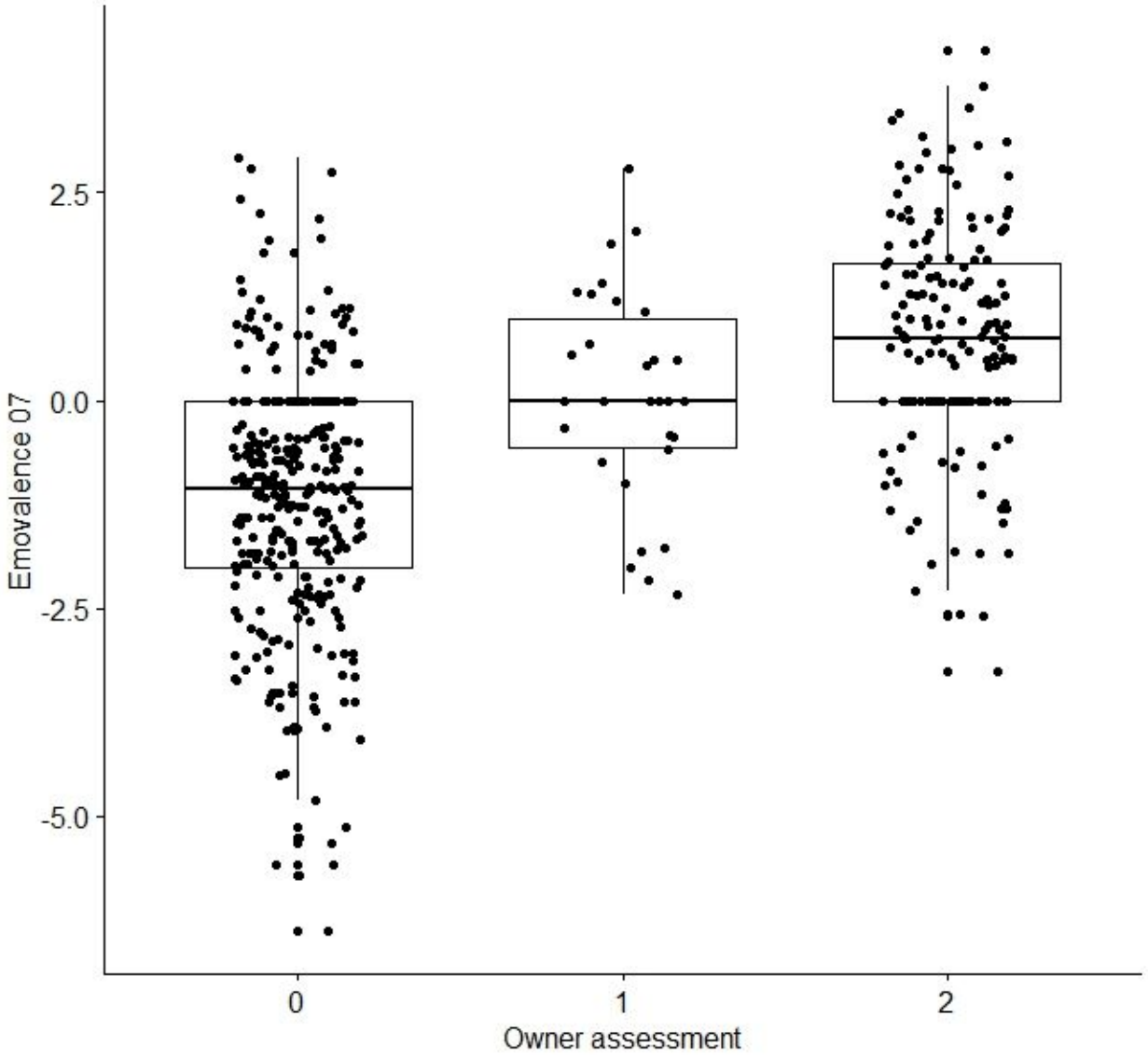


Figure 3. Pearson correlation with Emovalence 07 and owner assessment, resulting in a significant correlation. Assessment scale: 0 = Negative, 1 = Neutral, 2 = Positive.

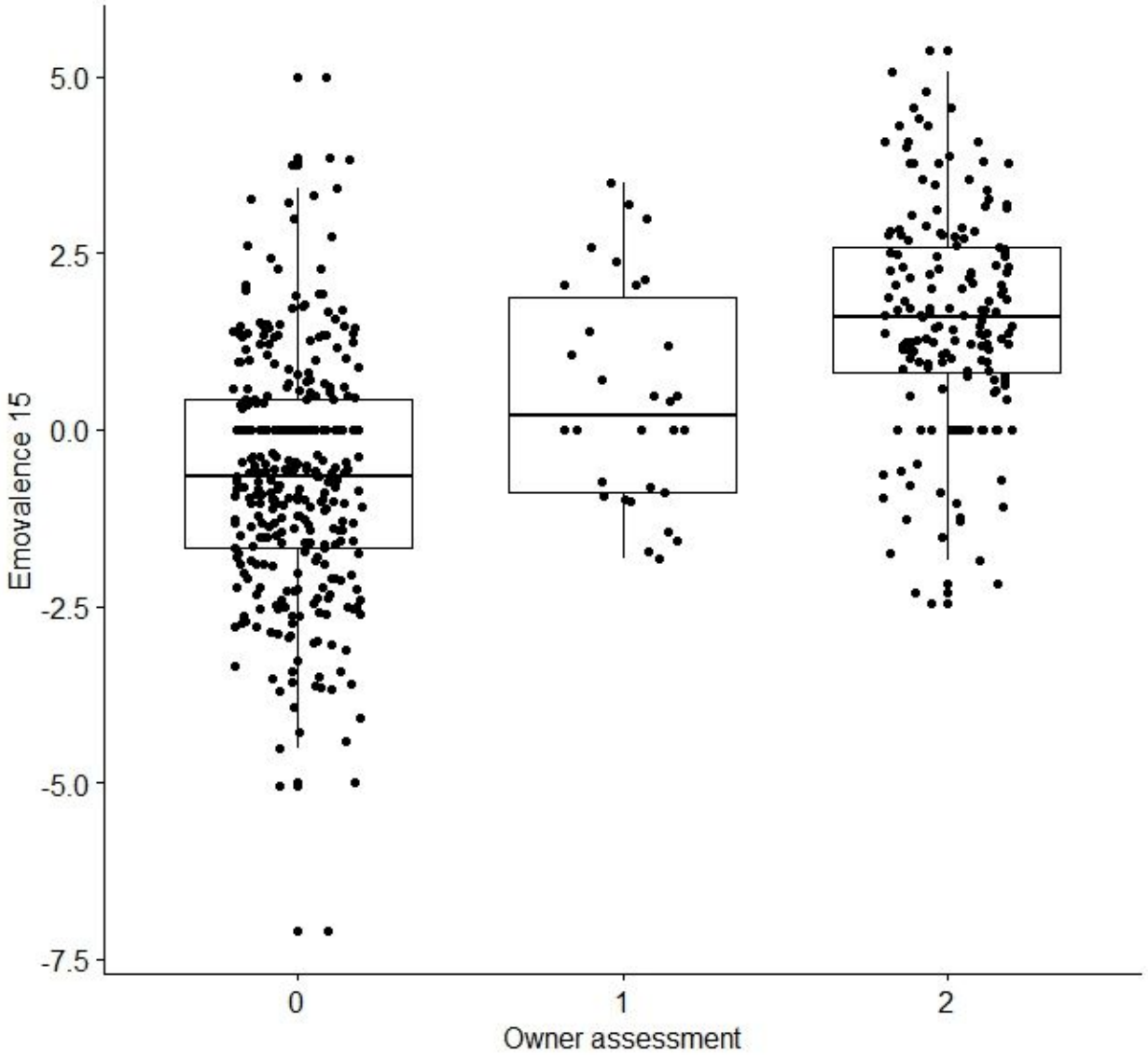


Figure 4. Pearson Correlation for the emotional valence, assessed by LIWC Dutch dictionary version 2007 and the memory owner. A strong correlation can be seen between Emovalence 15 and owner assessment. Assessment scale: 0 = Negative, 1 = Neutral, 2 = Positive.

The next two tests covered the correlation of the emotional valence of the dictionary version 2007/2015 and the participant assessment. The correlation between the assessment of version 2007 and the participants showed a high correlation, $r(511) = .495$, $p < .001$ (shown in Figure 5). The testing of emotional valence version 2015 and participant assessment resulted in a high correlation, $r(511) = .494$, $p < .001$ (Figure 6).

