

**Ruhr-Universität Bochum**

**Faculty of Psychology**

**Human Technology Interaction**

**Bachelor-Thesis 24.07.23**

**„An assessment of the reliability of self-report and behavioral measures of empathy“**

**Author: Markus Dieterich**

**B. Sc. Psychology**

**Matrikel-Nr.: 108019209422**

**Markus.Dieterich@rub.de**

**Supervisors: Dr. Ian Hussey; Prof. Dr. Malte Elson**

**Abstract**

# An assessment of the reliability of self-report and behavioral measures of empathy

Empathy plays a crucial role in everyday life. Without the ability to share the feelings of other persons or to know what they might feel and think, human interaction would be immensely hindered. The relevance of empathy has led to it being a subject in numerous fields of psychological research. In clinical psychology, the implications and treatments of psychopathological disorders that are associated with a lack of empathy are investigated. This includes autism spectrum disorders, psychopathy, alexithymia and antisocial, narcissistic and borderline personality disorders (Decety & Moriguchi, 2007). The function of empathy in altruism (Batson et al., 2014) and group identification (Miyazono & Inarimori, 2021) are typical topics of social psychology. Other exemplary fields are evolutionary (de Waal, 2008) and developmental (Mcdonald & Messinger, 2011) psychological research. There is also a multitude of educational programs aimed to enhance the empathy of specific groups, such as children, adolescents, parents and couples (Butters, 2010) or teachers, students and nurses (Lam et al., 2011).

## Definition of the term “empathy”

Despite the wide variety of fields of application, a consistently used definition of empathy does not exist (Cuff et al., 2016); (Coplan, 2011); (Neumann et al., 2015). In fact, it is common for literature on empathy to begin with a statement regarding this lack of uniformity (Eklund & Meranius, 2021). Possible consequences of the inconsistent use of the empathy term are a reduced comparability of study outcomes, mismatches between the way empathy is researched and applied in treatment or educational programs, and therapeutic difficulties (Cuff et al., 2016).

There have been efforts to synthesize different conceptualizations of empathy. For example, Eklund and Meranius (2021) investigated 52 reviews on empathy from the years 1980 – 2019. Their thematic analysis revealed four themes that were present in nearly all of the studies, which are: understanding, feeling, and sharing another person’s world while maintaining self-other differentiation. Cuff et al. (2016) examined 43 different definitions of empathy and identified eight themes present in most of them. Based on this, they defined empathy as following:

Empathy is an emotional response (affective), dependent upon the interaction between trait capacities and state influences. Empathic processes are automatically elicited but are also shaped by top-down control processes. The resulting emotion is similar to one’s perception (directly experienced or imagined) and understanding (cognitive empathy) of the stimulus emotion, with recognition that the source of

the emotion is not one’s own. (p. 150)

Establishing a comprehensive empathy definition might surely be helpful to unify the different views, and thus, lead to more comparable research outcomes. But, it remains unclear if these new definitions will indeed help to unify or will just add further definitions to the already existing ones.

The ambiguity of the empathy term might hamper the research of this concept, but it also demonstrates the importance of it. Coplan (2011) draws a suitable conclusion: “Given its central role in so many discussions and debates, it’s safe to conclude that whatever empathy is, it’s important”.

## Assessment of empathy

Ways to measure empathy include questionnaires (both self-report and observer ratings (Hall & Schwartz, 2019)), behavior observation and behavioral tasks (Melchers et al., 2015) and neuroscientific approaches (Neumann et al., 2015). Only self-report measures and behavioral tasks will be reviewed in this paper.

### ***Self-report measures of empathy***

Self-report measures are the most frequently used way to assess empathy (Baldner & McGinley, 2020; Gerdes et al., 2010). In the quantitative review of Hall and Schwartz (2019), 80% of 393 studies published in the years 2001 – 2013 used at least one measure of self-reported empathy. Similarly, in the meta-analysis by de Lima and de Lima Osório (2021), 21 of 23 psychometric studies of empathy measures published in the years 2009 – 2019 were focused on self-report scales. The ambiguity of the empathy term is also reflected in the high number of different scales: In the aforementioned works of Hall and Schwartz (2019), Neumann et al. (2015) and de Lima and de Lima Osório (2021), 32 different measures are mentioned (counting only instruments developed to assess empathy in the general population).

While having the advantage of being economical and easy to apply, self-report measures may be prone to problems such as motivational bias, response bias, presentation bias and social desirability (Neumann et al., 2015). Additionally, self-reported empathy should be interpreted with caution, as some studies report no correlation to behavioral outcomes (Sunahara et al., (2022); Murphy and Lilienfeld, (2019).

### ***Behavioral measures of empathy***

In the systematic literature review of de Lima and de Lima Osório (2021), no psychometric studies on behavioral empathy measures published in the years 2009 – 2019 are mentioned. While there is a variety of behavioral tasks which assess empathy (Murphy and Lilienfeld (2019) identified at least 14 behavioral tasks that were applied to measure cognitive empathy), these tests are far less used than self-report measures. Murphy and Lilienfeld (2019) also note that it is common for researchers to create new stimulus materials or to modify existing materials. One example of a consistently used behavioral task to measure empathy is the “Reading the mind in the Eyes Test” (S. Baron-Cohen et al., 2001) – although initially developed to measure Theory of Mind, a construct related to cognitive empathy.

### ***Reliability of empathy measures***

Reliability is an essential concept of psychological measurement and a frequently researched topic (Revelle & Condon, 2019). This also applies to the measurement of empathy. A multitude of studies aim to assess the reliability of empathy measurement instruments. The meta-analysis of de Lima and de Lima Osório (2021) names 50 psychometric studies published in the years 2009 – 2019, of which 44 report the internal consistency and 16 report the test-retest reliability of such instruments. While the overall internal consistency is described as “adequate” (with Cronbach’s α values between 0.61 and 0.86), there is no comment regarding the overall test-retest reliability. This might be due to the circumstance that there were fewer studies reporting test-retest reliability, and those that did varied in interval lengths and outcome units.

Internal consistency can be defined as “the coherence (or redundancy) of the components of a scale” and is commonly measured in Cronbach’s α (McCrae et al., 2011). Test-retest reliability “refers to the systematic examination of consistency, reproducibility, and agreement among two or more measurements of the same individual, using the same tool, under the same conditions” (Aldridge et al., 2017)

Given the importance of empathy measurement (e.g., in clinical assessment and treatment of psychopathological disorders), it is apparent that instruments should be as reliable as possible. Further, instruments with low reliability might add additional vagueness to the already disputed field of empathy research.

Problems in the determination of reliability can stem from different sources. There are multiple formulas to calculate test-retest reliability, such as Pearson’s *r*, Bland-Altman plots and the Intraclass correlation coefficient (ICC) – with the latter being recommended to use in test-retest -reliability studies (Weir, 2005; Koo & Li, 2016). As there are 10 different forms of the ICC, researchers have to cautiously select the one suited to their respective study plans.

In studies that use the ICC as outcome, a statement on which form of the ICC was calculated and what factors led to that choice is often missing. Additionally, confidence intervals are important for the interpretation of the ICC, but not always reported (Aldridge et al., 2017; Koo & Li, 2016). Moreover, it is rare for studies on test-retest reliability to take the estimated reliability of a measure into account when performing a power-analysis, thus possibly underestimating the needed sample size (Hedge et al., 2018).

Aldridge et al. (2017) recommend to state and justify what values of a test-retest difference will be considered as agreeable, before analyzing the data, in order to not bias the interpretation. When interpreting results, it must also be kept in mind that the ICC is influenced by the variability of the sample. The less variable a sample is, the lower the ICC will be (Hedge et al., 2018). This is important when, for example, comparing the results of a measurement instrument in two populations with different homogeneity, such as a clinical and a non-clinical population. The sample with higher homogeneity will produce a lower ICC, which does not mean that the instrument is less reliable for that group (Aldridge et al., 2017).

Watson (2004) discusses other problems that are common in reliability studies. Instability between two administrations of a psychological trait measurement can consist of measurement error and true change – so it is important to be able to distinguish them. This can be done by choosing a meaningful retest interval – with respect to what construct is being measured – that is long enough that memory effects can be ruled out, but short enough that no true change should be expected. With an interval like this, all instability should be attributable to measurement error. Based on Cattel’s definition of *dependability*, Chmielewski and Watson (2009) argue for an interval between two weeks and two months (longer intervals would be seen as *stability*, and contain both measurement error and true change). But, as the authors note, it would be unrealistic to assume that there is a single point of time that precisely differentiates short-term dependability from long-term stability.

Another recommendation of Watson (2004) is the use of large sample sizes – he suggests N = 300 as a minimum, to obtain confidence intervals that are narrow enough to analyze the conditions that influence stability. Also, the author calls for the use of benchmark scales, i.e. collecting retest data not only with one instrument, but also with others measuring the same construct and different constructs, to compare between those.

While the aforementioned considerations might help to produce more justified and reliable measures of reliability, it must be kept in mind that studies are typically limited by resource constraints. For example, Lakens (2022) names time, money and small numbers of possible participants as factors that affect the sample size. The design of a study is thus often the result of a trade-off between the costs of data collection and the value that this information might contain. The author also notes that sample sizes are often stated, but rarely justified in studies, making it difficult to evaluate how informative the study might be.

## Aim of the present study

The aim of the present study is to perform a reliability analysis of several self-report and behavioral empathy measures, assessing the internal consistency and test-retest reliability. As this is an observational approach, no hypothesis will be tested. However, the detected estimates (or their confidence intervals, respectively) will be compared to the estimates detected by published literature. The measures will be rank ordered by their test-retest reliability estimates. Recommendations will be made, regarding which of them are the most/least reliable to use.

# Method

## Selection of measures

To obtain a reasonable trade-off between high-quality information (i.e., a sample big enough to make statistical inferences on as many empathy measures as possible), financial resources and participant burden, an adequate number of measures for the reliability analysis was determined. The selection of possible measures was based on several considerations. The materials of the measures (such as questions, answer formats, pictures) had to be freely available. Also, the measures had to be applicable in an online survey without a lot of effort (e.g., measures that require feedback to the participants or use unusual answer formats were not seen as suitable). Additionally, the measures should be of significance for the current empathy research. This was ensured by taking the number of psychometric studies as reported by the meta-analysis of de Lima and de Lima Osório (2021) and the number of citations of the measures’ original papers (identified with Google Scholar) into account. Table 1 shows the identified measures.

| **Table 1** |  |  |
| --- | --- | --- |
| *Possible empathy measures* |  |  |
| Measure | No. of psychometric studies | No. of citations |
| Empathy Quotient  (Baron-Cohen & Wheelwright, 2004) | 11 | 5289 |
| Interpersonal Reactivity Index (Davis, 1983) | 10 | 13258 |
| Questionnaire of Cognitive and Affective Empathy  (Reniers et al., 2011) | 5 | 915 |
| Toronto Empathy Questionnaire  (Spreng et al., 2009) | 2 | 1251 |
| Basic Empathy Scale  (Jolliffe & Farrington, 2006) | 1 | 1936 |
| Pictorial Empathy Test (Lindeman et al., 2018) | 1 | 23 |
| Single Item Trait Empathy Scale (Konrath et al., 2018) | 1 | 69 |
| Emotion Specific Empathy Questionnaire (Olderbak et al., 2014) | 0 | 68 |
| Reading the mind in the eyes Test (Baron-Cohen et al., 2001) | 0 | 7212 |
| Faces Test  (Baron-Cohen et al., 1997) | 0 | 1436 |

*Note.* Number of psychometric studies in the years 2009 – 2019 as reported by de Lima and de Lima Osório (2021). Number of citations identified with Google Scholar on June 30th 2023.

The Pictorial Empathy Test (Lindeman et al., 2018), the Single Item Trait Empathy Scale (Konrath et al., 2018) and the Emotion Specific Empathy Questionnaire (Olderbak et al., 2014) show only one or no psychometric studies and just few citations. They were nonetheless included because of interesting approaches: The Pictorial Empathy Test measures affective empathy by asking the participant to report his arousal while watching emotional pictures. The Single Item Trait Empathy Scale aims to measure empathy with just a single self-report question. The Emotion Specific Empathy Questionnaire assesses cognitive and affective empathy separately for each of the six basic emotions. Also, the Reading The Mind In The Eyes Test (S. Baron-Cohen et al., 2001) and the Faces Test (Simon Baron-Cohen et al., 1997) are not mentioned in the meta-analysis (neither are any other behavioral tasks). They were included in the list as they were the only identified behavioral empathy measures that met the criteria.

To estimate the amount of time that it takes to complete each measure named in Table 1, a pilot study with 10 participants was carried out. The participants were recruited through the online research platform prolific.co. The survey was implemented on the online survey platform qualtrics.com. Exclusion criteria were an age lower than 18 years and not being a native English speaker. Two attention checks, consisting of items telling the participants exactly what to do (e.g. “Please choose the option 'Agree Strongly'. This is an attention check.”) were used in the survey. All 10 participants passed both attention checks. They were paid 9£ for their participation. Table 2 shows the completion times for each measure.

| **Table 2** |  |  |  |  |
| --- | --- | --- | --- | --- |
| *Duration of empathy measures* | | |  |  |
| Measure | Mean Completion  Time (s) | Mean Completion Time (min) | Lowest Completion  Time (min) | Highest Completion Time (min) |
| EQ | 506,55 | 8,44 | 3,37 | 15,87 |
| IRI | 294,25 | 4,90 | 1,99 | 8,54 |
| QCAE | 312,81 | 5,21 | 1,82 | 11,31 |
| TEQ | 124,43 | 2,07 | 0,90 | 4,01 |
| BES-A | 160,45 | 2,67 | 1,04 | 5,24 |
| PET | 69,55 | 1,16 | 0,50 | 2,62 |
| SITES | 13,21 | 0,22 | 0,06 | 0,57 |
| ESE | 596,33 | 9,94 | 4,32 | 20,50 |
| EYES | 413,70 | 6,90 | 3,31 | 18,49 |
| FACES | 118,30 | 1,97 | 1,09 | 4,96 |

*Note.* EQ = Empathy Quotient (Baron-Cohen & Wheelwright, 2004), IRI = Interpersonal Reactivity Index (Davis, 1983), QCAE = Questionnaire of Cognitive and Affective Empathy (Reniers et al., 2011), TEQ = Toronto Empathy Questionnaire (Spreng et al., 2009), BES = Basic Empathy Scale (Jolliffe & Farrington, 2006) PET = Pictorial Empathy Test (Lindeman et al., 2018), SITES = Single Item Trait Empathy Scale (Konrath et al., 2018), EYES = Reading the Mind in the Eyes Test (Baron-Cohen et al., 2001), FACES = Faces Test (Baron-Cohen et al., 1997)

Based on this, the measures Interpersonal Reactivity Index, Empathy Quotient, Single Item Trait Empathy Scale, Pictorial Empathy Test and Reading the Mind in the Eyes Test were selected, which have an approximately completion time of 22 minutes altogether.

