Test-retest reliability of empathy measures

Markus Dieterich, Malte Elson, & Ian Hussey

# 

[**Table of existing test-retest estimates and estimated completion times 1**](#_g4h8nku56o7)

[**Implementation of measures in our survey 3**](#_18muphcxzbue)

[**Details of measures 3**](#_188eyorwourr)

[Self-report measures 3](#_x7mhdh1av0ye)

[Interpersonal Reactivity Index 3](#_5q4zz6n1k1z4)

[Empathy Quotient 5](#_jgwbjbvh5hzk)

[Basic Empathy Scale 5](#_9xbu8c3l79ef)

[Toronto Empathy Questionnaire (TEQ) 8](#_696h9gm7h9y4)

[Questionnaire of Cognitive and Affective Empathy (QCAE) 9](#_ivvghzghc26w)

[Multidimensional Emotional Empathy Scale 10](#_mfvgkjdf1w7m)

[Adolescent Measure of Empathy and Sympathy (AMES) 13](#_xk5q1rkla9u6)

[Affective and Cognitive Measure of Empathy (ACME) 14](#_d77erkj3gozj)

[Emotion Specific Empathy Questionnaire 15](#_s5mjz1t42fcs)

[Pictorial Empathy Test (PET) 15](#_hceweyk14b0r)

[Emotional Empathic Drive Short Scale (EED) 15](#_dhid6zl2uivx)

[Single Item Trait Empathy Scale 18](#_h7myjytvitaq)

[Balanced Emotional Empathy Scale (Mehrabian, 1996) 20](#_724r3s6ljdtq)

[Feeling and Thinking Scale (Garton & Gringart, 2005) 20](#_d8oddqil2n5w)

[Griffith Empathy Measure (Dadds et al., 2008) 20](#_dopzd0b1d4tb)

[Meta-Analysis of measures from 2009-2019 20](#_54mzubgpju2n)

[Behavioral Measures 20](#_5w94hj6guoe1)

[Reading the mind in the eyes Test 20](#_vaike86g2gku)

[Faces Test 21](#_skfxfii4houl)

[The Cambridge Mindreading (CAM) Face-Voice Battery 22](#_rx3csarlwm0e)

[Reading the mind in the voice 23](#_slkf24je5lm6)

[Reading the mind in films 23](#_owmjuy3nd8wu)

[**Power analysis 23**](#_e0sskdguhf4w)

[**Consent form and information sheet 23**](#_942ax35nw4p6)

[From Nele’s study 23](#_ta17x7mgn0v7)

[Adaptations needed 27](#_e644ee2jz2u7)

[**Preregistration 27**](#_y669tg3fzybd)

[Study Information 27](#_zbpxugyzl67z)

[Design Plan 29](#_9smse6po1muy)

[Sampling Plan 31](#_ni52kgrind86)

[Variables 35](#_lcps5cwwlycy)

[Analysis Plan 38](#_b0fa1exh1099)

[Other 41](#_r3e1zkcuh4s)

# 

# 

# Table of existing test-retest estimates and estimated completion times

| **Measure** | **Retest-Reliability** | **Est. Time** |
| --- | --- | --- |
| Interpersonal Reactivity Index (Davis, 1983) | r = .62 -.71 (Davis, 1983)  r = .67-.89 (Fernández et al., 2011), 60 Days  ICC = .71-.86 (Gilet et al., 2013), 12 weeks | 4-6 min |
| Empathy Quotient (Baron-Cohen, Wheelwright, 2004) | r = .96 (Baron-Cohen, Wheelwright, 2004), 12 months  r = .835 (Lawrence et al., 2004), 10-12 months  r = .84 (Kim and Lee, 2010), 4 weeks  r = .76-.85 (Preti et al., 2011), 4 weeks  r = .94 (Kosonogov, 2014), 2 weeks  r = .948 (Kose et al., 2018), 2 weeks  ICC = .896 (Zhang et al., 2018), 1 week  ICC = .82 (Zhao et al., 2018), 2 weeks | 5-10 min |
| Basic Empathy Scale (Jolliffe,Farrington, 2006) | r = .6110 - .7980 (Carré et al., 2013), 7 weeks | 2-5 min |
| Toronto Empathy Questionnaire (Spreng et al., 2009) | r = .81 (Spreng et al., 2009), 66 days  r = .73 (Totan et al., 2012), 3 weeks | 2-3 min |
| Questionnaire of Cognitive and Affective Empathy | r = .76 (Liang et al. 2019), 4 weeks  r = .84 (Powell & Roberts, 2017), 3 weeks | 3-5 min |
| Pictorial Empathy Test (PET) | r = .77 (Lindeman et al., 2018), 7 months | 1-2 min |
| Emotion Specific Empathy Questionnaire | No test-retest reliability findable | 8-12 min |
| Single Item Trait Empathy Scale (SITES) | r = .57 ( 2 weeks), r = . 67 (3 weeks), r = .62 (6 months) (Konrath et al., 2018) | 30 sec |
| Reading the mind in the eyes Test | ICC = .63 (Fernández-Abascal et al., 2013), 1 year  ICC = .833 (Vellante et al., 2012), 1 month  ICC = .65 (Yıldırım, 2011), 2 weeks  ICC = .92 (Charernboon and Lerthattasilp, 2017), 2 weeks (thai version) | 6-10 min |
| Faces Test | ICC = .85 (Charernboon, 2017), 2-4 weeks (thai version) | 2 min |

# Implementation of measures in our survey

Order of presentation is currently:

1. Single Item Trait Empathy Scale (SITES)
2. Interpersonal Reactivity Index (IRI)
3. Empathy Quotient (EQ)
4. Basic Empathy Scale (BES-A)
5. Toronto Empathy Questionnaire (TEQ)
6. Questionnaire of Cognitive and Affective Empathy (QCAE)
7. Emotion Specific Empathy Questionnaire (ESE)
8. Pictorial Empathy Test (PET)
9. Reading the Mind in the Eyes Task (EYES)
10. Faces Task (FACES)

# Details of measures

## Self-report measures

### Interpersonal Reactivity Index

Measuring Individual Differences in Empathy: Evidence for a Multidimensional Approach (Davis, 1983)

Cited: 12601

28 items

4-6 Minutes

Retest reliability:

.62 -.71 (Davis, 1983)

Four subscales: (Fantasy-Empathy, Perspective Taking, Empathic Concern, Personal Distress). Two subscales assess the cognitive dimension: Perspective Taking (PT) and fantasy (FS). The affective dimension is also assessed by two subscales: Personal Distress (PD) and Empathic Concern (EC)

INTERPERSONAL REACTIVITY INDEX

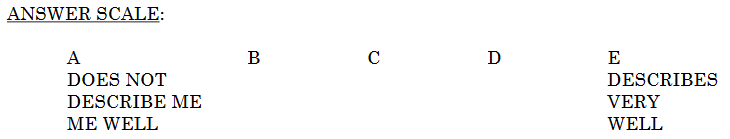
The following statements inquire about your thoughts and feelings in a variety of

situations. For each item, indicate how well it describes you by choosing the appropriate

letter on the scale at the top of the page: A, B, C, D, or E. When you have decided on your

answer, fill in the letter next to the item number. READ EACH ITEM CAREFULLY

BEFORE RESPONDING. Answer as honestly as you can. Thank you.



1. I daydream and fantasize, with some regularity, about things that might happen to me.

(FS)

2. I often have tender, concerned feelings for people less fortunate than me. (EC)

3. I sometimes find it difficult to see things from the "other guy's" point of view. (PT) (-)

4. Sometimes I don't feel very sorry for other people when they are having problems. (EC)

(-)

5. I really get involved with the feelings of the characters in a novel. (FS)

6. In emergency situations, I feel apprehensive and ill-at-ease. (PD)

7. I am usually objective when I watch a movie or play, and I don't often get completely

caught up in it. (FS) (-)

8. I try to look at everybody's side of a disagreement before I make a decision. (PT)

9. When I see someone being taken advantage of, I feel kind of protective towards them.

(EC)

10. I sometimes feel helpless when I am in the middle of a very emotional situation. (PD)

11. I sometimes try to understand my friends better by imagining how things look from

their perspective. (PT)

12. Becoming extremely involved in a good book or movie is somewhat rare for me. (FS) (-)

13. When I see someone get hurt, I tend to remain calm. (PD) (-)

14. Other people's misfortunes do not usually disturb me a great deal. (EC) (-)

15. If I'm sure I'm right about something, I don't waste much time listening to other

people's arguments. (PT) (-)

16. After seeing a play or movie, I have felt as though I were one of the characters. (FS)

17. Being in a tense emotional situation scares me. (PD)

18. When I see someone being treated unfairly, I sometimes don't feel very much pity for

them. (EC) (-)

19. I am usually pretty effective in dealing with emergencies. (PD) (-)

20. I am often quite touched by things that I see happen. (EC)

21. I believe that there are two sides to every question and try to look at them both. (PT)

22. I would describe myself as a pretty soft-hearted person. (EC)

23. When I watch a good movie, I can very easily put myself in the place of a leading

character. (FS)

24. I tend to lose control during emergencies. (PD)

25. When I'm upset at someone, I usually try to "put myself in his shoes" for a while. (PT)

26. When I am reading an interesting story or novel, I imagine how I would feel if the

events in the story were happening to me. (FS)

27. When I see someone who badly needs help in an emergency, I go to pieces. (PD)

28. Before criticizing somebody, I try to imagine how I would feel if I were in their place.

(PT)

### Empathy Quotient

The Empathy Quotient: An Investigation of Adults with Asperger Syndrome or High Functioning Autism, and Normal Sex Differences (Simon Baron-Cohen and Sally Wheelwright)

Cited: 4993

60 items (including 20 filler items)

5-10 minutes

Retest reliability:

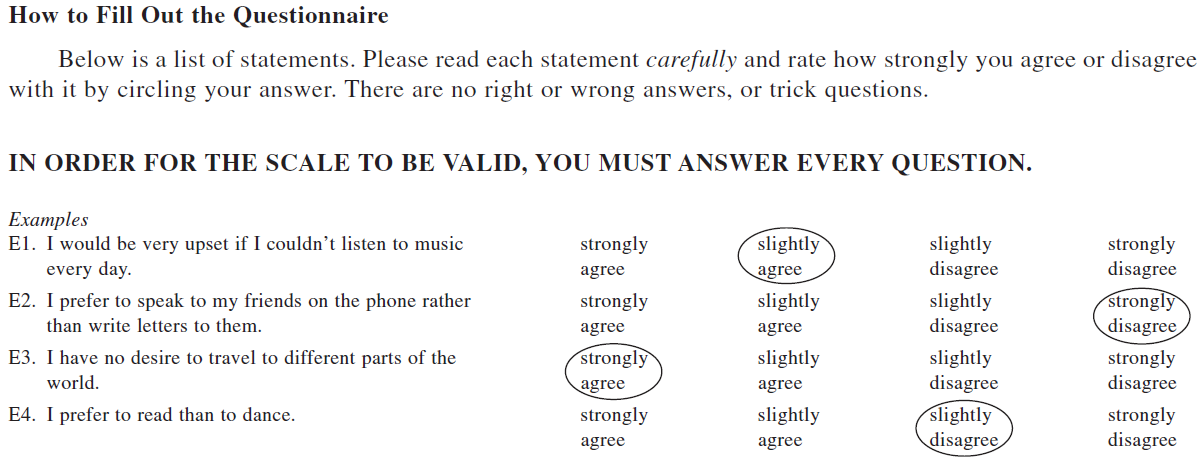
.96 12 months (Baron-Cohen, Wheelwright, 2004)