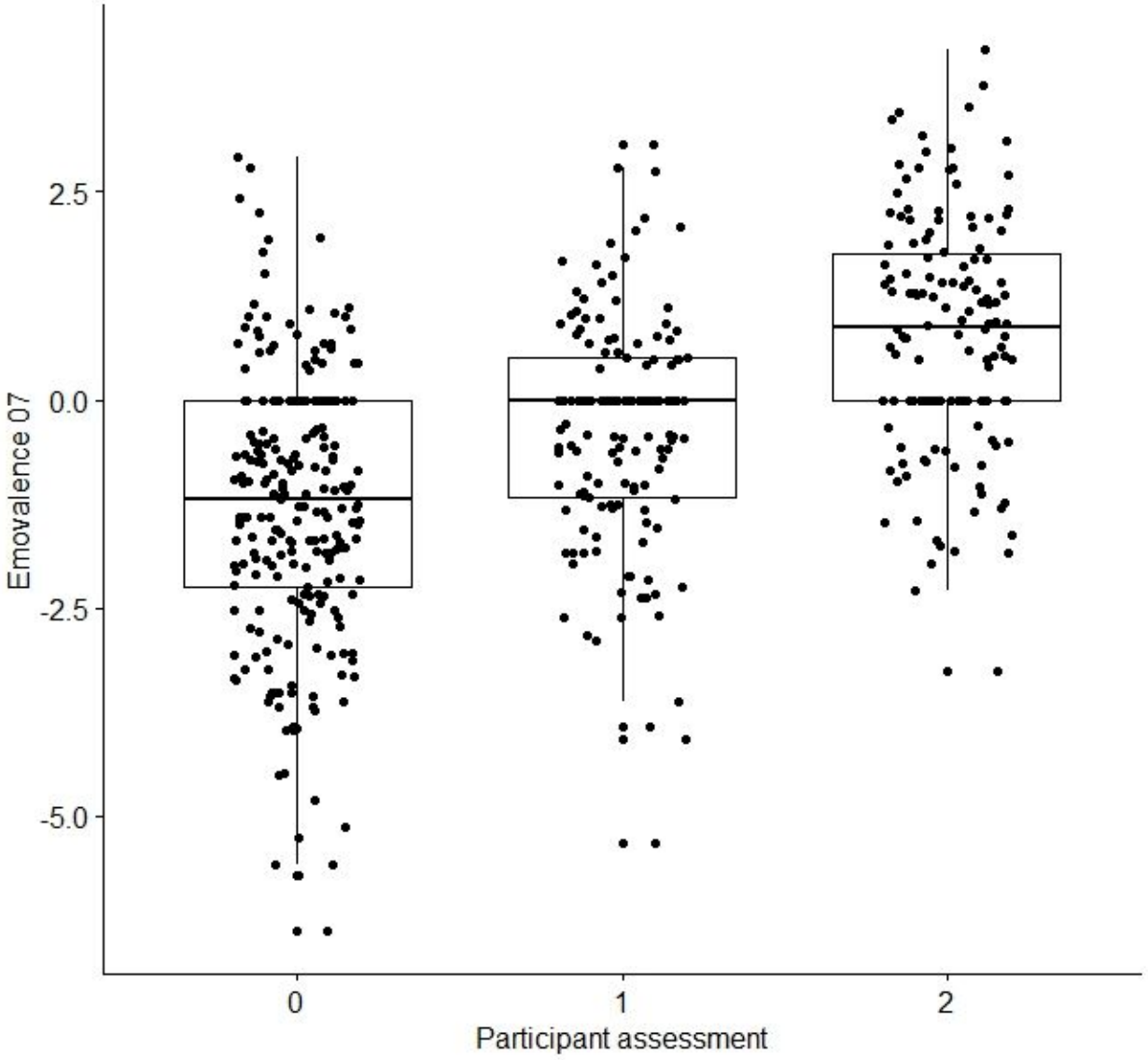


Figure 5. Pearson correlation for the emotional valence, assessed by LIWC dictionary version 2007 and independent participants. The result showed a significant correlation for Emovalence 07 and participant assessment. Assessment scale: 0 = Negative, 1 = Neutral, 2 = Positive.

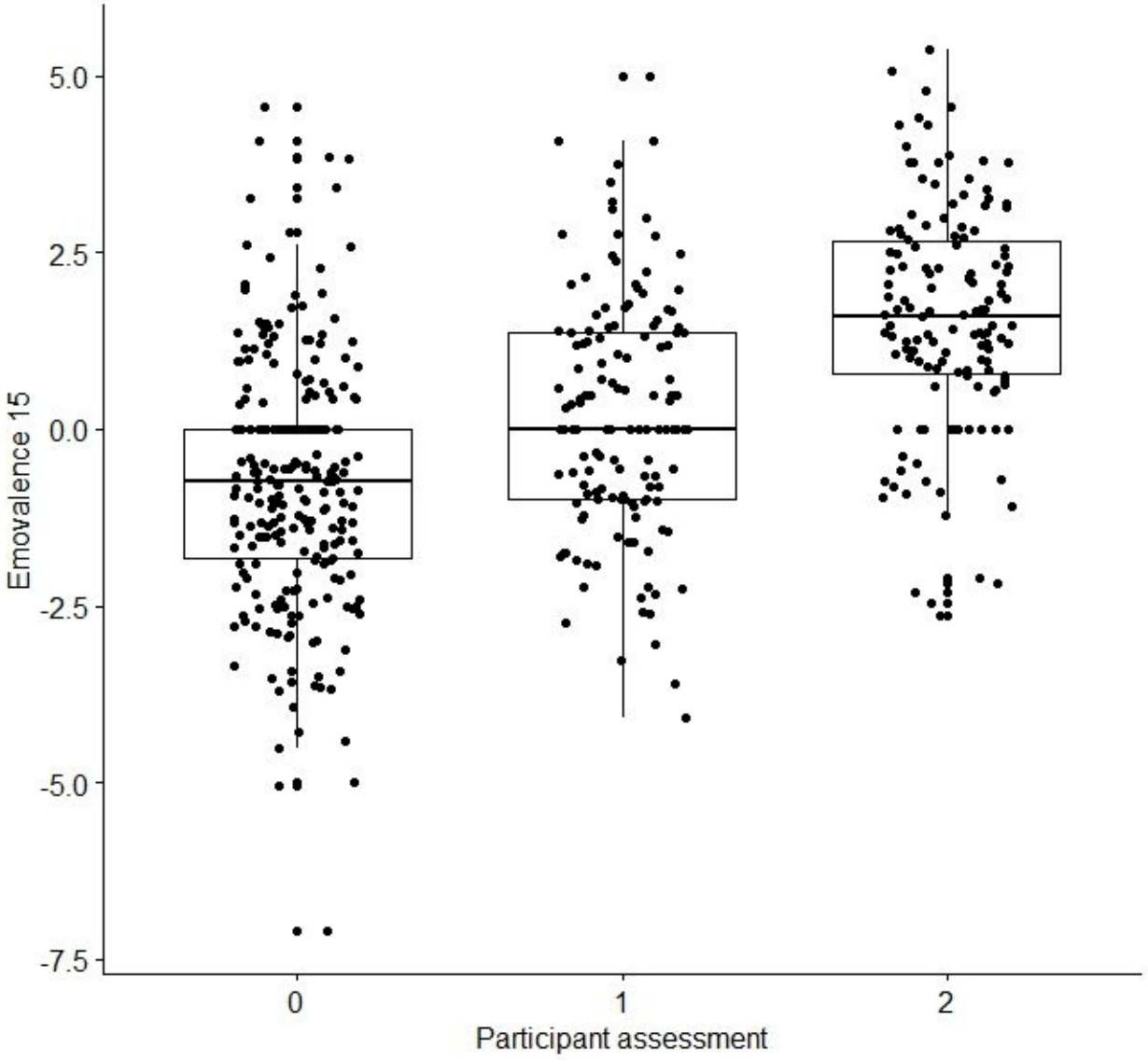


Figure 6. Pearson correlation with Emovalence 15 and participant assessment. The analysis resulted in a significant correlation. Assessment scale: 0 = Negative, 1 = Neutral, 2 = Positive.

The last test was done to examine the correlation of the assessments from LIWC version 2007 and 2015. The test resulted in a strong correlation between the dictionary version 2007 and 2015, $r(511) = .772$, $p < .001$ (shown in Figure 7).

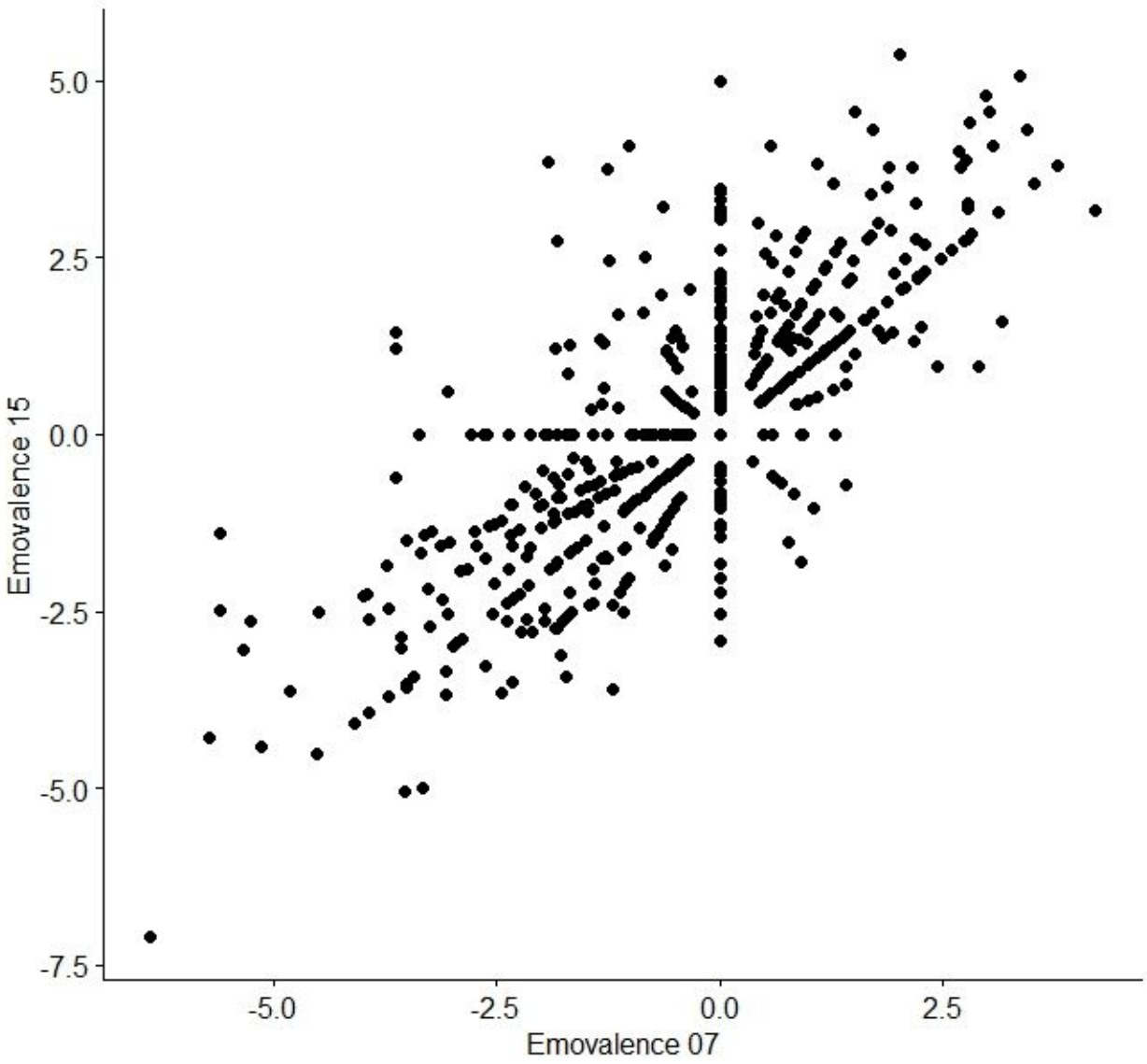


Figure 7. Pearson correlation for the emotional valence, assessed by LIWC dictionary version 2007 and 2015.

