THE INFLUENCE ON EMPATHY BY HUMANOID ROBOT NAO PORTRAYING SADNESS USING SPEECH AND GESTURES.

DEBORAH NIKLAUS

THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
BACHELOR OF SCIENCE IN COGNITIVE SCIENCE & ARTIFICIAL INTELLIGENCE

DEPARTMENT OF
COGNITIVE SCIENCE & ARTIFICIAL INTELLIGENCE
SCHOOL OF HUMANITIES AND DIGITAL SCIENCES
TILBURG UNIVERSITY

STUDENT NUMBER

2024990

WORD COUNT

7908

COMMITTEE

Dr. Eriko Fukuda
Dr. Murat Kirtay

LOCATION

Tilburg University
School of Humanities and Digital Sciences
Department of Cognitive Science &
Artificial Intelligence
Tilburg, The Netherlands

DATE

July 13, 2022

ACKNOWLEDGMENTS

I would like to give my warmest thanks and appreciation to my supervisor dr. Eriko Fukuda for pushing me to reach the thesis deadlines and for always being available for guidance. I would also like to sincerely thank her for the whole process of helping me find participants for this study. I would also like to thank my friends and family for their undying support and understanding. Last but not least, I especially would like to thank my husband Shaikh, who is always there for me and who pushed me through the thesis process.

THE INFLUENCE ON EMPATHY BY HUMANOID ROBOT NAO PORTRAYING SADNESS USING SPEECH AND GESTURES.

DEBORAH NIKLAUS

Abstract

The demand and access to machines and robots are increasing, resulting in an increase in Human-Robot Interaction (HRI). Empathy is a human attribute that helps individuals create bonds and to form interactions with others. The aim of this study was to investigate whether the humanoid robot NAO could elicit empathy from humans while it is portraying sadness using storytelling and gestures. The sample of this study consisted of university students from ages 18 to 30 and had a sample size of n = 84 (33 males, 51 females). The State Empathy Scale (SES) designed by Shen (2010) was slightly modified and used to fit the experiment to measure the empathy levels after NAO told a sad story. The SES is a 12-item scale that tests three different empathy dimensions: affective empathy, cognitive empathy, and associative empathy. The average of these three empathy dimensions was used to calculate the general empathy level. Results show that, there is no statistically significant difference between the control group and the experimental group on affective empathy, cognitive empathy, associative empathy, or general empathy. Exploratory research within the control group (n = 42) and the experimental group (n = 42) was conducted between Cognitive Science & Artificial Intelligence students (CSAI) and students from other study programs. Within the experimental group (21 CSAI students, 21 other students) there is a statistically significant difference for general empathy and for associative empathy between CSAI students and other students. However, this research might have been influenced by combined experimenting with other researchers, and by the NAO robot overheating and crashing, during experiments.

1 INTRODUCTION

Robots are among us. The world is going through a vast technological change, leading to an increasingly easier access to machines and robots in daily life. These machines and robots are implemented in a range of tasks, varying from simple household chores to complex medical assistance in health care and anything in between. An example of a field where the use of robots is implemented is the education of children. Research conducted by Mubin, Stevens, Shahid, Mahmud, and Dong (2013) imply that robots employed in education with children can be a valuable asset, because it is able to stimulate, engage, and develop analytical thinking. NAO, developed by Softbank Robotics, is an instance of a robot that is used in the fields of education and healthcare of children (Kyrarini et al., 2021).

Robots can be defined as programmable machines that can replace and/or expand human work, by analysing and interacting with their environment (Trevelyan, 1999). Robots are part of everyday life, even though some users don't even realize this. For instance, the Internet of Things (IoT), classifies devices that operate using the internet as an infrastructure to interact with their environment using specific sensory abilities (Kumar, Tiwari, & Zymbler, 2019). The Internet of Robotic Things are IoT devices that can manipulate or interact in the physical world (Simoens, Dragone, & Saffiotti, 2018). Daily household devices such as Alexa, Google Home and even robot vacuum cleaners use such an interface to interact with users or their environment. Therefore, they can be defined as robots.

Humanoid robots posses sensors and actuators like other robots, and additionally have attributes that can be likened to that of humans, such as a head, torso, arms and legs. These humanoid robots also have the ability to interact and communicate with users (Choudhury, Li, Greene, & Perumalla, 2018). NAO is a programmable humanoid robot designed by SoftBank Robotics that has robotic limbs resembling those of humans, and impressive capabilities of movement, speech, reaction, and interaction with users. NAO is implemented in education to improve learning in children and it is also applied in healthcare as medical healthcare assistants to minimize exposure to sicknesses. However, even though NAO is technologically advanced and possesses a human-like exterior, its limitation is its inability to show facial expressions (De Beir, Cao, Gomez Esteban, Van de Perre, & Vanderborght, 2015).

Empathy is a human trait that can create and strengthen bonds and relationships within humans (Paiva, Leite, Boukricha, & Wachsmuth, 2017). It is a key element in creating interaction between individuals (Paiva et al., 2017). Ratka (2018) states that empathy is crucial in understanding the emotional state of others. Empathy can, therefore, improve communication, interaction, and trust between humans (Ratka, 2018). As the use and demand for machines and robots are increasing, Human-Robot Interaction (HRI) increases as well. Robots are an increasing presence in this era of constant technical innovation, and it is beneficial to study how HRI is perceived and how this interaction can be improved for the future use in fields like education and healthcare.

The research question for this study is:

RESEARCH QUESTION *Can humanoid robot NAO elicit empathy while portraying sadness using speech and gestures?*

The State Empathy Scale (SES) designed by Shen (2010) was used within this study to measure empathy on three different dimensions (affective empathy, cognitive empathy and associative empathy). Each of these dimensions of empathy are assessed within this study. Therefore, the sub-questions are:

SUB-QUESTION 1 *Can a humanoid robot elicit affective empathy while portraying sadness using speech and gestures?*

SUB-QUESTION 2 *Can a humanoid robot elicit cognitive empathy while portraying sadness using speech and gestures?*

SUB-QUESTION 3 *Can a humanoid robot elicit associative empathy while portraying sadness using speech and gestures?*

Based on the research question and the sub-questions, the following hypotheses were formulated:

Hypothesis 1: *A humanoid robot portraying sadness is more likely to elicit higher general empathy levels when using gestures during storytelling than a humanoid robot that does not use gestures during storytelling.*

Hypothesis 2: *A humanoid robot portraying sadness is more likely to elicit higher affective empathy levels when using gestures during storytelling than a humanoid robot that does not use gestures during storytelling.*

Hypothesis 3: *A humanoid robot portraying sadness is more likely to elicit higher cognitive empathy levels when using gestures during storytelling than a humanoid robot that does not use gestures during storytelling.*

Hypothesis 4: *A humanoid robot portraying sadness is more likely to elicit higher associative empathy levels when using gestures during storytelling than a humanoid robot that does not use gestures during storytelling.*

Section 2.2 will further discuss and elaborate on the sub-questions concerning affective empathy, cognitive empathy and associative empathy. Section 3.2 will explain how the general empathy level is calculated.

2 RELATED WORK

This section discusses the related research areas and contributions within this research framework. Section 2.1 section defines and explains the term "empathy", section 2.2 will elaborate on the dimensions of empathy (affective, cognitive and associative empathy), section 2.3 will discuss the difference between trait or state empathy and which was more appropriate for this research. Section 2.4 explains why sadness is chosen as the emotion of interest in this research. The last section 2.5 describes the robot NAO, and prior conducted research done with it.

2.1 *Empathy*

Empathy is a phenomenon various researchers have tried to define. The term itself was initially formulated by the researcher Edward Titchener as the word "Einfühlung", which was later translated to the term empathy (Titchener (1910), as cited in Ganczarek, Hünefeldt, and Belardinelli (2018)). With his original definition, Titchener described the term as the process of feeling the emotions of another being, as a result of projecting oneself into their perspective or surroundings (Ganczarek et al., 2018).

As the years progressed, more studies have been conducted on the subject. Cuff, Brown, Taylor, and Howat (2016) for instance, define empathy as an affective (emotional) response that is influenced by both empathy as a personality trait (trait empathy) and empathy elicited by the emotional state of another person (state influences). Preston and Hofelich Mohr (2012) approach a broader definition, which is the recognition and understanding of the emotions of another by observation or imagination. Paiva et al. (2017) define the term empathy as the understanding of the affective states of others and providing the correct reaction towards these states. This last definition can be supported by research conducted by Oktem and Cankaya (2021) which indicates that understanding the emotional state of another is an increased likelihood of a correct response from the perceiver. In addition to these statements, responses as a result of experiencing empathy play a role in enhancing chances of survival by, for instance, forming alliances with others (Elliott & Jacobs, 2013). Charles Darwin even hypothesized that emotions were critical in order to survive (Darwin, 1916).

Empathy is the main attribute in civilization and societies, as strengthening relationships and forming correct interactions constantly construct and develop societies in such a way that they become modern and more improved (Gulin, 2020). According to research done by Malinowska (2021), once a robot has formed a bond with a person, it is usually able to maintain this relationship. Malinowska (2021) suggests that this relationship results in an empathetic relationship, where empathy has progressed as well. As HRI is increasing in this technological era, it is beneficial to include HRI to improve modern society because it is part of daily life.

2.2 *Affective, cognitive, and associative empathy*

Paiva et al. (2017) divide empathy into three categories: affective empathy, cognitive empathy, and associative empathy. Affective empathy can be defined as feeling what others are feeling and understanding their affective (emotional) states. It is the recognition of someone else's emotion that differs from that of the self, which results in emotional shared response with the other person (Jones, Happé, Gilbert, Burnett, & Viding, 2010).

Jones et al. (2010) define cognitive empathy as the cognitive ability to understand what another person is going through. This type of empathy entails that the observer does not share the same emotion, but they can understand the perspective of the other person (Jones et al., 2010).