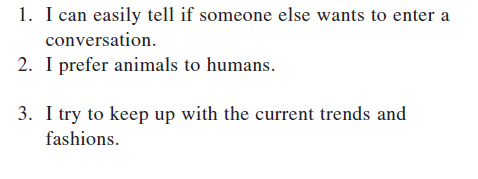
.835 10-12 months (Lawrence et al., 2004)

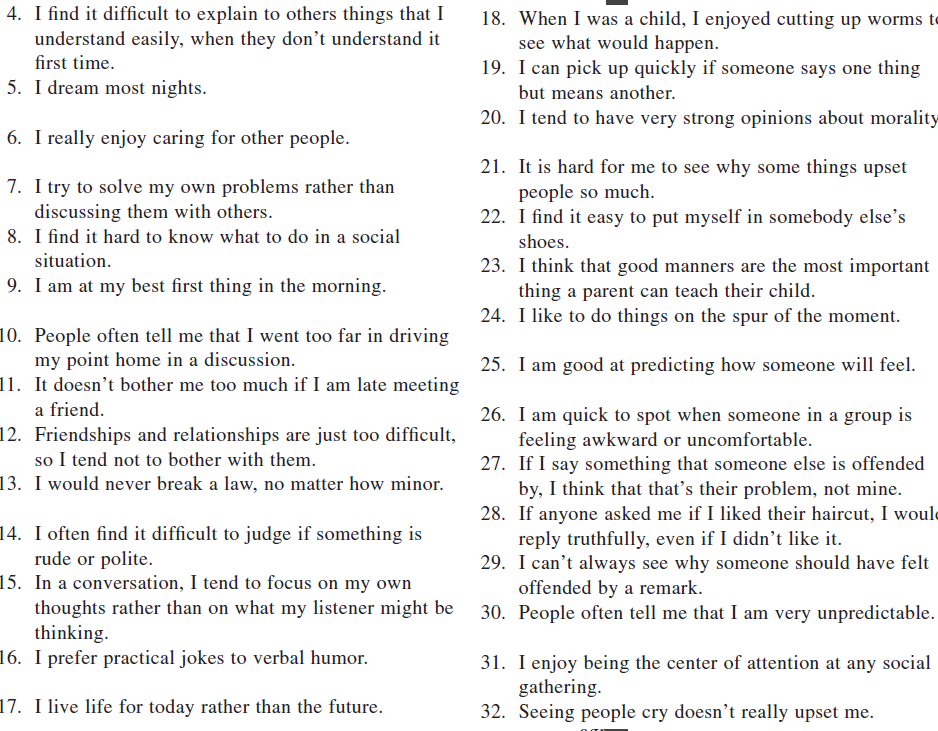
Short version: Empathy Quotient – 10 items (EQ-10)

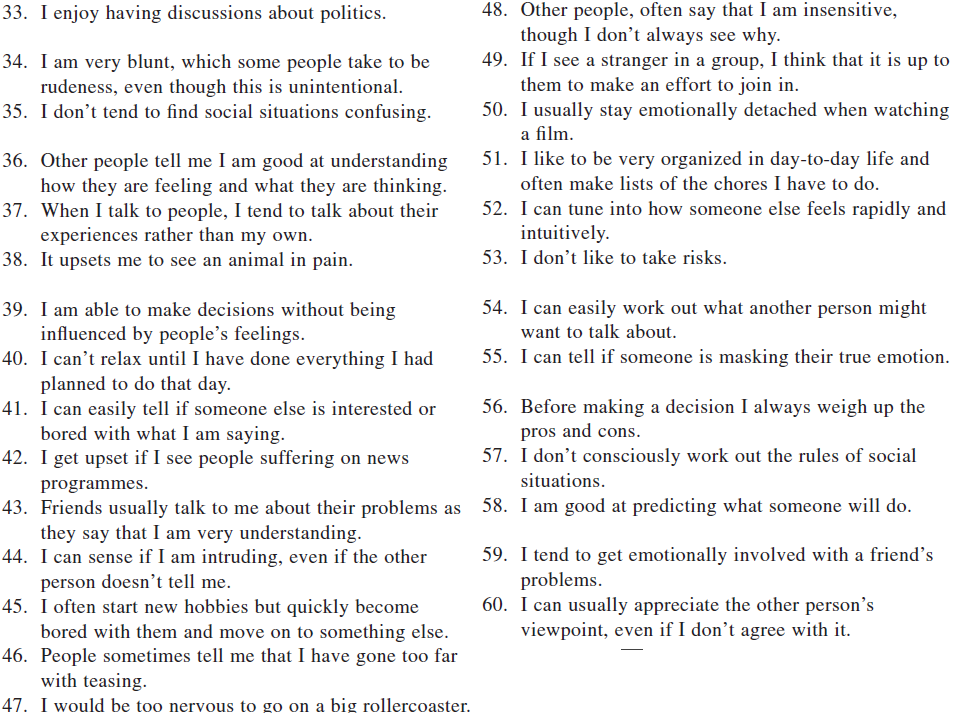
10 Items

1-2 Minutes









### Basic Empathy Scale

Development and validation of the Basic Empathy Scale (Darrick Jolliffe, David P. Farrington)

The Basic Empathy Scale in Adults (BES-A): Factor Structure of a Revised Form (CARRÉ ET AL.)

Cited: 1790

20 Items

2-5 minutes

Retest reliability:

.6110- .7980 7 weeks ( CARRÉ ET AL., 2013)

5-point Likert type scale (1 Strongly Disagree,2 Disagree, 3 Neither Agree nor Disagree, 4 Agree, 5 Strongly Agree).

Items of the Basic Empathy Scale (20 items)

1. My friends’ emotions don’t affect me much.

2. After being with a friend who is sad about something, I usually feel sad.

3. I can understand my friend’s happiness when she/he does well at something.

4. I get frightened when I watch characters in a good scary movie.

5. I get caught up in other people’s feelings easily.

6. I find it hard to know when my friends are frightened.

7. I don’t become sad when I see other people crying.

8. Other people’s feeling don’t bother me at all.

9. When someone is feeling ‘down’ I can usually understand how they feel.

10. I can usually work out when my friends are scared.

11. I often become sad when watching sad things on TV or in films.

12. I can often understand how people are feeling even before they tell me.

13. Seeing a person who has been angered has no effect on my feelings.

14. I can usually work out when people are cheerful.

15. I tend to feel scared when I am with friends who are afraid.

16. I can usually realize quickly when a friend is angry.

17. I often get swept up in my friends’ feelings.

18. My friend’s unhappiness doesn’t make me feel anything.

19. I am not usually aware of my friends’ feelings.

20. I have trouble figuring out when my friends are happy

### Toronto Empathy Questionnaire (TEQ)

The Toronto Empathy Questionnaire: Scale development and initial validation of a factor-analytic solution to multiple empathy measures (Spreng et al., 2009)

Cited: 1127

16 Items

2-3 Minutes

Retest reliability:

r = .81 (66.1 days)

Response options:

Toronto Empathy Questionnaire instructions

Below is a list of statements. Please read each statement carefully and rate how frequently you feel or act in the manner described. Circle your answer on the response form. There are no right or wrong answers or trick questions. Please answer each question as honestly as you can.

1. When someone else is feeling excited, I tend to get excited too

2. Other people’s misfortunes do not disturb me a great deal

3. It upsets me to see someone being treated disrespectfully

4. I remain unaffected when someone close to me is happy

5. I enjoy making other people feel better

6. I have tender, concerned feelings for people less fortunate than me

7. When a friend starts to talk about his\her problems, I try to steer the conversation

towards something else

8. I can tell when others are sad even when they do not say anything

9. I find that I am “in tune” with other people’s moods

10. I do not feel sympathy for people who cause their own serious illnesses

11. I become irritated when someone cries

12. I am not really interested in how other people feel

13. I get a strong urge to help when I see someone who is upset

14. When I see someone being treated unfairly, I do not feel very much pity for them

15. I find it silly for people to cry out of happiness

16. When I see someone being taken advantage of, I feel kind of protective towards him\her

Scoring Item responses are scored according to the following scale for positively

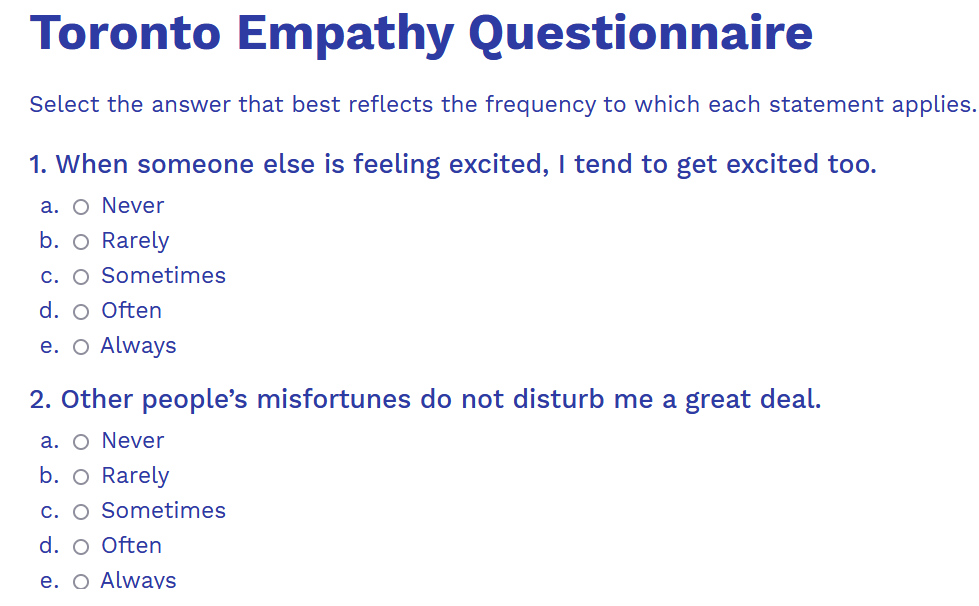
worded items 1, 3, 5, 6, 8, 9, 13, 16. Never = 0; Rarely = 1; Sometimes = 2; Often =

3; Always = 4. The following negatively worded items are reverse scored: 2, 4, 7, 10,