Discussion

The central research question of this study was if LIWC is a valid tool for assessing first childhood memory descriptions on their emotional valence. Based on the statistical tests the Dutch dictionaries of 2007 and 2015 might be seen as valid. Still, this outcome should be

regarded with particular caution, as the correlations for Posemo were quite low. Each dictionary version of LIWC was able to assess the emotional valence in a significant way but somewhat surprising was that the outcome of version 2007 came slightly closer to the assessment of the participants than the dictionary version 2015. The 2015 version came on the other hand closer to the evaluation of the owner, in comparison to the 2007 version. The comparison of the two versions also showed that the newer one seems to be more sensitive in categorizing positive emotions, but they were still heavily correlated with each other. Looking at the single Pos/Negemo variables, it showed that the version exceeded the older one in every correlation, except for Negemo 15 and participant assessment. Negemo 07 had a slightly higher correlation with the participant assessment, which should be the reason for the slight difference in the Emovalence comparison. The newer dictionary showed a higher capability of recognizing more words. The statistical tests also revealed that the independent participants were able in assessing the emotional valence of the described memories. The way of scoring the memories can explain the bigger differences between the participant assessment and LIWC's assessment.

In a contribution to the discussion, if positive or negative emotions are more prominent within first childhood memories, the current research found negative memories to be more prominent. This outcome was in line with the findings with Dudycha and Dudycha (1933), Cowan and Davidson (1984), Howes, Siegel, and Brown (1993), and Mullen (1994). Slight discrepancies were found for the age of onset as Draaisma (2005), Howes et al. (1993), Jack & Hayne (2007), Mullen (1994), Peterson, Grant, and Boland (2005), and Tusint and Hayne (2010) concluded that it was between the third and fourth year, the current outcomes suggest an onset above the fourth year, with an average of 4.25 years.

The statistical results indicate that LIWC is well capable of recognizing emotional valence in first childhood memories. This outcome was comparable to the results of Bantum and Owen (2005), Tausczik and Pennebaker (2010), Tov, Ng, Lin, & Qui (2013), Settanni, & Marengo (2015), as they also focused on validating the ability to assess emotions in different contexts. Both versions of LIWC correlated heavily with each other with a value of .77, which corresponds to an R^2 of .5. Regarding the single emotion variables, it was shown that Posemo 07 and 15 shared a correlation of .71 and Negemo 07/15 a correlation value of .76. This outcome was partly surprising, as the correlation of the English dictionary with the same variables has shown a correlation of .96 for either variable. The correlation value was given by the founders of LIWC and can be found on their official website. The website contains a whole list of all categories and how the dictionary version of 2007 is correlating with the 2015 version. Even though they showed a strong correlation, it was nowhere near .96. A possible explanation can be the translated dictionary, as such a translation is not that simple. Van Wissen and Boot (2017), compared their translated version with the English versions and reached a correlation of .77 for the 2007 version and .73 for the 2015 version. It might have come to more significant changes of categories within the Dutch version, as with more words, it is possible that they also represent more categories. An indication for this is the correlation of .96 for either Pos/Negemo variable of the English version. Such a high correlation indicates, that not that many changes were made for these categories. The problem is that a word in one language might have beside the meaning also another one, to which it might be more referring. Another point is the way the newer version got translated. The new way of translation, as it was done with a translation machine and a human

check afterward. Therefore the version of 2015 was still experimental, which might explain a slight difference within the correlation of 2007, 2015 and the English counterpart.

Comparing the dictionaries to the assessments of participants LIWC as well as the participants seemed fine in recognizing the emotional valence of the childhood memories. While the independent participants had a correlation of .7, which corresponds to an R^2 of .49, LIWC correlated .52 (2007) and .53 (2015), which corresponds to an R^2 of .27 and 28. The difference in the R^2 does not necessarily lead to the assumption that LIWC is not capable of judging the memories right, but it should be still kept in mind. Taking a look at the strength of LIWC, it can analyze a vast amount of texts in a reasonably short amount of times. As the investigated population rises, R^2 becomes less impactful as in smaller populations. For instance, with LIWC it is possible to analyze 10000 texts in a short amount of time and considerable small effort. Comparing this approach to a traditional qualitative analysis, it would take more time to analyze 100 texts, which would be more precise but still, the amount of data that is 'accurate' according to the statistics would be at around 49. LIWC, on the other hand, could analyze 10000 texts, of which 2700 - 2800 are statistically 'accurate'. Qualtrics was used in the current research, which gives the opportunity to look at the time that was needed to fill in the survey of the present study. The five participants required 9h 28 min to assess 500 memories. That makes an average of 113 minutes per participant. The participants merely needed this time. It is also necessary to take into account the time that is required to create such a survey, plus the time to recruit participants. In the end, it is even necessary to monitor the whole data gathering process, to ensure that everything proceeds as planned. All these steps take a considerable time, which might extend over several hours/days. The whole analysis with LIWC did not take longer than a minute. After

the collection of the memories, it was only necessary to select every text data to be analyzed and were then analyzed at once by LIWC. This research only examined 513 written memories, imagining how the working times expand at 10000 possible memories. These numbers show the immense advantage of text mining programs.

The single variables Posemo 07/15 and Negemo 07/15 showed exceptional correlation values with either owner and participant assessment. The correlation value of Posemo 07 and owner as well as participant assessment was relatively low, with .26 for owner and .23 for participant assessment. Such a value is considered as weak, but in the field of cognitive science lower correlation values are expected. A better correlation value achieved the version of 2015 with .33 with owner assessment, and .29 with participant assessment. The values of the 2015 version might be still considered as weak, but it shows significant improvements to the older version. This improvement might be the result of the greater pool of positive emotion words, which came with the newer version. At least for the Dutch version, as such a change was not noticeable within the English version. A further aspect was the distribution of positive and negative memories. More negative memories were collected than positive ones. Therefore a higher rate of assessing a memory as wrong seems potential. With a different research setup, it might be possible to achieve a higher correlations. The other setup should be able to assess the memories on both dimensions at once, which LIWC did. A memory might show both, positive and negative emotions, which resolves the three solid categories of negative, neutral, and positive. Instead of such categories it would enable a more fluid transition between these emotions, and would also allow multiple emotions. It could also be the case, that especially the positive emotion category suffers from the negation of positive words, which LIWC is not able

to recognize. These correlation numbers though indicate that it might be helpful to further improve the words within the positive emotion category. The Negemo variables scored crucially higher, Negemo 07 and owner correlated with a -.53, and with the participants -.498. Negemo 15 correlated with the owner with a value of -.54, and with the participants -.497. Such correlation values are seen as moderate, which for such research are already reasonably high. The numbers did not increase that much with the newer version, as the negative emotion category did not experience that many changes.

A further advantage of LIWC is a more precise assessment of emotional valence, as it is not bound to three categories. Instead, LIWC can judge the emotional valence based on the wording, and how many words fall into a specific category. This process makes it overall more objective in comparison with a human judge unless the human judge would get a list which ascribes each word towards a particular category. The human judge would be enabled to analyze the text in the same way as LIWC does, which in the end would take more resources, and might end up less precise. The human judge would instead assess the memory on its context and the overall tone of emotions, he or she might experience while reading the memory. This way of judging makes it difficult to introduce more categories, rather than negative, neutral, or positive. It might be possible to use a Likert-scale which could range from one to five for positive/negative emotions. The only problem would be a certain inconsistency between human judges, as each could categorize the same memories differently. A specific memory could score a five on positive emotions, while the next participant scored it as a three, and the equal participant might end up scoring it a week after differently than the first time. This example shows the advantage of LIWC as an objective and consistent tool to analyze texts. It can also be seen in the

results, as 137 memories were judged as neutral instead of positive or negative, while only 30 owners judged their memories as neutral. This outcome showed that participants already had struggled in many instances to categorize the memories in a right way. Adding more categories might lead to further confusion and overall worse results. The bias to assess more memories as neutral might be the outcome of missing information. The owner has more context about the event and can recall the emotions associated with the event, which does not inevitably lead to sufficient emotional expression in the wording.

By comparing the two versions of the dictionaries, it was possible to see that the sensitivity towards positive emotions has risen in comparison to the older version. While the average of the emotional valence for the older version was at -.48, the average emotional valence for the 2015 version was .13. That showed that the newer version added more words that get categorized as emotionally positive. This substantial difference might be due to the translation. As already discussed, it might be the case that a greater part of the new words that came with version 2015 increased the pool of words that categorize as positive emotional, due to the Dutch language, or the automatic translation. Nevertheless, the correlation between Emovalence 15 and owner assessment is slightly higher, than the 2007 version. The newer version still achieved an improvement, based on the difference. As the aim lays in understanding how the owners see their memories and what feelings are connected with their memories, it is most important to get as close to the owner assessment as possible. This outcome leads to the conclusion that the Dutch dictionary version of 2015 is slightly better in assessing early childhood memories.

The study was subject to specific limitations. It was intended to gather more human judges, which would assess the described memories, but to a lack of available participants, it was

not possible. This change would have enabled an interrater correlation, which would have been beneficial for the outcome. Therefore it is recommended to use more human judges in comparable studies. A further limitation was the composition of the participants, the owner as well as the judges. A significant part of the participants that had written down their memories were not Dutch. Every student was still capable of sufficient Dutch knowledge, but most of the first year students have not that much experience of the Dutch language, and therefore might have used wrong words, which might have influenced the outcome. It might have been possible to achieve more precise results if all of the participants would have been a native speaker. Looking at all the problems translating a dictionary for LIWC, due to different word meanings, it could have also affected the memories that were written down. This limitation leads to the recommendation of collecting data from native speakers or those capable of a higher level. Also, the judges were all Germans, which leads to the same conclusion as the limitation discussed beforehand. The human judges should also be in the best case with native speakers, to avoid more misinterpretations based on language differences. A further weakness was the gathering of data, in either case, the owner of the memories, as well as the participants, that judged the memories at a later stage. The limitation was the different methods that were used to get an assessment. A difference was made between the first and second year of memory collection, as in the second year of data collection, only one word was given to describe the emotional state, while the students in the first year could also use multiple descriptions of the emotional state. This difference led to an adjustment of the data set to bring it to a common ground. The participants were also asked by the researcher to assess the memories in on of three categories so that it was better compared with the owner assessment. This approach created the problem that it

was less suitable to compare with LIWC's ability to judge on multiple dimensions at the same time. Thus for further studies, it might be useful to use an even more comparable scale, with the possibility to be able to assess both tendencies. For instance, giving the judges two Likert-scales from zero to five, one for positive emotional and one for negative. Each participant had then to judge on either scale. Another interesting approach would be the stable judgment of a human judge, in the sense of multiple measurements with the same memories. This approach would enable more accurate analysis of how different human judges assess written memories, which might further strengthen the stable and objective usage of LIWC.