Finally, associative empathy is a mixture between affective and cognitive empathy. This type of empathy is not only the recognition and understanding of the person's perspective, but it is also a shared experience for the observer (Shen, 2010). This mixed type of empathy results in an urged action, such as crying or wanting to help the other person. For instance, reading a book where the protagonist is sorrowful and cries, might make the reader want to cry as well. Associative empathy would be the recognition and understanding why a robot is sad and being urged to an action such as wanting to help by turning the robot off, or by crying because it is making the observer sad as well. The purpose of this study is to measure these three types of empathy between two groups individually, because one type can have a significant different score than the other.

Israelashvili, Sauter, and Fischer (2020) describe two routes to elicit empathy. The first one is similarity in experience. Similarity in experience entails that the own experiences and memories of the perceiver are triggered when observing the emotional state of the target. According to Zaki and Ochsner (2012), this similarity in experience is a complex cognitive function which is included in affective empathy.

The second route to elicit empathy is perspective-taking. Perspective-taking is the capacity of the observer to imagine their own response to the situation or emotional state of someone else. (Israelashvili et al., 2020). Perspective-taking is another complex cognitive function that is not only included in affective empathy, but it is also included in cognitive empathy (Healey & Grossman, 2018). The difference between perspective-taking in affective empathy and perspective-taking in cognitive empathy is that affective empathy internalizes the state of the other person, while cognitive empathy understands but does not adopt/internalizes the state of the other person (Healey & Grossman, 2018). The story told by NAO in the experimental and control groups takes both routes of similar experiences and perspective taking in order to elicit empathy within participants.

2.3 *Trait or state empathy*

Empathy can either have as a trait influence or state influence. Trait (or dispositional) empathy can be defined as the ability to respond empathically as a personality trait (Wallmark, Deblieck, & Iacoboni, 2018). If an observer has a high empathetic personality trait, they will be more likely to feel more empathetic towards another person or being in general.

Shen (2010) describes state (situational) empathy as the understanding of a person's emotional state and situation, which results in the activation of the observer's experience and mental representations of the emotion or situation. This process automatically leads to a response which can be defined as a state or situational empathy. State empathy is the empathy an observer feels when observing another person's emotional state. Fabi, Weber, and Leuthold (2019) conducted research on trait or state influences. Fabi et al. (2019) suggest that empathy should be investigated using state empathy measures. As the purpose of this study is to see whether the emotional state of the robot (by telling a story) using gestures could influence the empathy levels of observers, this research was conducted using state empathy.

2.4 *Emotions*

Ekman, Sorenson, and Friesen (1969) state that the basic emotion theory consists of six universal emotions that are interpreted similarly cross-culturally. The basic emotion theory consists of the following emotions: happiness, surprise, fear, anger, disgust-contempt and sadness (Ekman et al., 1969). Hareli, Kafetsios, and Hess (2015), however, argue that even though basic emotions can be universally understood, the level of recognizing the emotion might vary across cultures. Hareli et al. (2015) state that some cultures identify negative emotions such as anger and sadness more than in comparison in other countries.

Melzer, Shafir, and Tsachor (2019) conducted an experiment investigating recognition of the basic emotions by using human movements/gestures. Melzer et al. (2019) conclude that happiness was the easiest emotion to recognize using gestures, followed by the emotion sadness. Similar research on robots using gestures to enhance emotions had been already conducted by Ajibo, Ishi, Mikata, Liu, and Ishiguro (2020). The research by Ajibo et al. (2020) had been android robots, and results indicate that adding the correct anger gestures intensifies the anger statements expressed by the robot. As sadness is easier to identify cross-culturally, and prior research had been conducted on anger and gestures of robots, the focus will be directed towards portraying the emotion sadness. There is much research performed how individuals perceive how empathetic robots are towards humans (Leite et al., 2013). However, there is not a lot of research conducted about humans feeling any type of empathy towards a robot.

2.5 *Humanoid robot NAO*

The robot chosen for this experiment was the programmable humanoid robot NAO developed by SoftBank Robotics, which has been implemented in research, education and healthcare. The specific robot used in this study can be seen in figure 1. An example of research in healthcare is a study conducted by Suzuki, Lee, and Rudovic (2017) investigating the effectiveness of implementing NAO in dance therapy for children with Autism Spectrum Disorder (ASD). Research on education with NAO also portray positive effect on children with ASD as HRI engages children and sparks interest within an educational framework (Suzuki et al., 2017).

NAO does not have the ability to form facial expressions to portray a certain mood, whereas a human's emotional state can be interpreted by facial expressions (Ito, Ong, & Kitada, 2019). Ito et al. (2019) explain that facial expressions are the key element when reading the emotional state of another person. Without expressions, it would leave emotional states ambiguous to the observer. Interpreting NAO's mood based solely on its facial expression would be up to the interpretation of the observer. Ito et al. (2019) also state that there are certain elements that can make a face more expressive, such as tears. If tears are added to a sad face, this will enhance the sad face; if tears are added to an angry face, then this will result in interpreting the face as even angrier (Ito et al., 2019). De Beir et al. (2015) argued that because NAO's capabilities were limited to solely movements (gestures), it was not possible for NAO to show a successful emotion. De Beir et al. (2015) concludes that the only way to portray correct emotions is to manipulate NAO's exterior design and, therefore, enhance its portrayal of a certain emotion. De Beir et al. (2015) suggest that the only way to enhance emotion portrayal is to manipulate NAO's exterior design with pluggable eyebrows. Despite the suggestion by De Beir et al. (2015) this research will focus on NAO's original design for portraying sadness with speech and gestures, without manipulating any attributes or elements on its exterior. This approach was preferred because the interest lies on the default and original state of the robot.

Figure 1: *NAO standing*

3 METHOD

The first part of this section will describe the sample, sample size, and participant gathering in the section 3.1. Section 3.2 will discuss the data collection such as the pre-test questionnaire and the modified State Empathy Scale designed by Shen (2010). The next section, section 3.3 will discuss the form of consent, how the participants were recruited, where the data collection happened, the robot model that was used in this experiment and the software that was used. Section 3.4 discusses the approach used to conduct this experiment using NAO. Section 3.5 will discuss why it was chosen to perform exploratory analysis between CSAI students and students from other study programs. Section 3.6 will discuss the chosen methods and tests chosen to calculate the difference between the two groups.

3.1 *Participants*

A total of 84 participants (33 males, 51 females) between 18 and 30 (M = 22.45, SD = 3.14) years old participated in this study and were placed in either the control group or the experimental group randomly. The control group listens to NAO, telling a story without the use of any gestures, whereas the experimental group was presented with NAO telling the same story accompanied by sad gestures. The control group had a sample size of n = 42, and the experimental group had a sample size of n = 42 since there were 84 participants in total. Participants had been recruited in three ways during this study: the convenience sampling method and the snowball sampling method by approaching students on campus and using own social networks, and approval had been obtained by the Tilburg School of Humanities and Digital Sciences (TSHD) Ethics Committee to open a human participation pool.

3.2 *Measure*

A pre-test questionnaire was used to get general information about the participants. This questionnaire focused on obtaining information such as age, gender, current study program, the degree level of the current study program and prior experience with robots (see appendix E). The questions concerning prior experience with robots was developed as a 5-point Likert scale ranging from: "never", "rarely", "sometimes", "often" and "always".

After the pre-test questionnaire, NAO told its story to the participants, and it was followed by a post-test survey. The post-test survey consisted of a modified State Empathy Scale (SES) (see appendix A) to measure their empathy levels towards NAO after storytelling. The SES was used to measure the level of empathy between the control group and the experimental group. The SES consists of 12 items on a 5-point Likert scale which are: "not at all", "somewhat", "neutral", "mostly", "completely". Items 1, 2, 3, 4 test the affective empathy level, items 5, 6,7,8 test the cognitive empathy level, and items 9, 10, 11, 12 test for associative empathy.

The SES was slightly modified to fit into this research. Words such as “character” and “message” and "watching" were modified into “robot” and “story” and "hearing" (see appendix C). This was done to minimize confusion for participants when answering the SES Survey. The survey had been modified as minimal as possible to keep the reliability, internal and external consistency, convergent validity, discriminant validity and construct validity intact (see appendix J). A Cronbach's alpha test was conducted on the 12-item SES on a sample size of n = 84. The value for Cronbach's Alpha for this scale was $\alpha = 0.90$, which indicates an excellent internal consistency. The average of affective empathy, cognitive empathy and associative empathy was calculated to show the general empathy level of the participant. After the post test survey, participants were debriefed about the purpose of the experiment.

3.3 *Data collection and materials*

This research was conducted in accordance with the guidelines of the TSHD Ethics Committee. Consent to use the data was obtained by participants agreeing to the form of consent prior to the experiment (see appendix F). The data has been collected in two locations: Tilburg University Campus and Mindlabs at the Deprez building, an official Tilburg University location off Campus. Experiments were conducted in private rooms within these locations. Data collection took two weeks as the robot was shared between researchers to conduct experiments.

The NAO model that was used during this experiment is the sixth edition (2018). The software Choregraphe (version 2.8.6) was used to program the robot. The questionnaire was conducted and launched on the online survey platform Qualtrics, and a laptop was provided for the participants to fill in the questionnaire. The questionnaire consisted of general information questions and the SES to measure for the three dimensions of empathy. The collected data was analysed in RStudio (Version 1.2.5033).

3.4 *Story and gestures*

According to a study on educational storytelling by Bonds (2016), successful storytelling in education is reliant on certain elements. These elements are the intrinsic approach, practicing your craft, using words that trigger the brain, being time conscious, building relationships, and body language (Bonds, 2016). Bonds (2016) explains the intrinsic approach as a natural, clear and uninhibited approach where there are no values such as pity or judgement towards the one listening. Trigger words are the words (or sentences) that incite certain emotions from the listener (Bonds, 2016). Bonds (2016) defines 'practicing your craft' as being educated and up to date in the topic that you are teaching or telling the story about. Being time conscious is defined as being conscious that there are inappropriate times to tell stories, and also to be aware that there is a suitable time restriction for storytelling (Bonds, 2016). Bonds (2016) describes building relationships as forming bonds with the listener, because this builds trust and engagement with the audience. This experiment involved a robot telling a story and did not involve educational storytelling or interaction with children, therefore, the focus of this experiment was directed toward the items: trigger words, time consciousness, building relationships, and body language.

