The Complexity of Capturing Empathy: Insights from an Exploratory Scale

Linette Anyiacha

Department of Psychology, University of Alberta

PSYCH 431: Theory and Practice of Psychometrics

Dr. Chris Westbury

December 9, 2024

Abstract

Empathy is a deep need within the soul of man. It is evident in our everyday interactions, yet it carries so many definitions with it. Considering the many constructs embedded in empathy, the current study explores the development and preliminary evaluation of an empathy scale designed to measure empathy in interpersonal contexts. The scale showed moderate validity (Pearson's r = 0.660) and reliability (Cronbach's alpha = 0.641), with room for improvement in internal consistency and item coherence. Classification accuracy was moderate, with a true positive rate of 58.6%, though a notable rate of false negatives was observed. These preliminary findings limit the scale's generalizability, but loosely support its utility as a general empathy assessment tool, necessitating further item refinement to enhance reliability, sensitivity and subtler consideration of the multidimensionality of empathy as a construct.

To feel or not to feel—is there really a question? Years of evolution of the human species has highlighted how we need competition to survive and to propagate our own, but there has long been something crucial to our survival: empathy. According to Riess (2017), if human existence was simple the result of survival of the fittest, we would be wired solely to dominate others, not respond to their suffering. But before diving further, what is empathy? Riess (2017) describes empathy as an emotional bridge that promotes pro-social behaviour, further distinguishing between cognitive and emotional empathy. Heyes (2018) distinguishes between two empathies: empathy 1 and empathy 2. Empathy 1 operates automatically and develops early in humans but can be found in other animals. It is often input for empathy 2 and provides quicker responses to emotional stimuli. Empathy 2 involves more controlled processing such as mediating goal-directed action and metacognitive and cognitive appraisal and is uniquely human. This distinction highlights empathy as both an instinctive reaction and a reflective process. Together, they form a dynamic system that enables humans to navigate complex social and emotional landscapes, showing that empathy is not just a feeling but a vital mechanism for communication and connection.

Smith (2015) describes empathy as "transparent fellow feeling," where the individual places a higher value on knowing that the other feels the same, and knowing that the other knows that one feels the same, rather than on the feeling itself. In that regard, empathy enables us to value feeling *with* others. To me, empathy entails actively sharing in another's emotional experience, experiencing it with them. To do so, one must first value the other to be able to identify with them. Batson (2007) states that this "valuing" of the other, of the person in need, is an important precursor of feeling empathy for that person. They saw that increased valuation of the other person led to increased perspective taking and empathic concern, which ultimately

increased the likelihood of them helping the person. If someone is in distress, wouldn't feeling distressed with them be harmful to the onlooker? Not exactly. We do "feel the pain" of others, but in an attenuated form (Riess, 2017). This attenuation makes it possible to empathize and not become overwhelmed with someone else's personal distress.

Common to the previous definitions are the embeddedness of constructs such as perspective taking, theory of mind, emotional regulation, emotional awareness, compassion, moral cognition and judgment, all of which are crucial for surviving and thriving in society. Emotional regulation ensures our responses to stimuli remain appropriate and constructive. Emotional awareness allows us to identify and process both our feelings and that of others, paving way for compassion. Compassion, in turn, motivates prosocial behavior, and moral cognition and judgment guide us in appraising and distinguishing right from wrong in social contexts. These constructs build upon one another and ultimately conceive empathy. Though they build on each other, the constructs themselves are implicit and as such need to be highlighted implicitly. How can one ask overtly about something that we often just do? How does one verbalize something we often just intuit? Something so ubiquitous but hard to capture? Those are exactly our efforts and rationalize for attempting to quantify empathy thus.