At last, it is possible to say, that the Dutch dictionary 07/15 from LIWC according to this study proved as a possible tool to analyze inscribed childhood memories on their emotional valence. Whereby improvements still have to be made, especially in recognizing positive emotions. The newer version achieved quite some improvements within this category. This difference clearly showed the superiority of the newer version. Therefore it is advised to use the newer version for future research, especially if it gets further refinements. This tool enables an easier way to objectively analyze memories, not only with less effort but also on much grander scales than with any qualitative research, which involves human judges.

References

Bantum, E. O. C., & Owen, J. E. (2009). Evaluating the validity of computerized content analysis programs for identification of emotional expression in cancer narratives. *Psychological assessment*, *21*(1), 79.

Boot, P., Zijlstra, H. & Geenen, R. (2017). The Dutch translation of the Linguistic Inquiry and Word Count (LIWC) 2007 dictionary. *Dutch Journal of Applied Linguistics*, 6(1).

Brockmeyer, T., Holtforth, M. G., Bents, H., Herzog, W., & Friederich, H. C. (2013). Lower body weight is associated with less negative emotions in sad autobiographical memories of patients with anorexia nervosa. *Psychiatry research*, *210*(2), 548-552.

Cohn, M. A., Mehl, M. R., & Pennebaker, J. W. (2004). Linguistic markers of psychological change surrounding September 11, 2001. *Psychological science*, *15*(10), 687-693.

Corkin, S. (1984). Lasting consequences of bilateral medial temporal lobectomy: Clinical course and experimental findings in HM. *In Seminars in Neurology Vol. 4*, No. 2, 249-259.

Cowan, N., & Davidson, G. (1984). Salient Childhood Memories. *The Journal of Genetic Psychology, 145*, 101-107.

Craik, F. I., Govoni, R., Naveh-Benjamin, M., & Anderson, N. D. (1996). The effects of divided attention on encoding and retrieval processes in human memory. *Journal of Experimental Psychology: General*, *125*(2), 159.

Dudycha, G. J., & Dudycha, M. M. (1933). Some Factors and Characteristics of Childhood Memories. *Child Development, 4,* 265-278.

Eich, J. E., Weingartner, H., Stillman, R. C., & Gillin, J. C. (1975). State-dependent accessibility

of retrieval cues in the retention of a categorized list. *Journal of Memory and Language*, *14*(4), 408.

Howes, M., Siegel, M., & Brown, F. (1993). Early Childhood Memories: Accuracy and Affect. *Cognition, 47*, 95-119.

Jack, F., & Hayne, H. (2007). Eliciting adults' earliest memories: Does it matter how we ask the question? *Memory, 15*, 647-663.

Kihlstrom, J. F., & Harackiewicz, J. M. (1982). The Earliest Recollection: A New Survey. *Journal of Personality, 50*, 134-148.

Lavenex, P., & Amaral, D. G. (2000). Hippocampal‐neocortical interaction: A hierarchy of associativity. *Hippocampus*, 10(4), 420-430.

Loftus, E. F. (2005). Planting misinformation in the human mind: A 30-year investigation of the malleability of memory. *Learning & Memory*, 12(4), 361-366.

McClelland, J. L., McNaughton, B. L., & O'reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, 102(3), 419.

Miles, C. (1895). A Study of Individual Psychology. *The American Journal of Psychology, 6*, 534-558.

Mullen, M. K. (1994). Earliest recollections of childhood: a demographic analysis. *Cognition, 52*, 55-79.

Nakazawa, K., Quirk, M. C., Chitwood, R. A., Watanabe, M., Yeckel, M. F., Sun, L. D., Kato, A., Carr, C. A., Johnston, D., Wilson, M. A., & Tonegawa, S., (2004). Requirement for Hippocampal Ca3 Nmda Receptors in Associative Memory Recall. *Year Book of*

Psychiatry & Applied Mental Health, *2004*(1), 336-337. Science, 297(5579), 211-218.

Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015). *The development and psychometric properties of LIWC2015*.

Peterson, C., Grant, V., & Boland, L. (2005). Childhood amnesia in children and adolescents: Their earliest memories. *Memory, 13*, 622-637.

Reyna, V. F., & Lloyd, F. (1997). Theories of false memory in children and adults. *Learning and individual differences*, *9*(2), 95-123.

Saunders, L. M. I., & Norcross, J. C. (1988). Earliest Childhood Memories: Relationship to Ordinal Position, Family Function, and Psychiatric Symptomatology. *Individual Psychology, 44*, 95-105.

Schacter, D. L. (1999). The seven sins of memory: Insights from psychology and cognitive neuroscience. *American psychologist*, *54*(3), 182.

Scoville, W. B., & Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. *Journal of neurology, neurosurgery, and psychiatry*, *20*(1), 11.

Settanni, M., & Marengo, D. (2015). Sharing feelings online: studying emotional well-being via automated text analysis of Facebook posts. *Frontiers in psychology*, *6*, 1045.

Smith, E. E. & Kosslyn, S. M., 1948- (2007). *Cognitive psychology : mind and brain* (1st ed). Prentice Hall, Harlow, 204-250.

Squire, L. R., Stark, C. E., & Clark, R. E. (2004). The medial temporal lobe. *Annual Reviews of Neuroscience*, 27, 279-306.

St-Laurent, M., Moscovitch, M., Jadd, R., & McAndrews, M. P. (2014). The perceptual richness

of complex memory episodes is compromised by medial temporal lobe damage. *Hippocampus*, *24*(5), 560-576.

Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of language and social psychology*, *29*(1), 24-54.

Tov, W., Ng, K. L., Lin, H., & Qiu, L. (2013). Detecting well-being via computerized content analysis of brief diary entries. *Psychological assessment*, *25*(4), 1069.

Tulving, E. (1972). Episodic and semantic memory. *Organization of memory*, 1, 381-403.

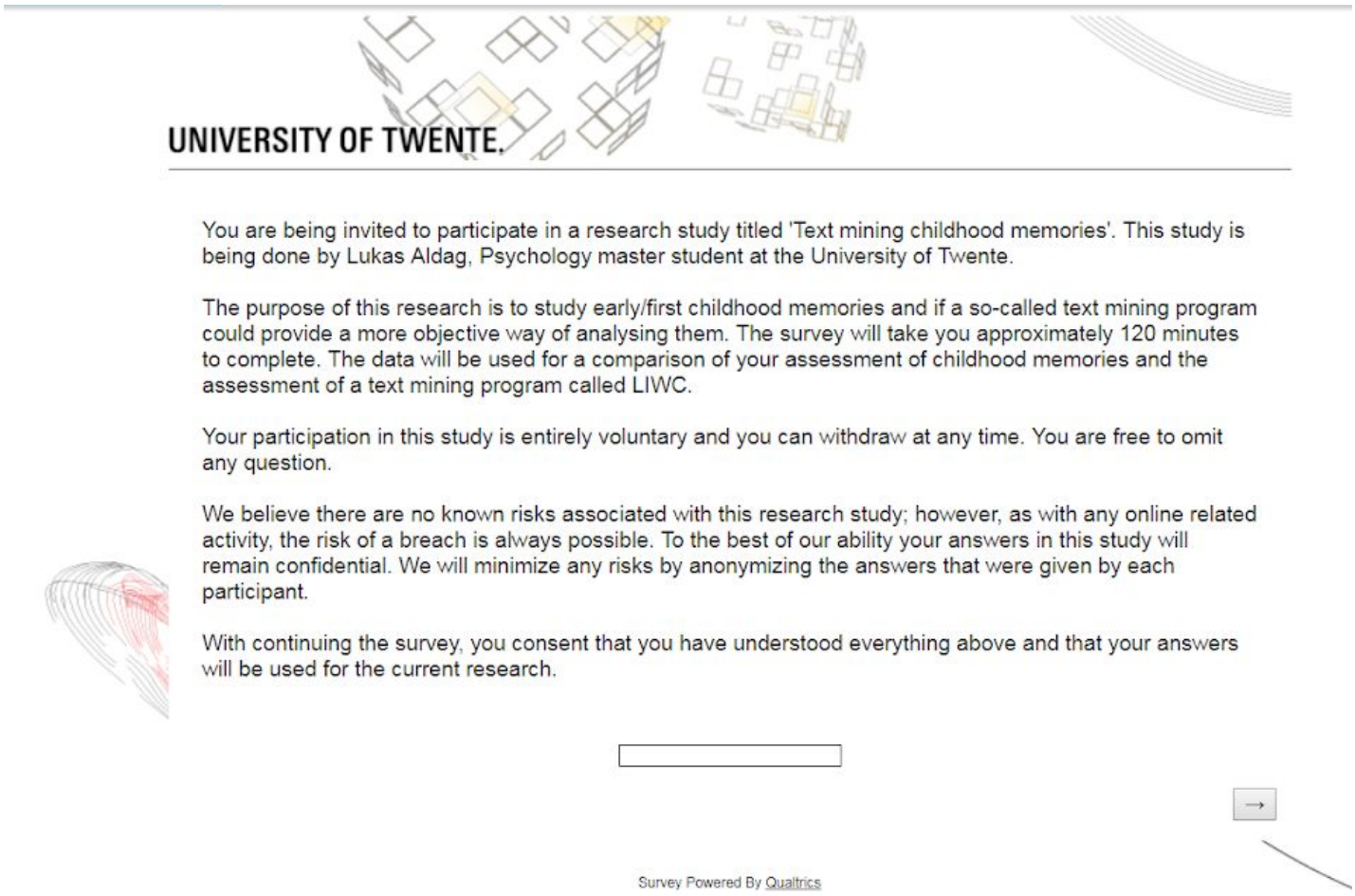
Tustin, K., & Hayne, H. (2010). Defining the Boundary: Age-Related Changes in Childhood Amnesia. *Developmental Psychology, 46*, 1049-1061.