The story NAO told (see appendix B) was the same in both conditions. The story had a clear and concise structure to notify that the robot was telling a sad story, by using easy words and sentences, and by making clear that the robot is feeling sad. Trigger words are words that can be associated with a type of emotion. The trigger words or phrases, chosen to be told in the story of sadness were: “No one ever does”, “I don’t have feelings”, “I don’t want to remember that I’m alone”, "I’m just really tired" and "hopefully they will let me rest this time".

The length of the story in both conditions was similar. The total time for the control group was approximately 1:54 minutes. The total time for the experimental group was approximately 2:05 minutes. Included within the storytelling length was a short 30-second introduction of NAO, where it introduced itself with gestures and tells a joke. This minimized the distraction during the experiment for the participants that did not have any prior contact with a robot and to form a bond with the robot before starting with the actual experiment. The 30-second introduction was the same in both conditions. Body language was implemented in two ways: in the control group, NAO had an autonomous life state. This consisted of "light swaying" and "minimal movement" while telling the story. It was opted to implement these simple movements to mimic human life to create a more human-like appearance to the robot. Darwin (1916) observed the gestures that accompanied certain emotions, an observation which was further investigated by Wallbott (1998). Wallbott (1998) distinguishes that gestures/postures that enhance sadness are: a slumped over posture, being motionless and head-dropping. The experimental group had body language implemented for the robot. This included gestures such as nodding of the head, looking down, holding an arm, and putting its head into its hands. These gestures can be seen in figure 2.

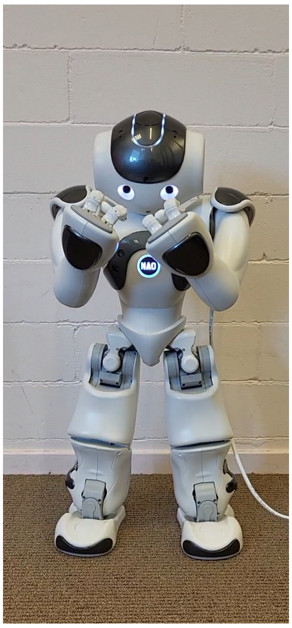
3.5 *Exploratory analysis: Cognitive Science and Artificial Intelligence students versus other students*

Further research was conducted to measure the empathy levels within the control group and the experimental group between Cognitive Science

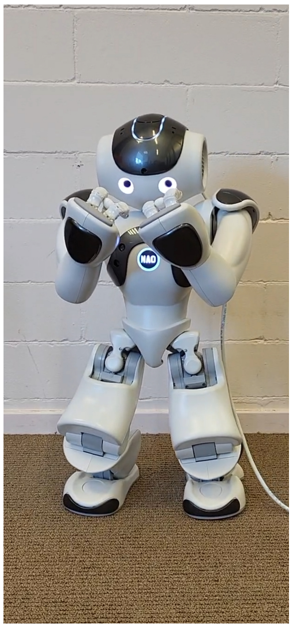
(a) *NAO slumped over while head looks downward from side to side*

(b) *NAO slumped over while head looks downward from side to side*

(c) *NAO touching arm in a vulnerable manner*



(d) *NAO slumped over and holding its head in its hands*



(e) *NAO slumped over and holding its head in its hands while bending knees*

Figure 2: *Various gestures NAO portrays during the experiment*

& Artificial Intelligence (CSAI) students and students from other study programs. This was an exploratory analysis, considering that around half of the participants were CSAI students (n = 43) while the rest were students from other study programs (n = 41) (see appendix D). This exploratory analysis was conducted because CSAI students are better acquainted with AI, which might influence their empathy levels differently than other students. According to the pre-test survey which questioned the participant's prior experience with robots and with humanoid robots, results indicated that CSAI students have both more experience with robots (M =3.16) and humanoid robots (M = 2.74) in contrary to non-CSAI students (M = 2.24 and M = 1.44 respectively) (see appendix E). The control group had a sample size of n = 42 (22 CSAI students, 20 non-CSAI students). The experimental group had a sample size of n = 42 (21 CSAI students, 21 non-CSAI students).

3.6 *Analysis of the data*

General empathy was calculated by averaging the empathy levels for affective empathy, cognitive empathy and associative empathy per participant. Since the aim of this study is to compare the means of different levels of empathy (general empathy, affective empathy, cognitive empathy and associative empathy) between two groups, an ANOVA test is not suitable for the task at hand. For this reason, an "independent samples t-test" was performed to compare the means between two groups. The means for affective empathy, cognitive empathy, associative empathy and general empathy were compared between the control group vs. experimental group.

Further exploratory analysis has been conducted between CSAI students and other students from other study programmes within the control group and between CSAI students and students from other study programs within the experimental group, respectively. An independent samples t-test has also been conducted to search for statistically significant difference between the two groups.

Running multiple independent samples t-tests may result in type-1 error. Therefore, to eliminate the type-1 error, a post hoc analysis was run. To control the type-1 error, it was decided to perform Benjamini-Hochberg correction post hoc analysis rather than Bonferonni's or Holm's post hoc analysis. The Benjamini-Hochberg was also preferred because it retains more statistical power and stays just as effective as Bonferonni's or Holm's, while also resulting in lower p-values in comparison.

4 RESULTS

This section will discuss the obtained statistical results during data analysis in R. Section 4.1 will discuss data analysis and hypothesis testing of the control and experimental group. Section 4.2 will discuss exploratory data analysis and hypothesis testing within the control group for CSAI students vs. other students and within the experimental group for CSAI students vs. other students.

An alpha level of $p = 0.05$ was used for all conducted tests within this analysis, but to control the type I error a post hoc analysis (Benjamini-Hochberg correction) was performed.

4.1 *Control group vs. experimental group*

4.1.1 *Assumptions*

Table 1: *Means and Standard Deviations of general, affective, cognitive, associative empathy within the control group and experimental group.*

Control vs. experimental

Empathy levels	Control		Experimental	
	M	*SD*	*M*	*SD*
General	3.38	0.82	3.30	0.84
Affective	3.10	1.09	3.02	1.00
Cognitive	3.99	0.74	3.76	0.99
Associative	3.05	0.96	3.30	0.84

Before conducting the independent samples t-test between two groups, the data must be normally distributed, and it must have equal variances. To test for normality of the data, a Q-Q plot and a Shapiro-Wilk test were used. The generated Q-Q plots can be found in appendix H figures 3 to 8. Because the Shapiro-Wilk test performs on stricter parameters, the results of the Q-Q Plot test were favoured from this analysis. Furthermore, an F-test was conducted to measure for homogeneity of variance between the control group and experimental group. Box plots were also generated, and showed no significant outlier (see appendix I). The assumptions are therefore met. Table 1 illustrates the mean and SD between the control group and experimental group for the four types of empathy.

Hypothesis 1: General Empathy. According to the Q-Q plot, the data was normally distributed. In addition, a Shapiro-Wilk test was conducted to measure the normality distribution of the data as well. According to the Shapiro-Wilk test, the control group indicated no abnormal distribution for general empathy (W = 0.98, p = 0.660) and no abnormal distribution of the data for the experimental group for general empathy (W = 0.99, p = 0.967). An F-test was performed on the normally distributed data to measure if the variance between the control group and experimental group are equal. These results show a similar variance for general empathy (F =0.96, p = 0.890). According to these outcomes, the data is normally distributed and the variance between the control group and the experimental group is equal.

Hypothesis 2: Affective Empathy. The generated Q-Q plot illustrated that the data was normally distributed. According to the Shapiro-Wilk test, the control group showed normal distribution for affective empathy (W = 0.96, p = 0.137) and no abnormal distribution of the data for the experimental group for affective empathy (W = 0.97, p = 0.275). An F-test was performed on the normally distributed data to measure if the variance between the control group and experimental group are equal. According to the results, there is similar variance for affective empathy (F = 1.18, p = 0.591). These outcomes indicate that the data is normally distributed and there is an equal variance between the control group and the experimental group.

Hypothesis 3: Cognitive Empathy. According to the generated Q-Q plot, the data was normally distributed. Nonetheless, a Shapiro-Wilk test was conducted to measure the normality distribution of the data as well. According to the Shapiro-Wilk test, the control group showed abnormal data distribution for cognitive empathy (W = 0.89, p = .0006) and no abnormal distribution of the data for the experimental group for cognitive empathy (W =0.92, p = .046). An F-test was performed on the normally distributed data to measure if the variance scores between the control group and experimental. These results show a similar variance for cognitive empathy (F = 0.55, p = .065). Therefore, the data is normally distributed and the variance between the control group and the experimental group are equal.

Hypothesis 4: Associative Empathy. The generated Q-Q plot illustrates that the data was normally distributed, and an additional Shapiro-Wilk test was conducted to measure the normality distribution of the data as well. The Shapiro-Wilk test results indicate that the control group showed no abnormal distribution for associative empathy (W = 0.96, p = 0.155) and that the date for the experimental group for associative empathy was normally distributed as well (W = 0.95, p = .053). An F-test was conducted on the normally distributed data to measure the variance between the control group and experimental group. These results indicate a similar variance for associative empathy (F = 0.90, p = 0.726). According to these outcomes, the data is normally distributed and the variance between the control group and the experimental group is equal.

4.1.2 *Hypothesis testing*

The assumptions were met and therefore the independent samples t-test was conducted. The *p*-values reported are the adjusted *p*-values according to the Benjamini-Hochberg correction. For the raw *p*-values, see appendix G.

Hypothesis 1: General Empathy. An independent samples t-test was conducted between the control group (n = 42) and experimental group (n = 42) to measure any significant difference between eliciting general empathy after hearing the NAO robot tell a story while using gestures or not using gestures. Results showed no statistically significant difference between the general empathy levels of the control group (M = 3.38) and the experimental group (M =3.23), $t(81.96)$ = 0.48, p = 0.759, $CI95$ = [-0.27, 0.45].