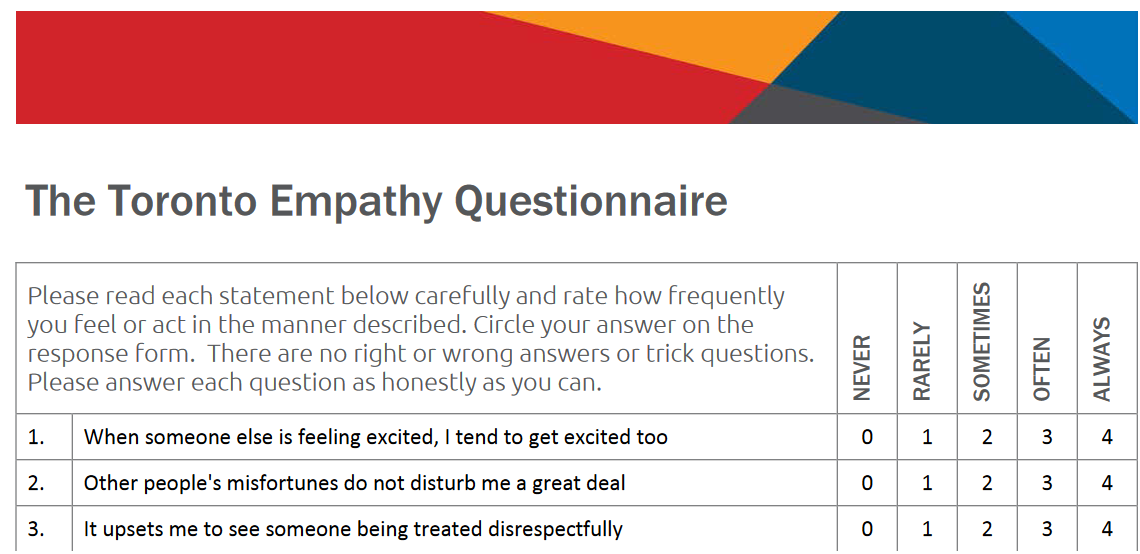
11, 12, 14, 15. Scores are summed to derive total for the Toronto Empathy

Questionnaire.

Examples of presentation:



Source: https://embrace-autism.com/toronto-empathy-questionnaire/

Source: <http://www.highdefpeople.com/wp-content/uploads/2018/07/The-Toronto-Empathy-Questionnaire.pdf>

### Questionnaire of Cognitive and Affective Empathy (QCAE)

The QCAE: A Questionnaire of Cognitive and Affective Empathy (Reniers et al., 2011)

Cited: 840

31 Items

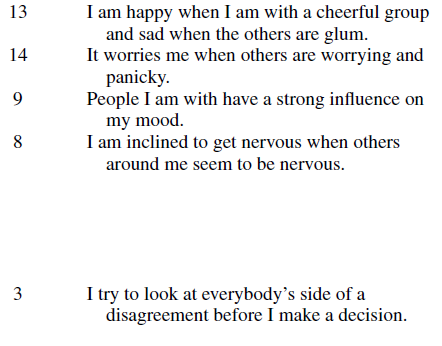
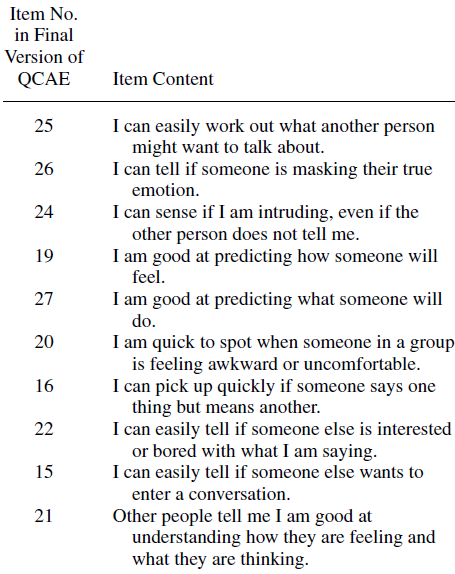
3-5 Minutes

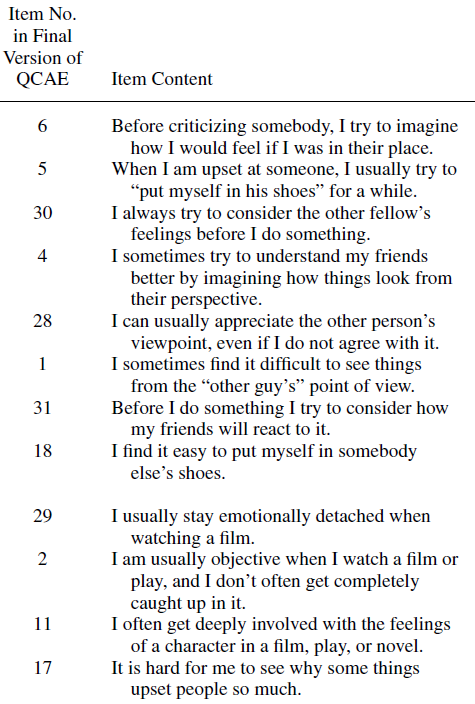
Retest reliability:

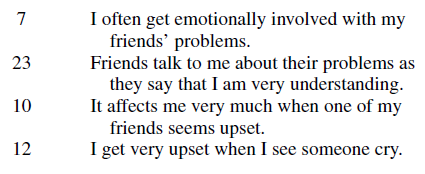
.84 (3 weeks) (Powell & Roberts, 2017)

Response options:

Items are rated on level of agreement using a 4-point Likert scale with the following response options: 4 (strongly agree), 3 (slightly agree), 2 (slightly disagree), and 1 (strongly disagree).







### Multidimensional Emotional Empathy Scale

A Measure of Emotional Empathy for Adolescents and Adults (Caruso & Mayer, 1998)

Cited: 226

30 Items

3-5 Minutes

1. I cry easily when watching a sad movie.

2. Certain pieces of music can really move me.

3. Seeing a hurt animal by the side of the road is very upsetting.

4-R. I don't give others' feelings much thought.

5. It makes me happy when I see people being nice to each other.

6. The suffering of others deeply disturbs me.

7. I always try to tune in to the feelings of those around me.

8 I get very upset when I see a young child who is being treated meanly.

9-R. Too much is made of the suffering of pets or animals.

10. If someone is upset I get upset, too.

11. When I'm with other people who are laughing I join in.

12. It makes me mad to see someone treated unjustly.

13-R. I rarely take notice when people treat each other warmly.

14. I feel happy when I see people laughing and enjoying themselves.

15. It's easy for me to get carried away by other people's emotions.

16-R. My feelings are my own and don’t reflect how others feel.

17. If a crowd gets excited about something so do I.

18. I feel good when I help someone out or do something nice for someone.

19. I feel deeply for others.

20-R. I don't cry easily.

21. I feel other people's pain.

22. Seeing other people smile makes me smile.

23. Being around happy people makes me feel happy, too.

24. TV or news stories about injured or sick children greatly upset me.

25. I cry at sad parts of the books I read.

26. Being around people who are depressed brings my mood down.

27-R. I find it annoying when people cry in public.

28. It hurts to see another person in pain.

29. I get a warm feeling for someone if I see them helping another person.

30. I feel other people's joy.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Note: R indicates a reverse-scored item. To score the scale, change the scoring on the reverse-scored items (1=5, 2=4, 3=3, 4=2, 5=1). Add all the scores for the Total score and divide by 30. Add the following items together for each scale, and divide by the number of items: Suffering ( 3, 5, 6, 8, 12, 18, 24, 28); Positive Sharing (14, 22, 23, 29, 30); Responsive Crying (1, 20, 25); Emotional Attention (4, 9, 13, 27); Feel for Others (10, 15, 16, 21); Emotional Contagion (11, 17). Take the mean of these sub-scales to compute a General Empathy scale.

### Adolescent Measure of Empathy and Sympathy (AMES)

Development of the Adolescent Measure of Empathy and Sympathy (AMES) (Vossen, H.G.M., Piotrowski, J.T., Valkenburg, P.M. (2015))

12 Items

1-3 Minutes

Retest reliability:

r = .56 for affective empathy, r = .66 for cognitive empathy, and

r = .69 for sympathy. (2 weeks)

Cited: 202

Factor 1: Cognitive empathy

1. I can often understand how people are feeling even before they tell me

2. I can tell when a friend is angry even if he/she tries to hide it

3. I can tell when someone acts happy, when they actually are not

4. I can easily tell how others are feeling

Factor 2: Affective empathy

5. When a friend is scared, I feel afraid

6. When my friend is sad, I become sad too

7. When a friend is angry, I feel angry too

8. When people around me are nervous, I become nervous too

Factor 3: Sympathy

9. I feel sorry for someone who is treated unfairly

10. I feel concerned for other people who are sick

11. I am concerned for animals that are hurt

12. I feel sorry for a friend who feels sad

### Affective and Cognitive Measure of Empathy (ACME)

Fixing the Problem With Empathy: Development and Validation of the Affective and Cognitive Measure of Empathy (Vacho, Lynam, 2016)

Cited: 181

36 Items

3-5 Minutes

Items and Scoring for the Affective and Cognitive Measure of Empathy (ACME).

1 I have a hard time reading people’s emotions

2 I think it’s fun to push people around once and a while

3 I can tell when someone is afraid

4 It’s obvious when people are pretending to be happy

5 I love watching people get angry

6 I enjoy seeing strangers get scared

7 It makes me feel good to help someone in need

8 I get excited to give someone a gift that I think they will enjoy

9 I usually understand why people feel the way they do

10 When my friends are having a good time I often get angry

11 People who are cheery disgust me

12 I don’t worry much about hurting people’s feelings

13 I don’t really care if other people feel happy

14 I have a hard time figuring out what someone else is feeling

15 I can tell when people are about to lose their temper

16 I can usually predict how someone will feel.

17 I don’t really care if people are feeling depressed

18 I like making other people uncomfortable

19 I get a kick out of making other people feel stupid

20 When my friends get angry I often feel like laughing

21 Sometimes I enjoy seeing people cry

22 Other people’s feelings don’t bother me at all

23 I feel awful when I hurt someone’s feelings

24 Other people’s misfortunes don’t bother me much

25 I can usually tell how people are feeling

26 Sometimes it’s funny to see people get humiliated

27 If I could get away with it, there are some people I would enjoy hurting

28 If I see that I am doing something that hurts someone, I will quickly stop

29 I often try to help people feel better when they are upset

30 I enjoy making others happy

31 I am not good at understanding other people’s emotions

32 People have told me that I’m insensitive

33 I can usually guess what’s making someone angry

34 People don’t have to tell me when they’re sad, I can see it in their faces

35 I find it hard to tell when someone is sad

36 I admit that I enjoy irritating other people

Cognitive Empathy (COG) = 1r, 3, 4, 9, 14r, 15, 16, 25, 31r, 33, 34, 35r

Affective Resonance (RES) = 7, 8, 12r, 13r, 17r, 22r, 23, 24r, 28, 29, 30, 32r

Affective Dissonance (DIS) = 2r, 5r, 6r, 10r, 11r, 18r, 19r, 20r, 21r, 26r, 27r, 36r

Note. r = reverse scored item (6—original score). The items are administered on 5-point Likert-type scale ranging from Strongly disagree (1) to Strongly agree (5). On all three scales (including DIS), high scores indicate greater empathy.