Waldfogel, S. (1948). The Frequency and Affective Character of Childhood Memories. *Psychological Monographs, 62*, Whole No. 291.

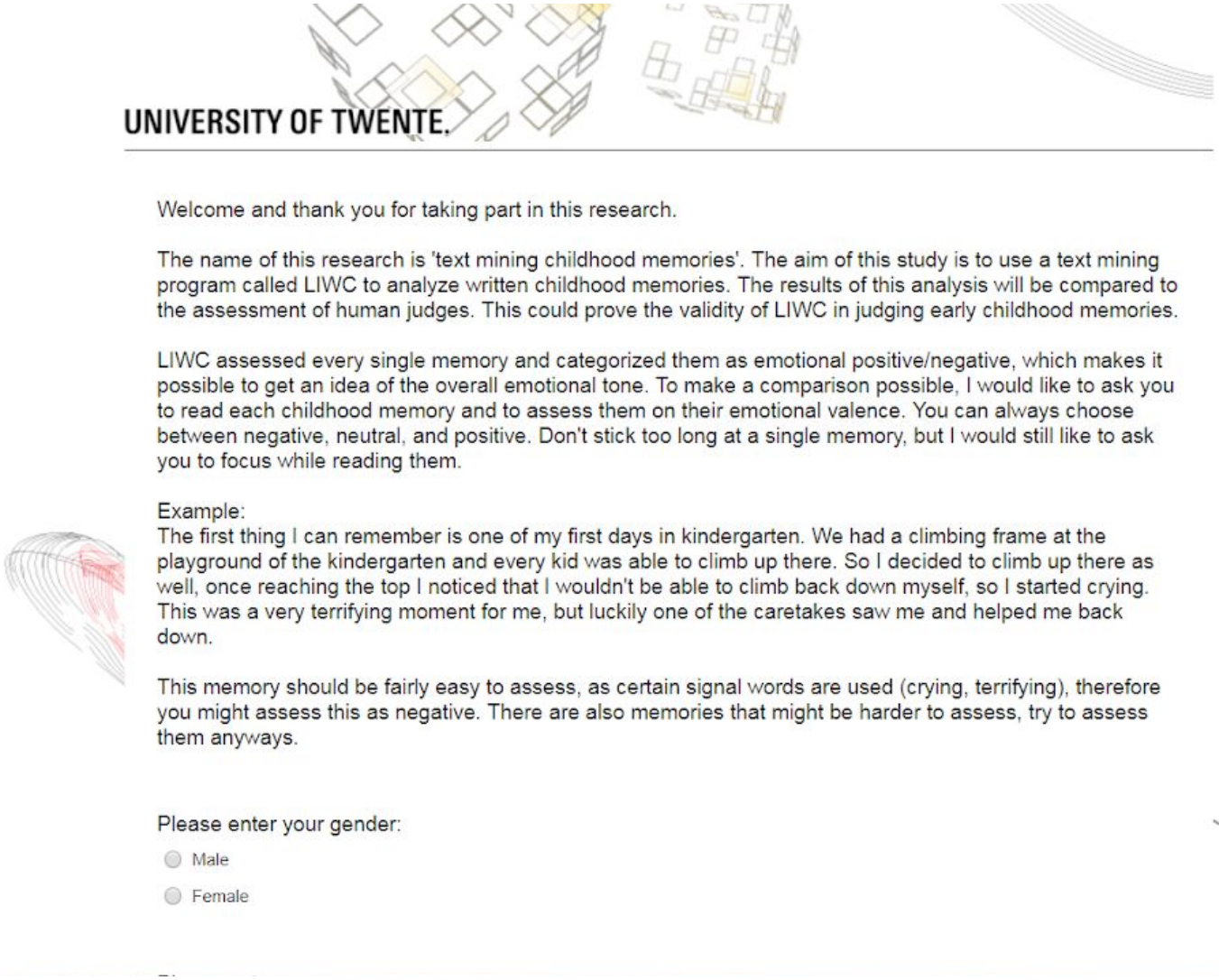
Van Wissen, L., & Boot, P. (2017). An Electronic Translation of the LIWC Dictionary into Dutch. In *Electronic lexicography in the 21st century: Proceedings of eLex 2017 conference* (pp. 703-715). Lexical Computing.

Appendix

Appendix A: First side of the Qualtrics questionnaire.



Appendix B: Second page of the Qualtrics survey.



UNIVERSITY OF TWENTE.

Welcome and thank you for taking part in this research.

The name of this research is 'text mining childhood memories'. The aim of this study is to use a text mining program called LIWC to analyze written childhood memories. The results of this analysis will be compared to the assessment of human judges. This could prove the validity of LIWC in judging early childhood memories.

LIWC assessed every single memory and categorized them as emotional positive/negative, which makes it possible to get an idea of the overall emotional tone. To make a comparison possible, I would like to ask you to read each childhood memory and to assess them on their emotional valence. You can always choose between negative, neutral, and positive. Don't stick too long at a single memory, but I would still like to ask you to focus while reading them.

Example:
The first thing I can remember is one of my first days in kindergarten. We had a climbing frame at the playground of the kindergarten and every kid was able to climb up there. So I decided to climb up there as well, once reaching the top I noticed that I wouldn't be able to climb back down myself, so I started crying. This was a very terrifying moment for me, but luckily one of the caretakes saw me and helped me back down.

This memory should be fairly easy to assess, as certain signal words are used (crying, terrifying), therefore you might assess this as negative. There are also memories that might be harder to assess, try to assess them anyways.

Please enter your gender:

- Male
- Female

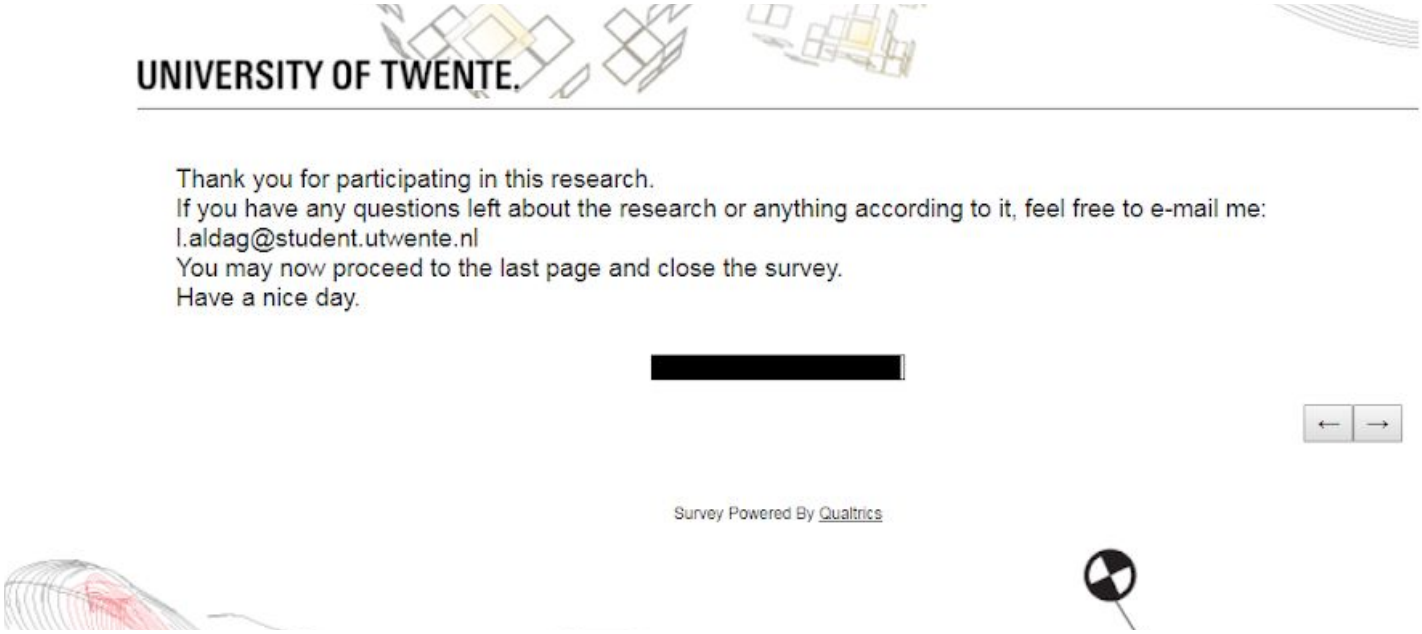
Please enter your age:

Enter your Sona number (for credit distribution):