### ***Interpersonal Reactivity Index***

The Interpersonal Reactivity Index (IRI(Davis, 1980) conceptualizes empathy as “a set of constructs, related in that they all concern responsivity to others but are also clearly discriminable from each other”. It consists of 28 items that are divided into four subscales with seven items each: The Perspective-Taking (PT) scale describes the “tendency to spontaneously adopt the psychological point of view of others”. The Fantasy (FS) scale “measures the tendency to transpose oneself into the feelings and actions of fictitious characters”. The other two subscales center around emotional reactions associated with empathy: The Empathic Concern (EC) scale assesses “the tendency to experience feelings of warmth, compassion, and concern for other people”, while the Personal Distress (PD) scale focuses on “one's own feelings of personal unease and discomfort in reaction to the emotions of others”. The 28 items describe behaviors and feelings and have to be answered on a five-point Likert scale ranging from “Does not describe me very well” to “Describes me very well”. For each answer, zero to four points are given, resulting in a total score with a possible range from zero to 112.

The test-retest reliability and internal consistency are seen as “satisfactory” by the author, with retest-reliability correlations ranging from .61 to .79 over an interval of 60 to 75 days, and internal consistencies ranging from .70 to .78 (Davis, 1980).

A more current analysis of the IRI’s psychometric properties found both “appropriate” test-retest reliability (r = .67 to .89 with an interval of 60 days) and internal consistency (Cronbach’s α of .70 to .76, Fernández-Abascal et al., 2013).

### ***Empathy Quotient***

The Empathy Quotient (EQ, (Simon Baron-Cohen & Wheelwright, 2004) was initially developed to compare the empathy ability of persons with Asperger Syndrome and high-functioning autism with those of healthy controls. The authors argue that empathy consists of an affective component (“Feeling an appropriate emotion triggered by seeing/learning of another’s emotion”) and a cognitive component (“Understanding and/or predicting what someone else might think, feel, or do”), which cannot be easily separated. This is reflected in the items of the EQ, which do not distinguish between affective/cognitive or other subgroups of empathy. The 60 items (of which 20 are filler items that do not affect the total score) describe behaviors and experiences that the respondent has to answer on a 4-point Likert scale with the options “strongly agree”, “slightly agree”, “slightly disagree” or “strongly disagree”. Each answer is scored between zero and two points, resulting in a total score with a possible range from zero to 80.

Internal consistency is mentioned to be “high” with a Cronbach’s α of .92, while the retest reliability is stated as r = .97 over an interval of 12 months. Kosonogov (2014) found similar results, reporting a Cronbach’s α of .85 and r = .94 retest-reliability with a two-week interval.

### ***Pictorial Empathy Test***

The Pictorial Empathy Test (PET, Lindeman et al., 2018) aims to measure only the affective components of empathy. The authors follow the view of Davis (2006), where these affective components are seen as “the sharing of other people’s emotions and the tendency to experience personal unease when witnessing the distress of other people in particular”. In that manner, the PET presents its respondents seven pictures of individuals in distress. For each picture, the respondents have to answer the question “How emotionally moving do you find the photograph?” on a 5-point Likert scale with the options “not at all”, “a little bit”, “it arouses some feelings”, “quite a lot” or “very much”. Each answer scores one to five points and the total PET score is obtained by calculating the mean of all the answers.

The authors of the PET report “high” internal consistency (Cronbach’s α = .9) and “good” test-retest reliability (r = .77 with a seven-months interval). The only other identified study that reports on the psychometric properties of the PET is the cultural adaption in the Spanish population by Baliyan et al. (2023). The authors report a Cronbach’s α of .77; a test-retest coefficient is not mentioned.

### ***Single Item Trait Empathy Scale***

The Single Item Trait Empathy Scale (SITES, Konrath et al., 2018) is characterized as a trait measure of empathy, in the sense that it “measures the extent to which people tend to vary in their empathic responses to others across situations”. It consists of only one item, which reads: “To what extent does the following statement describe you: ‘I am an empathetic person’”. Respondents use a five-point Likert scale that ranges from “Not very true of me” to “Very true of me”, or from one point to five points, respectively. The authors describe the SITES as reliable and valid and recommend the use when a longer empathy measure cannot be applied. They also emphasize that the SITES cannot assess different dimensions of empathy, and thus should only be used when a single dimension measure of empathy is sufficient.

The test-retest reliability of the SITES over an interval of two weeks is stated as r =.57, while it is r = .62 over an interval of 6 months. Studies reporting on psychometric properties of the scale other than the original paper of Konrath et al. (2018) could not be identified.

### ***Reading the Mind in the Eyes Test***

Similar to the EQ, the Reading the Mind in the Eyes Test (EYES, S. Baron-Cohen et al., 1997; S. Baron-Cohen et al., 2001) was developed to assess the Theory of Mind-abilities of individuals with Asperger Syndrome or high-functioning autism and compare it with those of healthy individuals. While Theory of Mind is not the same as empathy, there are overlaps between the two constructs, as the authors note. Similarly, Olderbak et al. (2015) state that Theory of Mind and cognitive empathy are often seen as conceptually similar or even equivalent, as both involve inferring the mental state of others.

In contrast to the aforementioned scales, the EYES is not a self-report measure. Participants are presented 36 black-and-white pictures of the eye region of humans. They have to choose the correct word (out of four options) that describes the mental state of the depicted individual. They score one point for each correct answer, resulting in a total score from zero to 36 points.

The authors of the EYES do not provide reliability estimates in their original paper.

While not stating specific values, Olderbak et al. (2015) report that according to published estimates, the EYES has typically poor internal consistency and acceptable test-retest reliability. As Vellante et al. (2013) state, information on reliability of the EYES is rarely reported and replication of these findings is lacking. In their systematic review of 23 studies, only four studies give detailed information on the test’s internal consistency (Cronbach’s α ranging from .58 to .70), and only study on the test-retest reliability (ICC =.65, 95% CI: .49 to .77, two-week interval). Vellante et al. (2013) also performed a reliability analysis of the Italian version of the EYES, finding internal consistency of Cronbach’s α = .605 and test-retest reliability of ICC = .833 (95%CI = .745 to .902, 30-day interval).

In general, the EYES should be used with caution when measuring cognitive empathy, as the EYES score and cognitive empathy are weakly correlated, both positive and negative (Olderbak et al., 2015).

## Sample size analysis

To determine the required sample size for an ICC analysis, the function “calculateIccSampleSize” of the R-package “ICC.Sample.Size” (Rathbone et. al., 2015) was used. Power was set to .80 and α to .05. Table 3 shows the different ICC values, with an error of ± 0.1 within a 95% confidence interval, and the respective number of participants that is needed to detect these values.

| **Table 3** |  |
| --- | --- |
| *Required sample sizes for different estimated ICC values* | |
| Estimated ICC (95% CI of ± 0.1) | Required N |
| 0.55 | 441 |
| 0.60 | 381 |
| 0.65 | 320 |
| 0.70 | 260 |
| 0.75 | 202 |
| 0.80 | 148 |
| 0.85 | 99 |
| 0.90 | 58 |
| 0.95 | 25 |

As the required sample size depends on the ICC value that is estimated to be found, reported reliability coefficients were taken into account. The reported reliability values of the selected empathy measures vary widely – from r = .57 (SITES) to r = .97 (EQ) – which means that a sample size between around 441 and 25 participants would be needed. As the budgeting of this study allowed for 134 participants to be included in two administrations of the measures, N = 134 was the targeted sample size.

## Selection of ICC-version

The guideline of Koo and Li (2016) was used to determine the appropriate version of the ICC for a test-retest analysis, which is “Two-way mixed effects, absolute agreement, single rater/measurement” using the McGraw and Wong (1996) convention (this is in contrast to the preregistration, where the appropriate ICC-version was mistakenly stated as “Two-way random effects, absolute agreement, single rater/measurement”). There is no corresponding ICC version in the Shrout and Fleiss convention (1979).

## Implementation of the study

For this observational study, the empathy measures IRI EQ, PET, SITES and EYES were implemented on the survey platform qualtrics.com. Introduction texts and answer formats of the measures were adapted to match the format of an online survey – e.g., “Circle your answer” was changed to “Click on the answer”. Unscored example items were implemented in the survey when they were described in the original papers. The word-definitions that are available as a handout in the pen-and-paper version of the EYES were written directly underneath the answer options for each question.

Participants were recruited through the online research platform prolific.co. Exclusion criteria were an age lower than 18 years and not being a native English speaker. They were paid 9£ for their participation in both time points. Two attention checks, consisting of items telling the participants exactly what to do (e.g. “Please choose the option 'Agree Strongly'. This is an attention check.”) were used in the survey. The second part of the survey was performed ten days after the first one.

# Results

In total, 100 participants completed both parts of the study. Of the initial 134 participants that completed the first administration (t1), 35 did not take part in the second administration (t2, around 26% drop-out). Two participants failed the attention checks, their data were excluded from further analysis. The final sample (n = 98) was 28.6% male and 1% self-identified. The mean age was 38.58 (SD = 12.94) years, with a range from 20 to 75 years.

The ICC of each measure was planned to be calculated with the function “ICC” of the R-package “psych”. But, since the outcomes did not clearly specify the results for the needed ICC-Version (“Two-way mixed effects, absolute agreement, single rater/measurement”), the function “icc” of the R-Package “irr” was used instead. Cronbach’s α of each measure was calculated with the function “alpha” of the R-package “psych”. The “raw\_alpha” values are reported here and confidence intervals are based on the “Feldt”-definition. Means, SDs and medians of the measures at t2 are reported. The R-Script can be found in the appendix. Ratings of ICC and Cronbach’s α (or their confidence intervals, respectively) are based on Koo and Li (2016), and George and Mallery (2007), respectively. Table 4 shows an overview of the results.

| **Table 4** |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| *Reliablity of empathy measures* | | |  |  |  |  |
| Measure | ICC  (95%-CI) | ICC rating | t1 Cronbach's α (95%-CI) | t1 Cronbach's α  rating | t2 Cronbach's α (95%-CI) | t2  Cronbach's α  rating |
| IRI | .90  (.86 - .93) | good -  excellent | .88  (.84 - .91) | good -  excellent | .88  (.84 - .91) | good -  excellent |
| IRI - PT | .81  (.73 - .87) | moderate -  good | .78  (.71 - .78) | acceptable | .80  (.73 - .85) | acceptable -  good |
| IRI - FS | .88  (.82 - .92) | good -  excellent | .83  (.77 - .87) | acceptable -  good | .85  (.80 - .89) | good |
| IRI - EC | .87  (.82 - .91) | good -  excellent | .85  (.80 - .89) | good | .89  (.85 - .92) | good -  excellent |
| IRI - PD | .88  (.82 - .92) | good -  excellent | .79  (.72 - .85) | acceptable -  good | .86  (.82 - .90) | good |
| EQ | .86  (.80 - .90) | good | .88  (.84 - .91) | good -  excellent | .88  (.84 - .91) | good -  excellent |
| PET | .86  (.80 - .90) | good | .92  (.89 - .94) | good -  excellent | .91  (.88 - .94) | good -  excellent |
| SITES | .77  (.67 - .84) | moderate -  good | - | - | - | - |
| EYES | .71  (.59 - .80) | moderate - good | .67  (.57 - .76) | poor -  acceptable | .76  (.69 - .82) | questionable -  good |

*Note*. IRI = Interpersonal Reactivity Index, PT = Perspective Taking, FS = Fantasy, EC = Empathic Concern, PD = Personal Distress, EQ = Empathy Quotient, PET = Pictorial Empathy Test SITES = Single Item Trait Empathy Scale EYES = Reading the Mind in the Eyes Test. t1 = First administration of the measures, t2 = second administration of the measures (10-day interval). ICC rating based on Koo and Li (2016), Cronbach’s α rating based on George and Mallery (2007).

## Interpersonal Reactivity Index (IRI)

The complete IRI with all four subscales showed an ICC of .90 (95%-CI: .86 - .93).

Cronbach’s α was .88 (95%-CI: .84 - .91) at t1 and t2. I.e., test-retest reliability was good or excellent, and internal consistency was also good or excellent. All of the IRI’s subscales showed good or excellent test-retest reliability, with exception of the Perspective-taking (PT) subscale, which showed moderate or good test-retest reliability. Internal consistency ranged from acceptable to excellent. Detailed reliabilities of the four subscales are presented in Table 4. Mean score of the total IRI was 70.18 (SD = 14.16) of a possible score from zero to 112. The median was 70.5. Mean scores of the subscales were: Perspective-Taking (PT): 19.48 (SD = 4.44), median = ;20 Fantasy (FS): 17.09 (SD = 5.83) median = 17; Empathic Concern (EC): 20.66 (SD = 5.11), median = 21 and Personal Distress (PD): 12.95 (SD = 5.84), median = 13. All subscales have a possible score from zero to 28.

## Empathy Quotient (EQ)

ICC of the EQ was .86 (95%-CI: .80 - .90), Cronbach’s α was .88 (95%-CI: .84 - .91) at both time points – indicating good test-retest reliability and good or excellent internal consistency. The mean score of the EQ was 44.84 (SD = 11.75) out of the possible range from zero to 80. The median was 46.