### Emotion Specific Empathy Questionnaire

An emotion-differentiated perspective on empathy with the emotion specific empathy questionnaire (Sally Olderbak et al., 2014)

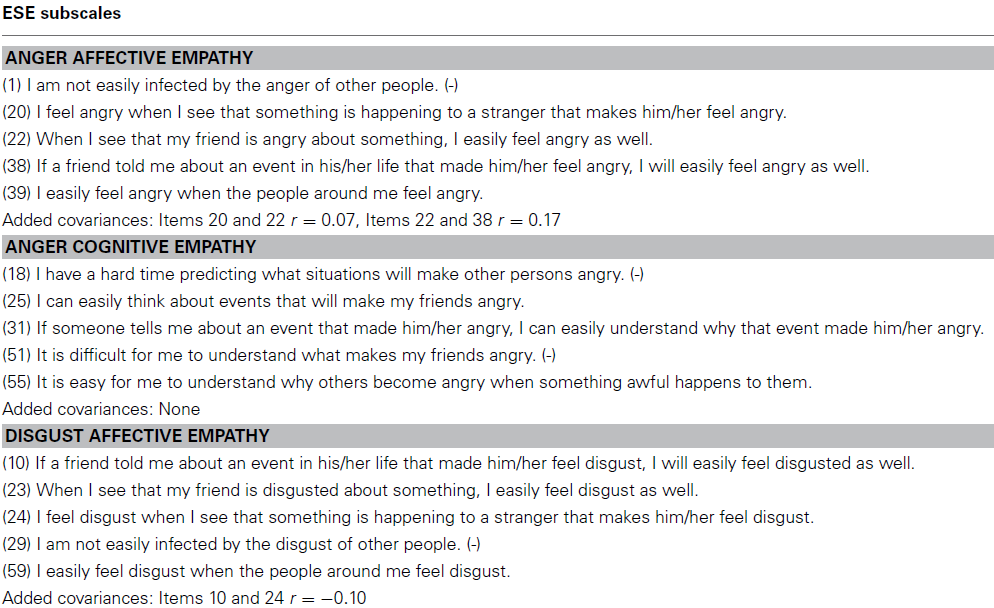
Cited: 59

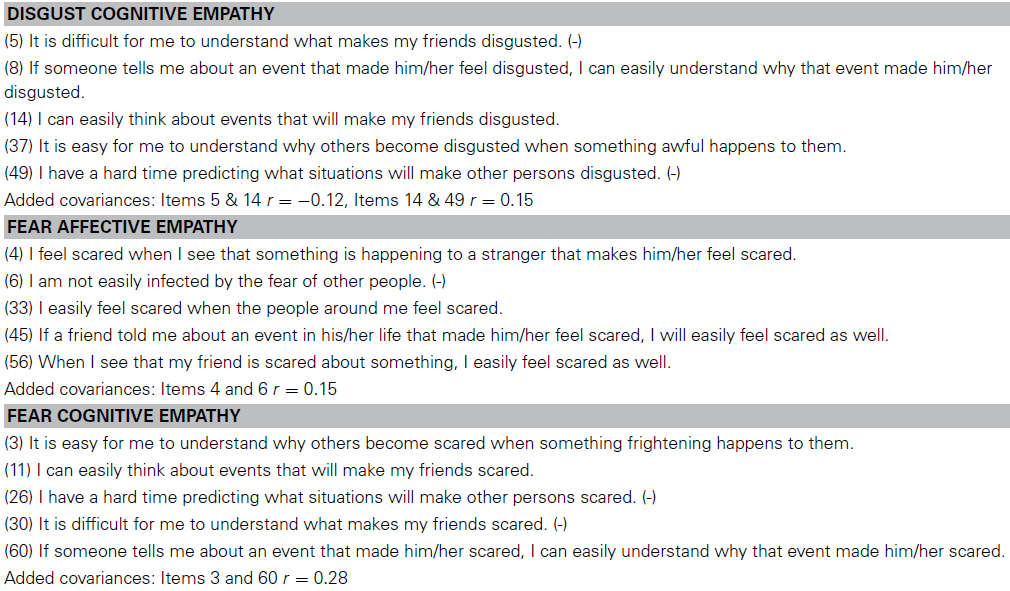
60 Items

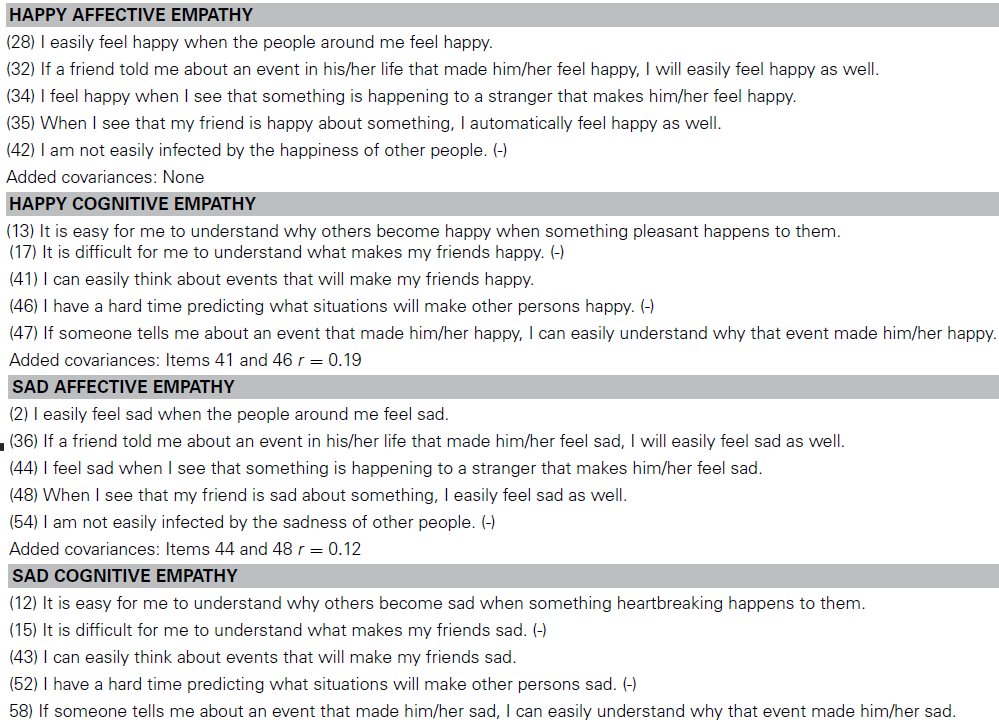
8-12 Minutes

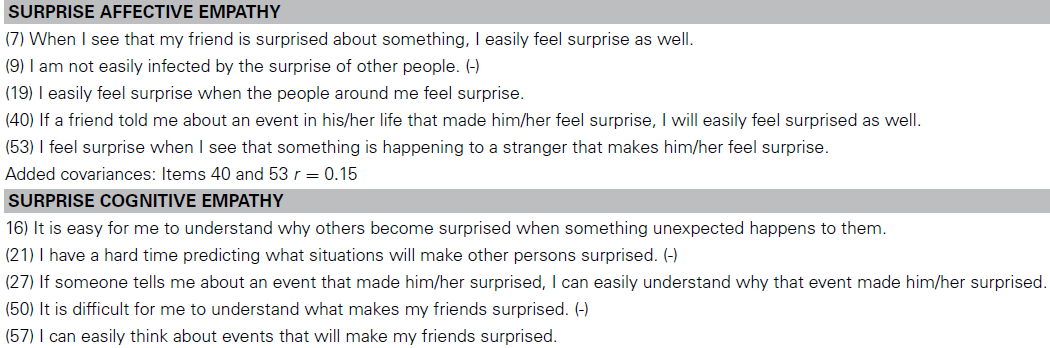
The measure uses a 7-point likertscale: -3,Disagree Strongly; -2,Disagree

Somewhat; -1,Disagree Slightly; 0,Neutral; 1, Agree Slightly; 2, Agree Somewhat;3, Agree Strongly.









### Pictorial Empathy Test (PET)

Pictorial Empathy Test (PET) An Easy-to-Use Method for Assessing Affective Empathic

Reactions (Lindeman et al., 2016)

Cited: No information on Google Scholar (0 times?)

7 Pictures (Question: “How emotionally moving do you find the photograph?”)

1-2 Minutes

Retest Reliability:

r =.77 (7 months)

“How emotionally moving do you find the photograph?”

1 = not at all, 2 = a little bit, 3 = it arouses some feelings, 4 = quite a lot, 5 = very much).





### Emotional Empathic Drive Short Scale (EED)

Emotional Empathic Drive Short Scale (EED). The first short scale measuring emotional empathy. (Karlstetter, W., 2017).

Cited: Not discoverable with Google Scholar

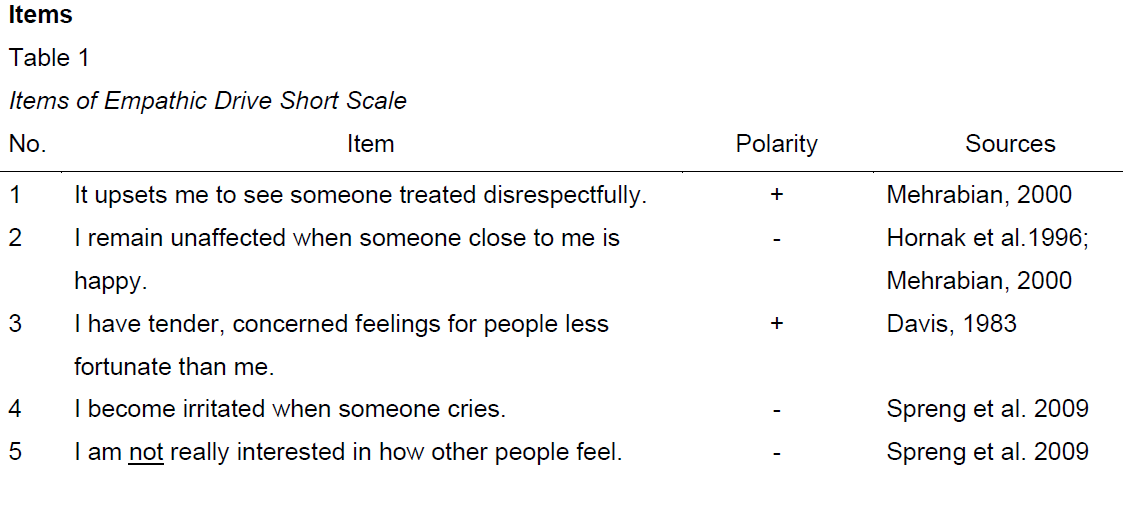
5 items

1 Minute

Retest reliability:  
r = .62 (61.1 days)

Instruction

Before the five items of the Empathic Drive Short Scale (EED) are presented, respondents read: “Below is a list of statements. Please read each statement carefully and rate how frequently you feel or act in the manner described. There are no right or wrong answers or trick questions. Please answer each question as honestly as you can.” The response categories for each of the five items are (0) never, (1) rarely, (2) sometimes, (3) often, (4) always.



