DO PICTURES SAY MORE THAN A THOUSAND WORDS?
A VALIDATION OF THE PICTORIAL EMPATHY TEST AND THE EFFECT OF TRAUMA.

Aantal woorden: 7606

Carmen Meeus
Studentnumber: 01508147
Promotor: Prof. dr. Sven Mueller
Supervisor: Anna Hudson

Master’s Thesis I submitted in order to obtain the degree of Master of Science in Psychology, option Clinical Psychology

Academic year: 2017 - 2018
Acknowledgements

Although it has been a long and bumpy road, I finally managed to bring this to a good end. First, I would like to thank my promotor ‘Sven Mueller’ for his patience and good advice throughout this journey. Besides that, I would like to thank my supervisor ‘Anna Hudson’ for the help with collecting the data and analysing the results. I am also grateful for the extensive skype sessions. Not to forget the support and the encouraging words of my family and friends helped me in this process. And at last, a special thanks to my parents for giving me the opportunity to realize this dream.
Abstract

**Objective:** Empathy is a significant aspect of interpersonal relationships and has an impact on many life domains. Many questionnaires have been developed to measure empathy, but all have their shortcomings. Because of a lack of a decent instrument to measure empathy, Koirikivi (2014) designed the Pictorial Empathy Test (PET). Using pictures provides many advantages and a chance to overcome shortcomings of questionnaires. Current research aims at finding further evidence for the PET as a reliable and valid instrument to measure affective empathy (hypothesis 1). Moreover, we want to investigate the effect childhood traumatic events has on affective empathy (hypothesis 2). Even though there is some evidence there are still uncertainties whether experiencing a trauma leads to alterations in empathy. **Methods:** In total 286 participants were collected via Prolific Academic and had to fill in a web-based questionnaire. To validate the PET, questionnaires that measure stress, anxiety and depression were also used in this study. Next to that, the Interpersonal Reactivity Index (IRI) was added to establish convergent validity. To answer the second hypothesis (n=196) concerning trauma, we used the Early Trauma Inventory Self Report-Short Form (ETISR-SF) which measures traumatic experiences. **Results:** Positive correlations between the PET and affective subscales of the IRI provided evidence for the reliability and validity of this instrument. Regarding the effect of childhood traumatic events, a significant difference was found between both groups on the affective empathy subscale of the IRI. Individuals who experienced child abuse, scored higher on affective empathy. **Conclusion:** This study confirmed that the PET can serve as an effective tool for assessing affective empathy. Besides that, experiencing childhood maltreatment led to higher affective empathy in this sample.
# Table of Contents

- Empathy ........................................................................................................... 1
- Self-report Measures of Empathy ................................................................. 2
- Behavioural Measures of Empathy ............................................................... 4
- Trauma ............................................................................................................... 5
- The Influence of Trauma on Empathy ........................................................... 6
- Present Research ............................................................................................ 7
- Methods ............................................................................................................ 9
  - Participants ..................................................................................................... 9
  - Measures ........................................................................................................ 10
  - Procedure ....................................................................................................... 12
  - Data Management and Analysis ................................................................. 12
- Results .............................................................................................................. 15
  - Confirmatory Analysis .................................................................................. 15
- Discussion ......................................................................................................... 21
- Results .............................................................................................................. 21
- Limitations ...................................................................................................... 23
- Suggestions for Future Research ................................................................. 23
- Conclusion ....................................................................................................... 25
- References ...................................................................................................... 26
Empathy

Empathy is derived from the German term “Einfühlung”. It was Theodor Lipps (1903) who started to use the term Einfühlung in the context of psychology at the beginning of the 20th century. This usage of Lipps’ terminology was translated into English by Titchener only a few years later (1909). Both scientists were convinced that empathy was caused by an inner imitation of the observed person’s emotional cues, today known as motor mimicry (Davis, 1996). According to de Vignemont and Singer (2006) there are as many definitions of empathy as there are researchers. Empathy can be defined as the reaction of a person towards the experiences of another and the ability to share and understand their feelings (Davis, 1983; Decety, & Moriguchi, 2007). In other words, being “empathic” means being able to put yourself in another’s place (Singer, 2006). Eisenberg and colleagues (2006), also added that these feelings should be similar to the feelings of the victim itself or should be logical in the context.

Empathy can be divided into cognitive and affective components (Davis, 1983). The cognitive components contain the ability to mentalize and understand people’s perspective, experiences and feelings. Furthermore, it helps with predicting the behaviour of others and facilitates communication (Davis, 1983; Staub, 1987). The cognitive component is closely related to what is known as “theory of mind”, which is the ability to attribute intentions, thoughts or feelings to others that are different from oneself (Spinella, 2005). Emotional empathy, on the other hand, is about the emotional reaction to another’s mental states. Whereas some researchers will emphasize the affective components in their definition, others will focus on the cognitive components (Davis, 1983; Spinella, 2005; Staub, 1987). There are also researchers who use a more multidimensional approach, which integrates both components in the definition and treats them as related constructs (Davis, 1983).

Empathy is a significant aspect of interpersonal relationships as it leads to a stimulation of prosocial behaviour and appropriate moral development (Spinella, 2005). Empathy can predict someone’s future actions and provides us more knowledge about other people. These two elements contribute to better and more harmonious interaction. It should not come as a surprise then that empathy constitutes a crucial component of social contact and of relationships between mother and child, and partners (Singer, 2006). The study of Konrath and colleagues (2014) shows that empathy is associated with a better recognition of one’s emotions. The higher the empathic ability, the more emotions the participants could guess. A recent study revealed that empathy is related to supportive behaviour, also known as helpful behaviour, in a romantic relationship.
Negative support (e.g. criticizing and giving useless advice) is lower in individuals who score higher on affective empathy. More instrumental support (e.g. providing advice) is observed when one scores higher on cognitive empathy (Verhofstadt et al., 2016). Furthermore, empathy discourages antisocial behaviours (Spinella, 2005). A review of Eisenberg and Miller (1987) showed that empathic ability is positively related to prosocial behaviour such as helping relatives. Research has shown that empathy has two opposing sides. Whereas in some circumstance it facilitates prosocial behaviour such as instrumental support, it also from time to time leads to immoral behaviour. People can use this ability for positive and negative purposes (Decety & Moriguchi, 2007). Blair (2010) found that a lack of empathy leads to more aggressive behaviour. Another ‘dark side’ of empathy is illustrated by Konrath and colleagues (2014), who found that the personality trait ‘exploitativeness’, which is an aspect of narcissism, is associated with better emotional recognition. A behavioural task was used to measure emotional recognition. Being able to recognize negative emotions can make it easier to manipulate and deceive people. Furthermore, a lack of empathy is related to many psychiatric disorders (e.g. autism spectrum disorder, antisocial and narcissistic personality disorder) (Decety & Moriguchi, 2007). For example, in a study of Ritter et al. (2010) was found that patients with narcissistic personality disorder (NPD) were characterized by a lack in emotional empathy. On the other hand, patients with Borderline Personality Disorder (BPD) showed a lack in cognitive empathy, but an increased level of affective empathy (Harari et al., 2010). Many studies show the relevance of empathy for human interactions. Therefore, a proper instrument to capture empathic responding is of great interest.