## Pictorial Empathy Task (PET)

The PET showed an ICC of .86 (95%-CI: .80 - .90) and Cronbach’s α of .92 (95%-CI: .89 - .94) at t1 and .91 (95%-CI: .88 - .94) at t1. Thus, test-retest reliability can be seen as good and internal consistency can be seen as good or excellent. The mean score of the PET – which is calculated as the mean of all seven items – was 3.76 (SD = .95) out of a possible score from one to five. The median was 4.

## Single Item Trait Empathy Scale (SITES)

The SITES showed an ICC of .77 (95%-CI: .67 - .84) – indicating moderate or good test-retest reliability. Cronbach’s α was not calculated, because the SITES consists of only one item. The mean score was 4.2 (SD = .72) of a maximum of 5. The median was 4.

## Reading the Mind in the Eyes Test (EYES)

The ICC of the EYES was .71 (95%-CI: .59 - .80), Cronbach’s α was .67 (95%-CI: .57 - .76) at t1 and .76 (95%-CI: .69 - .82) at t2. Test-retest reliability can thus be described as moderate or good, while the internal consistency was poor or acceptable at t1 and questionable or good at t2. On average, participants scored 26.12 (SD = 5.1) out of the possible 36 points. Median was 27.

# Discussion

# References

**References**

Aldridge, V. K., Dovey, T. M., & Wade, A. (2017). Assessing Test-Retest Reliability of Psychological Measures. *European Psychologist*, *22*(4), 207–218. https://doi.org/10.1027/1016-9040/a000298

Baldner, C., & McGinley, J. (2020). Self-report empathy scales lack consistency: Evidence from exploratory and confirmatory factor analysis. *TPM - Testing*, *27*(1), 103–128. https://doi.org/10.4473/TPM27.1.7

Baliyan, S., Cimadevilla, J. M., Bustillos, A., Escamilla, J. C., Leiman, M., Sandi, C., & Venero, C. (2023). Cultural Adaptation, Validation, and Psychometric Description of the Pictorial Empathy Test (PET) in the Spanish Population. *European Journal of Psychological Assessment*, *39*(2), 141–152. https://doi.org/10.1027/1015-5759/a000690

Baron-Cohen, S [S.], Jolliffe, T., Mortimore, C., & Robertson, M. (1997). Another advanced test of theory of mind: Evidence from very high functioning adults with autism or asperger syndrome. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *38*(7), 813–822. https://doi.org/10.1111/j.1469-7610.1997.tb01599.x

Baron-Cohen, S [S.], Wheelwright, S [S.], Hill, J., Raste, Y., & Plumb, I. (2001). The "Reading the Mind in the Eyes" Test revised version: A study with normal adults, and adults with Asperger syndrome or high-functioning autism. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *42*(2), 241–251. https://pubmed.ncbi.nlm.nih.gov/11280420/

Baron-Cohen, S [Simon], & Wheelwright, S [Sally] (2004). The empathy quotient: An investigation of adults with Asperger syndrome or high functioning autism, and normal sex differences. *Journal of Autism and Developmental Disorders*, *34*(2), 163–175. https://doi.org/10.1023/b:jadd.0000022607.19833.00

Baron-Cohen, S [Simon], Wheelwright, S [Sally], & Jolliffe, a. T. (1997). Is There a "Language of the Eyes"? Evidence from Normal Adults, and Adults with Autism or Asperger Syndrome. *Visual Cognition*, *4*(3), 311–331. https://doi.org/10.1080/713756761

Batson, C. D., Lishner, D. A., & Stocks, E. L. (2014). *The Empathy–Altruism Hypothesis*. https://academic.oup.com/edited-volume/28077/chapter/212115460?login=true https://doi.org/10.1093/oxfordhb/9780195399813.013.023

Butters, R. P. (2010). *Meta-analysis of empathy training programs for client populations*. https://www.researchgate.net/publication/50518340\_Meta-analysis\_of\_empathy\_training\_programs\_for\_client\_populations

Chmielewski, M., & Watson, D. (2009). What is being assessed and why it matters: The impact of transient error on trait research. *Journal of Personality and Social Psychology*, *97*(1), 186–202. https://doi.org/10.1037/a0015618

Coplan, A. (2011). *Understanding Empathy:Its Features and Effects*. https://academic.oup.com/book/7216/chapter/151858067 https://doi.org/10.1093/acprof:oso/9780199539956.003.0002

Cuff, B. M., Brown, S. J., Taylor, L., & Howat, D. J. (2016). Empathy: A Review of the Concept. *Emotion Review*, *8*(2), 144–153. https://doi.org/10.1177/1754073914558466

Davis, M. H. (1980). A Multidimensional Approach to Individual Differences in Empathy. *Journal of Personality and Social Psychology*. https://www.semanticscholar.org/paper/A-Multidimensional-Approach-to-Individual-in-Davis-Davis/c717eb4e913c3249eac18d0fba13a1aa02d60dad

de Lima, F. F., & de Lima Osório, F. (2021). Empathy: Assessment Instruments and Psychometric Quality - A Systematic Literature Review With a Meta-Analysis of the Past Ten Years. *Frontiers in Psychology*, *12*, 781346. https://doi.org/10.3389/fpsyg.2021.781346

de Waal, F. B. M. (2008). Putting the altruism back into altruism: The evolution of empathy. *Annual Review of Psychology*, *59*, 279–300. https://doi.org/10.1146/annurev.psych.59.103006.093625

Decety, J., & Moriguchi, Y. (2007). The empathic brain and its dysfunction in psychiatric populations: Implications for intervention across different clinical conditions. *BioPsychoSocial Medicine*, *1*(1), 22. https://doi.org/10.1186/1751-0759-1-22

Eklund, J. H., & Meranius, M. S. (2021). Toward a consensus on the nature of empathy: A review of reviews. *Patient Education and Counseling*, *104*(2), 300–307. https://doi.org/10.1016/j.pec.2020.08.022

Fernández-Abascal, E. G., Cabello, R., Fernández-Berrocal, P., & Baron-Cohen, S [Simon] (2013). Test-retest reliability of the 'Reading the Mind in the Eyes' test: A one-year follow-up study. *Molecular Autism*, *4*(1), 33. https://doi.org/10.1186/2040-2392-4-33

George, D., & Mallery, P. (2007). *SPSS for Windows step by step* (7th ed.). Allyn & Bacon. https://www.researchgate.net/publication/234827666\_SPSS\_for\_Windows\_Step-by-Step\_A\_Simple\_Guide\_and\_Reference\_140\_update\_7th\_Edition

Gerdes, K. E., Segal, E. A., & Lietz, C. A. (2010). Conceptualising and Measuring Empathy. *British Journal of Social Work*, *40*(7), 2326–2343. https://doi.org/10.1093/bjsw/bcq048

Hall, J. A., & Schwartz, R. (2019). Empathy present and future. *The Journal of Social Psychology*, *159*(3), 225–243. https://doi.org/10.1080/00224545.2018.1477442

Hedge, C., Powell, G., & Sumner, P. (2018). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods*, *50*(3), 1166–1186. https://doi.org/10.3758/s13428-017-0935-1

Jolliffe, D., & Farrington, D. P. (2006). Development and validation of the Basic Empathy Scale. *Journal of Adolescence*, *29*(4), 589–611. https://doi.org/10.1016/j.adolescence.2005.08.010

Konrath, S., Meier, B. P., & Bushman, B. J. (2018). Development and validation of the Single Item Trait Empathy Scale (SITES). *Journal of Research in Personality*, *73*, 111–122. https://doi.org/10.1016/j.jrp.2017.11.009

Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine*, *15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Kosonogov, V. (2014). The psychometric properties of the Russian version of the Empathy Quotient. *Psychology in Russia: State of the Art*, *7*(1), 96–104. https://doi.org/10.11621/pir.2014.0110

Lakens, D. (2022). Sample Size Justification. *Collabra: Psychology*, *8*(1), Article 33267. https://doi.org/10.1525/collabra.33267

Lam, T. C. M., Kolomitro, K., & Alamparambil, F. C. (2011). Empathy Training: Methods, Evaluation Practices, and Validity. *Journal of MultiDisciplinary Evaluation*, *7*(16), 162–200. https://doi.org/10.56645/jmde.v7i16.314

Lindeman, M., Koirikivi, I., & Lipsanen, J. (2018). Pictorial Empathy Test (PET). *European Journal of Psychological Assessment*, *34*(6), 421–431. https://doi.org/10.1027/1015-5759/a000353

McCrae, R. R., Kurtz, J. E., Yamagata, S., & Terracciano, A. (2011). Internal consistency, retest reliability, and their implications for personality scale validity. *Personality and Social Psychology Review : An Official Journal of the Society for Personality and Social Psychology, Inc*, *15*(1), 28–50. https://doi.org/10.1177/1088868310366253

Mcdonald, N. M., & Messinger, D. S. (2011). *The Development of Empathy: How, When, and Why*. https://www.researchgate.net/publication/267426505\_The\_Development\_of\_Empathy\_How\_When\_and\_Why

Melchers, M., Montag, C., Markett, S., & Reuter, M. (2015). Assessment of empathy via self-report and behavioural paradigms: Data on convergent and discriminant validity. *Cognitive Neuropsychiatry*, *20*(2), 157–171. https://doi.org/10.1080/13546805.2014.991781

Miyazono, K., & Inarimori, K. (2021). Empathy, Altruism, and Group Identification. *Frontiers in Psychology*, *12*, 749315. https://doi.org/10.3389/fpsyg.2021.749315

Murphy, B. A., & Lilienfeld, S. O. (2019). Are self-report cognitive empathy ratings valid proxies for cognitive empathy ability? Negligible meta-analytic relations with behavioral task performance. *Psychological Assessment*, *31*(8), 1062–1072. https://doi.org/10.1037/pas0000732

Neumann, D. L., Chan, R. C., Boyle, G. J., Wang, Y., & Rae Westbury, H. (2015). *Measures of Empathy*. https://reader.elsevier.com/reader/sd/pii/B9780123869159000103?token=261BD768D6A8DC72D5200DF611FC82CC9EDC8F0A84C4E86D88427012077C87F79288D0BC6719360A68F9055387F6AC38&originRegion=eu-west-1&originCreation=20221205111618 https://doi.org/10.1016/B978-0-12-386915-9.00010-3

Olderbak, S., Sassenrath, C., Keller, J., & Wilhelm, O. (2014). An emotion-differentiated perspective on empathy with the emotion specific empathy questionnaire. *Frontiers in Psychology*, *5*, 653. https://doi.org/10.3389/fpsyg.2014.00653

Olderbak, S., Wilhelm, O., Olaru, G., Geiger, M., Brenneman, M. W., & Roberts, R. D. (2015). A psychometric analysis of the reading the mind in the eyes test: Toward a brief form for research and applied settings. *Frontiers in Psychology*, *6*, 1503. https://doi.org/10.3389/fpsyg.2015.01503

Reniers, R. L. E. P., Corcoran, R., Drake, R., Shryane, N. M., & Völlm, B. A. (2011). The QCAE: A Questionnaire of Cognitive and Affective Empathy. *Journal of Personality Assessment*, *93*(1), 84–95. https://doi.org/10.1080/00223891.2010.528484

Revelle, W., & Condon, D. M. (2019). Reliability from α to ω: A tutorial. *Psychological Assessment*, *31*(12), 1395–1411. https://doi.org/10.1037/pas0000754

Spreng, R. N., McKinnon, M. C., Mar, R. A., & Levine, B. (2009). The Toronto Empathy Questionnaire: Scale development and initial validation of a factor-analytic solution to multiple empathy measures. *Journal of Personality Assessment*, *91*(1), 62–71. https://doi.org/10.1080/00223890802484381

Sunahara, C. S., Rosenfield, D., Alvi, T., Wallmark, Z., Lee, J., Fulford, D., & Tabak, B. A. (2022). Revisiting the association between self-reported empathy and behavioral assessments of social cognition. *Journal of Experimental Psychology: General*, *151*(12), 3304–3322. https://doi.org/10.1037/xge0001226

Vellante, M., Baron-Cohen, S [Simon], Melis, M., Marrone, M., Petretto, D. R., Masala, C., & Preti, A. (2013). The "Reading the Mind in the Eyes" test: Systematic review of psychometric properties and a validation study in Italy. *Cognitive Neuropsychiatry*, *18*(4), 326–354. https://doi.org/10.1080/13546805.2012.721728

Watson, D. (2004). Stability versus change, dependability versus error: Issues in the assessment of personality over time. *Journal of Research in Personality*, *38*(4), 319–350. https://doi.org/10.1016/j.jrp.2004.03.001

Weir, J. P. (2005). Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *Journal of Strength and Conditioning Research*, *19*(1), 231–240. https://doi.org/10.1519/15184.1

# Appendix

**R-code for the calculation of ICC and Cronbach’s α**

# Set working directory

setwd("D:/Studium/7. Semester/S Projektseminar MTI/Bachelorthesis/Timepoint 1 and 2 Qualtrics und R")

#

# Install needed packages

install.packages("dplyr")

library("dplyr")

install.packages("psych")

library("psych")

install.packages("irr")

library("irr")

install.packages("lme4")

library("lme4")

#

options(scipen=999)

#

# Read Data

timepoint1 <- read.csv("Test-Retest - 1 of 2\_July 10, 2023\_08.39.csv")

timepoint2 <- read.csv("Test-Retest - 2 of 2\_July 20, 2023\_23.39.csv")

#

dim(timepoint1)

names(timepoint1)

#

# Delete Tests and unnecessary rows

timepoint1 <- timepoint1[-c(1:7),]

timepoint2 <- timepoint2[-c(1:4),]

#

############## Filter variables of interest ###################

timepoint1.use <- timepoint1 %>% select(Consent\_question,Age\_1,Gender,Gender\_3\_TEXT,Prolific\_ID\_Entry,