### Single Item Trait Empathy Scale

Development and validation of the Single Item Trait Empathy Scale (SITES) (Konrath el al., 2018)

1 Item

30 Seconds

Retest reliability:

r = .57 ( 2 weeks), r = . 67 (3 weeks), r = .62 (6 months) (Konrath et al., 2018)  
  
To what extent does the following statement describe you: “I am an empathetic person,” rated using a scale that ranges from 1= Not very true of me to 5=Very true of me.

Other self-report scales:

### Balanced Emotional Empathy Scale (Mehrabian, 1996)

XXX

### Feeling and Thinking Scale (Garton & Gringart, 2005)

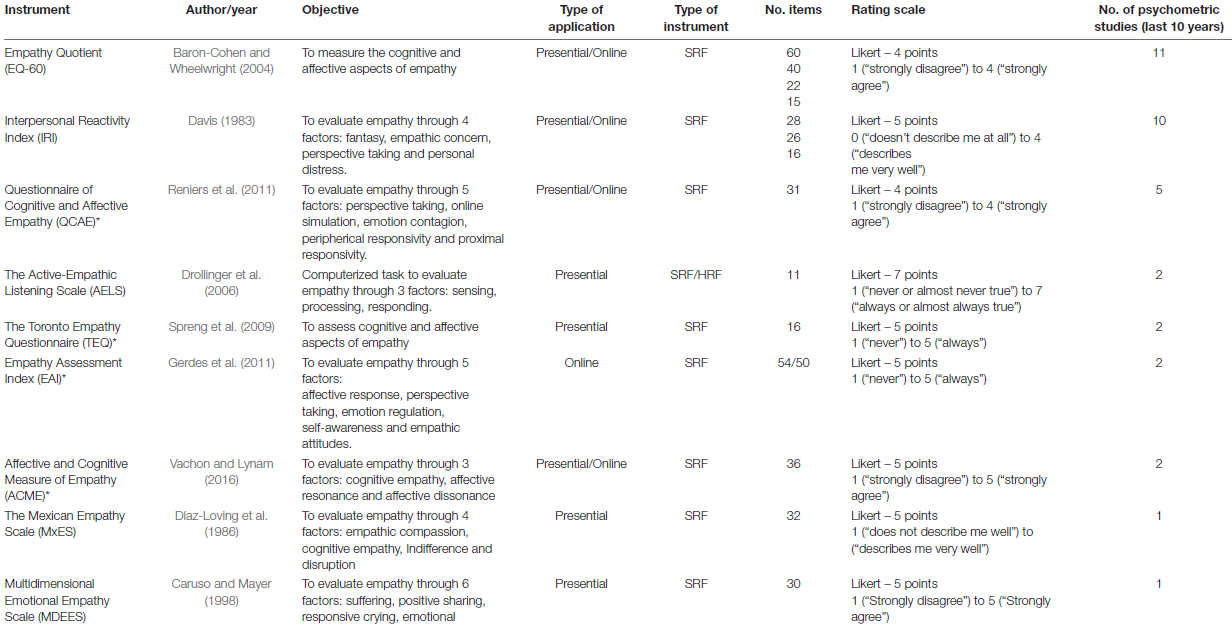
XXX

### Griffith Empathy Measure (Dadds et al., 2008)

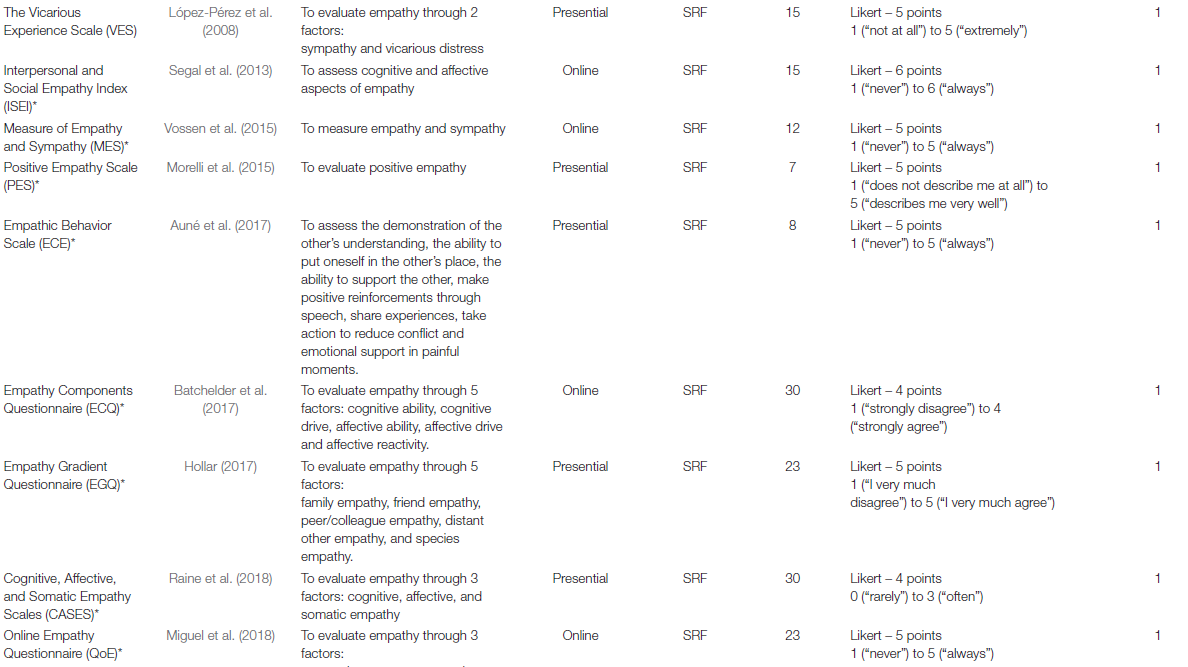
XXX

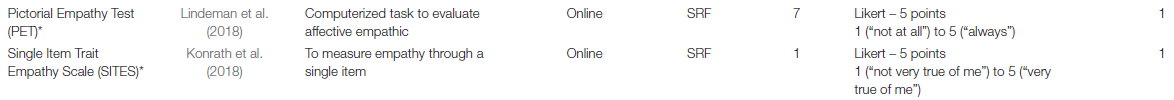
### Meta-Analysis of measures from 2009-2019

Empathy: Assessment Instruments and Psychometric Quality – A Systematic Literature Review With a Meta-Analysis of the Past Ten Years (de Lima, Osório, 2021)









## Behavioral Measures

### Reading the mind in the eyes Test

The "Reading the Mind in the Eyes" Test revised version: a study with normal adults, and adults with Asperger syndrome or high-functioning autism (Baron-Cohen, 2001)

Cited: 6848

36 Items (Pictures of faces showing emotions) with 4 options

6-10 Minutes

Retest Reliability:  
.63 (1 year, using ICC) Test-retest reliability of the ‘Reading the Mind in the Eyes’ test: a one-year follow-up study (Fernández-Abascal et al., 2013)

.833 (1 month, using ICC) The “Reading the Mind in the Eyes” test: Systematic review of psychometric properties and a validation study in Italy (Vellante et al., 2012)

.65 (2 weeks, using ICC) Investigation of the reliability of the "reading the mind in the eyes test" in a Turkish population (Yıldırım, 2011)

Standardization:

- Comes with instructions and A hand-out with word definitions (emotions). Feasible in an online survey?

- normally used as a pen and paper version

- 36 images with 4 answer options, should be easily installable in an online survey

### Faces Test

Is There a "Language of the Eyes"? Evidence from Normal Adults, and Adults with Autism or Asperger Syndrome (S. Baron-Cohen, S. Wheelwright and T. Jolliffe, (1997))

Cited: 1388

20 Items (10 pictures of a face showing basic emotions, 10 pictures of a face showing complex emotions) with two options

2 Minutes

Standardization:

- no special instructions for the participants

- was originally used as a pen and paper version

20 pictures with 2 answer option, should be easily installable in an online study

### The Cambridge Mindreading (CAM) Face-Voice Battery

The Cambridge Mindreading (CAM) Face-Voice Battery: Testing complex emotion recognition in adults with and without Asperger syndrome (Golan et al., 2006)

Cited: 416

The battery includes two tasks: emotion recognition in the face and emotion recognition in the voice. Each of these tasks has fifty questions, in which the participant is either watching 3–5 second silent clips of actors portraying an emotion (facial task), or listening to short sentences, spoken in a particular emotional intonation (vocal task). After watching the

clip/listening to the voice recording, the participant is presented with four adjectives and is asked to ‘‘choose the word that best describes how the person is feeling’’.

“The CAM took about 45 minutes to complete”

- probably too long to use in the study

### Reading the mind in the voice

The ‘Reading the Mind in the Voice’ Test-Revised: A Study of Complex Emotion Recognition in Adults with and Without Autism Spectrum Conditions (Golan et al., 2006)

Cited: 376

25 Items (sound clips of human voices) with 4 options

4-6 Minutes

Retest reliability:  
r = 0.8 (10 weeks)

Standardization:

- Comes with instructions and A hand-out with word definitions (emotions). Feasible in an online survey?

- participants in the original study were presented the study on a computer with DMDX

experimental software and used headphones

- the sound clips were played just one time

- 25 audio files (.WAV) with 4 answer options, should be installable in an online survey

### Reading the mind in films

The ‘‘Reading the Mind in Films’’ Task: Complex emotion recognition in adults with and without autism spectrum conditions (Golan et al., 2006)

Cited: 269

22 Items (short scenes from movies,5-30 seconds long, M=14.8, SD=9.2)

5-10 Minutes

- Comes with instructions and A hand-out with word definitions (emotions). Feasible in an online survey?