Self-report Measures of Empathy

Many questionnaires have been developed to measure empathy. First, we need to make a distinction between unidimensional and multidimensional empathy measurements. In unidimensional measurements, a common construct of empathy is assessed by the items, which focus either on the cognitive or the affective aspect (Spreng, McKinnon, Mar, & Levine, 2009). The aim of multidimensional measurements is to focus on more than one aspect of empathy (Davis, 1980). The construct validity is for most of the questionnaires poor and the length is rather variable. For example, the Toronto Empathy Questionnaire (TEQ) is a unidimensional, 16-item self-report questionnaire to assess emotional empathy by using a 5-point scale (Spreng et al., 2009). The questionnaire has a good internal consistency, convergent validity and test-
retest reliability. However, there is uncertainty about whether it contains all facets of empathy (Spreng et al., 2009).

Another unidimensional measurement for empathy is the *Hogan Empathy Scale* (HES), which consists of 64 items. It is used to measure cognitive empathy and contains four factors: social self-confidence, even-temperedness, sensitivity, and nonconformity (Hogan, 1969). However, it has a low test-retest reliability and internal consistency. The factor structure does not correspond to expectations and is uncorrelated. Some even suggest that this scale is rather a measure of social skill (Froman & Peloquin, 2001; Baron-Cohen & Wheelwright, 2004). Furthermore, it is not very varied since it only uses two-point true/false-scale (Jolliffe & Farrington, 2004).

The *Questionnaire Measure of Emotional Empathy* (QMEE) was developed by Mehrabian and Epstein (1972) and is used to assess emotional empathy. The test consists of 33 items and uses a 9-point scale. According to Mehrabian, Young and Sato (1988), the QMEE measures overall emotional arousal instead of empathy. Therefore, Mehrabian revised the QMEE in 2000, thus creating the Balanced Emotional Empathy Scale (BEES). However, this scale has not been published.

Finally, the *Interpersonal Reactivity Index* (IRI; Davis, 1980) is a multi-dimensional instrument to measure empathy. Davis (1980) was the first to design a self-report instrument to measuring the affective and cognitive aspect of empathy. This self-report questionnaire consists of 28 items including four subscales of empathy. The Perspective-Taking and Fantasy subscales measure cognitive empathy. Affective empathy on the other hand is measured by the Empathic concern and Personal Distress subscales. A five-point scale from 0 to 4 is used to indicate how much the participant identifies with the item. The IRI has proven high intrascale, test–retest reliability and convergent validity (Davis, 1980). However, it is debatable if the four subscales truly measure empathy. The Fantasy and Personal Distress subscales more likely measure imagination and emotional self-control (Baron-Cohen & Wheelwright, 2004). In a study of Ritter et al. (2010) was found that the cognitive empathy items in the IRI seem to measure the motivation instead of the ability. Items in the IRI start with “I try to…” or “I tend to” (Davis, 1980). This overview shows that many questionnaires of empathy are available, but all have their shortcomings.
Behavioural Measures of Empathy

Because of the disagreements between different empathy questionnaires as outlined above, other investigators tried alternative ways to measure the empathic ability, for example by using pictures. These tests were developed to overcome the shortcomings of questionnaires. The main reasons for developing an instrument with pictures is that Dziobek (2008), Lindeman, Koirikivi and Lipsanen (2016) wanted a shorter test and one that gives the opportunity to capture the feelings of empathy immediately as they occur. Thus, whereas questionnaires measure the perception of someone’s empathy, the pictorial form is supposed to be a direct way of measuring empathy. Furthermore, they wanted an instrument that is easier to translate into other languages and to use in different settings.

The Comic Strip Task (CST) developed by Völlm et al. (2006) is a non-verbal task, which uses a series of comic strips. This behavioural task was used in a study of brain activations. Participants need to estimate the mental state of the person illustrated in the strips. The task comprises four conditions: theory of mind, cognitive empathy, physical attribution with one character, and physical attribution with two characters. In the empathy condition, one must complete the story by choosing out of two or three options. In their choice, they must take into account the feelings of the main character. However, no information is available on internal consistency, test-retest reliability and convergent and concurrent validity (Völlm et al., 2006). Reid et al. (2012) describe the CST as a simplistic instrument and not as the best predictor for cognitive empathy. Besides that, the comic strips do not seem to reflect real-life situations. In another measure developed by Dziobek et al. (2008), namely the Multifaceted Empathy Test (MET), 26 pictures taken from the International Affective Picture System illustrate people in emotional situations. Participants need to describe the feelings of the subject on the picture to measure cognitive empathy and are given feedback about the accuracy of their answer. After the feedback, participants need to rate how they feel when they see the picture (affective empathy). The MET has a high ecological validity, compared to self-report questionnaires that measure empathy. Emotional empathy is measured implicitly and explicitly in the MET (Dziobek et al., 2008). An adapted version was developed by Kirchner, Hatri, Heekeren, and Dziobek, (2011) where participants are asked to estimate the emotion. Koirikivi (2014) believed it would be better to assess affective empathy with a separate test, since in the MET the participant is told about the distress of the person on the picture. In addition, the MET is not a good instrument to collect data from a large group, since it was developed to use in an experimental laboratory setting.
In response to the shortcomings of the MET, Koirikivi (2014) developed the *Pictorial Empathy Test* (PET), which is a test that measures affective empathy by using pictures. Seven pictures are used that display people in stressful situations. These pictures were taken from Wikimedia Commons by Koirikivi (2014) and have a Creative Commons license. Participants have to say how emotionally moving they find the photograph on a five-point scale ranging from *not at all* to *very much*. According to Lindeman, Koirikivi, and Lipsanen (2016) the PET has a high internal consistency and test-retest reliability. Next to that, their research shows evidence for convergent and discriminant validity.