Despite its ubiquity, empathy is apparently not coded in our genes (Heyes, 2018). Nativist theorists have located empathy mechanisms in the genes and in parts of the brain, but they do not explain them (de Waal et al., 2017). This lack of explanation forced researchers to appreciate what differences make a difference using sound scientific methods. Namely, learning methods. This is an advantage because it suggests that empathy can be taught and learned. In humans, this learning is done and consolidated primarily through mirroring (Heyes, 2018). With repeated exposure and with various experiences that reinforce and build upon its embedded constructs,

empathy can be taught and can play a role in drawing people toward various helping professions. In the world today, there is as much necessity for compassion and empathy as there is a need for people to provide it. With such a great need, it is crucial to have a way to measure it, allowing us to enhance individual levels of it. Focusing on the individual level is less daunting than attempting mass change all at once, which is where the necessity for an empathy scale arises. Consider the impact of empathy in personal relationships, the deep suffering that transcends words, or the bond between doctor and patient, and it is clear how empathy can deeply affect us all.

METHODS

Study Design

Our study was survey/questionnaire based, of which the participants were not randomly sampled, as our questionnaire was posted and sent, and participants could choose whether to complete it.

Participants

We posted the link to the questionnaire on various University of Alberta student forums on Discord and Reddit. We also sent the link to various friends and family members. As such, we recognize that our sample may not be fully representative of the general population.

We did not collect any demographic information such as age, gender or race. The questionnaire was open for three weeks, allowing ample time for responses and for an adequate sample size. The final sample size of analysis was 283 participants.

Materials

Items 2, 6, and 14 in our final question set come from the Narcissistic Personality Inventory (NPI) and the Empathy Quotient (EQ) scales (Baron-Cohen & Wheelwright, 2004; Raskin & Terry, 1988).

The final item set consists of 17 items. See the appendix for the question set.

Procedure

In making our questions, we considered and consulted various existing empathy scales and made questions that could be answered quantitatively to allow easy scoring and analysis. Our questions were weighted on a 9-point scale as follows: 1 (‘Strongly Disagree’), 2 ('Disagree'), 3 ('Moderately Disagree'), 4 ('Slightly Disagree'), 5 ('Undecided'), 6 ('Slightly Agree'), 7 ('Moderately Agree'), 8 ('Agree'), and 9 (‘Strongly Agree’).

We first considered various definitions of empathy and trying to isolate the various constructs within it. From there, we rationalized these constructs as overt actions and behaviours, considered how other researchers have studied it, and built created our statements was from there.

The questionnaire was built on Google Forms. The first portion detailed information about the project, ethics considerations, anticipated time commitment, potential risks and benefits, and the principal investigator’s contact information should they have had any questions. The first question directly asked for participants’ consent to participate. The instructions also explained the numerical value of all answer options. After the instructions, there were 52 questions for participants to complete.

The items were simple statements such as, “I appreciate when my friends open up to me about their struggles,” and the participants had to select the option that was most applicable to their situation.

We posted the questionnaire link several times on various forums and sent the link to friends and family, ensuring potential participants of utmost confidentiality and their anonymity. The form remained open for 3 weeks allowing ample time for participants to respond.

Data Analysis

Each question on the questionnaire was scored from 1 to 9 and the data was analyzed using both Microsoft Excel and JASP at different stages of analysis. Some questions were reversed coded and therefore were reversed again in the first part of the analysis to ensure that all the scores were going in the same direction. This was done using Excel. We also calculated sum scores, validity scores, and used correlation criteria to determine which questions didn’t provide much information. We used JASP to calculate and maximize Cronbach’s alpha, to determine standard error of measurement, and for factor analysis. Finally, we cut the test scores on Excel and determined the cut off based on the top 15% of our sample, as well as determined the empathy scale’s true positive rate.

Results

The purpose of this study was to develop and validate a scale that quantitatively investigates people’s empathy levels. Ultimately, our findings were meager and necessitate more vigorous validation procedures. Our analysis was mostly based on sum scores and validity scores. When the sum scores were first correlated with the validity scale, it yielded a very low correlation and necessitated further measures to improve. At that point, we considered that not all

the questions were correlated with the validity scale, and therefore needed to be removed. From there, the individual question scores were correlated with the validity scale, ordered from highest to lowest (i.e., best to worst). Those individual correlations were correlated with sum score correlations The first round of question selection was conducted in Excel using the correlation of sum score of questions ordered from best to worst with the validity scale (note that the validity scale was made up of seven questions that directly asked about empathy).