- participants in the original study were presented the study on a computer with DMDX

experimental software and used headphones

- 22 Video clips (.AVI) with 4 answer options

- video quality is not the best

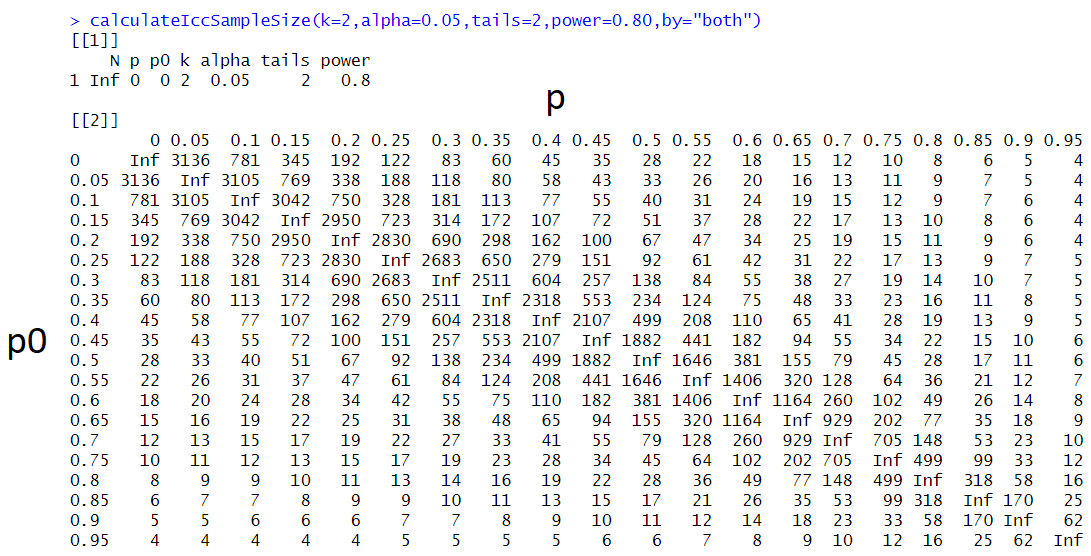
# Power analysis

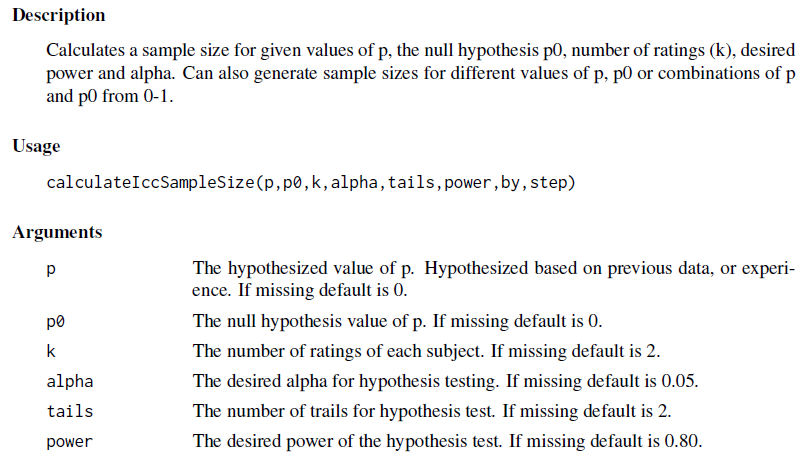
[see Rmd and html in slack for updated version of this with more accessible table]

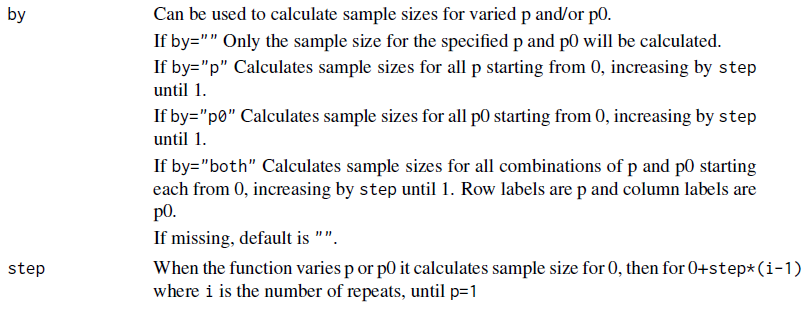
Package ‘ICC.Sample.Size’

calculateIccSampleSize

Function to calculate sample size required for studies where ICC is primary outcome.







# Consent form and information sheet

## From Nele’s study

Ruhr University Bochum, Faculty of Psychology

Psychology of Human-Technology Interaction

Jun.-Prof. Dr. Malte Elson

Contact person for any queries:

Larina Hillemann

larina.hillemann@rub.de

General Information for Participants

Study "XXXX"

In this study, we examine [study topic].

Participation requirements

To be eligible to participate in this study, you must be older than 18 years and be an English native speaker.

Procedure of the Study

The following study consists of a short online questionnaire. In total, participation takes about 15 minutes.

Your task is to answer the questions sincerely. Please answer all questions, even if you feel you have answered individual questions more than once. This is especially important for our statistical analysis.

If you still have questions, please contact the investigator (Larina Hillemann, larina.hillemann@rub.de).

Voluntariness and Anonymity

Participation in the study is voluntary. You can terminate your participation in this study at any time without giving reasons, without incurring any disadvantages.

The data and personal communications collected as part of this study and described above will be treated confidentially. Thus, those project staff members who have knowledge of personal data through direct contact with you are subject to the obligation of confidentiality. Furthermore, the publication of the results of the study will be in anonymous form, i.e. without your data being able to be assigned to your person.

Privacy

The collection of your personal data described above takes place without asking for your name. Your answers and results are stored under your personal Prolific-ID that only you yourself have access to. This means that it is not possible for anyone to associate your data with your name. After completion of the data evaluation, but no later than 30 June 2022, your ID will be deleted. Your data will then be anonymized. However, as long as the ID exists in the data, you can request the deletion of the data collected from you. To do so, you do not need to tell us your name, just your ID. Please store your personal Prolific-ID carefully so that you can request deletion of your data later if necessary. The anonymized data will be stored for at least 10 years.

According to the General Data Protection Regulation (GDPR), you have the following rights:

• Information on processing of personal data (Art 15)

• Withdrawal of consent given (Art 7)

• Correction (Art 16)

• Deletion (Art 17)

• Restriction of processing (Art 18)

In addition, you can contact the data protection officer of the Ruhr University (Dr. Kai-Uwe Loser, dsb@rub.de, +49 (234) 322-8720), and a data protection supervisory authority with questions at any time. The State Commissioner for Data Protection and Freedom of Information (ldi.nrw.de) is responsible for the Ruhr University.

Retention Period for Anonymized Data

The fully anonymized data are made publicly available via the research repository OSF. This procedure serves to ensure good scientific practice. Other researchers can thus, for example, reproduce the evaluation or test an alternative evaluation.

Compensation

You will receive compensation for participating in the study. The compensation will be paid to you via the "Prolific" platform. All related information is kept completely separate from the investigation data.

Privacy of Used Services

The services "Prolific" and "Qualtrics" are used to implement this study. Information on the privacy policies of these services can be found here on "Prolific" https://www.prolific.co/assets/docs/Prolific\_privacy-policy.pdf and here on "Qualtrics" https://www.qualtrics.com/privacy-statement/

Declaration of Consent

I have been informed in writing about the study and the experimental procedure. I agree to carefully complete the questionnaires. Insofar as I had questions about this intended study, they were answered completely and to my satisfaction by Larina Hillemann (larina.hillemann@rub.de).

In the case of studies, the legal basis for processing the personal data concerning you is your voluntary written consent in accordance with the DS-GVO (Art. 6 para. 1 letter a).

I agree to the described collection and processing of my demographic data and answers to questions about information security. The recording and evaluation of the data is done without giving my name in the Psychology of Human-Technology Interaction Group of the Ruhr University Bochum using my personal Prolific-ID that only I know. This means that it is not possible for anyone but me to associate my data with my name. I can revoke my consent to the storage of my data at any time with effect for the future without incurring any disadvantages. After completion of the data evaluation, but no later than 1 January 2022, my ID will be deleted. As long as the ID exists, I can request deletion of all my data.

I agree that my fully anonymized data may be used for research purposes. For this purpose, they will be kept for at least 10 years after data evaluation, or at least 10 years after publication of a paper on this study and made publicly available via the research repository the Open Science Framework (https://osf.io).

I have had sufficient time to make a decision and am willing to participate in the above study. I know that participation in the study is voluntary and that I can terminate participation at any time without giving reasons.

The participant information is part of this consent form.

I am at least 18 years old, I am an English native speaker, I have read the consent form, and agree to participate in the study.

[tickbox]

## Adaptations needed

* Nele says that her instructions aren’t completely in line with Malte’s recent paper, but she didn’t specify how. Could you check out Malte’s paper and see if you spot any changes needed? <https://psyarxiv.com/znb9m>
* Add a single instruction on the next page (because people dont read consent forms) that response options change between different pages of the survey and to pay attention.

# Preregistration

### **Study Information**

1. Title (required): An assessment of the reliability of self-report and behavioral measures of empathy

1.1. Provide the working title of your study. It may be the same title that you submit for publication of your final manuscript, but it is not a requirement.

1.2. **Example**: Effect of sugar on brownie tastiness.

1.3. **More info**: The title should be a specific and informative description of a project. Vague titles such as 'Fruit fly preregistration plan' are not appropriate.

1. Authors (required): Markus Dieterich, Malte Elson, Ian Hussey
2. Description (optional): Empathy is a crucial part of everyday human interaction. Many psychopathological disorders, such as autism spectrum disorders, psychopathy, alexithymia and antisocial, narcissistic and borderline personality disorders, are associated with a lack of empathy (Decety & Moriguchi, 2007). Also, empathy is a topic researched in the field of social psychology (e.g. intergroup empathy).  
   Thus, for the understanding and treatment of these disorders and the research of empathy, it is mandatory to use valid and reliable instruments.  
   The aim of this study is to check the test-retest reliability and internal consistency of several empathy measures. The following self-report scales and behavioral tasks will be collected in a pilot study with around 10 participants:

Single Item Trait Empathy Scale (SITES) (Konrath et al., 2018)

Interpersonal Reactivity Index (IRI) (Davis, 1983)

Empathy Quotient (EQ) (Simon Baron-Cohen & Wheelwright, 2004)

Basic Empathy Scale (BES-A) (Carré et al., 2013)

Toronto Empathy Questionnaire (TEQ) (Spreng et al., 2009)

Questionnaire of Cognitive and Affective Empathy (QCAE) (Reniers et al., 2011)

Emotion Specific Empathy Questionnaire (ESE) (Olderbak et al., 2014)

Pictorial Empathy Test (PET) (Lindeman et al., 2018)

Reading the Mind in the Eyes Task (EYES) (Baron-Cohen et al., 2001)

Faces Task (FACES) (Baron-Cohen et al., 1997)

The purpose of the pilot study is to estimate the amount of time that it takes to complete each measure. Based on this, a number of measures will be selected that take around 20-30 minutes to finish altogether, for the purposes of both limiting participant burden and in order to determine our sample size based on finite money to pay participants.

In the full study, participants will do the survey two times, with an interval of approximately two weeks. The test-retest reliability of each measure will be determined by calculating the intraclass-correlation-coefficient (ICC2) between the two time points. The internal consistency will be determined by calculating Cronbach’s Alpha for each measure.

3.1. Please give a brief description of your study, including some background, the purpose of the of the study, or broad research questions.

3.2. **Example**: Though there is strong evidence to suggest that sugar affects taste preferences, the effect has never been demonstrated in brownies. Therefore, we will measure taste preference for four different levels of sugar concentration in a standard brownie recipe to determine if the effect exists in this pastry.

3.3. **More info**: The description should be no longer than the length of an abstract. It can give some context for the proposed study, but great detail is not needed here for your preregistration.