Using this type of measure has several benefits. Photo-based measurements can be easily translated into other languages and used in different settings. Next to that, the feelings of empathy can be captured immediately, while text-based measures are rather retrospective. Compared to most of the text-based measures, the PET is short and therefore easy to complete (Koirikivi, 2014; Lindeman, Koirikivi & Lipsanen, 2016). A photo-based measure has a stronger ecological validity and leans closer to everyday situations compared to self-report questionnaires. Moreover, a pictorial form is less biased by social desirability due to the use of an implicit way of questioning (Dziobek et al., 2008). Questionnaires are longer and not easy to translate. Some people are not able to read and the reactions are not authentic (Lindeman, Koirikivi & Lipsanen, 2016). Using pictures is a more naturalistic way of measuring empathy. Furthermore, these instruments are easier to understand by people and can still be used on the web.

**Trauma**

One vulnerable group for whom empathy and empathic responding might be altered are individuals who have experienced a psychological trauma. Evidence to date, however, is inconsistent. The term “trauma” has Greek and Latin roots, which means “injury to the body”. Psychological trauma means there is an “injury to the mind”, which is characterized by alterations in cognitive processes and functions. The cause of a psychological trauma is external, but the consequences are mainly internal. However, a trauma can be physical (e.g. assault and rape) or psychological (e.g. emotional neglect). Physical trauma should not be confused with neurological trauma, which means an injury to the brain (Wilson & Keane, 2004). Some examples of traumatic events are: natural disasters, homicides, major car accidents, physical and sexual abuse. Wilson and Keane (2004) describe trauma as a combination of psychological and biological processes, which consists of brain, memory, cognition, emotion, perception etc. Traumatic experience can be seen as traumatic stressors, which have an influence on organismic functioning (Wilson & Keane, 2004). Exposure to trauma can develop into post-traumatic
stress disorder (PTSD); (Bonanno, 2004), which will not be further discussed in this study. Examples of trauma mentioned in the DSM-5 are exposure to actual or threatened death, serious injury, or sexual violation. This trauma can be experienced directly or witnessed. Besides that, hearing someone close to you being experienced to the trauma can be the cause or because of repeated exposure to aversive details of the trauma through work (American Psychiatric Association, 2013). The traumatic experience can be described as an intense stress with which one is not able to cope. Whether an event is experienced as a trauma is completely subjective. A trauma can be caused by a one-time occurrence or chronic, repetitive events (Giller, 1999) and is mostly understood as events happening to a person, such as sexual abuse, physical abuse. But a trauma can also be an omission of events that should have happened, such as neglect (Briere & Scott, 2006). A traumatic event can occur throughout your whole life. Childhood maltreatment is often linked to deficits in interpersonal skills during adulthood (Young & Widom, 2014).

**The Influence of Trauma on Empathy**

Understanding how empathy is altered after trauma is meaningful, since sensing other’s feelings’ is an important element in different life domains. As detailed above, empathy has an impact on close relationships and may lead to more satisfaction, emotional and instrumental support in the relationship (Feeney & Collins, 2001; Franzoi, Davis, & Young, 1985). Moreover, a correlation was found between compassionate goals and more closeness, trust and support in the relationship (Crocker & Cannevello, 2008).

Empathy is furthermore an important element in caring-related professional settings. For example, higher empathy in psychologists leads to better patient outcomes (Truax et al., 1966). There is some evidence for a decreasing effect. Research by Mrug, Madan, Cook and Wright (2015) revealed that empathy decreases when one is exposed to violence. Other researchers found that children who experienced violence, neglect and abuse showed less empathy (Hinchey & Gavelek, 1982; Main & George, 1985), while others did not find this effect. For example, research by McCloskey and Lichter (2003) showed no deficits in empathy for people who were exposed to martial violence. The study of Mrug et al. found a quadratic relation between real-life violence and empathy, which means that the level of empathy was the highest when confronted with medium levels of violence. When confronted with lower or higher levels of violence, the empathic ability decreased. Parlar et al. (2014) found that women who experienced a childhood trauma and have PTSD have changes in their empathic abilities. More specifically they have lower perspective taking and empathic concern and higher personal distress compared
to healthy controls (Parlar et al., 2014). On the other hand, for some people, experiencing a trauma can lead to hostile and aggressive behaviour. Next to that, their mental health and functioning can be affected. However, a study by Vollhardt and Staub (2008) found that people who have been victimized showed more feelings of empathy and felt more responsible to help victims of a natural disaster. This phenomenon is known as altruism born of suffering (ABS, Staub, 2003). As there are contradictions in the effect of trauma on empathy, further research is required.

Present Research

Research by Koirikivi (2014) on the PET has developed curiosity to further validate this instrument. Research showed that the PET has a high internal consistency and test-retest reliability. Next to that, research shows evidence for convergent and discriminant validity (Lindeman, Koirikivi & Lipsanen, 2016). These results are promising and need to be further examined. Moreover, this way of measuring empathy has many advantages and can be used in practice (e.g. health care or recruitment) as the PET is a brief questionnaire that can be easily translated and used in different settings. This test can help us overcome the shortcomings of questionnaires.

Further research is necessary since there are many inconsistent findings about the effect of trauma on empathy and no unambiguous results are available. Since empathy is a fundamental aspect of interpersonal relationships and societies (Spinella, 2005) and trauma perturbs one’s view and relationship to another person, research on the relationship between empathy and trauma is of great importance. Moreover, a lack of empathy can lead to many negative outcomes (Decety & Moriguchi, 2007; Blair, 2010) and would have to be considered a risk factor for development of psychopathology.