SITES,IRI\_1,IRI\_2,IRI\_3,IRI\_4,IRI\_5,IRI\_6,IRI\_7,IRI\_8,IRI\_9,IRI\_10,IRI\_11,IRI\_12,IRI\_13,IRI\_14,IRI\_15,

IRI\_16,IRI\_17,IRI\_18,IRI\_19,IRI\_20,IRI\_21,IRI\_22,IRI\_23,IRI\_24,IRI\_25,IRI\_26,IRI\_27,IRI\_28,IRI\_29,

EQ\_1,EQ\_2,EQ\_3,EQ\_4,EQ\_5,EQ\_6,EQ\_7,EQ\_8,EQ\_9,EQ\_10,EQ\_11,EQ\_12,EQ\_13,EQ\_14,EQ\_15,EQ\_16,EQ\_17,EQ\_18,

EQ\_19,EQ\_20,EQ\_21,EQ\_22,EQ\_23,EQ\_24,EQ\_25,EQ\_26,EQ\_27,EQ\_28,EQ\_29,EQ\_30,EQ\_31,EQ\_32,EQ\_33,EQ\_34,EQ\_35,

EQ\_36,EQ\_37,EQ\_38,EQ\_39,EQ\_40,EQ\_41,EQ\_42,EQ\_43,EQ\_44,EQ\_45,EQ\_46,EQ\_47,EQ\_48,EQ\_49,EQ\_50,EQ\_51,EQ\_52,

EQ\_53,EQ\_54,EQ\_55,EQ\_56,EQ\_57,EQ\_58,EQ\_59,EQ\_60,EQ\_61,

Q2,Q4,Q6,Q8,Q10,Q12,Q14,

EYES\_1,EYES\_2,EYES\_3,EYES\_4,EYES\_5,EYES\_6,EYES\_7,EYES\_8,EYES\_9,EYES\_10,EYES\_11,EYES\_12,EYES\_13,EYES\_14,

EYES\_15,EYES\_16,EYES\_17,EYES\_18,EYES\_19,EYES\_20,EYES\_21,EYES\_22,EYES\_23,EYES\_24,EYES\_25,EYES\_26,EYES\_27,EYES\_28,EYES\_29,

EYES\_30,EYES\_31,EYES\_32,EYES\_33,EYES\_34,EYES\_35,EYES\_36,PROLIFIC\_PID)

#

timepoint2.use <- timepoint2 %>% select(Consent\_question,Age\_1,Gender,Gender\_3\_TEXT,Prolific\_ID\_Entry,

SITES,IRI\_1,IRI\_2,IRI\_3,IRI\_4,IRI\_5,IRI\_6,IRI\_7,IRI\_8,IRI\_9,IRI\_10,IRI\_11,IRI\_12,IRI\_13,IRI\_14,IRI\_15,

IRI\_16,IRI\_17,IRI\_18,IRI\_19,IRI\_20,IRI\_21,IRI\_22,IRI\_23,IRI\_24,IRI\_25,IRI\_26,IRI\_27,IRI\_28,IRI\_29,

EQ\_1,EQ\_2,EQ\_3,EQ\_4,EQ\_5,EQ\_6,EQ\_7,EQ\_8,EQ\_9,EQ\_10,EQ\_11,EQ\_12,EQ\_13,EQ\_14,EQ\_15,EQ\_16,EQ\_17,EQ\_18,

EQ\_19,EQ\_20,EQ\_21,EQ\_22,EQ\_23,EQ\_24,EQ\_25,EQ\_26,EQ\_27,EQ\_28,EQ\_29,EQ\_30,EQ\_31,EQ\_32,EQ\_33,EQ\_34,EQ\_35,

EQ\_36,EQ\_37,EQ\_38,EQ\_39,EQ\_40,EQ\_41,EQ\_42,EQ\_43,EQ\_44,EQ\_45,EQ\_46,EQ\_47,EQ\_48,EQ\_49,EQ\_50,EQ\_51,EQ\_52,

EQ\_53,EQ\_54,EQ\_55,EQ\_56,EQ\_57,EQ\_58,EQ\_59,EQ\_60,EQ\_61,

Q2,Q4,Q6,Q8,Q10,Q12,Q14,

EYES\_1,EYES\_2,EYES\_3,EYES\_4,EYES\_5,EYES\_6,EYES\_7,EYES\_8,EYES\_9,EYES\_10,EYES\_11,EYES\_12,EYES\_13,EYES\_14,

EYES\_15,EYES\_16,EYES\_17,EYES\_18,EYES\_19,EYES\_20,EYES\_21,EYES\_22,EYES\_23,EYES\_24,EYES\_25,EYES\_26,EYES\_27,EYES\_28,EYES\_29,

EYES\_30,EYES\_31,EYES\_32,EYES\_33,EYES\_34,EYES\_35,EYES\_36,PROLIFIC\_PID)

#

########### Rename "Q" to PET, rename attention checks items #################

timepoint1.use <- timepoint1.use %>% rename(

PET\_1 = Q2,

PET\_2 = Q4,

PET\_3 = Q6,

PET\_4 = Q8,

PET\_5 = Q10,

PET\_6 = Q12,

PET\_7 = Q14,

ATTENTION\_CHECK\_1 = IRI\_24,

# Because of the attention check item, all IRI\_items after 24 have to be renamed

IRI\_24 = IRI\_25,

IRI\_25 = IRI\_26,

IRI\_26 = IRI\_27,

IRI\_27 = IRI\_28,

IRI\_28 = IRI\_29,

ATTENTION\_CHECK\_2 = EQ\_60,

# Because of the attention check item, EQ Item 61 has to be renamed

EQ\_60 = EQ\_61

)

timepoint2.use <- timepoint2.use %>% rename(

PET\_1 = Q2,

PET\_2 = Q4,

PET\_3 = Q6,

PET\_4 = Q8,

PET\_5 = Q10,

PET\_6 = Q12,

PET\_7 = Q14,

ATTENTION\_CHECK\_1 = IRI\_24,

# Because of the attention check item, all IRI\_items after 24 have to be renamed

IRI\_24 = IRI\_25,

IRI\_25 = IRI\_26,

IRI\_26 = IRI\_27,

IRI\_27 = IRI\_28,

IRI\_28 = IRI\_29,

ATTENTION\_CHECK\_2 = EQ\_60,

# Because of the attention check item, EQ Item 61 has to be renamed

EQ\_60 = EQ\_61

)

#

# Check Attention Checks

which(timepoint1.use$ATTENTION\_CHECK\_1 != "C")

which(timepoint1.use$ATTENTION\_CHECK\_2 != "strongly agree")

which(timepoint2.use$ATTENTION\_CHECK\_1 != "C")

which(timepoint2.use$ATTENTION\_CHECK\_2 != "strongly agree")

#

# Delete cases that failed the attention checks

timepoint1.use <- timepoint1.use[-c(88:88),]

timepoint2.use <- timepoint2.use[-c(27:27),]

#

############# Change all values to numeric values - Timepoint 1 ##############

# SITES

timepoint1.use <- timepoint1.use %>% mutate(

SITES.n = recode(SITES,

'1 Not very true of me' = 1,

'2' = 2,

'3' = 3,

'4' = 4,

'5 Very true of me.' = 5,

.default = 0

)

)

# Check if change worked

table(timepoint1.use$SITES, timepoint1.use$SITES.n)

#

# Interpersonal reactivity Index - normal scored items

timepoint1.use <- timepoint1.use %>% mutate(

IRI\_1.n = recode(IRI\_1,

'A - DOES NOT DESCRIBE ME VERY WELL' = 0,

'B' = 1,

'C' = 2,

'D' = 3,

'E - DESCRIBES ME VERY WELL' = 4,

.default = 0),

IRI\_2.n = recode(IRI\_2,

'A - DOES NOT DESCRIBE ME VERY WELL' = 0,

'B' = 1,

'C' = 2,

'D' = 3,

'E - DESCRIBES ME VERY WELL' = 4,

.default = 0),

IRI\_5.n = recode(IRI\_5,

'A - DOES NOT DESCRIBE ME VERY WELL' = 0,

'B' = 1,

'C' = 2,

'D' = 3,

'E - DESCRIBES ME VERY WELL' = 4,

.default = 0),

IRI\_6.n = recode(IRI\_6,

'A - DOES NOT DESCRIBE ME VERY WELL' = 0,

'B' = 1,

'C' = 2,

'D' = 3,

'E - DESCRIBES ME VERY WELL' = 4,

.default = 0),

IRI\_8.n = recode(IRI\_8,

'A - DOES NOT DESCRIBE ME VERY WELL' = 0,

'B' = 1,

'C' = 2,

'D' = 3,

'E - DESCRIBES ME VERY WELL' = 4,

.default = 0),

IRI\_9.n = recode(IRI\_9,

'A - DOES NOT DESCRIBE ME VERY WELL' = 0,

'B' = 1,

'C' = 2,

'D' = 3,

'E - DESCRIBES ME VERY WELL' = 4,

.default = 0),

IRI\_10.n = recode(IRI\_10,

'A - DOES NOT DESCRIBE ME VERY WELL' = 0,

'B' = 1,

'C' = 2,

'D' = 3,

'E - DESCRIBES ME VERY WELL' = 4,

.default = 0),

IRI\_11.n = recode(IRI\_11,

'A - DOES NOT DESCRIBE ME VERY WELL' = 0,

'B' = 1,

'C' = 2,

'D' = 3,

'E - DESCRIBES ME VERY WELL' = 4,

.default = 0),

IRI\_16.n = recode(IRI\_16,

'A - DOES NOT DESCRIBE ME VERY WELL' = 0,

'B' = 1,

'C' = 2,

'D' = 3,

'E - DESCRIBES ME VERY WELL' = 4,

.default = 0),

IRI\_17.n = recode(IRI\_17,

'A - DOES NOT DESCRIBE ME VERY WELL' = 0,

'B' = 1,

'C' = 2,

'D' = 3,

'E - DESCRIBES ME VERY WELL' = 4,

.default = 0),

IRI\_20.n = recode(IRI\_20,

'A - DOES NOT DESCRIBE ME VERY WELL' = 0,

'B' = 1,

'C' = 2,

'D' = 3,

'E - DESCRIBES ME VERY WELL' = 4,

.default = 0),

IRI\_21.n = recode(IRI\_21,

'A - DOES NOT DESCRIBE ME VERY WELL' = 0,

'B' = 1,

'C' = 2,

'D' = 3,

'E - DESCRIBES ME VERY WELL' = 4,

.default = 0),

IRI\_22.n = recode(IRI\_22,

'A - DOES NOT DESCRIBE ME VERY WELL' = 0,

'B' = 1,

'C' = 2,

'D' = 3,

'E - DESCRIBES ME VERY WELL' = 4,

.default = 0),

IRI\_23.n = recode(IRI\_23,

'A - DOES NOT DESCRIBE ME VERY WELL' = 0,

'B' = 1,

'C' = 2,

'D' = 3,

'E - DESCRIBES ME VERY WELL' = 4,

.default = 0),

IRI\_24.n = recode(IRI\_24,

'A - DOES NOT DESCRIBE ME VERY WELL' = 0,

'B' = 1,

'C' = 2,

'D' = 3,

'E - DESCRIBES ME VERY WELL' = 4,

.default = 0),

IRI\_25.n = recode(IRI\_25,

'A - DOES NOT DESCRIBE ME VERY WELL' = 0,

'B' = 1,

'C' = 2,

'D' = 3,

'E - DESCRIBES ME VERY WELL' = 4,

.default = 0),

IRI\_26.n = recode(IRI\_26,

'A - DOES NOT DESCRIBE ME VERY WELL' = 0,

'B' = 1,

'C' = 2,

'D' = 3,

'E - DESCRIBES ME VERY WELL' = 4,

.default = 0),

IRI\_27.n = recode(IRI\_27,

'A - DOES NOT DESCRIBE ME VERY WELL' = 0,

'B' = 1,

'C' = 2,

'D' = 3,

'E - DESCRIBES ME VERY WELL' = 4,

.default = 0),

IRI\_28.n = recode(IRI\_28,

'A - DOES NOT DESCRIBE ME VERY WELL' = 0,

'B' = 1,

'C' = 2,

'D' = 3,

'E - DESCRIBES ME VERY WELL' = 4,

.default = 0),

)

table(timepoint1.use$IRI\_20, timepoint1.use$IRI\_20.n)

#

# Interpersonal reactivity Index - reverse scored items

timepoint1.use <- timepoint1.use %>% mutate(

IRI\_3.n = recode(IRI\_3,

'A - DOES NOT DESCRIBE ME VERY WELL' = 4,

'B' = 3,

'C' = 2,

'D' = 1,

'E - DESCRIBES ME VERY WELL' = 0,

.default = 0),

IRI\_4.n = recode(IRI\_4,

'A - DOES NOT DESCRIBE ME VERY WELL' = 4,

'B' = 3,

'C' = 2,

'D' = 1,

'E - DESCRIBES ME VERY WELL' = 0,

.default = 0),

IRI\_7.n = recode(IRI\_7,

'A - DOES NOT DESCRIBE ME VERY WELL' = 4,

'B' = 3,

'C' = 2,

'D' = 1,

'E - DESCRIBES ME VERY WELL' = 0,

.default = 0),

IRI\_12.n = recode(IRI\_12,

'A - DOES NOT DESCRIBE ME VERY WELL' = 4,

'B' = 3,

'C' = 2,

'D' = 1,

'E - DESCRIBES ME VERY WELL' = 0,

.default = 0),

IRI\_13.n = recode(IRI\_13,

'A - DOES NOT DESCRIBE ME VERY WELL' = 4,

'B' = 3,

'C' = 2,

'D' = 1,

'E - DESCRIBES ME VERY WELL' = 0,

.default = 0),

IRI\_14.n = recode(IRI\_14,

'A - DOES NOT DESCRIBE ME VERY WELL' = 4,

'B' = 3,

'C' = 2,

'D' = 1,

'E - DESCRIBES ME VERY WELL' = 0,

.default = 0),

IRI\_15.n = recode(IRI\_15,

'A - DOES NOT DESCRIBE ME VERY WELL' = 4,

'B' = 3,

'C' = 2,

'D' = 1,

'E - DESCRIBES ME VERY WELL' = 0,

.default = 0),

IRI\_18.n = recode(IRI\_18,

'A - DOES NOT DESCRIBE ME VERY WELL' = 4,

'B' = 3,

'C' = 2,

'D' = 1,

'E - DESCRIBES ME VERY WELL' = 0,

.default = 0),

IRI\_19.n = recode(IRI\_19,

'A - DOES NOT DESCRIBE ME VERY WELL' = 4,

'B' = 3,

'C' = 2,

'D' = 1,

'E - DESCRIBES ME VERY WELL' = 0,

.default = 0),

)

table(timepoint1.use$IRI\_18, timepoint1.use$IRI\_18.n)