1. Hypotheses (required):

As the goal of this study is to estimate the test-retest reliability and internal consistency of empathy measures, no specific hypothesis will be tested. However, the detected estimates (or their confidence intervals, respectively) will be compared to the estimates detected by published literature (see the following table for an overview).

| **Measure** | **Retest-Reliability study** |
| --- | --- |
| Interpersonal Reactivity Index, (Davis, 1983) | ICC = .71-.86 (Gilet et al., 2013) |
| Empathy Quotient:(Baron-Cohen, Wheelwright, 2004) | ICC = .82 (Zhao et al., 2018) |
| Basic Empathy Scale (Jolliffe,Farrington, 2006) | r = .6110 - .7980 (Carré et al., 2013) |
| Toronto Empathy Questionnaire (Spreng et al., 2009) | r = .73 (Totan et al., 2012) |
| Questionnaire of Cognitive and Affective Empathy | r = .84 (Powell & Roberts, 2017) |
| Pictorial Empathy Test (PET) | r = .77 (Lindeman et al., 2018) |
| Emotion Specific Empathy Questionnaire | No test-retest reliabilty findable |
| Single Item Trait Empathy Scale (SITES) | r = .57 (Konrath et al., 2018) |
| Reading the mind in the eyes Test | ICC = .833 (Vellante et al., 2012) |
| Faces Test | ICC = .85 (Charernboon, 2017) |

4.1. List specific, concise, and testable hypotheses. Please state if the hypotheses are directional or non-directional. If directional, state the direction. A predicted effect is also appropriate here. If a specific interaction or moderation is important to your research, you can list that as a separate hypothesis.

4.2. **Example**: If taste affects preference, then mean preference indices will be higher with higher concentrations of sugar.

### **Design Plan**

In this section, you will be asked to describe the overall design of your study. Remember that this research plan is designed to register a single study, so if you have multiple experimental designs, please complete a separate preregistration.

1. Study type (required): Observational Study – Online-Survey

5.1. Experiment - A researcher randomly assigns treatments to study subjects, this includes field or lab experiments. This is also known as an intervention experiment and includes randomized controlled trials.

5.2. Observational Study - Data is collected from study subjects that are not randomly assigned to a treatment. This includes surveys, ñnatural experiments,î and regression discontinuity designs.

5.3. Meta-Analysis - A systematic review of published studies.

5.4. Other

1. Blinding (required): No blinding is involved in this study.

6.1. Blinding describes who is aware of the experimental manipulations within a study. Mark all that apply.

6.1.1. No blinding is involved in this study.

6.1.2. For studies that involve human subjects, they will not know the treatment group to which they have been assigned.

6.1.3. Personnel who interact directly with the study subjects (either human or non-human subjects) will not be aware of the assigned treatments. (Commonly known as “double blind”)

6.1.4. Personnel who analyze the data collected from the study are not aware of the treatment applied to any given group.

1. Is there any additional blinding in this study? No additional blinding is used.

1. Study design (required): There will be one single pool of participants, who will complete the self-report/behavioral measures two times.

8.1. Describe your study design. Examples include two-group, factorial, randomized block, and repeated measures. Is it a between (unpaired), within-subject (paired), or mixed design? Describe any counterbalancing required. Typical study designs for observation studies include cohort, cross sectional, and case-control studies.

8.2. Example: We have a between subjects design with 1 factor (sugar by mass) with 4 levels.

8.3. More info: This question has a variety of possible answers. The key is for a researcher to be as detailed as is necessary given the specifics of their design. Be careful to determine if every parameter has been specified in the description of the study design. There may be some overlap between this question and the following questions. That is OK, as long as sufficient detail is given in one of the areas to provide all of the requested information. For example, if the study design describes a complete factorial, 2 X 3 design and the treatments and levels are specified previously, you do not have to repeat that information.

1. Randomization (optional) No randomization will be used in this study.

9.1. If you are doing a randomized study, how will you randomize, and at what level?

9.2. Example: We will use block randomization, where each participant will be randomly assigned to one of the four equally sized, predetermined blocks. The random number list used to create these four blocks will be created using the web applications available at http://random.org.

9.3. More info: Typical randomization techniques include: simple, block, stratified, and adaptive covariate randomization. If randomization is required for the study, the method should be specified here, not simply the source of random numbers.

### **Sampling Plan**

In this section we’ll ask you to describe how you plan to collect samples, as well as the number of samples you plan to collect and your rationale for this decision. Please keep in mind that the data described in this section should be the actual data used for analysis, so if you are using a subset of a larger dataset, please describe the subset that will actually be used in your study.

1. Registration prior to creation of data.

1.1. Preregistration is designed to make clear the distinction between confirmatory tests, specified prior to seeing the data, and exploratory analyses conducted after observing the data. Therefore, creating a research plan in which existing data will be used presents unique challenges. Please select the description that best describes your situation. Please do not hesitate to contact us if you have questions about how to answer this question (prereg@cos.io).

1.1.1. Registration prior to creation of data: As of the date of submission of this research plan for preregistration, the data have not yet been collected, created, or realized.

1.1.2. Registration prior to any human observation of the data: As of the date of submission, the data exist but have not yet been quantified, constructed, observed, or reported by anyone - including individuals that are not associated with the proposed study. Examples include museum specimens that have not been measured and data that have been collected by non-human collectors and are inaccessible.

1.1.3. Registration prior to accessing the data: As of the date of submission, the data exist, but have not been accessed by you or your collaborators. Commonly, this includes data that has been collected by another researcher or institution.

1.1.4. Registration prior to analysis of the data: As of the date of submission, the data exist and you have accessed it, though no analysis has been conducted related to the research plan (including calculation of summary statistics). A common situation for this scenario when a large dataset exists that is used for many different studies over time, or when a data set is randomly split into a sample for exploratory analyses, and the other section of data is reserved for later confirmatory data analysis.

1.1.5. Registration following analysis of the data: As of the date of submission, you have accessed and analyzed some of the data relevant to the research plan. This includes preliminary analysis of variables, calculation of descriptive statistics, and observation of data distributions. Please see cos.io/prereg for more information.

1. Explanation of existing data (optional) No existing data will be used.

2.1. If you indicate that you will be using some data that already exist in this study, please describe the steps you have taken to assure that you are unaware of any patterns or summary statistics in the data. This may include an explanation of how access to the data has been limited, who has observed the data, or how you have avoided observing any analysis of the specific data you will use in your study.

2.2. **Example**: An appropriate instance of using existing data would be collecting a sample size much larger than is required for the study, using a small portion of it to conduct exploratory analysis, and then registering one particular analysis that showed promising results. After registration, conduct the specified analysis on that part of the dataset that had not been investigated by the researcher up to that point.

2.3. **More info**: An appropriate instance of using existing data would be collecting a sample size much larger than is required for the study, using a small portion of it to conduct exploratory analysis, and then registering one particular analysis that showed promising results. After registration, conduct the specified analysis on that part of the dataset that had not been investigated by the researcher up to that point.

1. Data collection procedures (required): Participants will be recruited through the survey panel “Prolific”. The survey will take place on the survey platform “Qualtrics”. They will be paid 10$ (?) after participating in both assessments of the study. They will be informed about the requirement to participate in both time points prior to participation. Participants must be at least 18 years old. They will be excluded from the analysis if they fail either of two attention checks. These consist of items telling the participant exactly what to do (e.g. “Please choose the option 'Agree Strongly'. This is an attention check.”

3.1. Please describe the process by which you will collect your data. If you are using human subjects, this should include the population from which you obtain subjects, recruitment efforts, payment for participation, how subjects will be selected for eligibility from the initial pool (e.g. inclusion and exclusion rules), and your study timeline. For studies that donÍt include human subjects, include information about how you will collect samples, duration of data gathering efforts, source or location of samples, or batch numbers you will use.

3.2. **Example**: Participants will be recruited through advertisements at local pastry shops. Participants will be paid $10 for agreeing to participate (raised to $30 if our sample size is not reached within 15 days of beginning recruitment). Participants must be at least 18 years old and be able to eat the ingredients of the pastries.

3.3. **More information**: The answer to this question requires a specific set of instructions so that another person could repeat the data collection procedures and recreate the study population. Alternatively, if the study population would be unable to be reproduced because it relies on a specific set of circumstances unlikely to be recreated (e.g., a community of people from a specific time and location), the criteria and methods for creating the group and the rationale for this unique set of subjects should be clear.

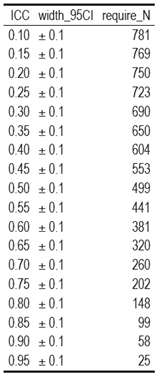
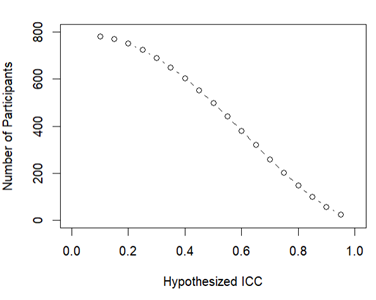
1. Sample size (required): The target sample size ranges from 25 to 381 participants for our desired level of precision, across different ICC2 values. The actual sample size that will be used in the study will depend on the selected measures.

4.1. Describe the sample size of your study. How many units will be analyzed in the study? This could be the number of people, birds, classrooms, plots, interactions, or countries included. If the units are not individuals, then describe the size requirements for each unit. If you are using a clustered or multilevel design, how many units are you collecting at each level of the analysis?

4.2. **Example**: Our target sample size is 280 participants. We will attempt to recruit up to 320, assuming that not all will complete the total task.

4.3. **More information**: For some studies, this will simply be the number of samples or the number of clusters. For others, this could be an expected range, minimum, or maximum number.

1. Sample size rationale (optional): The function “calculateIccSampleSize”of the R-package “ICC.Sample.Size” was used to perform a power analysis (with an estimated power of .8 and an alpha error probability of .95). The table and the power curve below show the different numbers of participants that are needed for different estimated ICCs (with a 95% confidence interval of ± .1). However, the targeted sample size does not only depend on the power analysis, but also on resource constraints (Lakens, 2022) and the call for large sample sizes in retest studies by Watson (2004). As a tradeoff between these factors, the used measures are limited to N per participant and the target sample size are N participants.



5.1. This could include a power analysis or an arbitrary constraint such as time, money, or personnel.

5.2. **Example**: We used the software program G\*Power to conduct a power analysis. Our goal was to obtain .95 power to detect a medium effect size of .25 at the standard .05 alpha error probability.

5.3. **More information**: This gives you an opportunity to specifically state how the sample size will be determined. A wide range of possible answers is acceptable; remember that transparency is more important than principled justifications. If you state any reason for a sample size upfront, it is better than stating no reason and leaving the reader to “fill in the blanks.” Acceptable rationales include: a power analysis, an arbitrary number of subjects, or a number based on time or monetary constraints.

1. Stopping rule (optional): Data collection will cease when we have collected the maximum possible number of participants within our budget of €2000..

6.1. If your data collection procedures do not give you full control over your exact sample size, specify how you will decide when to terminate your data collection.

6.2. **Example**: We will post participant sign-up slots by week on the preceding Friday night, with 20 spots posted per week. We will post 20 new slots each week if, on that Friday night, we are below 320 participants.

6.3. **More information**: You may specify a stopping rule based on p-values only in the specific case of sequential analyses with pre-specified checkpoints, alphas levels, and stopping rules. Unacceptable rationales include stopping based on p-values if checkpoints and stopping rules are not specified. If you have control over your sample size, then including a stopping rule is not necessary, though it must be clear in this question or a previous question how an exact sample size is attained.