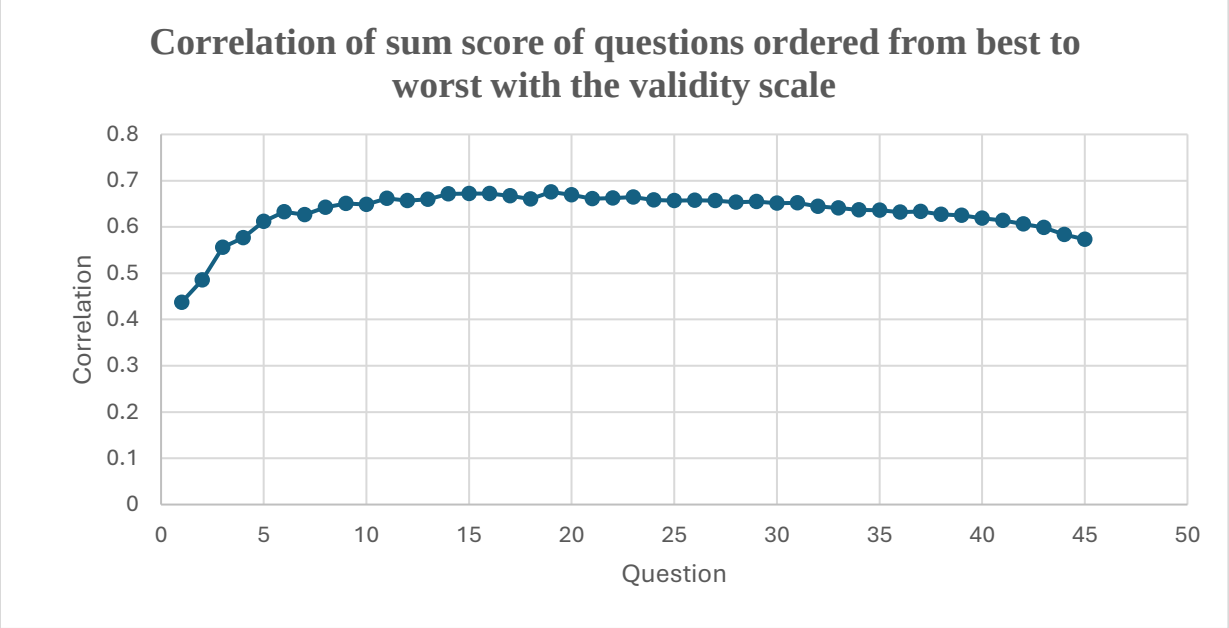

Figure 1: a graph depicting the ordering of questions, where the first highest peak was determined to be the cutoff, i.e., the point at which the questions no longer provided information.

Based on figure 1, we cut the scale off at 19 questions because from that point, the correlation was simply going back to a previously attained value of 0.67 and would only decrease or stagnate from there. Further inspection of those questions past the cutoff point showed that the items were either convoluted in their wording, or were possibly too general in their application, and therefore could be justifiably left out. Due to the limitations in our context, our validity scale consisted of questions that directly asked about people's empathy levels. These

validity scores were correlated with the sum scores, yielding a Pearson's value of $r = 0.660$, $p < 0.001$.

In the second phase of analysis, as we looked at the Cronbach's Alpha, we dropped one more question from the set, maximizing Cronbach's Alpha at $\alpha = 0.641$. Looking at the item itself, we saw that the question was not worded well, and again justified leaving it out of the final item list. Upon looking at other questions that were suggested to be dropped, the projected alpha values were not much different (i.e., were not differences that made a difference) and the questions directly hit at our construct, empathy, so we left the final question set at 17 items. A Cronbach's Alpha value of 0.641 is in the range of questionable internal consistency. While this value is below the accepted 0.7 threshold, empathy is a multidimensional construct, that though we attempted to capture its essence with our questions, did not. The low number of items on the item list could have also contributed to this alpha value. Ultimately, it suggests that what little items there are on the scale are not very strongly interrelated and a more robust scale is required to pinpoint empathy.