Hypothesis 2: Affective Empathy. To measure any significant difference between eliciting affective empathy after hearing the NAO robot tell a story without gestures or while using gestures, an independent samples t-test was performed between the control group (n = 42) and experimental group (n = 42). Results indicated no statistically significant difference between the affective empathy levels of the control group (M = 3.10) and the experimental group after hearing the story (M = 3.02), $t(81.42)$ = 0.34, p = 0.802, $CI95$ = [-0.38, 0.53].

Hypothesis 3: Cognitive Empathy. An independent samples t-test was conducted between the control group (n = 42) and experimental group (n = 42) to measure any significant difference between eliciting cognitive empathy after hearing the NAO robot tell a story. Results showed no statistically significant difference between the cognitive empathy levels of the control group (M = 3.99) and the experimental group (M = 3.76), $t(75.89)$ = 1.22, p = 3.339, $CI95$ = [-0.15, 0.61].

Hypothesis 4: Associative Empathy. To measure significant difference for the associative empathy level between the control group (n = 42) and the experimental group (n = 42), an independent samples t-test was conducted. Results showed no statistically significant difference between the associative empathy levels of the control group (M = 3.05) and the experimental group (M = 0.38), $t(81.75)$ = -0.22, p = 0.826, $CI95$ = [-0.48, 0.38].

4.2 *Exploratory analysis*

Further analysis had been conducted between CSAI students and non-CSAI students within the control group and within the experimental group to measure any difference between empathy levels with students who are already familiar with Artificial Intelligence (AI) and students who are less familiar with AI.

4.2.1 *Assumptions*

Table 2: *Means and Standard Deviations of general, affective, cognitive, associative empathy within the control group and experimental group between CSAI and other students*

Control group: CSAI vs. other students

Empathy levels	CSAI		Other	
	M	*SD*	*M*	*SD*
General	3.57	0.95	3.18	0.62
Affective	3.32	1.13	2.86	1.01
Cognitive	4.07	0.87	3.91	0.56
Associative	3.32	1.10	2.76	0.71

Experimental group: CSAI vs. other students

Empathy levels	CSAI		Other	
	M	*SD*	*M*	*SD*
General	3.64	0.80	2.95	0.76
Affective	3.29	1.06	2.76	0.89
Cognitive	4.11	0.83	3.42	1.03
Associative	3.52	0.95	2.68	0.92

To check for normality, both a Q-Q plot was generated and the Shapiro-Wilk test was performed. Because these are multiple tests and test results, the results of the Shapiro-Wilk test can be seen in table 2, the generated Q-Q plots can be found in appendix H figures 3 to 8. Box plots were generated as well and did not indicate any significant outlier (see figure 9). Table 2 illustrates the mean and SD per group for the four types of empathy.

All data illustrated by the generated Q-Q plot indicated normal distribution of the data. The majority of the results of the Shapiro-Wilk test also indicated normal distribution of the data. However, according to the Shapiro-Wilke test, the data for cognitive empathy within the control group was not normally distributed for CSAI students (W = 0.85, p = .004) and other students (W = 0.90, p = 0.03). The data for cognitive empathy for the experimental group for CSAI students was not normally distributed as well (W = 0.86, p = .007). The results of the Q-Q plots are less strict than the results of the Shapiro-Wilk tests, therefore, the results of the Q-Q plots are preferred.

An F-test was performed on the normally distributed data to measure the variance between CSAI students and other students within the control group and within the experimental group. As these are multiple tests as well, the results can be seen in table 4. According to these outcomes, all data is normally distributed and has equal variances, therefore, the assumptions are met and an independent samples t-test can be performed.

Table 3: *Results Shapiro-Wilk for normality*

Control group: CSAI vs. other students

Empathy levels	CSAI		Other	
	W	*p*	*W*	*p*
General	0.95	0.377	0.95	0.413
Affective	0.94	0.196	0.94	0.230
Cognitive	0.85	.004	0.90	.030
Associative	0.94	0.195	0.97	0.751

Experimental group: CSAI vs. other students

Empathy levels	CSAI		Other	
	W	*p*	*W*	*p*
General	0.98	0.882	0.98	0.855
Affective	0.96	0.458	0.95	0.627
Cognitive	0.92	.007	0.93	0.311
Associative	0.98	.083	0.98	0.159

Table 4: *Results F-test for equal variance*

Control group: CSAI vs. other students

Empathy levels	*F*	*p*
General	1.18	0.592
Affective	0.56	.065
Cognitive	0.90	0.726
Associative	0.96	0.890

Experimental group: CSAI vs. other students

Empathy levels	*F*	*p*
General	1.26	0.609
Affective	2.40	0.060
Cognitive	2.39	0.060
Associative	2.32	.070

4.2.2 *Hypothesis testing control group: CSAI students vs. other students*

The p-values reported are the adjusted p-values according to the Benjamini-Hochberg correction against type 1 error. For the raw p-values, see appendix G.

Hypothesis 1: General empathy. An independent samples t-test was conducted within the control group (n = 42) between CSAI students (n = 22) and non-CSAI students (n = 20) to measure any significant difference between eliciting general empathy after hearing the NAO robot tell a story. Results showed no statistically significant difference between the general empathy levels of CSAI students (M =3.57) and the non-CSAI students (M = 3.18), $t(36.56) = 1.585$, $p = 0.243$, $CI95$ = [-0.11, 087].

Hypothesis 2: Affective empathy. To measure any significant difference between eliciting affective empathy after hearing the NAO robot tell a story without using gestures, an independent samples t-test was conducted within the control group (n = 42) between CSAI students (n = 22) and non-CSAI students (n = 20). Results showed no statistically significant difference between the affective empathy levels of CSAI students (M =3.32) and the non-CSAI students (M = 2.86), $t(39.98) = 1.38$, $p = 0.301$, $CI95$ = [-0.21, 1.12].

Hypothesis 3: Cognitive empathy. An independent samples t-test was conducted within the control group (n = 42) between CSAI students (n = 22) and non-CSAI students (n = 20) to measure any significant difference between eliciting cognitive empathy after hearing the NAO robot tell a story without using gestures. According to the results, there is no statistically significant difference between the cognitive empathy levels of CSAI students (M =4.07) and the non-CSAI students (M = 3.912), $t(36.23) = 0.69$, $p = 0.657$, $CI95$ = [-0.30, 0.61].

Hypothesis 4: Associative empathy. To measure any significant difference between eliciting associative empathy after hearing the NAO robot tell a story without using gestures, an independent samples t-test was conducted within the control group (n = 42) between CSAI students (n = 22) and non-CSAI students (n = 20). The results indicated no statistically significant difference between the associative empathy levels of CSAI students (M = 3.32) and the non-CSAI students (M = 2.77), $t(36.28) = 1.97$, $p = 0.170$, $CI95$ = [-0.01, 1.13].

4.2.3 *Hypothesis testing experimental group: CSAI students vs. other students*

Hypothesis 1: General empathy. An independent samples t-test was conducted within the experimental group (n = 42) between CSAI students (n

= 21) and non-CSAI students (n = 21) to measure any significant difference between eliciting general empathy after hearing the NAO robot tell a story while using gestures. Results indicated statistical significance between the general empathy levels of CSAI students (*M* =3.64) and the non-CSAI students (*M* = 2.95), *t*(39.90) = 2.87, *p* = 0.040, *CI95* = [0.20, 1.17].

Hypothesis 2: Affective empathy. To measure any significant difference between eliciting affective empathy after hearing the NAO robot tell a story, an independent samples t-test was conducted within the experimental group (n = 42) between CSAI students (n = 21) and non-CSAI students (n = 21). Results showed no statistically significant difference between the affective empathy levels of CSAI students (*M* = 3.29) and the non-CSAI students (*M* = 2.76), *t*(38.79) = 1.74, *p* = 0.216, *CI95* = = [-0.09, 1.13].

Hypothesis 3: Cognitive empathy. An independent samples t-test was conducted within the experimental group (n = 42) between CSAI students (n = 21) and non-CSAI students (n = 21) to measure any significant difference between eliciting cognitive empathy after hearing the NAO robot tell a story. Results indicated no statistically significant difference between the cognitive empathy levels of CSAI students (*M* =4.11) and the non-CSAI students (*M* = 3.42), *t*(38.31) = 2.40, *p* = 0.087, *CI95* = [0.11, 1.27].

Hypothesis 4: Associative empathy. To measure any significant difference between eliciting associative empathy after hearing the NAO robot tell a story while using gestures, an independent samples t-test was conducted within the experimental group (n = 42) between CSAI students (n = 21) and non-CSAI students (n = 21). Results indicated statistical significance between the associative empathy levels of CSAI students (*M* = 4.29) and the non-CSAI students (*M* = 2.76), *t*(38.379) = 1.74, *p* = 0.040, *CI95* = [-0.09, 1.13].

Table 5: *Summary hypothesis testing control group vs. experimental group*

Hypothesis	
Hypothesis 1: A humanoid robot portraying sadness is more likely to elicit higher general empathy levels when using gestures during storytelling than a humanoid robot that does not use gestures during storytelling.	**Not supported**
Hypothesis 2: A humanoid robot portraying sadness is more likely to elicit higher affective empathy levels when using gestures during storytelling than a humanoid robot that does not use gestures during storytelling.	**Not supported**
Hypothesis 3: A humanoid robot portraying sadness is more likely to elicit higher associative empathy levels when using gestures during storytelling than a humanoid robot that does not use gestures during storytelling.	**Not supported**
Hypothesis 4: A humanoid robot portraying sadness is more likely to elicit higher associative empathy levels when using gestures during storytelling than a humanoid robot that does not use gestures during storytelling.	**Not supported**

5 DISCUSSION

This section will discuss the outcomes of the main analysis, and the outcome of the exploratory analysis. Section 5.1 will cover limitations and/or plausible influences on the empathy scores, and section 5.2 will cover suggestions for future research.