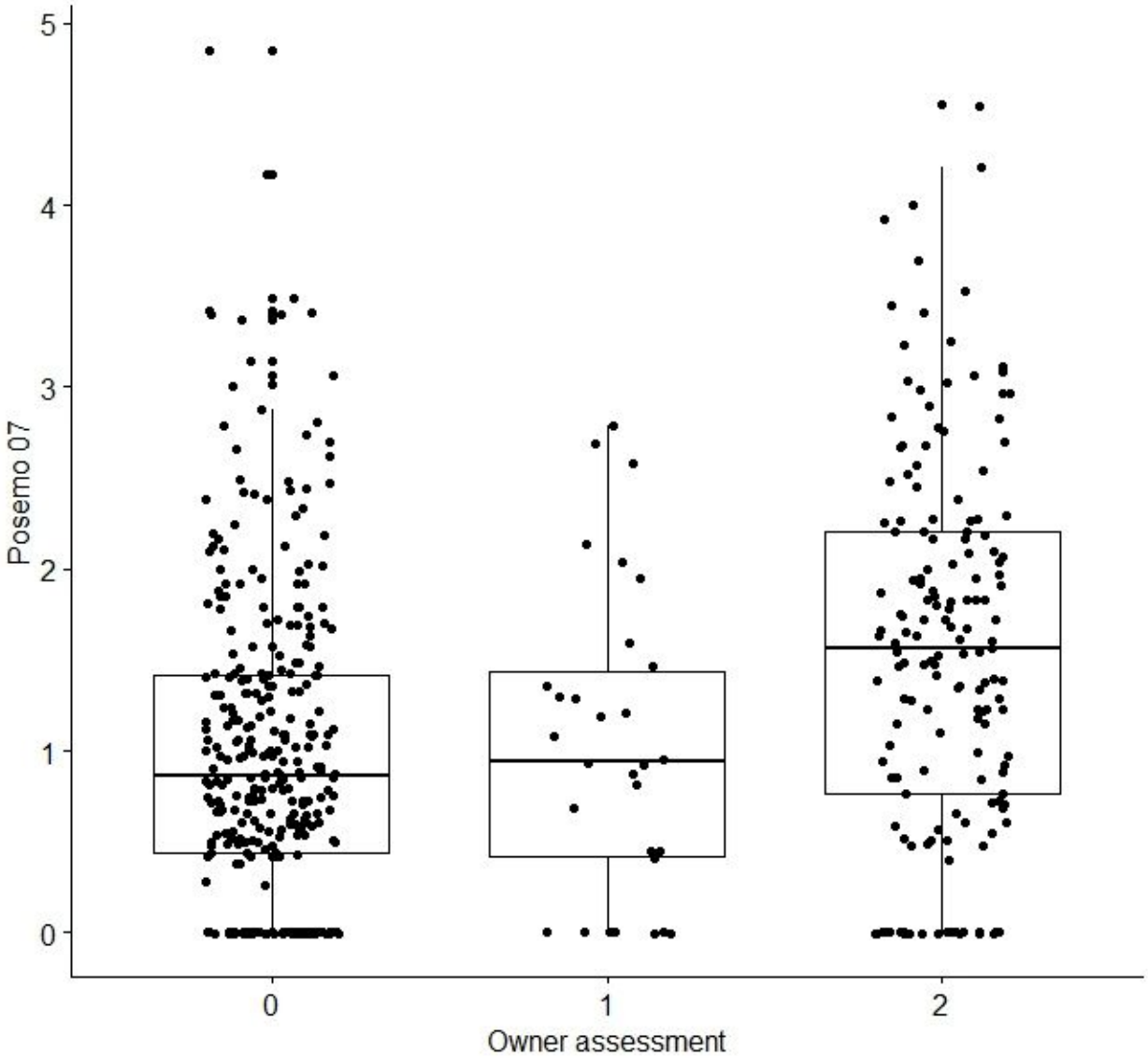
With the next page, you will start the actual survey. As this is a fairly long survey you should take a break of five minutes, after each page. Even if the survey might take somewhat longer, I would ask you to take this survey seriously and try to stay focused, as it could influence the actual outcome. If you feel exhausted, take a longer break and assess the memories at a later stage.

Survey Powered By Qualtrics

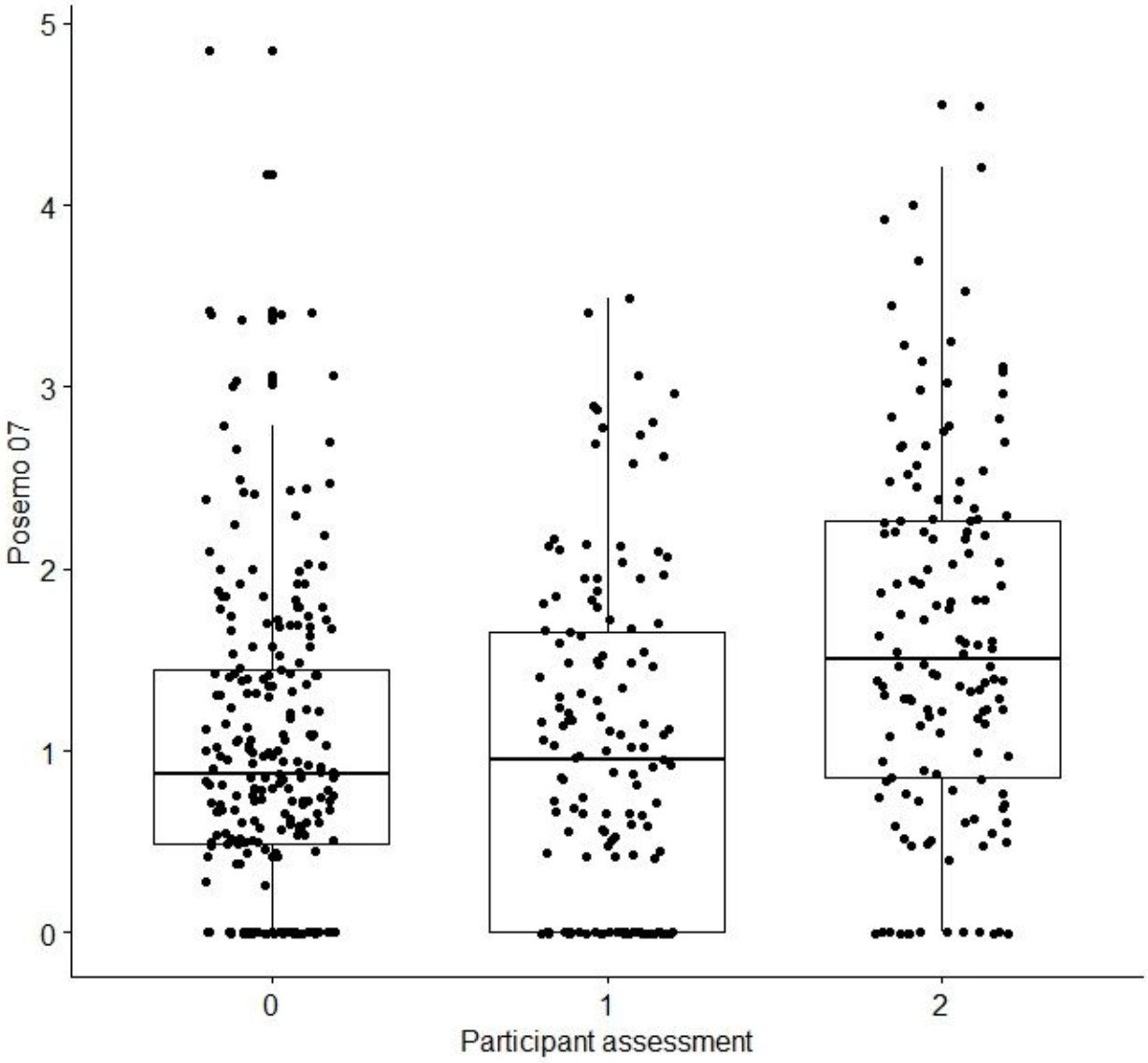
Appendix C: Last page of the Qualtrics survey.



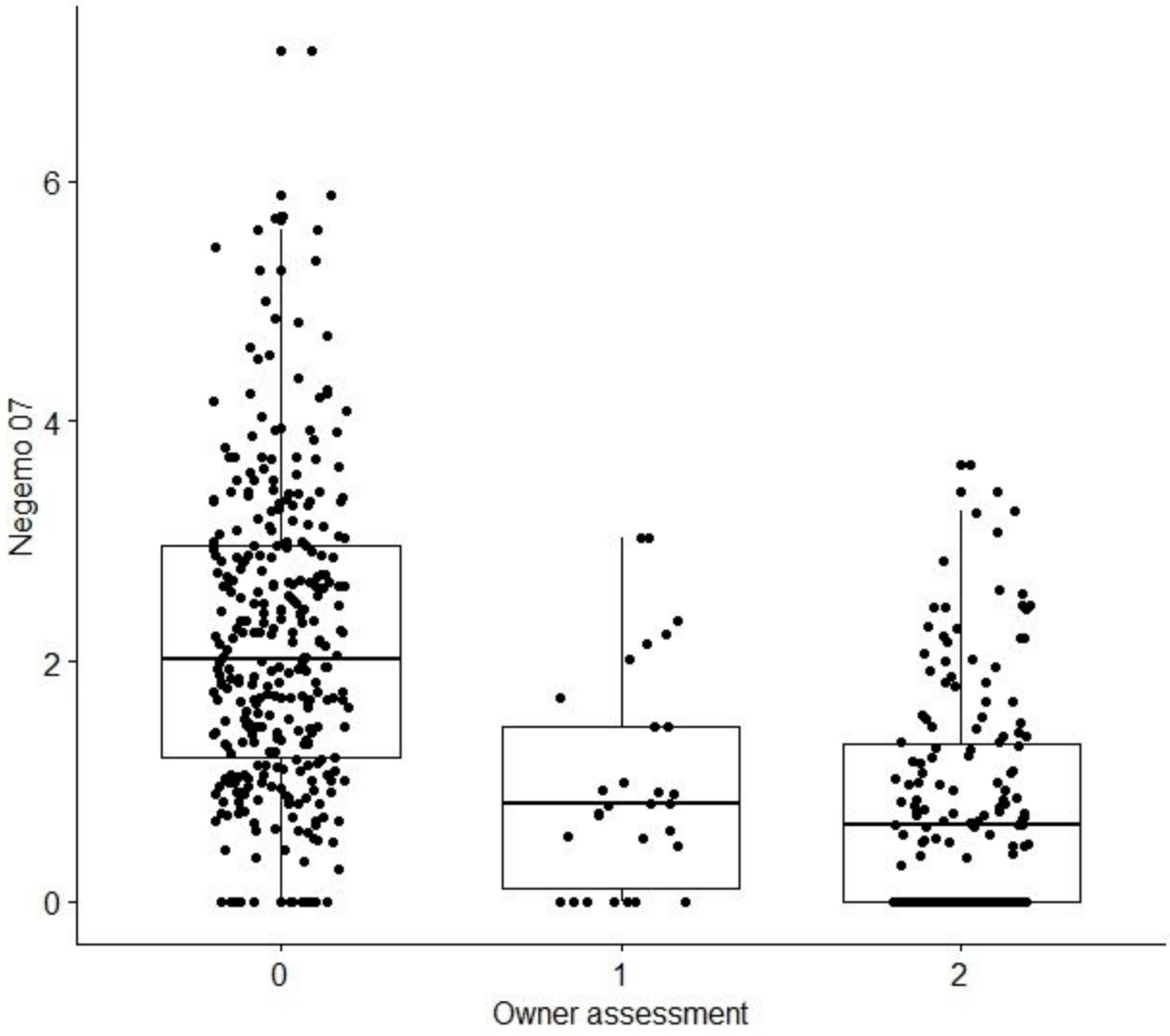
Appendix D: Spearman correlations for Pos/Negemo variables with owner/participant assessment variables.