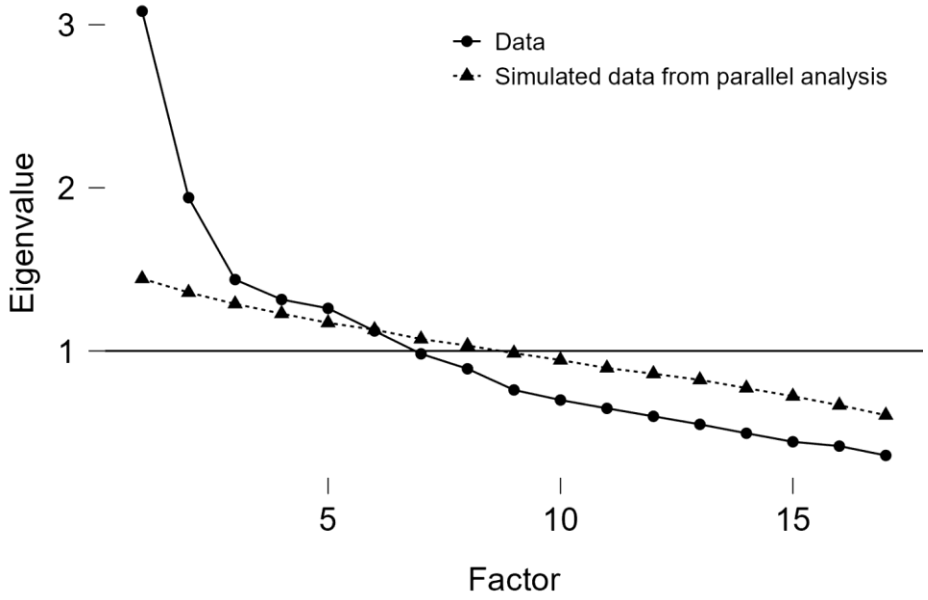

Figure 2: A scree plot after factor analysis on 17 items. Based on the plot, there are 3 factors across the scale.

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Uniqueness
q002	0.723						0.454
rq043	0.510						0.704
q030	0.424						0.670
q028		0.706					0.405
q027		0.636					0.473
q011			0.986				0.005
q018			0.443				0.727
q017				0.843			0.281
q016				0.561			0.613
q025					0.648		0.544
q029					0.580		0.618
q044					0.522		0.707
q034						0.729	0.397
q035						0.644	0.565
q039							0.808
q007							0.865
q013							0.937

Note. Applied rotation method is varimax.

A

Factor Characteristics

	Eigenvalues
Factor 1	3.083
Factor 2	1.940
Factor 3	1.438
Factor 4	1.315
Factor 5	1.261
Factor 6	1.122

B

Table 1A and B: Table A depicts the statistical factor loadings of across 17 empathy scale items, and table B shows the factor characteristics in terms of eigenvalues.

Exploratory factor analysis was conducted on the remaining 17 items and yielded three factors, shown in figure 2. Further inspection of the factor loadings table grouped the questions according to arbitrary factors, seen in table 1A. Of course, there weren't that many factors. Looking at the eigenvalues in table 1B, there appear to be 3 factors, as the eigenvalue from the third is not much different from the fourth. Based on these groupings, we manually considered the questions according to their semantic coherence with each other and proposed three factors as follows: valuing others, helping others, and feeling emotion. The 17 items were not evenly divided across the three factors. Feeling emotion had 5 items (Items 6, 10, 12, 13, 15) valuing others had 3 items (Items 9, 11, 14), and helping others had the remaining 9 items (Items 1, 2, 3, 4, 5, 7, 8, 16, 17). We also tried to do confirmatory factor analysis, but it did not work because all the items are supposedly too correlated with each other. Considering that factors are explanatory

evidence for the purposes of investigating a concept of interest, we grouped our ‘factors’ semantically, rather than statistically.