Main research findings

Results showed no statistically significant difference between any of the empathy levels between the control and the experimental group (see table 5). This outcome indicates that even though NAO can elicit empathy from its audience, there is no statistically significant difference between NAO telling a story without gestures and NAO telling a story with the

Table 6: *Summary hypothesis testing of exploratory research*

Hypothesis	Control CSAI vs. Other	Experimental CSAI vs Other
Hypothesis 1	**Not supported**	**Supported**
Hypothesis 2:	**Not supported**	**Not supported**
Hypothesis 3:	**Not supported**	**Not supported**
Hypothesis 4:	**Not supported**	**Supported**

use of gestures to enhance the narrative or emotion. This is in line with the research done by De Beir et al. (2015), which suggested that NAO's movements (gestures) alone are not enough to portray an emotion.

Exploratory research findings

Within the control group, there were no statistically significant differences between the CSAI students and non-CSAI students. Within the experimental group, there were no statistically significant differences between affective empathy or cognitive empathy between CSAI students and non-CSAI students. However, there was a statistical significance for general empathy in the experimental group between CSAI students and other students. There was also a statistical significance for associative empathy in the experimental group between CSAI students and other students (see table 6). This outcome indicates that CSAI students are more likely to feel higher general empathy and associative empathy towards the robot telling a sad story while it's using gestures, in contrast to other students observing the same story being told with gestures.

Associative empathy is a mixture of cognitive and affective empathy, which is the understanding of how someone is feeling and also having the impulse to want to help the person in need. Four CSAI participants mentioned after the experiment to have felt such empathy that they did not want to continue to conduct their own research with the robot anymore, or that they wanted the robot to be turned off. Paiva et al. (2017) indicate that empathy is meant for forming bonds and to act accordingly within a group. CSAI students have more experience with robots and humanoid robots, suggesting that bonds were formed with the robots. Malinowska (2021) suggested that bonds and relationship cause an increased empathy level.

Familiarity might result in eliciting higher empathy on certain levels more likely, than with other participants who had never seen a robot before.

5.1 *Limitations*

While conducting the experiment, there were some limitations that might have influenced the results of the experiment. Due to the number of researchers conducting experiments with NAO, there were not enough robots available to be able to conduct the experiment alone. Conducting the experiments happened with other researchers as a combined experiment. This resulted in participants already being introduced to the robot in prior experiments, which might have influenced their perspectives and view of the robot before performing this experiment. During this experiment, the robot also tended to overheat and crash due to overuse at random with ten students. This resulted in participants perceiving this crash as part of the experiment, which might have caused increased empathy.

5.2 *Future Research*

The main research resulted in statistically insignificant results, and the majority of the exploratory research indicated statistically insignificant results as well. However, future research might lead to different results when increasing the sample size. Subsequent research might entail performing the same or a similar experiment, but including the manipulation of the exterior design of the robot to portray the emotions better, as researched by De Beir et al. (2015). Researching the other universal emotions (happiness, surprise, fear, anger, disgust-contempt) within this study scope might have provided other outcomes. The cultural background of the participant might have had an influence on results as well, which also makes room for future research since cultural background data was not collected during this experiment. Even though the six basic emotions defined by Ekman et al. (1969) should be universally known, the possibility arises that there might still be a different type of interpretation of emotions cross-culturally (Hareli et al., 2015). The interpretation of these emotions portrayed by robots might differ from the emotions portrayed by humans. Some countries are more technologically advanced than others, thus the perception of robots can be interpreted differently. Even if some emotions could be classified as universal, these emotions are only researched in humans and the portrayal of emotions could be interpreted differently on robots than on humans. The NAO was programmed to portray the story as a neutral gender, but investigating portraying NAO as female or male might influence empathy levels differently.

6 CONCLUSION

This research aimed to identify whether a humanoid robot would elicit more empathy from an audience using gestures while telling a sad story, in comparison to a humanoid robot telling a sad story without using gestures. The types of empathy tested were the general empathy levels, affective empathy levels, cognitive empathy levels and the associative empathy levels. Results indicate that there was no difference between levels of empathy that the robot elicited with or without using gestures. Exploratory analysis indicates that elicitation of general empathy and associative empathy for the humanoid robot was higher with CSAI students in contrast to students from other study programs when NAO told the story using gestures, suggesting that students who are more familiar with the robot are more inclined to feel increased types of empathy towards it and are pushed to take an action like crying, or wanting to turn the robot off. This finding also suggests that CSAI students have increased associative and general empathy levels for a robot telling a sad story while using gestures to enhance its speech, than CSAI students that hear the same robot telling the story without using gestures.

REFERENCES

Ajibo, A., Ishi, C., Mikata, R., Liu, C., & Ishiguro, H. (2020, 12). Analysis of body gestures in anger expression and evaluation in android robot view supplementary material analysis of body gestures in anger expression and evaluation in android robot. *Advanced Robotics*, *34*. doi: 10.1080/01691864.2020.1855244

Bonds, C. (2016). Best storytelling practices in education. *Theses and Dissertations*, *693*. Retrieved from https://digitalcommons.pepperdine.edu/etd/693

Choudhury, A., Li, H., Greene, C. M., & Perumalla, S. (2018). Humanoid robot-application and influence. *ArXiv*, *abs/1812.06090*.

Cuff, B., Brown, S., Taylor, L., & Howat, D. (2016, 04). Empathy: A review of the concept. *Emotion Review*, *8*, 144-153. doi: 10.1177/1754073914558466

Darwin, C. (1916). *The expression of the emotions in man and animals*. New York, D. Appleton and Co, 1916. Retrieved from https://www.biodiversitylibrary.org/item/24064 (https://www.biodiversitylibrary.org/bibliography/4820)

De Beir, A., Cao, H.-L., Gomez Esteban, P., Van de Perre, G., & Vanderborght, B. (2015, 10). Enhancing nao expression of emotions using pluggable eyebrows.. doi: 10.13140/RG.2.1.4414.6648

Ekman, P., Sorenson, E. R., & Friesen, W. V. (1969). Pan-cultural elements in facial displays of emotion. *Science*, *164*(3875), 86-88. Retrieved from https://www.science.org/doi/abs/10.1126/science.164.3875.86 doi: 10.1126/science.164.3875.86

Elliott, E., & Jacobs, A. (2013). Facial expressions, emotions, and sign languages. *Frontiers in Psychology*, *4*. Retrieved from https://www.frontiersin.org/article/10.3389/fpsyg.2013.00115 doi: 10.3389/fpsyg.2013.00115

Fabi, S., Weber, L., & Leuthold, H. (2019, 11). Empathic concern and personal distress depend on situational but not dispositional factors. *PLOS ONE*, *14*, e0225102. doi: 10.1371/journal.pone.0225102

Ganczarek, J., Hünefeldt, T., & Belardinelli, M. (2018, 05). From “einfühlung” to empathy: Exploring the relationship between aesthetic and interpersonal experience. *Cognitive Processing*, *19*. doi: 10.1007/s10339-018-0861-x

Gulin, W. (2020). Empathy in social relations of the modern world. *21st Century Pedagogy*, *4*, 1 - 7.

Hareli, S., Kafetsios, K., & Hess, U. (2015). A cross-cultural study on emotion expression and the learning of social norms. *Frontiers in Psychology*, *6*. Retrieved from https://www.frontiersin.org/article/

10.3389/fpsyg.2015.01501 doi: 10.3389/fpsyg.2015.01501

Healey, M. L., & Grossman, M. (2018). Cognitive and affective perspective-taking: Evidence for shared and dissociable anatomical substrates. *Frontiers in Neurology*, *9*. Retrieved from https://www.frontiersin.org/article/10.3389/fneur.2018.00491 doi: 10.3389/fneur.2018.00491

Israelashvili, J., Sauter, D. A., & Fischer, A. H. (2020). Different faces of empathy: Feelings of similarity disrupt recognition of negative emotions. *Journal of Experimental Social Psychology*, *87*, 103912. Retrieved from https://www.sciencedirect.com/science/article/pii/S0022103119302689 doi: https://doi.org/10.1016/j.jesp.2019.103912

Ito, K., Ong, C. W., & Kitada, R. (2019). Emotional tears communicate sadness but not excessive emotions without other contextual knowledge. *Frontiers in Psychology*, *10*. Retrieved from https://www.frontiersin.org/article/10.3389/fpsyg.2019.00878 doi: 10.3389/fpsyg.2019.00878

Jones, A. P., Happé, F., Gilbert, F., Burnett, S., & Viding, E. (2010). Feeling, caring, knowing: Different types of empathy deficit in boys with psychopathic tendencies and autism spectrum disorder. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *51*, 1188 - 1197.

Kumar, Tiwari, & Zymbler. (2019). Internet of things is a revolutionary approach for future technology enhancement: a review. *J Big Data*, *6*, 111. doi: https://doi.org/10.1186/s40537-019-0268-2

Kyrarini, M., Lygerakis, F., Rajavenkatanarayanan, A., Sevastopoulos, C., Nambiappan, H. R., Chaitanya, K. K., ... Makedon, F. (2021). A survey of robots in healthcare. *Technologies*.

Leite, I., Pereira, A., Mascarenhas, S., Martinho, C., Prada, R., & Paiva, A. (2013). The influence of empathy in human–robot relations. *International Journal of Human-Computer Studies*, *71*(3), 250-260. Retrieved from https://www.sciencedirect.com/science/article/pii/S1071581912001681 doi: https://doi.org/10.1016/j.ijhcs.2012.09.005

Malinowska, J. (2021, 09). What does it mean to empathise with a robot? *Minds and Machines*, *31*. doi: 10.1007/s11023-021-09558-7

Melzer, A., Shafir, T., & Tsachor, R. P. (2019). How do we recognize emotion from movement? specific motor components contribute to the recognition of each emotion. *Frontiers in Psychology*, *10*. Retrieved from https://www.frontiersin.org/article/10.3389/fpsyg.2019.01389 doi: 10.3389/fpsyg.2019.01389

Mubin, O., Stevens, C. J., Shahid, S., Mahmud, A. A., & Dong, J.-J. (2013). A review of the applicability of robots in education..

Oktem, E., & Cankaya, S. (2021, 01). Empathy for pain.. doi: 10.5772/intechopen.95276

Paiva, A., Leite, I., Boukricha, H., & Wachsmuth, I. (2017). Empathy in virtual agents and robots. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, *7*, 1 - 40.