#

# Empathy Quotient

timepoint1.use <- timepoint1.use %>% mutate(

EQ\_1.n = recode(EQ\_1,

'strongly agree' = 2,

'slightly agree' = 1,

'slightly disagree' = 0,

'strongly disagree' = 0,

.default = 0),

EQ\_4.n = recode(EQ\_4,

'strongly agree' = 0,

'slightly agree' = 0,

'slightly disagree' = 1,

'strongly disagree' = 2,

.default = 0),

EQ\_6.n = recode(EQ\_6,

'strongly agree' = 2,

'slightly agree' = 1,

'slightly disagree' = 0,

'strongly disagree' = 0,

.default = 0),

EQ\_8.n = recode(EQ\_8,

'strongly agree' = 0,

'slightly agree' = 0,

'slightly disagree' = 1,

'strongly disagree' = 2,

.default = 0),

EQ\_10.n = recode(EQ\_10,

'strongly agree' = 0,

'slightly agree' = 0,

'slightly disagree' = 1,

'strongly disagree' = 2,

.default = 0),

EQ\_11.n = recode(EQ\_11,

'strongly agree' = 0,

'slightly agree' = 0,

'slightly disagree' = 1,

'strongly disagree' = 2,

.default = 0),

EQ\_12.n = recode(EQ\_12,

'strongly agree' = 0,

'slightly agree' = 0,

'slightly disagree' = 1,

'strongly disagree' = 2,

.default = 0),

EQ\_14.n = recode(EQ\_14,

'strongly agree' = 0,

'slightly agree' = 0,

'slightly disagree' = 1,

'strongly disagree' = 2,

.default = 0),

EQ\_15.n = recode(EQ\_15,

'strongly agree' = 0,

'slightly agree' = 0,

'slightly disagree' = 1,

'strongly disagree' = 2,

.default = 0),

EQ\_18.n = recode(EQ\_18,

'strongly agree' = 0,

'slightly agree' = 0,

'slightly disagree' = 1,

'strongly disagree' = 2,

.default = 0),

EQ\_19.n = recode(EQ\_19,

'strongly agree' = 2,

'slightly agree' = 1,

'slightly disagree' = 0,

'strongly disagree' = 0,

.default = 0),

EQ\_21.n = recode(EQ\_21,

'strongly agree' = 0,

'slightly agree' = 0,

'slightly disagree' = 1,

'strongly disagree' = 2,

.default = 0),

EQ\_22.n = recode(EQ\_22,

'strongly agree' = 2,

'slightly agree' = 1,

'slightly disagree' = 0,

'strongly disagree' = 0,

.default = 0),

EQ\_25.n = recode(EQ\_25,

'strongly agree' = 2,

'slightly agree' = 1,

'slightly disagree' = 0,

'strongly disagree' = 0,

.default = 0),

EQ\_26.n = recode(EQ\_26,

'strongly agree' = 2,

'slightly agree' = 1,

'slightly disagree' = 0,

'strongly disagree' = 0,

.default = 0),

EQ\_27.n = recode(EQ\_27,

'strongly agree' = 0,

'slightly agree' = 0,

'slightly disagree' = 1,

'strongly disagree' = 2,

.default = 0),

EQ\_28.n = recode(EQ\_28,

'strongly agree' = 0,

'slightly agree' = 0,

'slightly disagree' = 1,

'strongly disagree' = 2,

.default = 0),

EQ\_29.n = recode(EQ\_29,

'strongly agree' = 0,

'slightly agree' = 0,

'slightly disagree' = 1,

'strongly disagree' = 2,

.default = 0),

EQ\_32.n = recode(EQ\_32,

'strongly agree' = 0,

'slightly agree' = 0,

'slightly disagree' = 1,

'strongly disagree' = 2,

.default = 0),

EQ\_34.n = recode(EQ\_34,

'strongly agree' = 0,

'slightly agree' = 0,

'slightly disagree' = 1,

'strongly disagree' = 2,

.default = 0),

EQ\_35.n = recode(EQ\_35,

'strongly agree' = 2,

'slightly agree' = 1,

'slightly disagree' = 0,

'strongly disagree' = 0,

.default = 0),

EQ\_36.n = recode(EQ\_36,

'strongly agree' = 2,

'slightly agree' = 1,

'slightly disagree' = 0,

'strongly disagree' = 0,

.default = 0),

EQ\_37.n = recode(EQ\_37,

'strongly agree' = 2,

'slightly agree' = 1,

'slightly disagree' = 0,

'strongly disagree' = 0,

.default = 0),

EQ\_38.n = recode(EQ\_38,

'strongly agree' = 2,

'slightly agree' = 1,

'slightly disagree' = 0,

'strongly disagree' = 0,

.default = 0),

EQ\_39.n = recode(EQ\_39,

'strongly agree' = 0,

'slightly agree' = 0,

'slightly disagree' = 1,

'strongly disagree' = 2,

.default = 0),

EQ\_41.n = recode(EQ\_41,

'strongly agree' = 2,

'slightly agree' = 1,

'slightly disagree' = 0,

'strongly disagree' = 0,

.default = 0),

EQ\_42.n = recode(EQ\_42,

'strongly agree' = 2,

'slightly agree' = 1,

'slightly disagree' = 0,

'strongly disagree' = 0,

.default = 0),

EQ\_43.n = recode(EQ\_43,

'strongly agree' = 2,

'slightly agree' = 1,

'slightly disagree' = 0,

'strongly disagree' = 0,

.default = 0),

EQ\_44.n = recode(EQ\_44,

'strongly agree' = 2,

'slightly agree' = 1,

'slightly disagree' = 0,

'strongly disagree' = 0,

.default = 0),

EQ\_46.n = recode(EQ\_46,

'strongly agree' = 0,

'slightly agree' = 0,

'slightly disagree' = 1,

'strongly disagree' = 2,

.default = 0),

EQ\_48.n = recode(EQ\_48,

'strongly agree' = 0,

'slightly agree' = 0,

'slightly disagree' = 1,

'strongly disagree' = 2,

.default = 0),

EQ\_49.n = recode(EQ\_49,

'strongly agree' = 0,

'slightly agree' = 0,

'slightly disagree' = 1,

'strongly disagree' = 2,

.default = 0),

EQ\_50.n = recode(EQ\_50,

'strongly agree' = 0,

'slightly agree' = 0,

'slightly disagree' = 1,

'strongly disagree' = 2,

.default = 0),

EQ\_52.n = recode(EQ\_52,

'strongly agree' = 2,

'slightly agree' = 1,

'slightly disagree' = 0,

'strongly disagree' = 0,

.default = 0),

EQ\_54.n = recode(EQ\_54,

'strongly agree' = 2,

'slightly agree' = 1,

'slightly disagree' = 0,

'strongly disagree' = 0,

.default = 0),

EQ\_55.n = recode(EQ\_55,

'strongly agree' = 2,

'slightly agree' = 1,

'slightly disagree' = 0,

'strongly disagree' = 0,

.default = 0),

EQ\_57.n = recode(EQ\_57,

'strongly agree' = 2,

'slightly agree' = 1,

'slightly disagree' = 0,

'strongly disagree' = 0,

.default = 0),

EQ\_58.n = recode(EQ\_58,

'strongly agree' = 2,

'slightly agree' = 1,

'slightly disagree' = 0,

'strongly disagree' = 0,

.default = 0),

EQ\_59.n = recode(EQ\_59,

'strongly agree' = 2,

'slightly agree' = 1,

'slightly disagree' = 0,

'strongly disagree' = 0,

.default = 0),

EQ\_60.n = recode(EQ\_60,

'strongly agree' = 2,

'slightly agree' = 1,

'slightly disagree' = 0,

'strongly disagree' = 0,

.default = 0),

)

table(timepoint1.use$EQ\_1, timepoint1.use$EQ\_1.n)

table(timepoint1.use$EQ\_50, timepoint1.use$EQ\_50.n)

#

# Pictorial Empathy Test

timepoint1.use <- timepoint1.use %>% mutate(

PET\_1.n = recode(PET\_1,

'not at all' = 1,

'a little bit' = 2,

'it arouses some feelings' = 3,

'quite a lot' = 4,

'very much' = 5,

.default = 0),

PET\_2.n = recode(PET\_2,

'not at all' = 1,

'a little bit' = 2,

'it arouses some feelings' = 3,

'quite a lot' = 4,

'very much' = 5,

.default = 0),

PET\_3.n = recode(PET\_3,

'not at all' = 1,

'a little bit' = 2,

'it arouses some feelings' = 3,

'quite a lot' = 4,

'very much' = 5,

.default = 0),

PET\_4.n = recode(PET\_4,

'not at all' = 1,

'a little bit' = 2,

'it arouses some feelings' = 3,

'quite a lot' = 4,

'very much' = 5,

.default = 0),

PET\_5.n = recode(PET\_5,

'not at all' = 1,

'a little bit' = 2,

'it arouses some feelings' = 3,

'quite a lot' = 4,

'very much' = 5,

.default = 0),

PET\_6.n = recode(PET\_6,

'not at all' = 1,

'a little bit' = 2,

'it arouses some feelings' = 3,

'quite a lot' = 4,

'very much' = 5,

.default = 0),

PET\_7.n = recode(PET\_7,

'not at all' = 1,

'a little bit' = 2,

'it arouses some feelings' = 3,

'quite a lot' = 4,

'very much' = 5,

.default = 0)

)

table(timepoint1.use$PET\_6, timepoint1.use$PET\_6.n)

#

# Reading the Mind in the Eyes

timepoint1.use <- timepoint1.use %>% mutate(

EYES\_1.n = recode(EYES\_1,

'playful' = 1,

.default = 0),

EYES\_2.n = recode(EYES\_2,

'upset' = 1,

.default = 0),

EYES\_3.n = recode(EYES\_3,

'desire' = 1,

.default = 0),

EYES\_4.n = recode(EYES\_4,

'insisting' = 1,

.default = 0),

EYES\_5.n = recode(EYES\_5,

'worried' = 1,

.default = 0),

EYES\_6.n = recode(EYES\_6,

'fantasizing' = 1,

.default = 0),

EYES\_7.n = recode(EYES\_7,

'uneasy' = 1,

.default = 0),

EYES\_8.n = recode(EYES\_8,

'despondent' = 1,

.default = 0),

EYES\_9.n = recode(EYES\_9,

'preoccupied' = 1,

.default = 0),

EYES\_10.n = recode(EYES\_10,

'cautious' = 1,

.default = 0),

EYES\_11.n = recode(EYES\_11,

'regretful' = 1,

.default = 0),

EYES\_12.n = recode(EYES\_12,

'sceptical' = 1,

.default = 0),

EYES\_13.n = recode(EYES\_13,

'anticipating' = 1,

.default = 0),

EYES\_14.n = recode(EYES\_14,

'accusing' = 1,

.default = 0),

EYES\_15.n = recode(EYES\_15,

'contemplative' = 1,

.default = 0),

EYES\_16.n = recode(EYES\_16,

'thoughtful' = 1,

.default = 0),

EYES\_17.n = recode(EYES\_17,

'doubtful' = 1,

.default = 0),

EYES\_18.n = recode(EYES\_18,

'decisive' = 1,

.default = 0),

EYES\_19.n = recode(EYES\_19,

'tentative' = 1,

.default = 0),

EYES\_20.n = recode(EYES\_20,

'friendly' = 1,

.default = 0),

EYES\_21.n = recode(EYES\_21,

'fantasizing' = 1,

.default = 0),

EYES\_22.n = recode(EYES\_22,

'preoccupied' = 1,

.default = 0),

EYES\_23.n = recode(EYES\_23,

'defiant' = 1,

.default = 0),

EYES\_24.n = recode(EYES\_24,

'pensive' = 1,

.default = 0),

EYES\_25.n = recode(EYES\_25,

'interested' = 1,

.default = 0),

EYES\_26.n = recode(EYES\_26,

'hostile' = 1,

.default = 0),

EYES\_27.n = recode(EYES\_27,

'cautious' = 1,

.default = 0),

EYES\_28.n = recode(EYES\_28,

'interested' = 1,

.default = 0),

EYES\_29.n = recode(EYES\_29,

'reflective' = 1,

.default = 0),

EYES\_30.n = recode(EYES\_30,

'flirtatious' = 1,

.default = 0),

EYES\_31.n = recode(EYES\_31,

'confident' = 1,

.default = 0),

EYES\_32.n = recode(EYES\_32,

'serious' = 1,

.default = 0),

EYES\_33.n = recode(EYES\_33,

'concerned' = 1,

.default = 0),

EYES\_34.n = recode(EYES\_34,

'distrustful' = 1,

.default = 0),

EYES\_35.n = recode(EYES\_35,

'nervous' = 1,

.default = 0),

EYES\_36.n = recode(EYES\_36,

'suspicious' = 1,

.default = 0)

)

table(timepoint1.use$EYES\_27, timepoint1.use$EYES\_27.n)

#

#

############# Change all values to numeric values - Timepoint 2 ##############

# SITES

timepoint2.use <- timepoint2.use %>% mutate(

SITES.n = recode(SITES,

'1 Not very true of me' = 1,

'2' = 2,

'3' = 3,

'4' = 4,

'5 Very true of me.' = 5,

.default = 0

)

)

# Check if change worked

table(timepoint2.use$SITES, timepoint2.use$SITES.n)