The standard error of measurement (SEM) for our empathy scale is 8.086. This is a rather large value and is completely warranted, considering our low reliability value and relatively short scale of only 17 items. This means that an observed score may deviate from a participant’s true score by approximately ±8.086 points, on average.

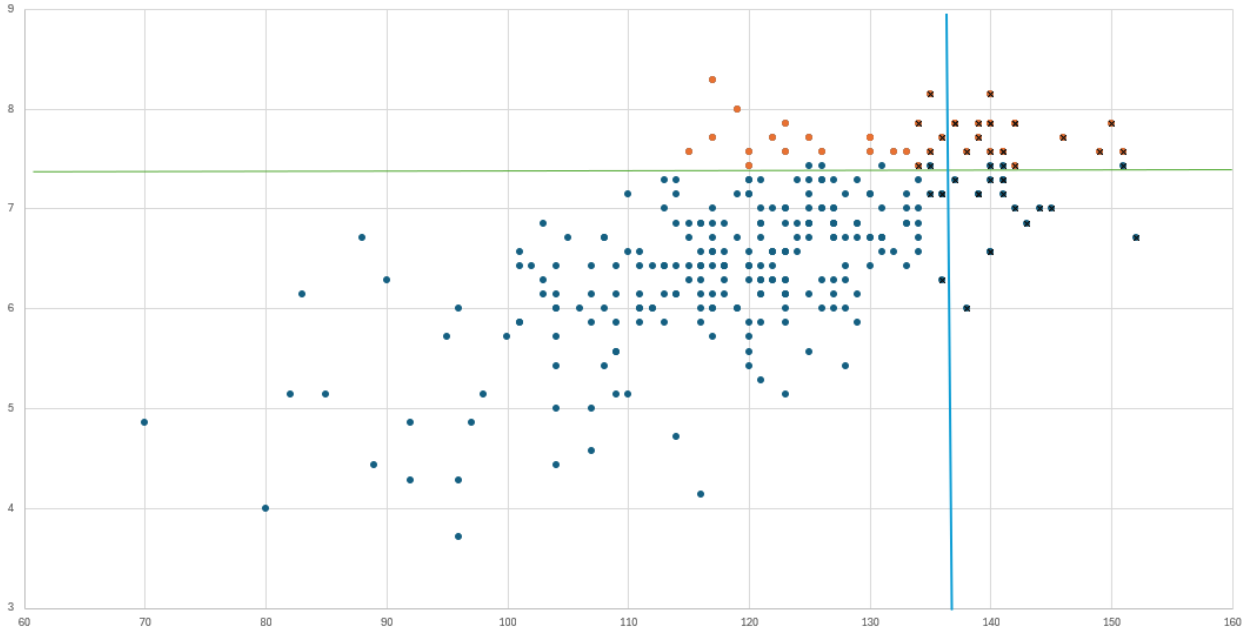

Figure 3: A graph depicting validity scores as a function of sum score to determine scoring cutoff, with sum score cutoffs determined based on the top 15% of the sample. The upper right quadrant represents the ‘diagnosed empaths’ in the sample.

The empathy scores ranged from 70 to 152 points, with a cutoff sum score of 137, based on the top 15% of the sample, plus 4 participants to account for score ties. The lowest scores potentially reflect a deficiency in empathy, and perhaps identifies a need for additional support to increase empathy, while those on the higher end reflect proficiency in empathy and their

likelihood to help others. The cutoff for the validity scores was determined to be 7.428. Scores with a validity above this range were categorized as ‘true empaths.’ Those who scored above the 137 range on sum scores and above 7.428 are categorized as ‘diagnosed empaths,’ pictured in figure 3. Ultimately, the true positive rate is 58.6%, indicating that the scale has potential to be a diagnostic tool, with lots of room for improvement through further calibration and item refinement.