The aim of this study was to validate the PET as a valid and reliable instrument for assessing affective empathy. Next to that, we wanted to clarify if the experience of a traumatic event during childhood has an influence on empathy. Based on the research by Mrug et al. (2015), we expected to see a decrease in affective empathy after being exposed to child maltreatment. The focus of this study was on the experience of a traumatic event and not traumatization.

Hypothesis 1: if the PET correlates with the subscales of the IRI that measure affective empathy and measures empathy, then participants who score high on the IRI subscales ‘empathic concern’ and ‘personal distress’ will also score high on the PET.
Hypothesis 2: participants who have previously experienced a traumatic event during childhood will score significantly lower on affective empathy than people who did not experience a traumatic event. To validate the PET, questionnaires that measure stress, anxiety and depression were used in this study. Next to that, the IRI was added to establish convergent validity. We used the IRI as it is a well-validated empathy questionnaire and a world-wide used instrument. To answer the second hypothesis concerning trauma, this study used the Early Trauma Inventory Self Report-Short Form (ETISR-SF) which measures traumatic experiences.
Methods

Participants

A total of 372 participants, of whom 144 male, 172 female and 3 genderqueer, have completed the online survey on Prolific Academic in exchange for £1.67 each. The average duration of this survey was 700.37 seconds (SD=505.18). The preset eligibility criteria of this survey were an age of 18-50 years old, a maximum of 30 previous Prolific studies and a minimum of 90% approval rating on Prolific for previous studies. Only participants with English as a native language, or who were fluent or near-fluent in English were included to ensure the questions would be understood correctly. Furthermore, individuals needed to complete the whole survey to be part of this study. This study used for the time of completion a cut off of 360 seconds. People who completed the survey in less than 360 seconds were excluded. Besides that, participants who failed more than one of the trick questions were excluded. The sample for testing the first hypothesis is different from the second hypothesis. For the second hypothesis, individuals reporting a psychiatric disorder were excluded. Taking these criteria into account, 286 out of the 372 participants were included in the study of the first hypothesis (mean age 29.23, SD=7.90, range 18–49 years old). Of whom 45.8% were male, 53.5% female and 0.7% genderqueer (table 1).

For the second hypothesis, 60.8% of the 286 participants were included. In total 112 individuals were excluded for reporting that they had been diagnosed with a current psychological condition. However, individuals who reported a condition in the past, were included in this study. In total 195 individuals were included for the second part. In the second hypothesis, 49.7% of participants were male and 50.3% female. The age varied between 18 and 49 and the mean age was 29.43 (SD = 8.06). The majority of this group (90.2%) experienced a minimum of one traumatic event ($M=6.05$, SD= 4.57). However, only 35.9% considered themselves to have suffered from a traumatic event. Child abuse was reported by 79.0% of the participants. The mean number of child abuse events in the first sample was 4.52 (SD=3.83) and in the second 4.04 (SD=3.55). Physical punishment ($M=1.92$, SD=1.63) was reported more than emotional abuse ($M=1.46$, SD=1.78) and sexual events ($M=.66$, SD=1.29).

This study was approved by the ethics committee of the Faculty of Psychology and Educational Sciences, Ghent University.
Table 1

Summary demographic data of participants.

<table>
<thead>
<tr>
<th>Hypothesis 1 (n = 286)</th>
<th>Hypothesis 2 (n = 195)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>131 45.8</td>
</tr>
<tr>
<td>Female</td>
<td>153 53.5</td>
</tr>
<tr>
<td>Genderqueer</td>
<td>2 0.7</td>
</tr>
<tr>
<td>No traumatic event</td>
<td>20 7.0</td>
</tr>
<tr>
<td>Traumatic event</td>
<td>266 93.0</td>
</tr>
<tr>
<td>No Child Abuse (NCA)</td>
<td>55 19.2</td>
</tr>
<tr>
<td>Child Abuse (CA)</td>
<td>231 80.8</td>
</tr>
<tr>
<td>Physical punishment</td>
<td>204 71.3</td>
</tr>
<tr>
<td>Emotional abuse</td>
<td>163 57.0</td>
</tr>
<tr>
<td>Sexual events</td>
<td>92 32.2</td>
</tr>
<tr>
<td>Perception no trauma</td>
<td>154 53.8</td>
</tr>
<tr>
<td>Perception trauma</td>
<td>132 46.2</td>
</tr>
</tbody>
</table>

Measures

To measure empathy in a more behavioural manner, we used the *Pictorial Empathy Test* (PET; Lindeman & Koirikivi, 2015) which measures affective empathy. In this measure, seven photographs of people in stressful situations were rated by participants. The pictures were visible for as long as the participants wanted and were presented in a fixed order. Participants had to say how emotionally moving they find the photograph, ranging from *not at all* to *very much*. The score was obtained by taking the mean score of the seven photographs (ranging from 1 to 5). The PET has a high internal consistency and test-retest reliability. Next to that research showed evidence for convergent and discriminant validity (Lindeman, Koirikivi, & Lipsanen, 2016).

Self-reported empathy was measured using the *Interpersonal Reactivity Index* (IRI; Davis, 1980). The IRI is a self-report questionnaire consisting of 28 items including four subscales of empathy. The ‘fantasy’ scale tells us more about the urge to identify with fictional characters. The second subscale, the ‘perspective-taking’ scale, gives insight into the attempts to take other peoples’ perspectives. Compassion and worry for the individual was measured with the ‘empathic concern’ scale. At last the ‘personal distress’ scale contains items about feelings of fear and discomfort while observing the other person. A five-point scale from 0 to 4 was used to indicate how much the participant
identified with the item. The total score can vary from 0 to 112. Higher scores indicate a higher degree of empathy. The IRI has proven high intrascale, test–retest reliability and convergent validity (Davis, 1980). Cronbach’s alpha for the different scales ranges from .83 to .87, which is similar to what Davis (1980) obtained.