#

# Interpersonal reactivity Index - normal scored items

timepoint2.use <- timepoint2.use %>% mutate(

IRI\_1.n = recode(IRI\_1,

'A - DOES NOT DESCRIBE ME VERY WELL' = 0,

'B' = 1,

'C' = 2,

'D' = 3,

'E - DESCRIBES ME VERY WELL' = 4,

.default = 0),

IRI\_2.n = recode(IRI\_2,

'A - DOES NOT DESCRIBE ME VERY WELL' = 0,

'B' = 1,

'C' = 2,

'D' = 3,

'E - DESCRIBES ME VERY WELL' = 4,

.default = 0),

IRI\_5.n = recode(IRI\_5,

'A - DOES NOT DESCRIBE ME VERY WELL' = 0,

'B' = 1,

'C' = 2,

'D' = 3,

'E - DESCRIBES ME VERY WELL' = 4,

.default = 0),

IRI\_6.n = recode(IRI\_6,

'A - DOES NOT DESCRIBE ME VERY WELL' = 0,

'B' = 1,

'C' = 2,

'D' = 3,

'E - DESCRIBES ME VERY WELL' = 4,

.default = 0),

IRI\_8.n = recode(IRI\_8,

'A - DOES NOT DESCRIBE ME VERY WELL' = 0,

'B' = 1,

'C' = 2,

'D' = 3,

'E - DESCRIBES ME VERY WELL' = 4,

.default = 0),

IRI\_9.n = recode(IRI\_9,

'A - DOES NOT DESCRIBE ME VERY WELL' = 0,

'B' = 1,

'C' = 2,

'D' = 3,

'E - DESCRIBES ME VERY WELL' = 4,

.default = 0),

IRI\_10.n = recode(IRI\_10,

'A - DOES NOT DESCRIBE ME VERY WELL' = 0,

'B' = 1,

'C' = 2,

'D' = 3,

'E - DESCRIBES ME VERY WELL' = 4,

.default = 0),

IRI\_11.n = recode(IRI\_11,

'A - DOES NOT DESCRIBE ME VERY WELL' = 0,

'B' = 1,

'C' = 2,

'D' = 3,

'E - DESCRIBES ME VERY WELL' = 4,

.default = 0),

IRI\_16.n = recode(IRI\_16,

'A - DOES NOT DESCRIBE ME VERY WELL' = 0,

'B' = 1,

'C' = 2,

'D' = 3,

'E - DESCRIBES ME VERY WELL' = 4,

.default = 0),

IRI\_17.n = recode(IRI\_17,

'A - DOES NOT DESCRIBE ME VERY WELL' = 0,

'B' = 1,

'C' = 2,

'D' = 3,

'E - DESCRIBES ME VERY WELL' = 4,

.default = 0),

IRI\_20.n = recode(IRI\_20,

'A - DOES NOT DESCRIBE ME VERY WELL' = 0,

'B' = 1,

'C' = 2,

'D' = 3,

'E - DESCRIBES ME VERY WELL' = 4,

.default = 0),

IRI\_21.n = recode(IRI\_21,

'A - DOES NOT DESCRIBE ME VERY WELL' = 0,

'B' = 1,

'C' = 2,

'D' = 3,

'E - DESCRIBES ME VERY WELL' = 4,

.default = 0),

IRI\_22.n = recode(IRI\_22,

'A - DOES NOT DESCRIBE ME VERY WELL' = 0,

'B' = 1,

'C' = 2,

'D' = 3,

'E - DESCRIBES ME VERY WELL' = 4,

.default = 0),

IRI\_23.n = recode(IRI\_23,

'A - DOES NOT DESCRIBE ME VERY WELL' = 0,

'B' = 1,

'C' = 2,

'D' = 3,

'E - DESCRIBES ME VERY WELL' = 4,

.default = 0),

IRI\_24.n = recode(IRI\_24,

'A - DOES NOT DESCRIBE ME VERY WELL' = 0,

'B' = 1,

'C' = 2,

'D' = 3,

'E - DESCRIBES ME VERY WELL' = 4,

.default = 0),

IRI\_25.n = recode(IRI\_25,

'A - DOES NOT DESCRIBE ME VERY WELL' = 0,

'B' = 1,

'C' = 2,

'D' = 3,

'E - DESCRIBES ME VERY WELL' = 4,

.default = 0),

IRI\_26.n = recode(IRI\_26,

'A - DOES NOT DESCRIBE ME VERY WELL' = 0,

'B' = 1,

'C' = 2,

'D' = 3,

'E - DESCRIBES ME VERY WELL' = 4,

.default = 0),

IRI\_27.n = recode(IRI\_27,

'A - DOES NOT DESCRIBE ME VERY WELL' = 0,

'B' = 1,

'C' = 2,

'D' = 3,

'E - DESCRIBES ME VERY WELL' = 4,

.default = 0),

IRI\_28.n = recode(IRI\_28,

'A - DOES NOT DESCRIBE ME VERY WELL' = 0,

'B' = 1,

'C' = 2,

'D' = 3,

'E - DESCRIBES ME VERY WELL' = 4,

.default = 0),

)

table(timepoint2.use$IRI\_20, timepoint2.use$IRI\_20.n)

#

# Interpersonal reactivity Index - reverse scored items

timepoint2.use <- timepoint2.use %>% mutate(

IRI\_3.n = recode(IRI\_3,

'A - DOES NOT DESCRIBE ME VERY WELL' = 4,

'B' = 3,

'C' = 2,

'D' = 1,

'E - DESCRIBES ME VERY WELL' = 0,

.default = 0),

IRI\_4.n = recode(IRI\_4,

'A - DOES NOT DESCRIBE ME VERY WELL' = 4,

'B' = 3,

'C' = 2,

'D' = 1,

'E - DESCRIBES ME VERY WELL' = 0,

.default = 0),

IRI\_7.n = recode(IRI\_7,

'A - DOES NOT DESCRIBE ME VERY WELL' = 4,

'B' = 3,

'C' = 2,

'D' = 1,

'E - DESCRIBES ME VERY WELL' = 0,

.default = 0),

IRI\_12.n = recode(IRI\_12,

'A - DOES NOT DESCRIBE ME VERY WELL' = 4,

'B' = 3,

'C' = 2,

'D' = 1,

'E - DESCRIBES ME VERY WELL' = 0,

.default = 0),

IRI\_13.n = recode(IRI\_13,

'A - DOES NOT DESCRIBE ME VERY WELL' = 4,

'B' = 3,

'C' = 2,

'D' = 1,

'E - DESCRIBES ME VERY WELL' = 0,

.default = 0),

IRI\_14.n = recode(IRI\_14,

'A - DOES NOT DESCRIBE ME VERY WELL' = 4,

'B' = 3,

'C' = 2,

'D' = 1,

'E - DESCRIBES ME VERY WELL' = 0,

.default = 0),

IRI\_15.n = recode(IRI\_15,

'A - DOES NOT DESCRIBE ME VERY WELL' = 4,

'B' = 3,

'C' = 2,

'D' = 1,

'E - DESCRIBES ME VERY WELL' = 0,

.default = 0),

IRI\_18.n = recode(IRI\_18,

'A - DOES NOT DESCRIBE ME VERY WELL' = 4,

'B' = 3,

'C' = 2,

'D' = 1,

'E - DESCRIBES ME VERY WELL' = 0,

.default = 0),

IRI\_19.n = recode(IRI\_19,

'A - DOES NOT DESCRIBE ME VERY WELL' = 4,

'B' = 3,

'C' = 2,

'D' = 1,

'E - DESCRIBES ME VERY WELL' = 0,

.default = 0),

)

table(timepoint2.use$IRI\_18, timepoint2.use$IRI\_18.n)

#

# Empathy Quotient

timepoint2.use <- timepoint2.use %>% mutate(

EQ\_1.n = recode(EQ\_1,

'strongly agree' = 2,

'slightly agree' = 1,

'slightly disagree' = 0,

'strongly disagree' = 0,

.default = 0),

EQ\_4.n = recode(EQ\_4,

'strongly agree' = 0,

'slightly agree' = 0,

'slightly disagree' = 1,

'strongly disagree' = 2,

.default = 0),

EQ\_6.n = recode(EQ\_6,

'strongly agree' = 2,

'slightly agree' = 1,

'slightly disagree' = 0,

'strongly disagree' = 0,

.default = 0),

EQ\_8.n = recode(EQ\_8,

'strongly agree' = 0,

'slightly agree' = 0,

'slightly disagree' = 1,

'strongly disagree' = 2,

.default = 0),

EQ\_10.n = recode(EQ\_10,

'strongly agree' = 0,

'slightly agree' = 0,

'slightly disagree' = 1,

'strongly disagree' = 2,

.default = 0),

EQ\_11.n = recode(EQ\_11,

'strongly agree' = 0,

'slightly agree' = 0,

'slightly disagree' = 1,

'strongly disagree' = 2,

.default = 0),

EQ\_12.n = recode(EQ\_12,

'strongly agree' = 0,

'slightly agree' = 0,

'slightly disagree' = 1,

'strongly disagree' = 2,

.default = 0),

EQ\_14.n = recode(EQ\_14,

'strongly agree' = 0,

'slightly agree' = 0,

'slightly disagree' = 1,

'strongly disagree' = 2,

.default = 0),

EQ\_15.n = recode(EQ\_15,

'strongly agree' = 0,

'slightly agree' = 0,

'slightly disagree' = 1,

'strongly disagree' = 2,

.default = 0),

EQ\_18.n = recode(EQ\_18,

'strongly agree' = 0,

'slightly agree' = 0,

'slightly disagree' = 1,

'strongly disagree' = 2,

.default = 0),

EQ\_19.n = recode(EQ\_19,

'strongly agree' = 2,

'slightly agree' = 1,

'slightly disagree' = 0,

'strongly disagree' = 0,

.default = 0),

EQ\_21.n = recode(EQ\_21,

'strongly agree' = 0,

'slightly agree' = 0,

'slightly disagree' = 1,

'strongly disagree' = 2,

.default = 0),

EQ\_22.n = recode(EQ\_22,

'strongly agree' = 2,

'slightly agree' = 1,

'slightly disagree' = 0,

'strongly disagree' = 0,

.default = 0),

EQ\_25.n = recode(EQ\_25,

'strongly agree' = 2,

'slightly agree' = 1,

'slightly disagree' = 0,

'strongly disagree' = 0,

.default = 0),

EQ\_26.n = recode(EQ\_26,

'strongly agree' = 2,

'slightly agree' = 1,

'slightly disagree' = 0,

'strongly disagree' = 0,

.default = 0),

EQ\_27.n = recode(EQ\_27,

'strongly agree' = 0,

'slightly agree' = 0,

'slightly disagree' = 1,

'strongly disagree' = 2,

.default = 0),

EQ\_28.n = recode(EQ\_28,

'strongly agree' = 0,

'slightly agree' = 0,

'slightly disagree' = 1,

'strongly disagree' = 2,

.default = 0),

EQ\_29.n = recode(EQ\_29,

'strongly agree' = 0,

'slightly agree' = 0,

'slightly disagree' = 1,

'strongly disagree' = 2,

.default = 0),

EQ\_32.n = recode(EQ\_32,

'strongly agree' = 0,

'slightly agree' = 0,

'slightly disagree' = 1,

'strongly disagree' = 2,

.default = 0),

EQ\_34.n = recode(EQ\_34,

'strongly agree' = 0,

'slightly agree' = 0,

'slightly disagree' = 1,

'strongly disagree' = 2,

.default = 0),

EQ\_35.n = recode(EQ\_35,

'strongly agree' = 2,

'slightly agree' = 1,

'slightly disagree' = 0,

'strongly disagree' = 0,

.default = 0),

EQ\_36.n = recode(EQ\_36,

'strongly agree' = 2,

'slightly agree' = 1,

'slightly disagree' = 0,

'strongly disagree' = 0,

.default = 0),

EQ\_37.n = recode(EQ\_37,

'strongly agree' = 2,

'slightly agree' = 1,

'slightly disagree' = 0,

'strongly disagree' = 0,

.default = 0),

EQ\_38.n = recode(EQ\_38,

'strongly agree' = 2,

'slightly agree' = 1,

'slightly disagree' = 0,

'strongly disagree' = 0,

.default = 0),

EQ\_39.n = recode(EQ\_39,

'strongly agree' = 0,

'slightly agree' = 0,

'slightly disagree' = 1,

'strongly disagree' = 2,

.default = 0),

EQ\_41.n = recode(EQ\_41,

'strongly agree' = 2,

'slightly agree' = 1,

'slightly disagree' = 0,

'strongly disagree' = 0,

.default = 0),

EQ\_42.n = recode(EQ\_42,

'strongly agree' = 2,

'slightly agree' = 1,

'slightly disagree' = 0,

'strongly disagree' = 0,

.default = 0),

EQ\_43.n = recode(EQ\_43,

'strongly agree' = 2,

'slightly agree' = 1,

'slightly disagree' = 0,

'strongly disagree' = 0,

.default = 0),

EQ\_44.n = recode(EQ\_44,

'strongly agree' = 2,

'slightly agree' = 1,

'slightly disagree' = 0,

'strongly disagree' = 0,

.default = 0),

EQ\_46.n = recode(EQ\_46,

'strongly agree' = 0,

'slightly agree' = 0,

'slightly disagree' = 1,

'strongly disagree' = 2,

.default = 0),

EQ\_48.n = recode(EQ\_48,

'strongly agree' = 0,

'slightly agree' = 0,

'slightly disagree' = 1,

'strongly disagree' = 2,

.default = 0),

EQ\_49.n = recode(EQ\_49,

'strongly agree' = 0,

'slightly agree' = 0,

'slightly disagree' = 1,

'strongly disagree' = 2,

.default = 0),

EQ\_50.n = recode(EQ\_50,

'strongly agree' = 0,

'slightly agree' = 0,

'slightly disagree' = 1,

'strongly disagree' = 2,

.default = 0),

EQ\_52.n = recode(EQ\_52,

'strongly agree' = 2,

'slightly agree' = 1,

'slightly disagree' = 0,

'strongly disagree' = 0,

.default = 0),

EQ\_54.n = recode(EQ\_54,

'strongly agree' = 2,

'slightly agree' = 1,

'slightly disagree' = 0,

'strongly disagree' = 0,

.default = 0),

EQ\_55.n = recode(EQ\_55,

'strongly agree' = 2,

'slightly agree' = 1,

'slightly disagree' = 0,

'strongly disagree' = 0,

.default = 0),

EQ\_57.n = recode(EQ\_57,

'strongly agree' = 2,

'slightly agree' = 1,

'slightly disagree' = 0,

'strongly disagree' = 0,

.default = 0),

EQ\_58.n = recode(EQ\_58,

'strongly agree' = 2,

'slightly agree' = 1,

'slightly disagree' = 0,

'strongly disagree' = 0,

.default = 0),

EQ\_59.n = recode(EQ\_59,

'strongly agree' = 2,

'slightly agree' = 1,

'slightly disagree' = 0,

'strongly disagree' = 0,

.default = 0),

EQ\_60.n = recode(EQ\_60,

'strongly agree' = 2,

'slightly agree' = 1,

'slightly disagree' = 0,

'strongly disagree' = 0,

.default = 0),

)

table(timepoint2.use$EQ\_1, timepoint2.use$EQ\_1.n)

table(timepoint2.use$EQ\_50, timepoint2.use$EQ\_50.n)