Discussion

Given the exploratory nature of the proposed empathy scale, it is unsurprising that it yielded such moderate or low validity and reliability measures. First, we tried to capture empathy by considering all the possible constructs that could be embedded therein. However, our explicit attempts were potentially counterintuitive for concepts that are inherently, and only implicitly, born in the grander construct empathy. Rather than viewing the three semantic factors we categorized after exploratory factor analysis disparately, we can see them as building upon each other. The unequal distribution of questions across these supposed factors was not ideal, and therefore would indicate that helping others weighed the most heavily amongst the three, loading on nine questions. The starting ‘factor’ must necessarily be feeling emotion. As one feels the emotion, they can identify it in themselves and in others. From there, they can respond to those in need. The latter two factors described could then potentially follow the sequence outlined by Batson (2007), where increased valuation of another increases the likelihood of helping that ‘other.” Many of the questions that did not make it to the final question set were heavily focused on emotion, and perhaps that was why they did not end up giving us much information. And while there is an emotional component to empathy, it could be so implicit that asking about it directly only acts as a distraction. All these factors, and potentially more, may be so entwined

that our attempts to neatly pull them apart were futile, and that may have posed as the biggest problem in our scale.

A Pearson's r of 0.660 demonstrates moderate correlation between the sum scores and the validity scores, suggesting somewhat reasonable construct validity, although since validation was done on questions that directly asked about empathy, it's hard to pull apart whether participant responses were honest. The internal consistency of the scale, Cronbach's α, was below the acceptable 0.70 threshold, potentially indicating that our scale items are not strongly coherent as a unidimensional measure of empathy. Trying to account for all possible concepts in empathy may have set us in the wrong direction, consequently impacting reliability. The Standard Error of Measurement (SEM) of 8.086 indicates that observed sum scores may deviate from participants' true empathy levels by approximately ±8.086 points. This lessens our scale's precision and limits its utility as a tool for individual assessment and would require further item refinement and item addition to the scale. The standard deviation of the summed scores (SD = 13.190) suggests a reasonable spread in empathy levels among participants. Taken together with a true positive rate of 58.6%, we see that there's still a moderate rate of false negatives at 41.4%. This trade-off suggests that the current scale may be more effective as a screening tool rather than a definitive diagnostic measure.

The exploratory nature of this empathy scale and its unsurprisingly low validity and reliability score measures thereby limit it in its generalizability and utility as a tool for measuring levels of empathy and even as a tool for diagnosing the need for more empathy. But seeing as empathy is necessary in our daily lives, this gives us hope that even those scoring on the lower end of the scale can be taught to increase their empathy levels, especially considering those who scored on the lower end of the scale. Empathy is a worldwide, necessary currency. It is widely

spoken, yet often without a precise vocabulary to describe it. It transcends words, manifesting as an innate compulsion to care, to connect, and to respond. Empathy is a skill we observe, learn, and practice until it becomes intuitive, like the fluency of a native speaker. Still, it is a language spoken quietly but powerfully through actions rather than words. While our scale faced challenges in quantifying and operationalizing such a complex construct, the preliminary results provide a starting point for future consideration.

Though demographic data were not collected in this study, as the primary aim was to explore overall empathy levels, future work could enhance the scale's internal consistency and sensitivity by expanding the item pool to better reflect the multidimensionality of empathy, revising items, and collecting a more diverse sample with defined demographic criteria. A closer look at how demographic variables influence empathy could provide richer insights and help tailor the scale for specific populations.

References

Baron-Cohen, S., & Wheelwright, S. (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*, 163-175.