Traumatic experiences were measured with the Early Trauma Inventory Self Report-Short Form (ETISR-SF). It is a 27-item self-report questionnaire which was developed by Bremner, Bolus and Mayer (2007). The ETISR-SF contains four domains: general trauma, physical, emotional and sexual abuse. Whereof general trauma measures adulthood trauma after the age of 18 and the other domains measure childhood trauma before the age of 18. Each domain consists of questions where the responder needs to answer with ‘yes’ or ‘no’. Besides that, there are two additional items at the end of the questionnaire for people who answered positive on at least one of the previous questions. These questions explore whether they have felt intense fear, which is one of the criteria for PTSD in the DSM-IV (American Psychiatric Association, 2000), or depersonalization during the event. A higher score indicates a higher experience of traumatic events. The ETISR-SF has determined an acceptable validity and internal consistency (Bremner, Bolus, & Mayer, 2007). Cronbach’s alpha for the different domains ranges from .69 to .85, which is acceptable, but lower than what Bremner, Bolus and Mayer (2007) obtained. Although the domain of general trauma has a Cronbach’s alpha below .70, it does not differ extremely.

Depression and anxiety were measured with the Mood and Anxiety Symptom Questionnaire (MASQ; Watson & Clark, 1991). The original MASQ is a 90-item self-report questionnaire. For this study the 30-item short adaptation of the MASQ (MASQ-D30) is being used. On a five-point scale, participants can indicate how strong they experienced the symptoms during the week before, ranging from not at all to extremely. The total score varies from 30 to 150. Higher scores indicate a greater degree of symptomatology (Watson et al., 1995). It was constructed in order to test the Tripartite Model of Watson and Clark (1991). It is a reliable and valid instrument with a good internal consistency and an acceptable convergent validity (Wardenaar et al., 2010). Reliability for the MASQ in this study was good (α=.93).

The Perceived Stress Scale (PSS; Cohen, Kamarck, & Mermelstein, 1983) measures the perception of stress, more specifically the extent to which people experience situations as stressful. Feelings and thoughts about the last month are questioned in the PSS. On a five-point scale, respondents can specify how often they have particular
feelings, ranging from never to very often (Cohen et al., 1983). The original scale consists of 14 items. The total score can vary from 0 to 56. A higher score indicates a higher perception of stress (Cohen & Williamson, 1988). This scale has demonstrated high internal and test-retest reliability (Cohen et al., 1983). Reliability for the PSS in this study was good (α=.87), as in Cohen et al. (1983).

Procedure
The questionnaire was programmed in Qualtrics (Qualtrics, Provo, UT, USA) and shared on Prolific Academic, which is a secure questionnaire tool, where everyone could voluntarily and anonymously participate. The participants were asked to fill in an informed consent giving permission to use the results. By clicking a little box, the participants agreed to voluntarily participate in this experiment. The answers were processed confidentially. Besides that, they needed to type in their name to sign the consent form. To receive the payment, one had to fill in their Prolific Academic ID at the beginning of the study. Participants had the ability to stop with the questionnaire at any time.

The questionnaire started with some questions concerning exclusion criteria, namely their level of English and their age. When the participants did not meet the criteria, they were redirected to another page which stated that they are ineligible to take part in this experiment. After that, a description was given about the experiment and the informed consent needed to be signed. The PET was the first questionnaire and randomly followed by the MASQ, IRI, PSS and ETI to avoid order effects. Within the questionnaires some trick questions were added to avoid response bias e.g. ‘answer this question by clicking on the second answer option’. After the questionnaires, the participants had to answer some demographic questions e.g. age, sex, native language, etc. and had the opportunity to write a comment. At the end there was a debriefing about the experiment and contact details of the researchers were added.

Data Management and Analysis
SPSS Statistics 25 was used to carry out the data analysis. Before the use of parametric tests, Shapiro-Wilk tests and Levene’s tests were performed to assess normality and homogeneity of variance. For the first hypothesis the Pearson correlation was used since the sample consists of 286 subjects (n > 30), which assumes a normal distribution. The Pearson correlation analysis was used to analyse correlations between the PET and the IRI. More specifically a partial correlation was used to control the effect of depression, anxiety and stress. Also, a partial correlation between the PET and the
affective empathy subscales of the IRI was carried out since the PET is a measurement for affective empathy. All the correlations were also completed without controlling for these variables. Next, the correlation coefficients were compared by using the Steiger’s Z test (Steiger, 1980) to explore if certain correlations differed significantly. Thereafter, Pearson correlations were executed between the MASQ, the PSS and the PET. These analyses were applied to establish convergent validity of the PET. The analyses are two tailed and the alpha level was set at p<0.05 for all analyses. For the internal consistency, the Cronbach’s alpha was used. This determines whether the scales are reliable. Also, the ‘Cronbach’s Alpha If Item is deleted’ is relevant to see whether one of the items should be removed.

Sex differences in affective empathy were analysed by using independent samples t-test. For this analysis, the sample of the second study (n = 193) was used to avoid effects of psychological conditions. Both males and females were normally distributed for the two affective subscales of the IRI and the PET. Next to that, a partial Pearson correlation for both genders was used between the affective subscales of the IRI and the PET while controlling for the MASQ and PSS. These correlations could provide evidence whether the scores differed significantly on explicit and implicit measures of affective empathy between males and females. Thereafter, the correlation coefficients were compared by using a formula taken from Cohen and his colleagues (2003) for independent groups.