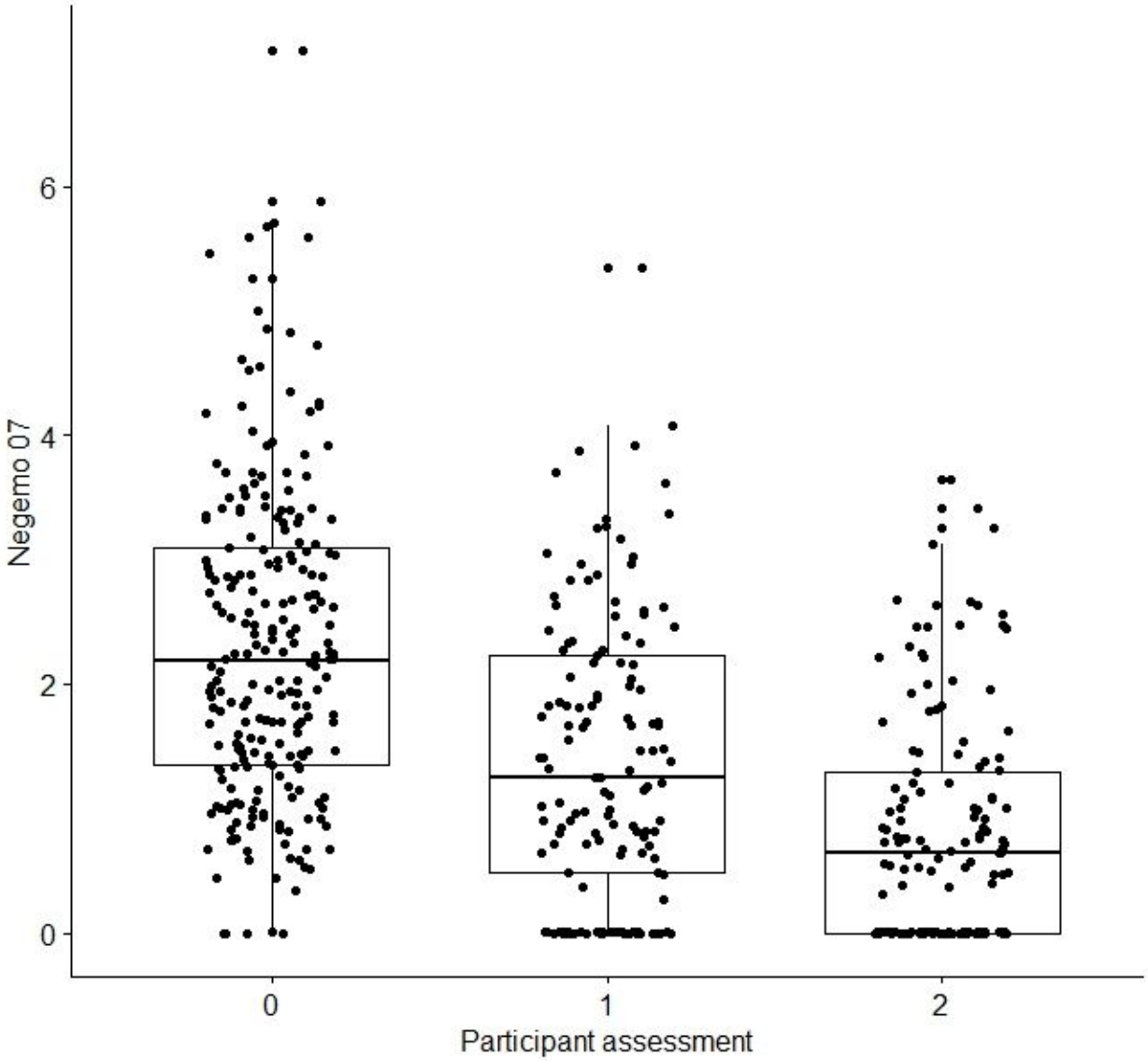
Spearman correlation for Posemo 07 and owner assessment, resulting in significant correlation. Assessment scale: 0 = Negative, 1 = Neutral, 2 = Positive.



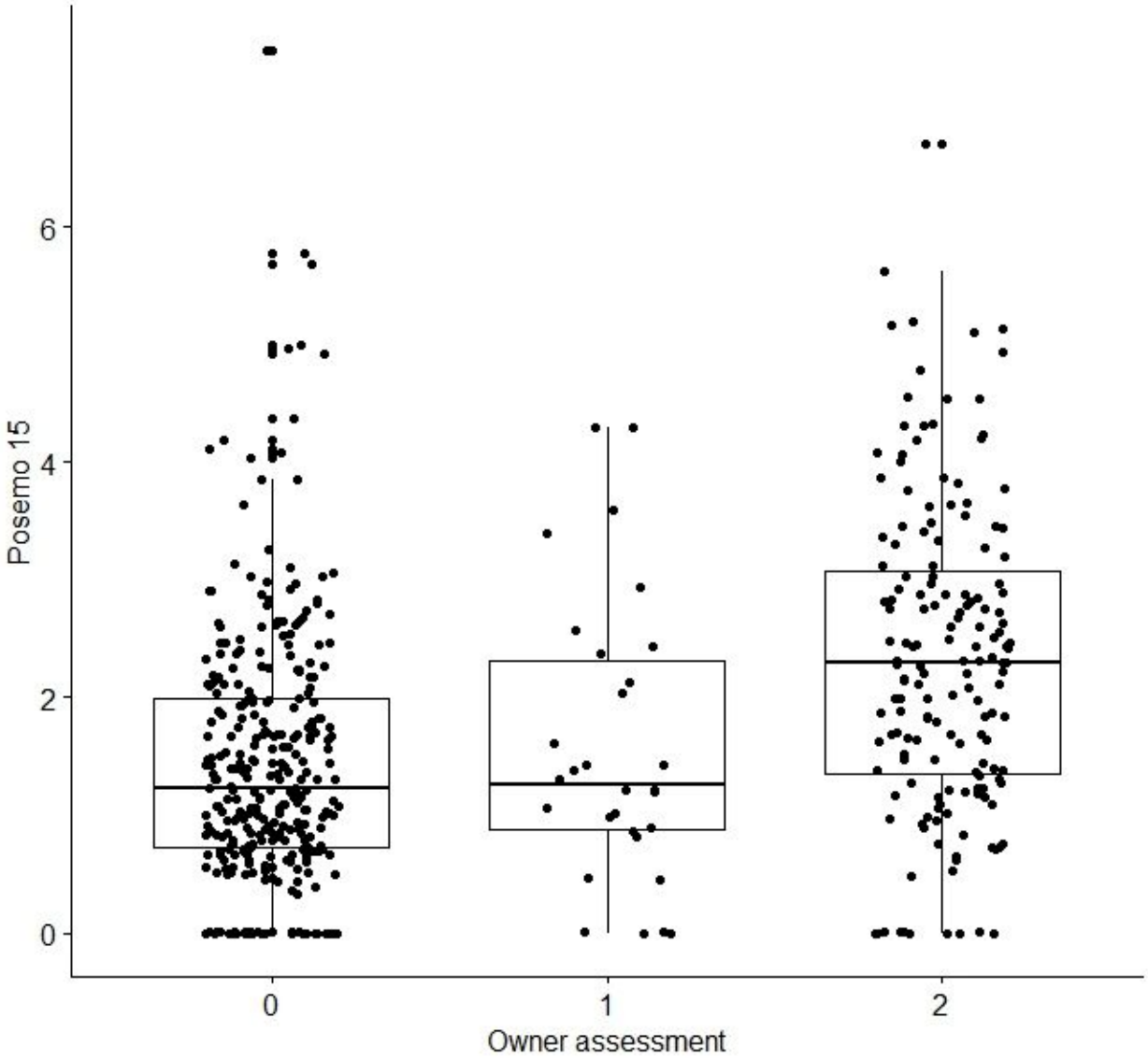
Spearman correlation for Posemo 07 and Participant assessment, resulting in significant correlation. Assessment scale: 0 = Negative, 1 = Neutral, 2 = Positive.



Spearman correlation for Negemo 07 and owner assessment, resulting in a significant negative correlation. Assessment scale: 0 = Negative, 1 = Neutral, 2 = Positive.