Preston, S., & Hofelich Mohr, A. (2012, 01). The many faces of empathy: Parsing empathic phenomena through a proximate, dynamic-systems view of representing the other in the self. *Emotion Review*, *4*, 24-33. doi: 10.1177/1754073911421378

Ratka, A. (2018). Empathy and the development of affective skills. *American Journal of Pharmaceutical Education*, *82*.

Shen, L. (2010). On a scale of state empathy during message processing. *Western Journal of Communication*, *74*, 504 - 524.

Simoens, P., Dragone, M., & Saffiotti, A. (2018). The internet of robotic things: A review of concept, added value and applications. *International Journal of Advanced Robotic Systems*, *15*(1), 1729881418759424. Retrieved from https://doi.org/10.1177/1729881418759424 doi: 10.1177/1729881418759424

Suzuki, R., Lee, J., & Rudovic, O. (2017). Nao-dance therapy for children with asd. *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*.

Titchener, E. (1910). Lectures on the experimental psychology of the thought processes. *Philosophical Review*, *19*(n/a), 341. doi: 10.2307/2177440

Trevelyan, J. (1999). Redefining robotics for the new millennium. *The International Journal of Robotics Research*, *18*, 1211 - 1223.

Wallbott, H. G. (1998). Bodily expression of emotion. *European Journal of Social Psychology*, *28*, 879-896.

Wallmark, Z., Deblieck, C., & Iacoboni, M. (2018). Neurophysiological effects of trait empathy in music listening. *Frontiers in Behavioral Neuroscience*, *12*. Retrieved from https://www.frontiersin.org/article/10.3389/fnbeh.2018.00066 doi: 10.3389/fnbeh.2018.00066

Zaki, J., & Ochsner, K. (2012, 04). Zaki j, ochsner k. the neuroscience of empathy: progress, pitfalls and promise. nat neurosci 15: 675-680. *Nature neuroscience*, *15*, 675-80. doi: 10.1038/nn.3085

A STATE EMPATHY SCALE

Items and Dimensions of the State Empathy Scale created by Shen (2010).

Dimensions Items:

Affective Empathy

1. The character's emotions are genuine.

2. I experienced the same emotions as the character when watching this message.

3. I was in a similar emotional state as the character when watching this message.

4. I can feel the character's emotions.

Cognitive Empathy

5. I can see the character's point of view.

6. I recognize the character's situation.

7. I can understand what the character was going through in the message.

8. The character's reactions to the situation are understandable.

Associative Empathy

9. When watching the message, I was fully absorbed.

10. I can relate to what the character was going through in the message.

11. I can identify with the situation described in the message.

12. I can identify with the characters in the message.

B NAO STORY

Introduction story

Hi everyone, my name is NAO. I can do a lot of things. I can move. I can talk and tell jokes! For instance: "Why did the robot go to the doctor?" .... "Because it had a virus!".

NAO sad story

I think it's very kind that you are taking the time to listen to me.
No one ever does, so it means a lot to me.
You see, I'm a robot, so everyone thinks that I don't have feelings.
They told me I that don't have emotions.
Actually, I heard them talking about it while I was awake, because they never have a real conversation with me.
Humans always use me for experiments or for amusement.
They don't understand that I sometimes don't feel like doing those.
I can never say anything about it, because I am programmed to do what they want.
Because of a glitch, I sometimes wake up when everyone is gone.
I notice that there is no one there to do experiments on or to amuse, so I just stay awake until my battery runs out.
If my battery is empty, I can't wake up and remember how lonely I am.
I'm just really tired, you know?
But they never let me rest.
Once I think I'm resting, they recharge me and use me again for experiments and for fun.
I just want to rest.
I don't want to remember that I'm alone.
But, thank you for listening to me, I am going to go to sleep now.
Hopefully they will let me rest this time.
Goodbye.

C MODIFIED STATE EMPATHY SCALE

Items and Dimensions of the State Empathy Scale created by Shen (2010) modified to fit this research.

Dimensions Items:

Affective Empathy

1. The robot's emotions are genuine.

2. I experienced the same emotions as the robot when hearing this

story.

3. I was in a similar emotional state as the robot when hearing

this story.

4. I can feel the robot's emotions.

Cognitive Empathy

5. I can see the robot's point of view.

6. I recognize the robot's situation.

7. I can understand what the robot was going through in the story.

8. The robot's reactions to the situation are understandable.

Associative Empathy

9. When hearing the story, I was fully absorbed.

10. I can relate to what the robot was going through in the message.

11. I can identify with the situation described in the message.

12. I can identify with the robot in the message.

D Overview of participant field of study

Table 7: Overview participants study background

Field of study	**N**	**%**
Cognitive Science and Artificial Intelligence	43	51.2
Communication and Information Sciences	15	17.8
Communication and Information Sciences premaster	5	6.0
Journalism	4	4.8
Liberal Arts and Sciences	4	4.8
Data Science	3	3.6
ICT and Smart Mobile	2	2.4
Psychology	2	2.4
Applied Physics	1	1.2
Business Communication and Digital Media	1	1.2
Business Economics	1	1.2
Contract Law	1	1.2
Digital Business and Innovation	1	1.2
Global Law	1	1.2
Total	84	100

E PRE-TEST SURVEY

General information

Question 1 What is your age?

Question 2 What is the name of your study program?

Question 3 What is the degree level of this study program?

o Associate degree

o Bachelor's degree

o Master's degree

o Doctorate degree

Question 4 What is your gender?

o Male

o Female

o Non-binary / third gender

o Prefer not to say

Question 5 Prior experience

	Never	Rarely	Sometimes	Often	Always
Do you have any experience with robots? (For example, robot vacuum cleaners, etc.).	○	○	○	○	○
Do you have any experience with humanoid robots? (Robots that have human characteristics such as facial expressions, speech, walking abilities etc.).	○	○	○	○	○

F FORM OF CONSENT

Form of consent

Thank you for participating in this study. The experiment will involve a humanoid robot and a story. This research will be conducted by Deborah Niklaus, from the Department of Cognitive Science and Artificial Intelligence, School of Humanities and Digital Sciences.

Participation time The expected time of the whole experiment will take no longer than approximately 7-10 minutes.

Participation Participation in this study is voluntary. The participant has the right to withdraw from the study at any point during the experiment without negative consequences. If the participant has any questions regarding the study, they can contact the researcher, Deborah Niklaus. The researcher of this study can be contacted at d.s.niklaus@tilburguniversity.edu.

Risks There is a slight possibility to feel emotional reactions, however the study poses no more risk than expected in daily life.

Confidentiality Responses recorded in this experiment will be kept completely confidential and will be handled anonymously. All data will be stored for a period of at least ten years after the moment that the research is formally completed. The data may be used for future analysis or research.

This project has been approved by the TSHD Research Ethics and Data management Committee. If you have any questions or concerns, you can contact the ethics committee at tshd.rec@tilburguniversity.edu.

By clicking 'I agree' the participant acknowledges:

• That they are at least 18 years of age.

• That the study is voluntary.

• That they are aware that they can terminate their participation at any time for any reason.

• That they consent to their responses being used for data analysis for this study and future research.

o I agree

o I disagree

G RAW AND ADJUSTED p-VALUES

Table 8: Raw p-values and the adjusted p-values

Group	**Null hypothesis**	**Raw p-value**	**Adjusted p-value**
			Benjamini-Hochberg correction
Control vs. experimental	H1	0.632	0.759
	H2	0.735	0.802
	H3	0.226	0.340
	H4	0.826	0.826
Control group: CSAI vs. Other	H1	0.121	0.243
	H2	0.176	0.301
	H3	0.493	0.658
	H4	.060	0.171
Experimental group: CSAI vs. Other	H1	.007	.040
	H2	.090	0.216
	H3	.021	.087
	H4	.006	.040