Also, for the second part, the Shapiro-Wilk normality test was used to check whether the data follows a normal distribution. The null hypothesis that the data came from a normally distributed population could not be rejected, since the p-value was greater than the chosen alpha level. In other words, the data of both groups were from a normally distributed population (Shapiro-Wilk, p > .05). Next to that, the Q-Q plot showed that the data points were close to the diagonal line. However, some outliers were observed in the data. Therefore, the analyses were completed with and without the outliers. This means an independent samples t-test was used to assess group differences in affective empathy between participants who experienced childhood abuse (CA) and participants who did not (No Child Abuse = NCA). This test compared the means on affective empathy scores between both groups to determine whether the differences are significant. Furthermore, the Spearman correlation analysis was used between the number of child abuse events and the affective empathy scores. Spearman correlations were also executed between the number of child abuse and the MASQ or PSS to control their effects.
Also, independent samples t-tests were completed separately for the three domains of childhood abuse, namely physical punishment, emotional abuse and sexual events. The mean differences were examined between these three domains and all the affective empathy scores. Next, sex differences in child abuse were examined by using the independent samples t-test. The means of these domains were compared between both sexes. In addition, the two additional questions of the ETISR-SF, concerning intense fear and depersonalization, were included in the analysis by using the independent samples t-test. In the ‘child abuse’ group (n=154), the participants who answered yes to both questions were compared with individuals who answered yes to maximum one. Finally, the effect of the perception of a trauma on empathy was included in this study. Independent samples t-tests were used to compare these two groups. This test will be executed twice. Once with the outliers and once without the outliers to avoid wrong conclusions.
Results

Confirmatory Analysis

Hypothesis 1: In table 2 the results of the Pearson correlations between the PET and the IRI are displayed. There seems to be a positive correlation between the PET and the IRI \((r(283)=.377, \ p<.01)\). Also, between the PET and the affective subscales a positive correlation was found \((r(283)=.410, \ p<.01)\); (figure 1). Furthermore, a positive correlation was found between the PET and the cognitive subscales \((r(283)=.202, \ p<.01)\). The PET and the affective subscales of the IRI had the strongest correlation. More specifically, it had the highest correlation with the ‘empathic concern’ scale (figure 2). The positive correlation with the ‘personal distress’ scale is less high (figure 3). Partial correlations controlling for the MASQ and the PSS show similar results. Except for the ‘personal distress’ scale the correlation was lower when controlled for the PSS. A comparison of the correlation coefficients showed that the correlation of the PET with the affective and the cognitive subscales differed significantly \(Z=3.298, \ p<.01\). Also, the correlation between the PET and ‘empathic concern’ differed significantly from the PET and ‘personal distress' scale \(Z=3.983, \ p<.01\). However, the correlation of the ‘personal distress’ scale was not significantly higher than the ‘fantasy’ scale \(Z=0.159, \ p>.05\) and the ‘perspective-taking’ scale \(Z=0.743, \ p>.05\). Furthermore, there was no meaningful correlation between the PET and the MASQ \((r(286)=.016, \ p=.789)\) and neither the PSS \((r(286)=.039, \ p=.509)\). Cronbach's alpha was .877, which indicates a high level of internal consistency for the PET with this specific sample. The removal of any question would result in a lower Cronbach's alpha. Therefore, none of the questions should be removed.

Table 2
Pearson Correlations between the IRI and the PET.

<table>
<thead>
<tr>
<th></th>
<th>PET</th>
<th>PET (controlling MASQ)</th>
<th>PET (controlling PSS)</th>
</tr>
</thead>
<tbody>
<tr>
<td>IRI</td>
<td>.377**</td>
<td>.379***</td>
<td>.377***</td>
</tr>
<tr>
<td>Affective subscales</td>
<td>.410**</td>
<td>.423***</td>
<td>.420***</td>
</tr>
<tr>
<td>Empathic concern</td>
<td>.458**</td>
<td>.458***</td>
<td>.457***</td>
</tr>
<tr>
<td>Personal distress</td>
<td>.193**</td>
<td>.204**</td>
<td>.192**</td>
</tr>
<tr>
<td>Cognitive subscales</td>
<td>.202**</td>
<td>.203**</td>
<td>.203**</td>
</tr>
<tr>
<td>Fantasy</td>
<td>.181**</td>
<td>.180**</td>
<td>.178**</td>
</tr>
<tr>
<td>Perspective-taking</td>
<td>.130*</td>
<td>.134*</td>
<td>.138*</td>
</tr>
</tbody>
</table>

\*\(p<0.05\), \**\(p<0.01\), \***\(p<0.001\)
Figure 1. Positive correlation between the PET and the affective subscales of the IRI.

Figure 2. Positive correlation between the PET and the subscale 'empathic concern'.
Figure 3. Positive correlation between the PET and the affective subscale 'personal distress'.

When observing the independent samples t-tests for sex differences, a significant difference on the PET scores was found between males ($M=3.49$, $SD=0.86$) and females ($M=3.72$, $SD=0.77$); $t(193)=-2.00$, $p=.047$ (figure 4). There was also a difference in means on the 'empathic concern' scale between males ($M=16.79$, $SD=6.02$) and females ($M=19.39$, $SD=4.85$); $t(183.83)=-3.311$, $p=.001$. On the 'personal distress' scale, males ($M=10.40$, $SD=5.36$) and females ($M=12.15$, $SD=5.55$) also scored different; $t(193)=-2.241$, $p=.026$. The partial correlation between affective subscales and PET while controlling for the MASQ ($r(131)=.439$, $p<0.001$) and PSS ($r(131)=.475$, $p<0.001$) was within males larger than for females controlling for the MASQ ($r(153)=.324$, $p<0.001$) and PSS ($r(153)=.336$, $p<0.001$). Also, for the correlations without controlling, the males' ($r(131)=.484$, $p<0.001$) correlation was higher than the females' ($r(153)=.307$, $p<0.001$). A comparison of the correlation coefficients while controlling for the MASQ showed that these correlations did not differ significantly $Z=1.121$, $p>.05$. When controlling for the PSS the correlations also did not differ significantly $Z=1.387$, $p>.05$. 

---

**Figure 3.** Positive correlation between the PET and the affective subscale 'personal distress'.
Figure 4: Comparison of the affective empathy means between males and females.

Hypothesis 2: The independent samples t-test showed that there was no significant difference on any of the affective empathy scores between participants who experienced childhood abuse and participants who did not. After deleting the outliers, only the affective subscales showed a significant difference; \( t(187) = -1.975, p = .05 \) (table 3). Participants who experienced childhood maltreatment, scored higher on the affective subscales than participants who did not. Spearman correlations were used since the data was not normally distributed. For these correlations the data without the outliers was used. Child abuse correlated significantly with the affective subscales of the IRI \( (r(189) = .149, p = .041) \), the ‘personal distress’ scale \( (r(189) = .155, p = .034) \), the MASQ \( (r(189) = .315, p = .01) \) and the PSS \( (r(189) = .210, p = .004) \). The PET \( (r(189) = -.042, p = .565) \) and the ‘empathic concern’ scale \( (r(189) = .073, p = .315) \) did not correlate meaningful with child abuse.
Table 3
Independent samples t-test between child abuse and affective empathy scores.