Batson, C. D., Eklund, J. H., Chermok, V. L., Hoyt, J. L., & Ortiz, B. G. (2007). An additional antecedent of empathic concern: Valuing the welfare of the person in need. *Journal of Personality and Social Psychology*, *93*(1), 65–74. https://doi.org/10.1037/0022-3514.93.1.65

de Waal, F. B., & Preston, S. D. (2017). Mammalian empathy: Behavioural manifestations and neural basis. *Nature Reviews Neuroscience*, *18*(8), 498–509. https://doi.org/10.1038/nrn.2017.72

Heyes, C. (2018). Empathy is not in our genes. *Neuroscience & Biobehavioral Reviews*, *95*, 499–507. https://doi.org/10.1016/j.neubiorev.2018.11.001

Raskin, R., & Terry, H. (1988). A principal-components analysis of the Narcissistic Personality Inventory and further evidence of its construct validity. *Journal of personality and social psychology*, *54*(5), 890.

Riess, H. (2017). The Science of Empathy. *Journal of Patient Experience*, *4*(2), 74–77. https://doi.org/10.1177/2374373517699267

Smith, J. (2015). What is empathy for? *Synthese*, *194*(3), 709–722. https://doi.org/10.1007/s11229-015-0771-8

Appendix

Empathy Scale Item Set[i]

1. I appreciate when my friends open up to me about their struggles.
2. If someone trips and falls (injures themselves) I feel concerned for them.
3. I will help a stranger if they look like they need help.
4. I will do things for other people even if I don't get anything out of it.
5. I try to cheer someone up right away when they are upset.
6. I find it easy to put myself in somebody else's shoes.
7. I offer a solution right away when someone is overwhelmed.
8. I feel guilty when I think someone has misunderstood me.
9. When I see someone in need I am only interested if I have the same problem.
10. I am comfortable sharing my feelings openly with those around me.
11. I keep weak boundaries to avoid upsetting others.
12. I often get emotional and I don't know why.
13. I will apologize if I think I did something wrong.
14. I would never break a law, no matter how minor.
15. I am able to predict how others are feeling.
16. I am upset when I see animals in pain.
17. If I could, I would bring home a stray dog or cat.
18. I donate old clothes to places like Goodwill.
19. I will tell a stranger that their shirt is inside out.
20. I will put an insect outside instead of killing it.
21. I give money to homeless people.

22. I agree that people deserve the bad things that happen to them.
23. If I'm cut off in traffic, I assume the person is in a rush.
24. When someone makes a funny joke at a stranger's expense, I laugh.
25. I would call someone out for being wrong about something in public.
26. I would call the police if I saw a dog in an unattended car.
27. In a conversation, I tend to focus on my own thoughts rather than on what the other person might be thinking.
28. I have very strong morals.
29. If someone asked me if I liked their haircut, I would answer truthfully even if I didn't like it.
30. I struggle with differentiating between if someone is being mean or if they are being kind.
31. I have been called insensitive but struggled to understand why.
32. When making important decisions it is important to me to consider other people's emotions.*
33. I avoid media that make me upset (news, movies, TV shows, etc.).
34. I always try to understand where someone is coming from even if I don't agree with them.*
35. If a friend is telling me about something that is upsetting them, I validate their emotions.*
36. When someone is going through a tough time, I ask how they're feeling.
37. I try to match the emotional tone of someone sharing their feelings before responding.*
38. I find it easy to stay emotionally present when someone is in distress.
39. I try to change the topic of emotional conversations that make me uncomfortable.

40. In emotional conversations, I focus on the other person's experience without bringing mine into the conversation.
41. I can sit with my feelings without trying to distract myself until they pass.
42. When I feel anxious, I can name the specific emotion I'm experiencing.
43. I don't think about how my emotions affect me.
44. I think about the impact of my emotions.
45. When someone upsets me, I can get over it quickly.
46. I am easily upset.
47. I can sit with my emotions for a short time but then look for distractions.
48. It is hard for me to sit quietly when someone is in distress.
49. I tend to get emotionally involved with a friend's problems.*
50. Tuning into how someone else feels is intuitive for me.*
51. I enjoy caring for other people.*
52. When I watch movies, I often resonate with the antagonist's motives.

[i] The validation questions are marked with * and the items after item analysis were items 1 to 17.