#

# Pictorial Empathy Test

timepoint2.use <- timepoint2.use %>% mutate(

PET\_1.n = recode(PET\_1,

'not at all' = 1,

'a little bit' = 2,

'it arouses some feelings' = 3,

'quite a lot' = 4,

'very much' = 5,

.default = 0),

PET\_2.n = recode(PET\_2,

'not at all' = 1,

'a little bit' = 2,

'it arouses some feelings' = 3,

'quite a lot' = 4,

'very much' = 5,

.default = 0),

PET\_3.n = recode(PET\_3,

'not at all' = 1,

'a little bit' = 2,

'it arouses some feelings' = 3,

'quite a lot' = 4,

'very much' = 5,

.default = 0),

PET\_4.n = recode(PET\_4,

'not at all' = 1,

'a little bit' = 2,

'it arouses some feelings' = 3,

'quite a lot' = 4,

'very much' = 5,

.default = 0),

PET\_5.n = recode(PET\_5,

'not at all' = 1,

'a little bit' = 2,

'it arouses some feelings' = 3,

'quite a lot' = 4,

'very much' = 5,

.default = 0),

PET\_6.n = recode(PET\_6,

'not at all' = 1,

'a little bit' = 2,

'it arouses some feelings' = 3,

'quite a lot' = 4,

'very much' = 5,

.default = 0),

PET\_7.n = recode(PET\_7,

'not at all' = 1,

'a little bit' = 2,

'it arouses some feelings' = 3,

'quite a lot' = 4,

'very much' = 5,

.default = 0)

)

table(timepoint2.use$PET\_6, timepoint2.use$PET\_6.n)

#

# Reading the Mind in the Eyes

timepoint2.use <- timepoint2.use %>% mutate(

EYES\_1.n = recode(EYES\_1,

'playful' = 1,

.default = 0),

EYES\_2.n = recode(EYES\_2,

'upset' = 1,

.default = 0),

EYES\_3.n = recode(EYES\_3,

'desire' = 1,

.default = 0),

EYES\_4.n = recode(EYES\_4,

'insisting' = 1,

.default = 0),

EYES\_5.n = recode(EYES\_5,

'worried' = 1,

.default = 0),

EYES\_6.n = recode(EYES\_6,

'fantasizing' = 1,

.default = 0),

EYES\_7.n = recode(EYES\_7,

'uneasy' = 1,

.default = 0),

EYES\_8.n = recode(EYES\_8,

'despondent' = 1,

.default = 0),

EYES\_9.n = recode(EYES\_9,

'preoccupied' = 1,

.default = 0),

EYES\_10.n = recode(EYES\_10,

'cautious' = 1,

.default = 0),

EYES\_11.n = recode(EYES\_11,

'regretful' = 1,

.default = 0),

EYES\_12.n = recode(EYES\_12,

'sceptical' = 1,

.default = 0),

EYES\_13.n = recode(EYES\_13,

'anticipating' = 1,

.default = 0),

EYES\_14.n = recode(EYES\_14,

'accusing' = 1,

.default = 0),

EYES\_15.n = recode(EYES\_15,

'contemplative' = 1,

.default = 0),

EYES\_16.n = recode(EYES\_16,

'thoughtful' = 1,

.default = 0),

EYES\_17.n = recode(EYES\_17,

'doubtful' = 1,

.default = 0),

EYES\_18.n = recode(EYES\_18,

'decisive' = 1,

.default = 0),

EYES\_19.n = recode(EYES\_19,

'tentative' = 1,

.default = 0),

EYES\_20.n = recode(EYES\_20,

'friendly' = 1,

.default = 0),

EYES\_21.n = recode(EYES\_21,

'fantasizing' = 1,

.default = 0),

EYES\_22.n = recode(EYES\_22,

'preoccupied' = 1,

.default = 0),

EYES\_23.n = recode(EYES\_23,

'defiant' = 1,

.default = 0),

EYES\_24.n = recode(EYES\_24,

'pensive' = 1,

.default = 0),

EYES\_25.n = recode(EYES\_25,

'interested' = 1,

.default = 0),

EYES\_26.n = recode(EYES\_26,

'hostile' = 1,

.default = 0),

EYES\_27.n = recode(EYES\_27,

'cautious' = 1,

.default = 0),

EYES\_28.n = recode(EYES\_28,

'interested' = 1,

.default = 0),

EYES\_29.n = recode(EYES\_29,

'reflective' = 1,

.default = 0),

EYES\_30.n = recode(EYES\_30,

'flirtatious' = 1,

.default = 0),

EYES\_31.n = recode(EYES\_31,

'confident' = 1,

.default = 0),

EYES\_32.n = recode(EYES\_32,

'serious' = 1,

.default = 0),

EYES\_33.n = recode(EYES\_33,

'concerned' = 1,

.default = 0),

EYES\_34.n = recode(EYES\_34,

'distrustful' = 1,

.default = 0),

EYES\_35.n = recode(EYES\_35,

'nervous' = 1,

.default = 0),

EYES\_36.n = recode(EYES\_36,

'suspicious' = 1,

.default = 0)

)

table(timepoint2.use$EYES\_27, timepoint2.use$EYES\_27.n)

#

#

############## Calculate total scores - Timepoint 1 #####################

# IRI

timepoint1.use$IRI\_FS.n <- timepoint1.use$IRI\_1.n + timepoint1.use$IRI\_5.n + timepoint1.use$IRI\_7.n +

timepoint1.use$IRI\_12.n + timepoint1.use$IRI\_16.n + timepoint1.use$IRI\_26.n + timepoint1.use$IRI\_23.n

#

timepoint1.use$IRI\_EC.n <- timepoint1.use$IRI\_2.n + timepoint1.use$IRI\_4.n + timepoint1.use$IRI\_9.n +

timepoint1.use$IRI\_14.n + timepoint1.use$IRI\_18.n + timepoint1.use$IRI\_20.n + timepoint1.use$IRI\_22.n

#

timepoint1.use$IRI\_PT.n <- timepoint1.use$IRI\_3.n + timepoint1.use$IRI\_8.n + timepoint1.use$IRI\_11.n +

timepoint1.use$IRI\_15.n + timepoint1.use$IRI\_21.n + timepoint1.use$IRI\_25.n + timepoint1.use$IRI\_28.n

#

timepoint1.use$IRI\_PD.n <- timepoint1.use$IRI\_6.n + timepoint1.use$IRI\_10.n + timepoint1.use$IRI\_13.n +

timepoint1.use$IRI\_17.n + timepoint1.use$IRI\_19.n + timepoint1.use$IRI\_24.n + timepoint1.use$IRI\_27.n

#

timepoint1.use$IRI\_TOTAL.n <- timepoint1.use$IRI\_FS.n + timepoint1.use$IRI\_EC.n +

timepoint1.use$IRI\_PT.n + timepoint1.use$IRI\_PD.n

#

# EQ

timepoint1.use$EQ\_TOTAL.n <- timepoint1.use$EQ\_1.n + timepoint1.use$EQ\_4.n + timepoint1.use$EQ\_6.n +

timepoint1.use$EQ\_8.n + timepoint1.use$EQ\_10.n + timepoint1.use$EQ\_11.n + timepoint1.use$EQ\_12.n +

timepoint1.use$EQ\_14.n + timepoint1.use$EQ\_15.n + timepoint1.use$EQ\_18.n + timepoint1.use$EQ\_19.n +

timepoint1.use$EQ\_21.n + timepoint1.use$EQ\_22.n + timepoint1.use$EQ\_25.n + timepoint1.use$EQ\_26.n +

timepoint1.use$EQ\_27.n + timepoint1.use$EQ\_28.n + timepoint1.use$EQ\_29.n + timepoint1.use$EQ\_32.n +

timepoint1.use$EQ\_34.n + timepoint1.use$EQ\_35.n + timepoint1.use$EQ\_36.n + timepoint1.use$EQ\_37.n +

timepoint1.use$EQ\_38.n + timepoint1.use$EQ\_39.n + timepoint1.use$EQ\_41.n + timepoint1.use$EQ\_42.n +

timepoint1.use$EQ\_43.n + timepoint1.use$EQ\_44.n + timepoint1.use$EQ\_46.n + timepoint1.use$EQ\_48.n +

timepoint1.use$EQ\_49.n + timepoint1.use$EQ\_50.n + timepoint1.use$EQ\_52.n + timepoint1.use$EQ\_54.n +

timepoint1.use$EQ\_55.n + timepoint1.use$EQ\_57.n + timepoint1.use$EQ\_58.n + timepoint1.use$EQ\_59.n +

timepoint1.use$EQ\_60.n

#

# PET - the mean score is calculated

timepoint1.use$PET\_TOTAL.n <- timepoint1.use$PET\_1.n + timepoint1.use$PET\_2.n + timepoint1.use$PET\_3.n +

timepoint1.use$PET\_4.n + timepoint1.use$PET\_5.n + timepoint1.use$PET\_6.n + timepoint1.use$PET\_7.n

timepoint1.use$PET\_TOTAL.n <- timepoint1.use$PET\_TOTAL.n / 7

#

# EYES

#

timepoint1.use$EYES\_TOTAL.n <- timepoint1.use$EYES\_1.n + timepoint1.use$EYES\_2.n + timepoint1.use$EYES\_3.n +

timepoint1.use$EYES\_4.n + timepoint1.use$EYES\_5.n + timepoint1.use$EYES\_6.n + timepoint1.use$EYES\_7.n +

timepoint1.use$EYES\_8.n + timepoint1.use$EYES\_9.n + timepoint1.use$EYES\_10.n + timepoint1.use$EYES\_11.n +

timepoint1.use$EYES\_12.n + timepoint1.use$EYES\_13.n + timepoint1.use$EYES\_14.n + timepoint1.use$EYES\_15.n +

timepoint1.use$EYES\_16.n + timepoint1.use$EYES\_17.n + timepoint1.use$EYES\_18.n + timepoint1.use$EYES\_19.n +

timepoint1.use$EYES\_20.n + timepoint1.use$EYES\_21.n + timepoint1.use$EYES\_22.n + timepoint1.use$EYES\_23.n +

timepoint1.use$EYES\_24.n + timepoint1.use$EYES\_25.n + timepoint1.use$EYES\_26.n + timepoint1.use$EYES\_27.n +

timepoint1.use$EYES\_28.n + timepoint1.use$EYES\_29.n + timepoint1.use$EYES\_30.n + timepoint1.use$EYES\_31.n +

timepoint1.use$EYES\_32.n + timepoint1.use$EYES\_33.n + timepoint1.use$EYES\_34.n + timepoint1.use$EYES\_35.n # +

timepoint1.use$EYES\_36n

#

############## Calculate total scores - Timepoint 2 #####################

# IRI

timepoint2.use$IRI\_FS.n <- timepoint2.use$IRI\_1.n + timepoint2.use$IRI\_5.n + timepoint2.use$IRI\_7.n +

timepoint2.use$IRI\_12.n + timepoint2.use$IRI\_16.n + timepoint2.use$IRI\_26.n + timepoint2.use$IRI\_23.n

#

timepoint2.use$IRI\_EC.n <- timepoint2.use$IRI\_2.n + timepoint2.use$IRI\_4.n + timepoint2.use$IRI\_9.n +

timepoint2.use$IRI\_14.n + timepoint2.use$IRI\_18.n + timepoint2.use$IRI\_20.n + timepoint2.use$IRI\_22.n

#

timepoint2.use$IRI\_PT.n <- timepoint2.use$IRI\_3.n + timepoint2.use$IRI\_8.n + timepoint2.use$IRI\_11.n +

timepoint2.use$IRI\_15.n + timepoint2.use$IRI\_21.n + timepoint2.use$IRI\_25.n + timepoint2.use$IRI\_28.n

#

timepoint2.use$IRI\_PD.n <- timepoint2.use$IRI\_6.n + timepoint2.use$IRI\_10.n + timepoint2.use$IRI\_13.n +

timepoint2.use$IRI\_17.n + timepoint2.use$IRI\_19.n + timepoint2.use$IRI\_24.n + timepoint2.use$IRI\_27.n