<table>
<thead>
<tr>
<th></th>
<th>CA</th>
<th>NCA</th>
<th>Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>PET</td>
<td>n = 153</td>
<td>n = 41</td>
<td>(t(192) = .269, p = .788)</td>
</tr>
<tr>
<td></td>
<td>3.61 (.81)</td>
<td>3.65 (.80)</td>
<td></td>
</tr>
<tr>
<td>Affective subscales</td>
<td>n = 149</td>
<td>n = 40</td>
<td>(t(187) = -1.975, p = .05)</td>
</tr>
<tr>
<td></td>
<td>30.42 (7.69)</td>
<td>27.60 (9.19)</td>
<td></td>
</tr>
<tr>
<td>Empathic concern</td>
<td>n = 153</td>
<td>n = 41</td>
<td>(t(192) = -1.435, p = .153)</td>
</tr>
<tr>
<td></td>
<td>18.46 (5.28)</td>
<td>17.07 (6.30)</td>
<td></td>
</tr>
<tr>
<td>Personal distress</td>
<td>n = 154</td>
<td>n = 41</td>
<td>(t(193) = -1.652, p = .100)</td>
</tr>
<tr>
<td></td>
<td>11.62 (5.42)</td>
<td>10.02 (5.73)</td>
<td></td>
</tr>
</tbody>
</table>

The independent samples t-tests were executed separately for each of the child abuse domains (n=195). The analyses showed that in none of the groups, affective empathy differed significantly (p>.05). Comparisons between the different sexes presented a significant difference for sexual events \(t(176.66) = -2.696, p = .008\). Females \((M=91, SD=1.46)\) reported more sexual abuse than males \((M=42, SD=1.05)\). Physical punishment and emotional abuse did not differ between the two groups.

Table 4
Independent samples t-test between sex and child abuse domains.

<table>
<thead>
<tr>
<th></th>
<th>Male (n = 96)</th>
<th>Female (n = 98)</th>
<th>Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physical punishment</td>
<td>2.10 (1.62)</td>
<td>1.72 (1.62)</td>
<td>(t(192) = 1.631, p = .105)</td>
</tr>
<tr>
<td>Emotional abuse</td>
<td>1.21 (1.60)</td>
<td>1.70 (1.93)</td>
<td>(t(186.66) = -1.950, p = .053)</td>
</tr>
<tr>
<td>Sexual events</td>
<td>.42 (1.05)</td>
<td>.91 (1.46)</td>
<td>(t(176.66) = -2.696, p = .008)</td>
</tr>
</tbody>
</table>

Moreover, independent samples t-tests were completed to examine differences between participants who answered both or maximum one of the extra questions concerning trauma. There was no significant difference on the results of the PET between people who experienced both of the items \((M=3.48, SD=.83)\) and individuals who did not \((M=3.68, SD=.83)\); \(t(152) = 1.444, p = .151\). Also for the ‘personal distress’ scale no significant difference was found between individuals who answered both \((M=10.68, SD=5.11)\) and the ones who did not \((M=12.30, SD=5.57); t(152) = 1.854, p = .066\). Further, there was a significant difference for the affective subscales of the IRI.
between people who experienced the two items (\(M=27.71, SD=8.18\)) and people who did not (\(M=31.65, SD=8.46\)); \(t(152)=2.898, p=.004\). Similarly, the ‘empathic concern’ scale differed significantly for the two groups (\(M=17.03, SD=5.69\); \(M=19.35, SD=4.97\)); \(t(152)=2.687, p=.008\). After deleting the outliers for three of the four scales, the difference for the PET was still not significant \(t(151)=1.184, p=.238\) (figure 5). The two groups still differ significantly for the ‘empathic concern’ scale \(t(151)=2.483, p=.014\). Also, the affective subscales still differ significantly for the two groups \(t(146)=2.184, p=.031\). The ‘personal distress’ scale had no outliers.

Figure 5: Comparison of the affective empathy means after answering the extra trauma questions.

Finally, when looked at the independent samples t-test for perception of a trauma, there was no significant difference in PET scores between those who consider themselves to have suffered from a trauma (\(M=3.59, SD=0.83\)) and participants who did not (\(M=3.62, SD=0.83\)); \(t(193)=.205, p=0.838\). Also, for the ‘empathic concern’ \(t(193)=-1.610, p=0.109\) and ‘personal distress’ scale \(t(193)=.832, p=0.406\) no meaningful differences were found.
Discussion

Results

The current study had two main hypotheses: 1. if the PET correlates with the subscales of the IRI that measure affective empathy and measures empathy, then participants who score high on the IRI subscales ‘empathic concern’ and ‘personal distress’ will also score high on the PET. Hypothesis 2: participants who have previously experienced a childhood trauma will score significantly lower on affective empathy than people who have not experienced a childhood trauma. With regards to hypothesis 1, positive significant correlations were found between the PET and the affective subscales of the IRI and mainly the ‘empathic concern’ scale. Even though the PET also correlated with the cognitive subscale, the affective subscale showed a significant higher correlation. Regarding hypothesis 2, the expectation was to see a decrease in affective empathy after being exposed to a childhood traumatic event. The results of the independent samples t-test suggested that the experience of childhood abuse increased the amount of affective empathy on the IRI.

These findings are consistent with previous research by Lindeman, Koirikivi and Lipsanen (2016) who found evidence for convergent validity. The instrument seems to measure the same general construct. However, the ‘personal distress’ scale did not correlate significantly higher than the cognitive subscales of the IRI (‘fantasy’ and ‘perspective-taking’). This could mean, the PET assesses only one of the two dimensions of affective empathy. Empathic concern is oriented on the unfortunate other and measures feelings of compassion and worry, while personal distress is self-oriented and measures feelings of anxiety and stress (Davis, 1980). The correlations found, were not very high, so it is possible that the perception of the individual’s empathy deviates from the actual degree of empathy. Self-report measures can be affected by perception errors or social desirability (Duan & Hill, 1996). Ritter et al. (2010) found that people overestimated their capacity for emotional empathy. However, these were patients with a narcissistic personality disorder. A low correlation does not immediately mean the instruments do not measure the same construct. Further research is of great importance.