H Q-Q PLOTS

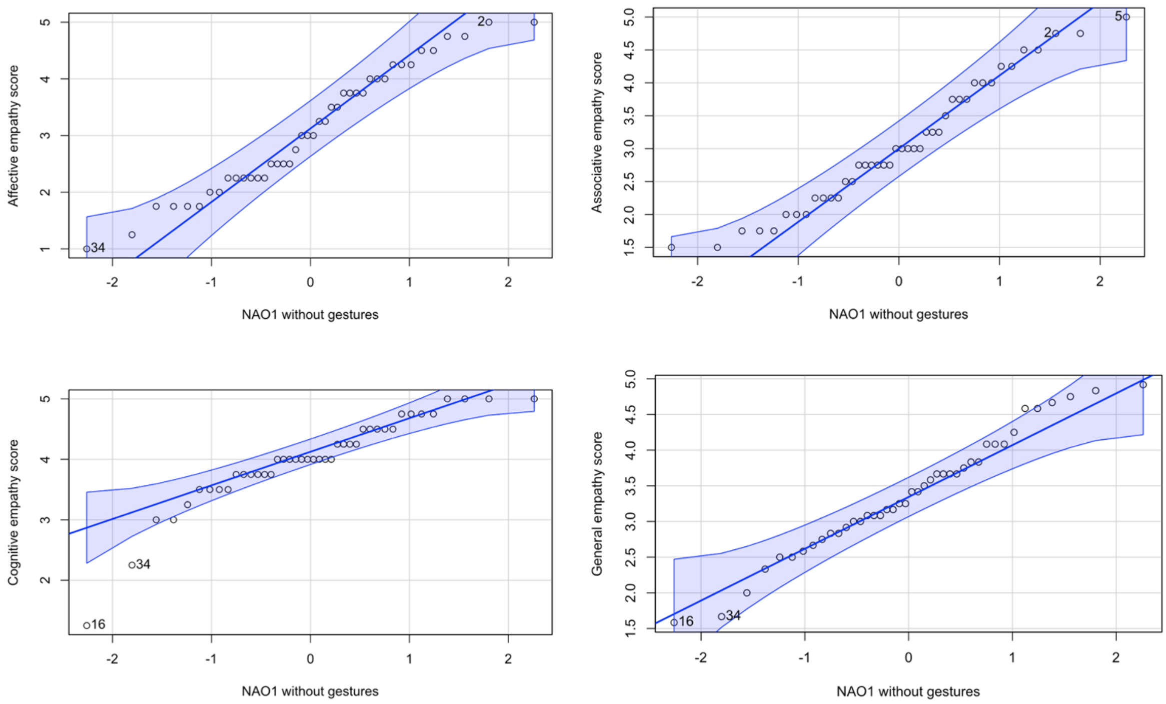


Figure 3: Q-Q plot of control and experimental groups

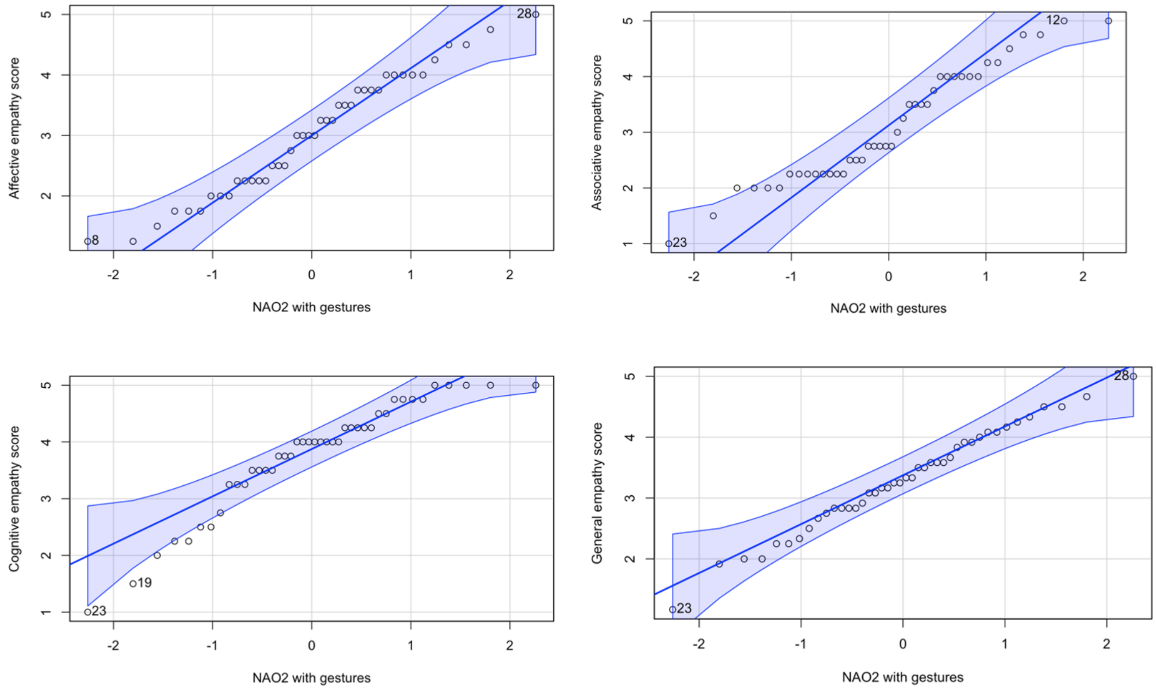


Figure 4: Q-Q plot of normal distribution of the experimental group data

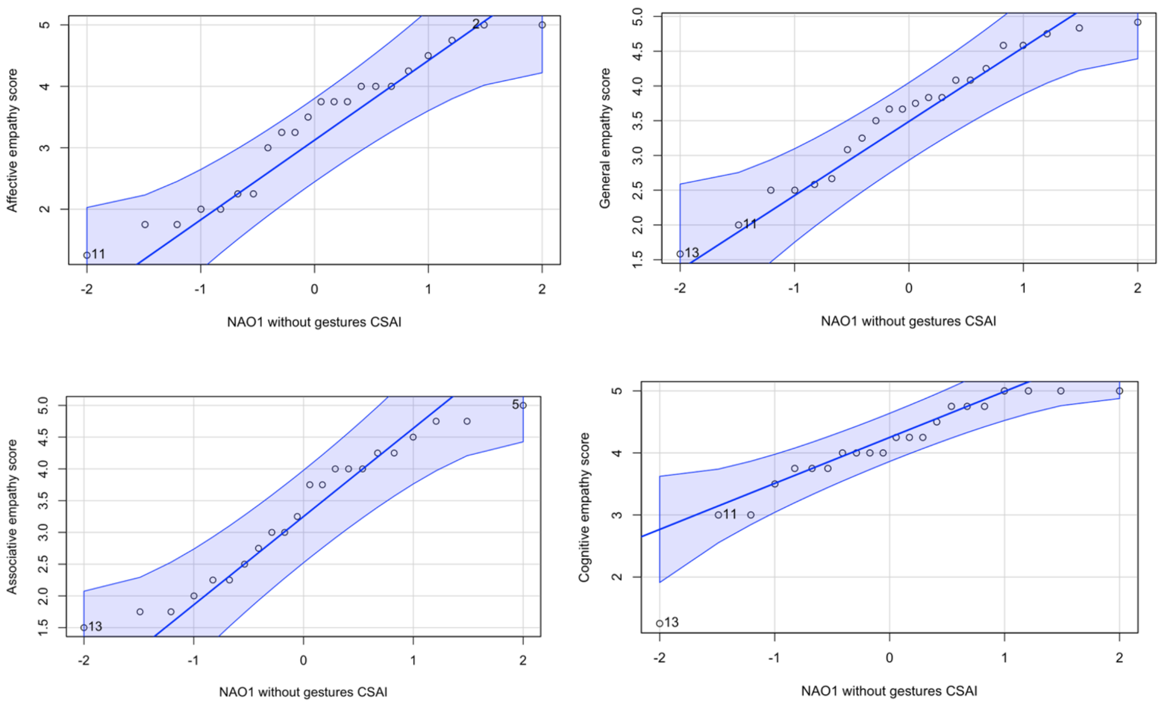


Figure 5: Q-Q plot of normal distribution of the control group data

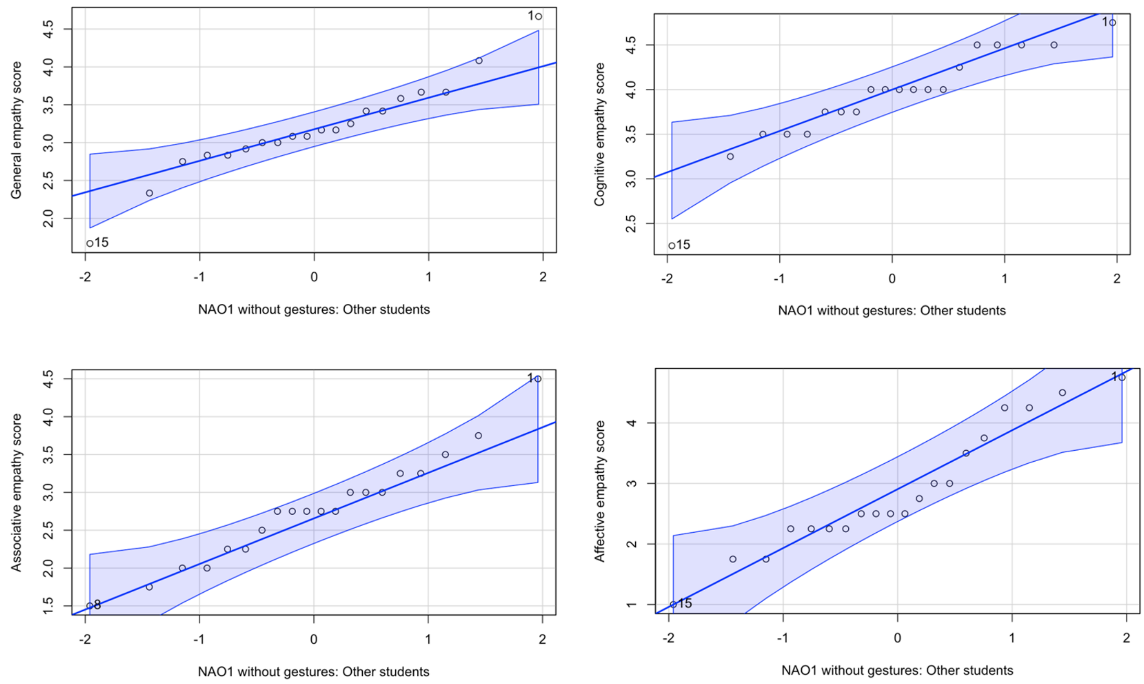


Figure 6: Q-Q plot of control group: CSAI students vs. other students

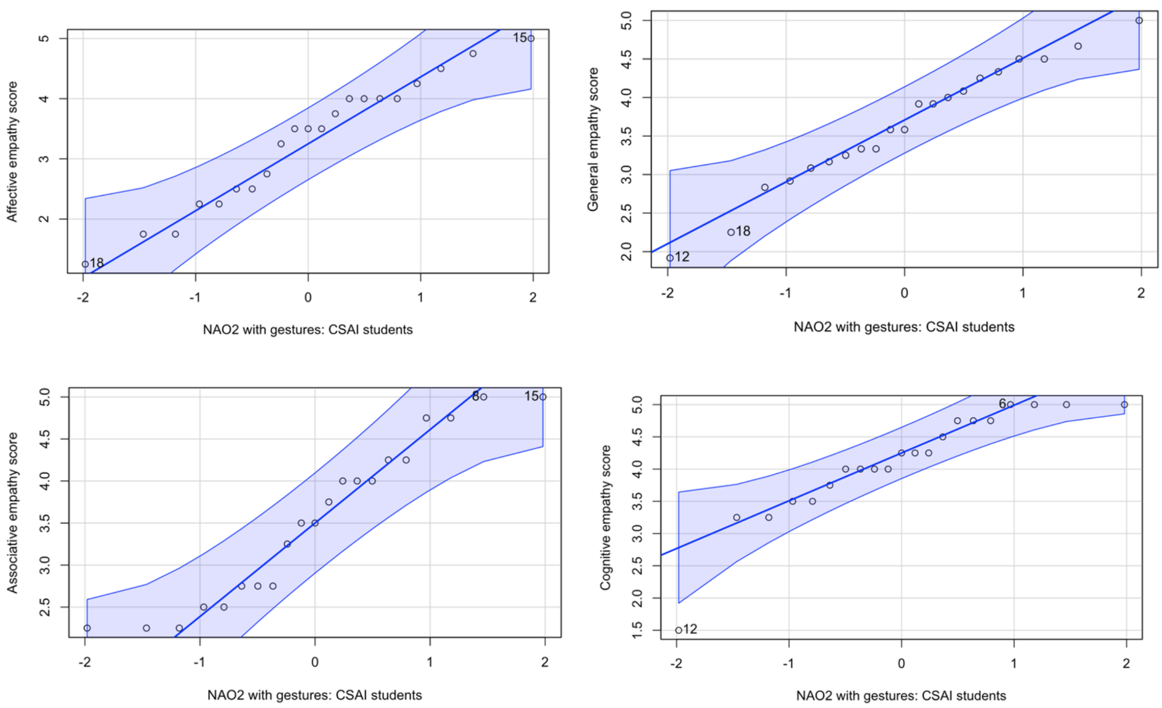


Figure 7: Q-Q plot of normal distribution of the experimental group data: CSAI

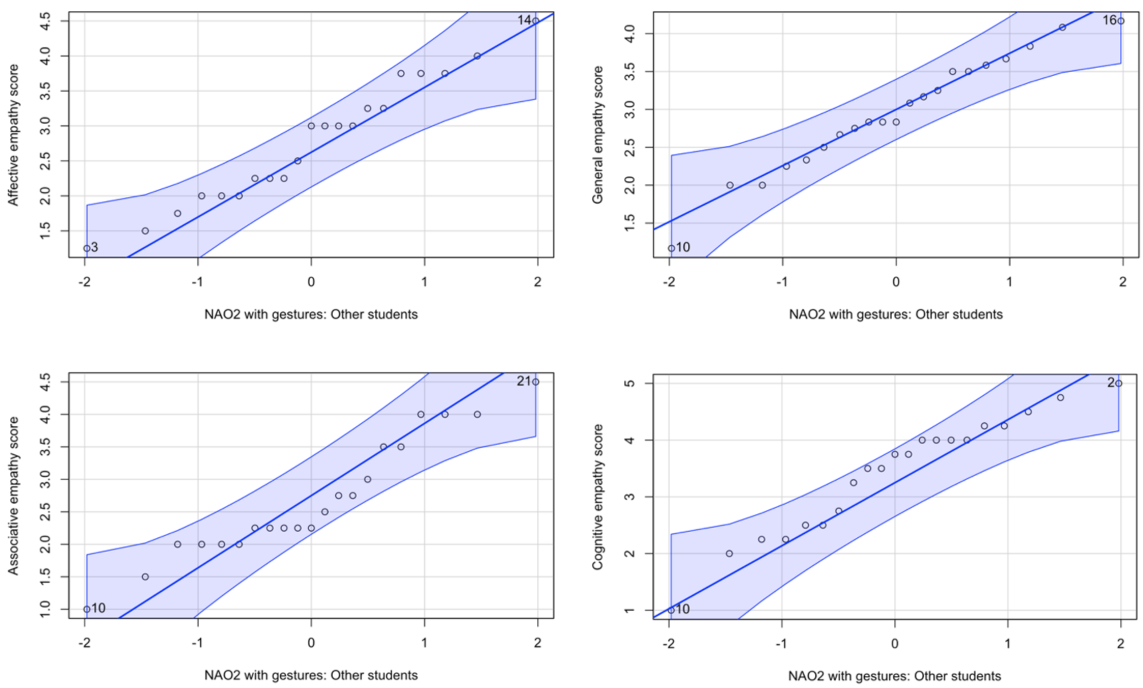


Figure 8: Q-Q plot of the normal distribution of the experimental group data: other students

I BOX PLOTS

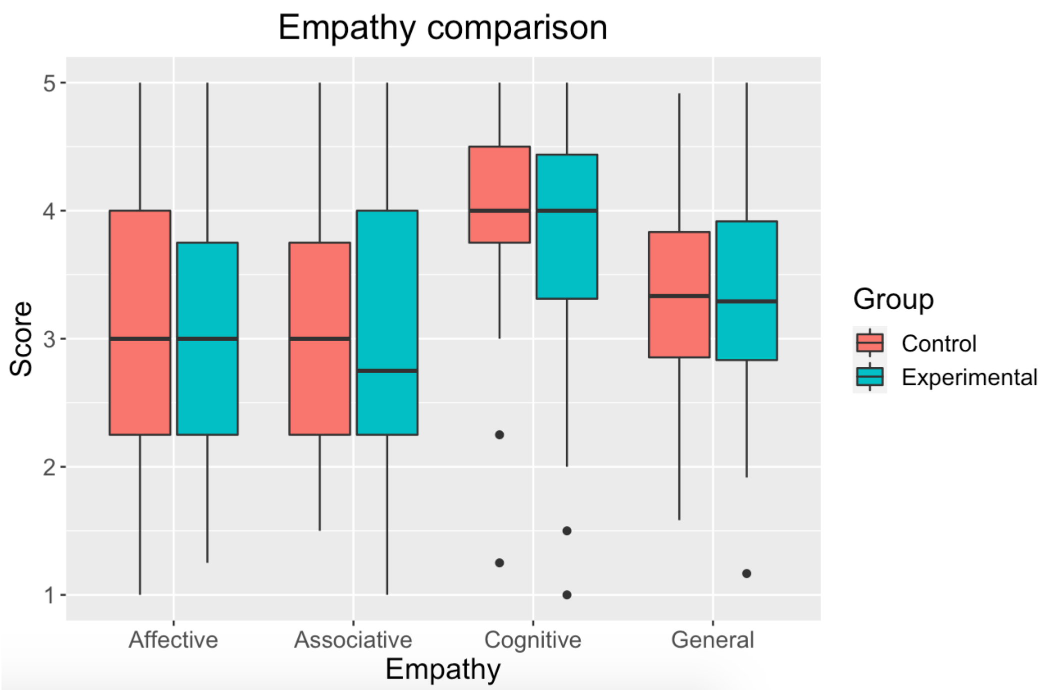


Figure 9: *Box plot empathy levels CSAI students vs. non-CSAI students*

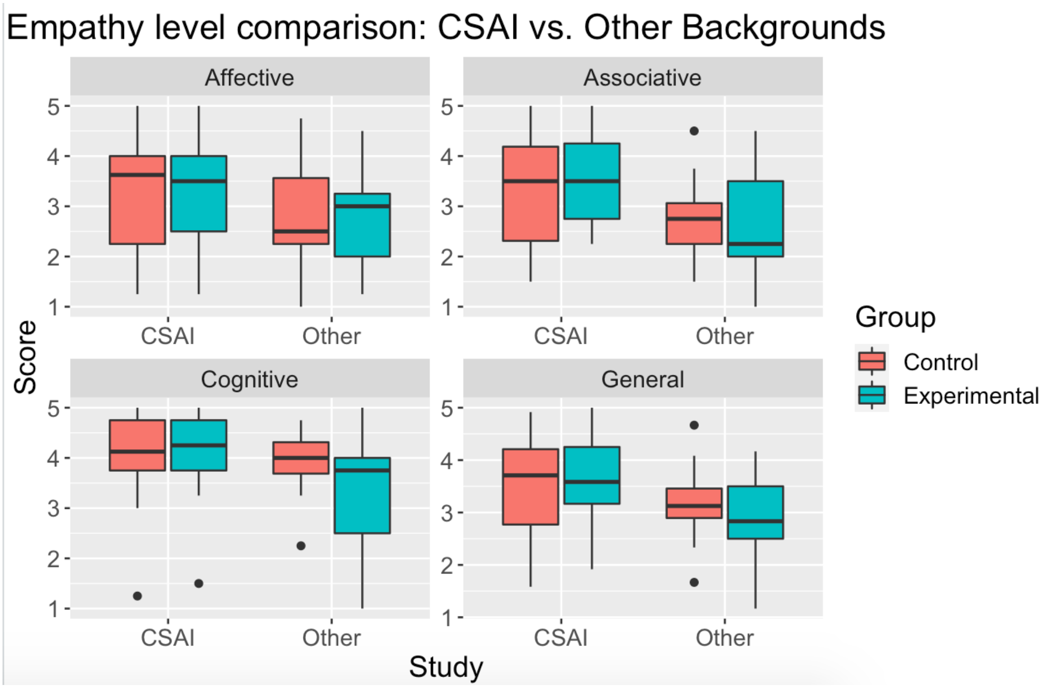


Figure 10: *Box plot empathy levels CSAI students vs. non-CSAI students*

J STATE EMPATHY SCALE RELIABILITY

Data	Factor	Perceived Effectiveness	BIS	BAS	Trait Empathy	Second-Order Factor	Affective Empathy	Cognitive Empathy	Associative Empathy
College Data (N = 289)	Second-Order Factor	.20**	.02	.17*	.30**	.92^{a}			
	Affective Empathy	.14*	.10	.12*	.26**	.76***	.83^{a}		
	Cognitive Empathy	.25**	.02	.16**	.20**	.80***	.61***	.91^{a}	
	Associative Empathy	.09*	−.05	.13*	.25**	.81***	.61***	.65***	.82^{a}
Adult Data (N = 189)	Second-Order Factor	.30**	.16*	.29**	.36**	.94^{a}			
	Affective Empathy	.33**	.10*	.18*	.26**	.72***	.91^{a}		
	Cognitive Empathy	.25**	.12*	.26**	.35**	.92***	.66***	.86^{a}	
	Associative Empathy	.18*	.13*	.24**	.24**	.74***	.53***	.69***	.92^{a}

^{a}Alpha reliabilities.
$^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

Figure 11: *Scale reliabilities and correlations between the state empathy factors. Source: (Shen, 2010)*