#

timepoint2.use$IRI\_TOTAL.n <- timepoint2.use$IRI\_FS.n + timepoint2.use$IRI\_EC.n +

timepoint2.use$IRI\_PT.n + timepoint2.use$IRI\_PD.n

#

# EQ

timepoint2.use$EQ\_TOTAL.n <- timepoint2.use$EQ\_1.n + timepoint2.use$EQ\_4.n + timepoint2.use$EQ\_6.n +

timepoint2.use$EQ\_8.n + timepoint2.use$EQ\_10.n + timepoint2.use$EQ\_11.n + timepoint2.use$EQ\_12.n +

timepoint2.use$EQ\_14.n + timepoint2.use$EQ\_15.n + timepoint2.use$EQ\_18.n + timepoint2.use$EQ\_19.n +

timepoint2.use$EQ\_21.n + timepoint2.use$EQ\_22.n + timepoint2.use$EQ\_25.n + timepoint2.use$EQ\_26.n +

timepoint2.use$EQ\_27.n + timepoint2.use$EQ\_28.n + timepoint2.use$EQ\_29.n + timepoint2.use$EQ\_32.n +

timepoint2.use$EQ\_34.n + timepoint2.use$EQ\_35.n + timepoint2.use$EQ\_36.n + timepoint2.use$EQ\_37.n +

timepoint2.use$EQ\_38.n + timepoint2.use$EQ\_39.n + timepoint2.use$EQ\_41.n + timepoint2.use$EQ\_42.n +

timepoint2.use$EQ\_43.n + timepoint2.use$EQ\_44.n + timepoint2.use$EQ\_46.n + timepoint2.use$EQ\_48.n +

timepoint2.use$EQ\_49.n + timepoint2.use$EQ\_50.n + timepoint2.use$EQ\_52.n + timepoint2.use$EQ\_54.n +

timepoint2.use$EQ\_55.n + timepoint2.use$EQ\_57.n + timepoint2.use$EQ\_58.n + timepoint2.use$EQ\_59.n +

timepoint2.use$EQ\_60.n

#

# PET - the mean score is calculated

timepoint2.use$PET\_TOTAL.n <- timepoint2.use$PET\_1.n + timepoint2.use$PET\_2.n + timepoint2.use$PET\_3.n +

timepoint2.use$PET\_4.n + timepoint2.use$PET\_5.n + timepoint2.use$PET\_6.n + timepoint2.use$PET\_7.n

timepoint2.use$PET\_TOTAL.n <- timepoint2.use$PET\_TOTAL.n / 7

#

# EYES

#

timepoint2.use$EYES\_TOTAL.n <- timepoint2.use$EYES\_1.n + timepoint2.use$EYES\_2.n + timepoint2.use$EYES\_3.n +

timepoint2.use$EYES\_4.n + timepoint2.use$EYES\_5.n + timepoint2.use$EYES\_6.n + timepoint2.use$EYES\_7.n +

timepoint2.use$EYES\_8.n + timepoint2.use$EYES\_9.n + timepoint2.use$EYES\_10.n + timepoint2.use$EYES\_11.n +

timepoint2.use$EYES\_12.n + timepoint2.use$EYES\_13.n + timepoint2.use$EYES\_14.n + timepoint2.use$EYES\_15.n +

timepoint2.use$EYES\_16.n + timepoint2.use$EYES\_17.n + timepoint2.use$EYES\_18.n + timepoint2.use$EYES\_19.n +

timepoint2.use$EYES\_20.n + timepoint2.use$EYES\_21.n + timepoint2.use$EYES\_22.n + timepoint2.use$EYES\_23.n +

timepoint2.use$EYES\_24.n + timepoint2.use$EYES\_25.n + timepoint2.use$EYES\_26.n + timepoint2.use$EYES\_27.n +

timepoint2.use$EYES\_28.n + timepoint2.use$EYES\_29.n + timepoint2.use$EYES\_30.n + timepoint2.use$EYES\_31.n +

timepoint2.use$EYES\_32.n + timepoint2.use$EYES\_33.n + timepoint2.use$EYES\_34.n + timepoint2.use$EYES\_35.n +

timepoint2.use$EYES\_36.n

#

# Merge both datasets to contain only participants that completed both time points

timepointboth <- merge(timepoint1.use, timepoint2.use, by = "PROLIFIC\_PID")

#

# Number of participants, age, gender

#

# N of both timepoints

nrow(timepointboth)

# Gender

table(timepointboth$Gender.y)

prop.table(table(timepointboth$Gender.y))

round(100 \* prop.table(table(timepointboth$Gender.y)), digits= 1)

# Age

timepointboth$Age\_1.y.n <- as.numeric(timepointboth$Age\_1.y)

describe(timepointboth$Age\_1.y.n)

range(timepointboth$Age\_1.y.n)

median(timepointboth$Age\_1.y.n)

#

# Create new data-sets for ICC calculation

sites.icc <- timepointboth %>% select(SITES.n.x, SITES.n.y)

iri\_fs.icc <- timepointboth %>% select(IRI\_FS.n.x, IRI\_FS.n.y)

iri\_ec.icc <- timepointboth %>% select(IRI\_EC.n.x, IRI\_EC.n.y)

iri\_pt.icc <- timepointboth %>% select(IRI\_PT.n.x, IRI\_PT.n.y)

iri\_pd.icc <- timepointboth %>% select(IRI\_PD.n.x, IRI\_PD.n.y)

iri\_total.icc <- timepointboth %>% select(IRI\_TOTAL.n.x, IRI\_TOTAL.n.y)

eq\_total.icc <- timepointboth %>% select(EQ\_TOTAL.n.x, EQ\_TOTAL.n.y)

pet\_total.icc <- timepointboth %>% select(PET\_TOTAL.n.x, PET\_TOTAL.n.y)

eyes\_total.icc <- timepointboth %>% select(EYES\_TOTAL.n.x, EYES\_TOTAL.n.y)

#

# Create new data-sets for Cronbach's Alpha calculation

# SITES has just one item, Cronbach's Alpha not possible

# Timepoint 1

iri\_fs.alpha1 <- timepointboth %>% select(IRI\_1.n.x, IRI\_5.n.x,IRI\_7.n.x,IRI\_12.n.x,

IRI\_16.n.x, IRI\_26.n.x, IRI\_23.n.x)

iri\_ec.alpha1 <- timepointboth %>% select(IRI\_2.n.x,IRI\_4.n.x, IRI\_9.n.x, IRI\_14.n.x, IRI\_18.n.x,

IRI\_20.n.x, IRI\_22.n.x)

iri\_pt.alpha1 <- timepointboth %>% select(IRI\_3.n.x, IRI\_8.n.x, IRI\_11.n.x, IRI\_15.n.x,

IRI\_21.n.x, IRI\_25.n.x, IRI\_28.n.x)

iri\_pd.alpha1 <- timepointboth %>% select(IRI\_6.n.x, IRI\_10.n.x, IRI\_13.n.x, IRI\_17.n.x,

IRI\_19.n.x, IRI\_24.n.x, IRI\_27.n.x)

iri\_total.alpha1 <- cbind(iri\_fs.alpha, iri\_ec.alpha, iri\_pt.alpha, iri\_pd.alpha)

eq.alpha1 <- timepointboth %>% select(EQ\_1.n.x, EQ\_6.n.x, EQ\_4.n.x, EQ\_8.n.x, EQ\_10.n.x,

EQ\_11.n.x, EQ\_12.n.x, EQ\_14.n.x, EQ\_15.n.x, EQ\_18.n.x,

EQ\_19.n.x, EQ\_21.n.x, EQ\_22.n.x, EQ\_25.n.x, EQ\_26.n.x,

EQ\_27.n.x,EQ\_28.n.x, EQ\_29.n.x, EQ\_32.n.x, EQ\_34.n.x,

EQ\_35.n.x, EQ\_36.n.x, EQ\_37.n.x, EQ\_38.n.x,EQ\_39.n.x,

EQ\_41.n.x, EQ\_42.n.x, EQ\_43.n.x, EQ\_44.n.x, EQ\_46.n.x,

EQ\_48.n.x, EQ\_49.n.x, EQ\_50.n.x, EQ\_52.n.x, EQ\_54.n.x,

EQ\_55.n.x, EQ\_57.n.x, EQ\_58.n.x, EQ\_59.n.x, EQ\_60.n.x)

pet.alpha1 <- timepointboth %>% select(PET\_1.n.x, PET\_2.n.x, PET\_3.n.x, PET\_4.n.x,

PET\_5.n.x, PET\_6.n.x, PET\_7.n.x)

eyes.alpha1 <- timepointboth %>% select(EYES\_1.n.x, EYES\_2.n.x, EYES\_3.n.x, EYES\_4.n.x,

EYES\_5.n.x, EYES\_6.n.x, EYES\_7.n.x,EYES\_8.n.x,

EYES\_9.n.x, EYES\_10.n.x, EYES\_11.n.x, EYES\_12.n.x,

EYES\_13.n.x, EYES\_14.n.x, EYES\_15.n.x, EYES\_16.n.x,

EYES\_17.n.x, EYES\_18.n.x, EYES\_19.n.x, EYES\_20.n.x,

EYES\_21.n.x, EYES\_22.n.x, EYES\_23.n.x, EYES\_24.n.x,

EYES\_25.n.x, EYES\_26.n.x, EYES\_27.n.x, EYES\_28.n.x,

EYES\_29.n.x, EYES\_30.n.x, EYES\_31.n.x, EYES\_32.n.x,

EYES\_33.n.x, EYES\_34.n.x, EYES\_35.n.x, EYES\_36.n.x)

# Timepoint 2

iri\_fs.alpha <- timepointboth %>% select(IRI\_1.n.y, IRI\_5.n.y,IRI\_7.n.y,IRI\_12.n.y,

IRI\_16.n.y, IRI\_26.n.y, IRI\_23.n.y)

iri\_ec.alpha <- timepointboth %>% select(IRI\_2.n.y,IRI\_4.n.y, IRI\_9.n.y, IRI\_14.n.y, IRI\_18.n.y,

IRI\_20.n.y, IRI\_22.n.y)

iri\_pt.alpha <- timepointboth %>% select(IRI\_3.n.y, IRI\_8.n.y, IRI\_11.n.y, IRI\_15.n.y,

IRI\_21.n.y, IRI\_25.n.y, IRI\_28.n.y)

iri\_pd.alpha <- timepointboth %>% select(IRI\_6.n.y, IRI\_10.n.y, IRI\_13.n.y, IRI\_17.n.y,

IRI\_19.n.y, IRI\_24.n.y, IRI\_27.n.y)

iri\_total.alpha <- cbind(iri\_fs.alpha, iri\_ec.alpha, iri\_pt.alpha, iri\_pd.alpha)

eq.alpha <- timepointboth %>% select(EQ\_1.n.y, EQ\_6.n.y, EQ\_4.n.y, EQ\_8.n.y, EQ\_10.n.y,

EQ\_11.n.y, EQ\_12.n.y, EQ\_14.n.y, EQ\_15.n.y, EQ\_18.n.y,

EQ\_19.n.y, EQ\_21.n.y, EQ\_22.n.y, EQ\_25.n.y, EQ\_26.n.y,

EQ\_27.n.y,EQ\_28.n.y, EQ\_29.n.y, EQ\_32.n.y, EQ\_34.n.y,

EQ\_35.n.y, EQ\_36.n.y, EQ\_37.n.y, EQ\_38.n.y,EQ\_39.n.y,

EQ\_41.n.y, EQ\_42.n.y, EQ\_43.n.y, EQ\_44.n.y, EQ\_46.n.y,

EQ\_48.n.y, EQ\_49.n.y, EQ\_50.n.y, EQ\_52.n.y, EQ\_54.n.y,

EQ\_55.n.y, EQ\_57.n.y, EQ\_58.n.y, EQ\_59.n.y, EQ\_60.n.y)

pet.alpha <- timepointboth %>% select(PET\_1.n.y, PET\_2.n.y, PET\_3.n.y, PET\_4.n.y,

PET\_5.n.y, PET\_6.n.y, PET\_7.n.y)

eyes.alpha <- timepointboth %>% select(EYES\_1.n.y, EYES\_2.n.y, EYES\_3.n.y, EYES\_4.n.y,

EYES\_5.n.y, EYES\_6.n.y, EYES\_7.n.y,EYES\_8.n.y,

EYES\_9.n.y, EYES\_10.n.y, EYES\_11.n.y, EYES\_12.n.y,

EYES\_13.n.y, EYES\_14.n.y, EYES\_15.n.y, EYES\_16.n.y,

EYES\_17.n.y, EYES\_18.n.y, EYES\_19.n.y, EYES\_20.n.y,

EYES\_21.n.y, EYES\_22.n.y, EYES\_23.n.y, EYES\_24.n.y,

EYES\_25.n.y, EYES\_26.n.y, EYES\_27.n.y, EYES\_28.n.y,

EYES\_29.n.y, EYES\_30.n.y, EYES\_31.n.y, EYES\_32.n.y,

EYES\_33.n.y, EYES\_34.n.y, EYES\_35.n.y, EYES\_36.n.y)

#

# Analysis of each measure

#

### IRI ###

icc(iri\_fs.icc, model = "twoway", type = "agreement", unit = "single",

r0 = 0, conf.level = 0.95)

icc(iri\_ec.icc, model = "twoway", type = "agreement", unit = "single",

r0 = 0, conf.level = 0.95)

icc(iri\_pt.icc, model = "twoway", type = "agreement", unit = "single",

r0 = 0, conf.level = 0.95)

icc(iri\_pd.icc, model = "twoway", type = "agreement", unit = "single",

r0 = 0, conf.level = 0.95)

icc(iri\_total.icc, model = "twoway", type = "agreement", unit = "single",

r0 = 0, conf.level = 0.95)

# Alpha timepoint 1

alpha(iri\_fs.alpha1)

alpha(iri\_ec.alpha1)

alpha(iri\_pt.alpha1)

alpha(iri\_pd.alpha1)

alpha(iri\_total.alpha1)

# Alpha timepoint 2

alpha(iri\_fs.alpha)

alpha(iri\_ec.alpha)

alpha(iri\_pt.alpha)

alpha(iri\_pd.alpha)

alpha(iri\_total.alpha)

describe(iri\_fs.icc)

describe(iri\_ec.icc)

describe(iri\_pt.icc)

describe(iri\_pd.icc)

describe(iri\_total.icc)

### EQ ###

icc(eq\_total.icc, model = "twoway", type = "agreement", unit = "single",

r0 = 0, conf.level = 0.95)

alpha(eq.alpha1) # timepoint 1

alpha(eq.alpha) # timepoint 2

describe(eq\_total.icc)

#

### PET ###

icc(pet\_total.icc, model = "twoway", type = "agreement", unit = "single",

r0 = 0, conf.level = 0.95)

alpha(pet.alpha1) # timepoint 1

alpha(pet.alpha) # timepoint 2

describe(pet\_total.icc)

#

### SITES ###

icc(sites.icc, model = "twoway", type = "agreement", unit = "single",

r0 = 0, conf.level = 0.95)

# Cronbach's Alpha not possible for SITES; just one item

describe(sites.icc)

range(sites.icc)

which.max(table(sites.icc$SITES.n.y))

prop.table(table(sites.icc$SITES.n.y))

round(100 \* prop.table(table(sites.icc$SITES.n.y)), digits= 1)

median(sites.icc$SITES.n.y)

#

### EYES ###

icc(eyes\_total.icc, model = "twoway", type = "agreement", unit = "single",

r0 = 0, conf.level = 0.95)

alpha(eyes.alpha1) # timepoint 1

alpha(eyes.alpha) # timepoint 2

describe(eyes\_total.icc)