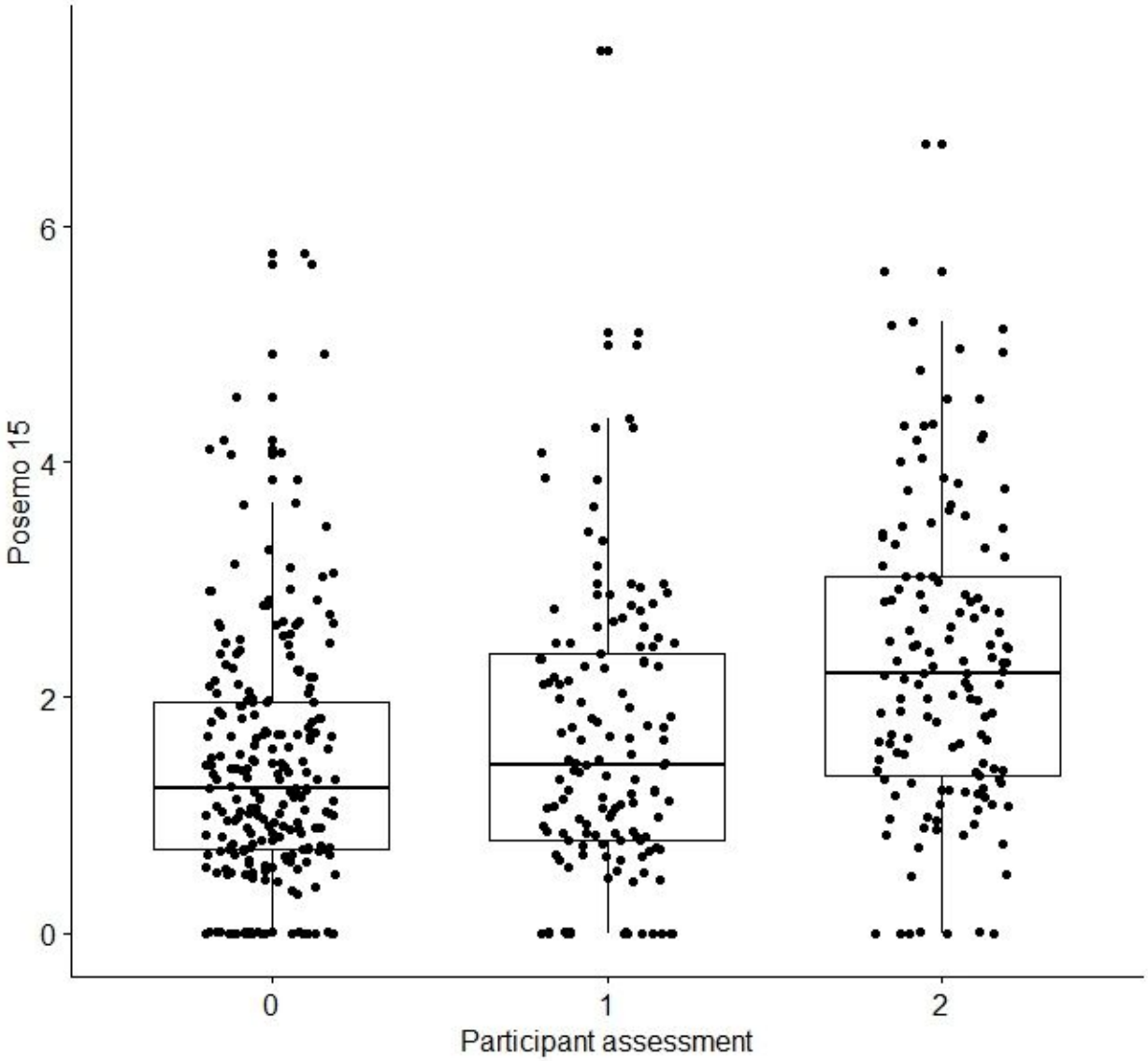


Spearman correlation for Negemo 07 and participant assessment, resulting in a significant negative correlation. Assessment scale: 0 = Negative, 1 = Neutral, 2 = Positive.

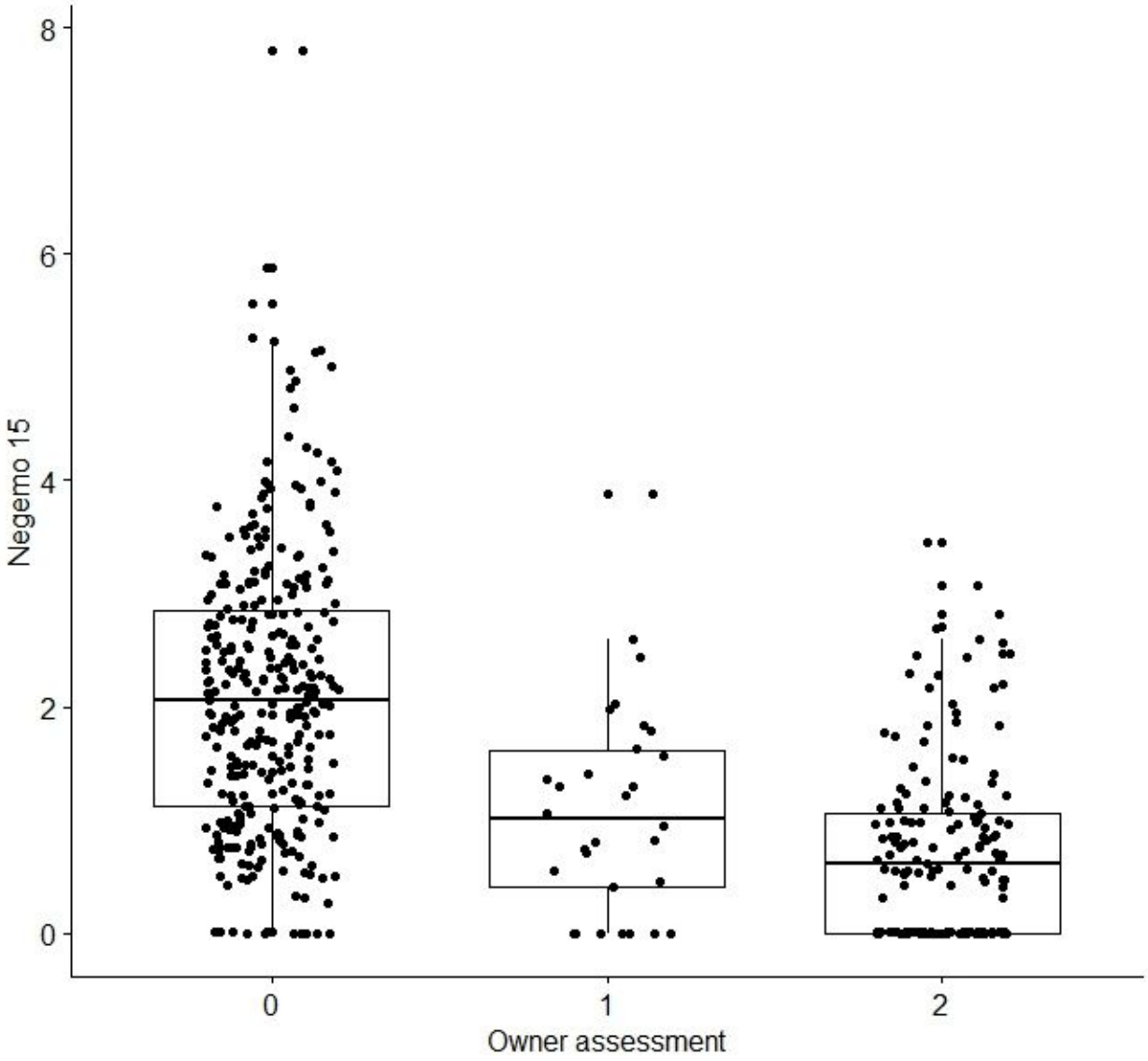


Spearman correlation for Posemo 15 and owner assessment,resulting in a significant correlation.

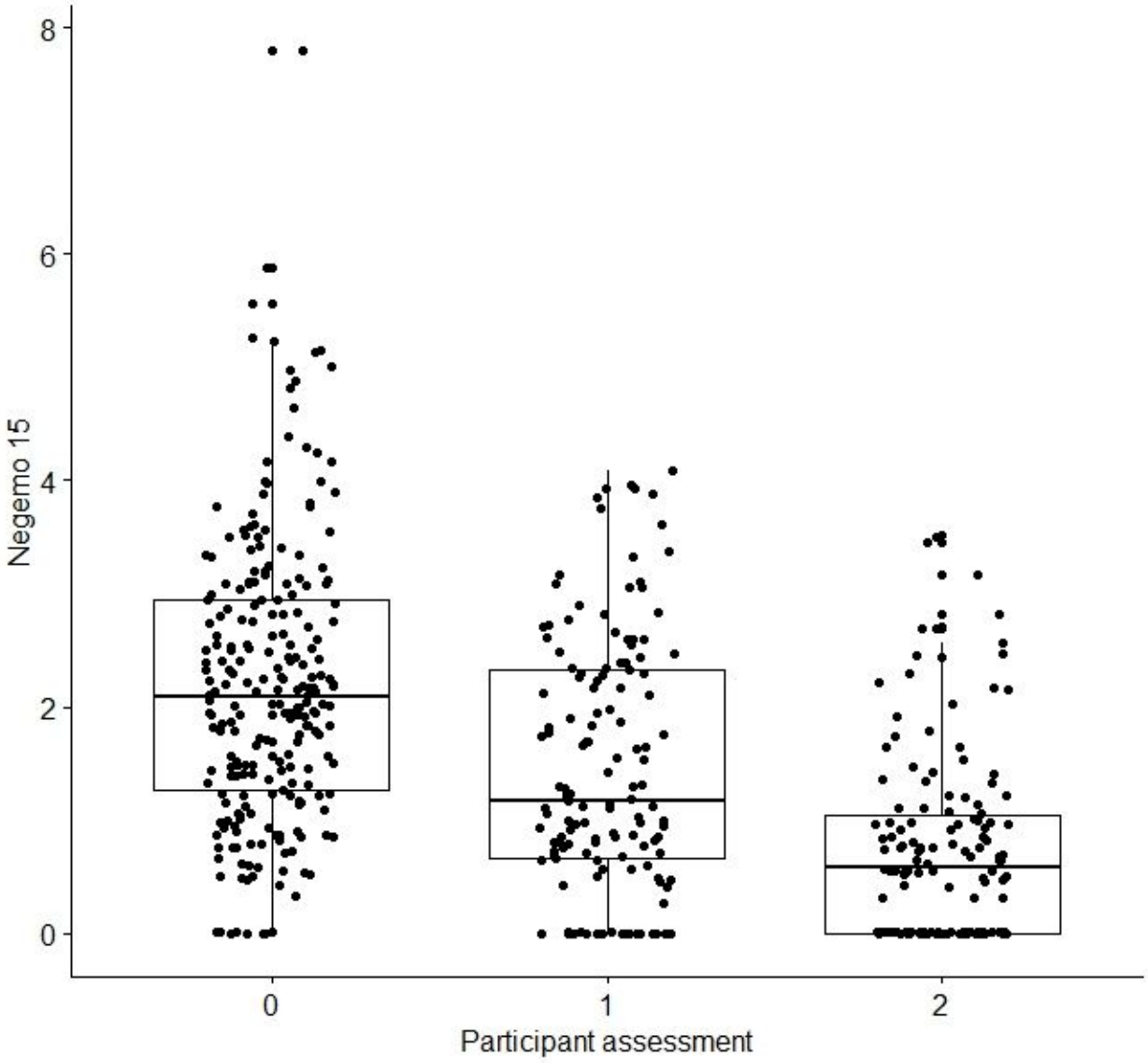
Assessment scale: 0 = Negative, 1 = Neutral, 2 = Positive.



Spearman correlation for Posemo 15 and participant assessment, resulting in a significant correlation. Assessment scale: 0 = Negative, 1 = Neutral, 2 = Positive.



Spearman correlation for Negemo 15 and owner assessment, resulting in a significant negative correlation. Assessment scale: 0 = Negative, 1 = Neutral, 2 = Positive.



Spearman correlation for Negemo and participant assessment, resulted in a significant correlation. Assessment scale: 0 = Negative, 1 = Neutral, 2 = Positive.