For the second hypothesis, child abuse led to increased empathy scores on the affective empathy of the IRI. However, these results were not found with the PET and affective subscales ‘empathic concern’ and ‘personal distress’. We would expect these differences in minimum one of the two subscales as well. Furthermore, child abuse correlated positively with the affective subscale of the IRI and the ‘personal distress’
scale. These findings were in contrast with previous research by Mrug, Cook and Wright (2015), who found that empathy decreases when one is exposed to violence. Also, other researchers found similar results (Hinchey & Gavelek, 1982; Main & George, 1985). However, this cannot be stated without examining it. In our study, ‘personal distress’ seemed to correlate positively with child abuse. This is in line with findings by Parlar et al. (2014), who found that women who experienced childhood trauma and had PTSD scored higher on ‘personal distress’. It is possible that the participants who experienced child abuse, felt more responsible towards others (Vollhardt & Staub, 2008).

Also, the effects of the specific child abuse domains were examined. None of the domains (physical punishment, emotional abuse and sexual events) showed an effect on affective empathy. In other words, empathic ability was not different in the specific domains. What we see is that females reported significantly more sexual events than males. Which is in line with the findings by Tolin and Foa (2006), which showed that females experienced more sexual abuse in their childhood compared to males. Other child maltreatment was the same in males and females. The findings by Thompson, Kingree and Desai (2004), who found that males had more chance to experience physical punishment (e.g. pushed, bitten and kicked), were not found in our study. Participants who reported both intense fear and depersonalization, showed impaired empathy on the affective subscale and ‘empathic concern’ scale. This effect was not found on the PET and ‘personal distress’ scale.

Furthermore, the effect of perception of a trauma was examined on empathic ability. Independent samples t-tests compared individuals who indicated to have experienced a trauma and participants who did not. The results revealed no relation between the perception of a trauma and affective empathy scores. One’s perception is not the same as truly experiencing a trauma. Perception errors and differences in defining trauma can be a possible cause (Duan & Hill, 1996). Finally, the study showed sex differences in affective empathy. Females scored significantly higher on empathy than males. This is consistent with previous findings of Baron-Cohen and Wheelwright (2004) and Lindeman, Koirikivi and Lipsanen (2016). Both studies found that females scored significantly higher on empathy than males. Assumedly both culture and biology played a role in this relation (Christov-Moore et al., 2014). Pearson correlations were used to compare the coherence between explicit and implicit measures for each gender. These results showed less coherence between both measures for males but was not significantly lower than for females. Sex did not seem to influence the degree of coherence between the two empathy measures in this sample.
Limitations

The sample sizes for participants who did not experience child abuse was rather small compared to the ones who did. This could have influenced the results, since there were not many people to compare with. It is questionable if the results are representative enough. Further, it is not sure if all results were reliable, since some people on Prolific could have answered less serious. However, that is why some preset exclusion criteria were set, but there is still a chance that some of these people participated in the study. After setting a cut off for the average time spend on the test, many people still seemed to have fill in the questionnaire fast. It could be that some of the unreliable participants were still included in this study.

It is questionable if the participants experienced the events as traumatic. Some of the items ask about events that are common. For example: ‘Were you ever pushed or shoved?’ This seems like an event that many people would experience throughout lifetime. In addition, the person responsible for the event could make a difference. Being punched by a parent or sibling can be different in the degree of trauma. Almost all of the participants experienced one of the traumatic events. However, the number of people who reported feeling traumatized were a lot less. It is doubtful if the ETISR-SF is the right measure for assessing traumatic experiences. The perception should always be taken into account. Finally, it is important to acknowledge that we did not compare the group who experienced intense fear with the group who experienced depersonalization. This test could have given us valuable information. Both groups could differ significantly from each other.

Suggestions for Future Research

Further research is recommended, since the PET could be a good alternative for self-report measures. The PET had the highest correlation with the ‘empathic concern’ scale, but not with the ‘personal distress’ scale. There was no significant difference between the ‘personal distress’ scale and the cognitive subscales. There is still ambiguity about which domains the PET exactly captures.

When participants answered both questions (e.g. intense fear and depersonalization) positive, they seemed to show less empathy scores. However, after reporting childhood trauma, the empathy scores seemed to increase. These findings are rather contrasting. A possible explanation could be the presence of possible curvilinear relationships. This effect was also found in the study of Mrug, Cook and Wright (2015). The empathic ability could depend on the severity of the traumatic event.
Schreiter, Pijnenborg and Aan Het Rot (2013) suggest using more implicit and more ecologically valid measures of empathy to research the relation between empathy, depression and gender. If more evidence is found for the reliability and validity, the PET could be a good alternative to overcome the shortcomings of most studies. It seems relevant to combine self-report, behavioural and physiological measurements (e.g. skin conductance and heart rate) to avoid errors in perception and social desirability.
Conclusion

In summary, the present study tried to find evidence for the reliability and validity of the Pictorial Empathy Test (PET) as an instrument to assess affective empathy. Results were in line with our expectations, that the PET would correlate positively with the affective subscales of the Interpersonal Reactivity Index (IRI). These findings indicated that the PET can serve as an effective tool for assessing affective empathy. However, the instrument seems to measure mainly one of the two constructs of affective empathy, namely ‘empathic concern’. One should be attentive that the test may not measure all aspects of affective empathy. Besides, the effect of experiencing childhood trauma on affective empathy was examined. Findings were in contrast with our expectations to see a decrease in affective empathy. Individuals who suffered from child abuse did not score lower, but higher on affective empathy. This was only found for the IRI but not the PET. The number of child abuse events was associated with higher scores on the affective subscales of the IRI, mainly ‘personal distress’. Further research will be of interest, to find more evidence for the PET in different populations. Also, further exploration of the relation between child abuse and affective empathy is recommended.
References