### **Variables**

In this section you can describe all variables (both manipulated and measured variables) that will later be used in your confirmatory analysis plan. In your analysis plan, you will have the opportunity to describe how each variable will be used. If you have variables which you are measuring for exploratory analyses, you are not required to list them, though you are permitted to do so.

1. Manipulated variables (optional): No variables will be manipulated in this study.

7.1. Describe all variables you plan to manipulate and the levels or treatment arms of each variable. This is not applicable to any observational study.

7.2. **Example:** We manipulated the percentage of sugar by mass added to brownies. The four levels of this categorical variable are: 15%, 20%, 25%, or 40% cane sugar by mass.

7.3. **More information**: For any experimental manipulation, you should give a precise definition of each manipulated variable. This must include a precise description of the levels at which each variable will be set, or a specific definition for each categorical treatment. For example, “loud or quiet,” should instead give either a precise decibel level or a means of recreating each level. 'Presence/absence' or 'positive/negative' is an acceptable description if the variable is precisely described.

1. Measured variables (required): The items of the following empathy self-report scales / behavioral tasks will be measured:

* Single Item Trait Empathy Scale (SITES) (Konrath et al., 2018)
* Interpersonal Reactivity Index (IRI) (Davis, 1983)
* Empathy Quotient (EQ) (Simon Baron-Cohen & Wheelwright, 2004)
* Basic Empathy Scale (BES-A) (Carré et al., 2013)
* Toronto Empathy Questionnaire (TEQ) (Spreng et al., 2009)
* Questionnaire of Cognitive and Affective Empathy (QCAE) (Reniers et al., 2011)
* Emotion Specific Empathy Questionnaire (ESE) (Olderbak et al., 2014)
* Pictorial Empathy Test (PET) (Lindeman et al., 2018)
* Reading the Mind in the Eyes Task (EYES) (Baron-Cohen et al., 2001)
* Faces Task (FACES) (Baron-Cohen et al., 1997)

In the pilot study, all of the measures will be applied. In the following test-retest study, only a selection of the measures will be used and further analyzed.

8.1. Describe each variable that you will measure. This will include outcome measures, as well as any predictors or covariates that you will measure. You do not need to include any variables that you plan on collecting if they are not going to be included in the confirmatory analyses of this study.

8.2. **Example**: The single outcome variable will be the perceived tastiness of the single brownie each participant will eat. We will measure this by asking participants ‘How much did you enjoy eating the brownie’ (on a scale of 1-7, 1 being ‘not at all’, 7 being ‘a great deal’) and ‘How good did the brownie taste’ (on a scale of 1-7, 1 being ‘very bad’, 7 being ‘very good’).

8.3. **More information**: Observational studies and meta-analyses will include only measured variables. As with the previous questions, the answers here must be precise. For example, 'intelligence,' 'accuracy,' 'aggression,' and 'color' are too vague. Acceptable alternatives could be 'IQ as measured by Wechsler Adult Intelligence Scale' 'percent correct,' 'number of threat displays,' and 'percent reflectance at 400 nm.'

1. Indices (optional): Item level responses will be converted to mean scores.

9.1. If any measurements are going to be combined into an index (or even a mean), what measures will you use and how will they be combined? Include either a formula or a precise description of your method. If your are using a more complicated statistical method to combine measures (e.g. a factor analysis), you can note that here but describe the exact method in the analysis plan section.

**9.2.** **Example**: We will take the mean of the two questions above to create a single measure of ‘brownie enjoyment.’

9.3. **More information**: If you are using multiple pieces of data to construct a single variable, how will this occur? Both the data that are included and the formula or weights for each measure must be specified. Standard summary statistics, such as “means” do not require a formula, though more complicated indices require either the exact formula or, if it is an established index in the field, the index must be unambiguously defined. For example, “biodiversity index” is too broad, whereas “Shannon’s biodiversity index” is appropriate.

### **Analysis Plan**

You may describe one or more confirmatory analysis in this preregistration. Please remember that all analyses specified below must be reported in the final article, and any additional analyses must be noted as exploratory or hypothesis generating.

A confirmatory analysis plan must state up front which variables are predictors (independent) and which are the outcomes (dependent), otherwise it is an exploratory analysis. You are allowed to describe any exploratory work here, but a clear confirmatory analysis is required.

1. Statistical models (required):

This will be an observational study.

The test-retest correlation will be determined by calculating the ICC and its 95% CIs for the two assessments of the empathy measures. The version of the ICC will be “Two-way random effects, absolute agreement, single rater/measurement” (ICC2, using the McGraw and Wong (1996) convention), or the “ICC (2,1)” (using the Shrout and Fleiss (1979) convention). The selection of the version was based on the guideline by Koo and Li (2016).

Internal consistency will be estimated for each scale at each timepoint by calculating Cronbach’s alpha and its 95% CIs. The R packages psych and lavaan will be used.

10.1. What statistical model will you use to test each hypothesis? Please include the type of model (e.g. ANOVA, multiple regression, SEM, etc) and the specification of the model (this includes each variable that will be included as predictors, outcomes, or covariates). Please specify any interactions, subgroup analyses, pairwise or complex contrasts, or follow-up tests from omnibus tests. If you plan on using any positive controls, negative controls, or manipulation checks you may mention that here. Remember that any test not included here must be noted as an exploratory test in your final article.

**10.2.** **Example**: We will use a one-way between subjects ANOVA to analyze our results. The manipulated, categorical independent variable is 'sugar' whereas the dependent variable is our taste index.

10.3. **More information**: This is perhaps the most important and most complicated question within the preregistration. As with all of the other questions, the key is to provide a specific recipe for analyzing the collected data. Ask yourself: is enough detail provided to run the same analysis again with the information provided by the user? Be aware for instances where the statistical models appear specific, but actually leave openings for the precise test. See the following examples:

10.3.1.1. If someone specifies a 2x3 ANOVA with both factors within subjects, there is still flexibility with the various types of ANOVAs that could be run. Either a repeated measures ANOVA (RMANOVA) or a multivariate ANOVA (MANOVA) could be used for that design, which are two different tests.

10.3.1.2. If you are going to perform a sequential analysis and check after 50, 100, and 150 samples, you must also specify the p-values you’ll test against at those three points.

1. Transformations (optional): Reverse scored items will be recoded to match the direction of the other items. Apart from that, no transformations of the data are required.

11.1. If you plan on transforming, centering, recoding the data, or will require a coding scheme for categorical variables, please describe that process.

11.2. **Example**: The “Effect of sugar on brownie tastiness” does not require any additional transformations. However, if it were using a regression analysis and each level of sweet had been categorically described (e.g. not sweet, somewhat sweet, sweet, and very sweet), ‘sweet’ could be dummy coded with ‘not sweet’ as the reference category.

11.3. **More information**: If any categorical predictors are included in a regression, indicate how those variables will be coded (e.g. dummy coding, summation coding, etc.) and what the reference category will be.

1. Inference criteria (optional): The found estimates (or their confidence intervals, respectively) of retest reliability and internal consistency for each measure will be compared to published estimates. Also, common guidelines will be used to describe the estimates: An Interclass Correlation Coefficient of > .9 indicating excellent, of > .75 good, of > .5 moderate and of < .5 poor retest reliability (Koo & Li, 2016). A Cronbach's Alpha of > .9 indicating excellent, of > .8 good, of > .7 acceptable, of > .6 questionable, of > .5: poor, and of < .5 unacceptable internal consistency (George and Mallery, 2003) The measures will be rank ordered by their retest reliability estimates. Recommendations will be made, regarding which of them are the most/least reliable to use.

12.1. What criteria will you use to make inferences? Please describe the information youÍll use (e.g. p-values, bayes factors, specific model fit indices), as well as cut-off criterion, where appropriate. Will you be using one or two tailed tests for each of your analyses? If you are comparing multiple conditions or testing multiple hypotheses, will you account for this?

12.2. **Example**: We will use the standard p<.05 criteria for determining if the ANOVA and the post hoc test suggest that the results are significantly different from those expected if the null hypothesis were correct. The post-hoc Tukey-Kramer test adjusts for multiple comparisons.

12.3. **More information:** P-values, confidence intervals, and effect sizes are standard means for making an inference, and any level is acceptable, though some criteria must be specified in this or previous fields. Bayesian analyses should specify a Bayes factor or a credible interval. If you are selecting models, then how will you determine the relative quality of each? In regards to multiple comparisons, this is a question with few “wrong” answers. In other words, transparency is more important than any specific method of controlling the false discovery rate or false error rate. One may state an intention to report all tests conducted or one may conduct a specific correction procedure; either strategy is acceptable.

1. Data exclusion (optional): Two attention checks will be used in the survey. They consist of items telling the participant exactly what to do (e.g. “Please choose the option 'Agree Strongly'. This is an attention check.” If participants fail these attention checks, their data will be excluded from the analysis. No exclusions of outliers will be performed.

13.1. How will you determine what data or samples, if any, to exclude from your analyses? How will outliers be handled? Will you use any awareness check?

13.2. **Example**: No checks will be performed to determine eligibility for inclusion besides verification that each subject answered each of the three tastiness indices. Outliers will be included in the analysis.

13.3. **More information**: Any rule for excluding a particular set of data is acceptable. One may describe rules for excluding a participant or for identifying outlier data.

1. Missing data (optional): All answers of the survey are mandatory. The participants will not be able to progress and finish the survey without answering every question. If, in a hypothetical case of technical error, a data set should be incomplete, it will not be included in the analysis.

14.1. How will you deal with incomplete or missing data?

14.2. **Example**: If a subject does not complete any of the three indices of tastiness, that subject will not be included in the analysis.

14.3. **More information**: Any relevant explanation is acceptable. As a final reminder, remember that the final analysis must follow the specified plan, and deviations must be either strongly justified or included as a separate, exploratory analysis.

1. Exploratory analysis (optional): Does not apply.

15.1. If you plan to explore your data set to look for unexpected differences or relationships, you may describe those tests here. An exploratory test is any test where a prediction is not made up front, or there are multiple possible tests that you are going to use. A statistically significant finding in an exploratory test is a great way to form a new confirmatory hypothesis, which could be registered at a later time.

15.2. **Example**: We expect that certain demographic traits may be related to taste preferences. Therefore, we will look for relationships between demographic variables (age, gender, income, and marital status) and the primary outcome measures of taste preferences.

### **Other**

1. Other (Optional):

References

1. 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/

2. 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

3. 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

4. Carré, A., Stefaniak, N., D'Ambrosio, F., Bensalah, L., & Besche-Richard, C. (2013). The Basic Empathy Scale in adults (BES-A): Factor structure of a revised form. *Psychological Assessment*, *25*(3), 679–691. https://doi.org/10.1037/a0032297

5. Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology*, *44*(1), 113–126. https://doi.org/10.1037/0022-3514.44.1.113

6. 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

7. 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

8. 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

9. 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

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

11. 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

12. 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

13. 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

14. 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

15. 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

16.1. If there is any additional information that you feel needs to be included in your preregistration, please enter it here. Literature cited, disclosures of any related work such as replications or work that uses the same data, or other context that will be helpful for future readers would be appropriate here.