Virtual Character Design and its potential to foster Empathy, Immersion, and 21st Century Learning Skills in Video Games and Virtual Reality Simulations

Sierra Rativa, Alexandra

Publication date:
2022

Document Version
Publisher's PDF, also known as Version of record

Virtual Character Design

— and —

its potential to foster Empathy, Immersion, and 21st Century Learning Skills in Video Games and Virtual Reality Simulations

ALEXANDRA SIERRA RATIVA

Virtual Character Design and its potential to foster Empathy, Immersion, and 21st Century Learning Skills in Video Games and Virtual Reality Simulations

Alexandra Sierra Rativa

Virtual Character Design and its potential to foster Empathy, Immersion, and 21st Century Learning Skills in Video Games and Virtual Reality Simulations
Alexandra Sierra Rativa
PhD Thesis
Tilburg University, 2022

Financial Support was received from the Administrative Department of Science, Technology and Innovation, also known as Colciencias (MinCiencias) from Colombia, Grant Doctorado Exterior N. 679.

SIKS Dissertation Series No. 2022-22.
The research reported in this thesis has been carried out under the auspices of SIKS, the Dutch Research School for Information and Knowledge Systems.

ISBN: 9789464690088
Cover design: Mercedes Benjaminse | ProefschriftMaken
Inside design: Bregje Jaspers | ProefschriftOntwerp.nl
Print: ProefschriftMaken | www.proefschriftmaken.nl

Virtual Character Design and its potential to foster Empathy, Immersion, and 21st Century Learning Skills in Video Games and Virtual Reality Simulations

Proefschrift ter verkrijging van de
graad van doctor aan Tilburg University
op gezag van de rector magnificus,
prof. dr. W.B.H.J. van de Donk, in het openbaar te
verdedigen ten overstaan van een door het college
voor promoties aangewezen commissie in de Aula
van de Universiteit op vrijdag 16 september 2022
om 10.00 uur door Alexandra Sierra Rativa,
geboren te Bogota, Colombia.

door

Alexandra Sierra Rativa

geboren op 20 december 1985 te Bogota, Colombia.

Promotors

Prof. dr. M. M. Louwerse, Tilburg University.
Dr. M. Postma, Tilburg University.

Co-promotor

Prof. dr. M. M. van Zaanen, South African Centre for Digital Language Resources, North-West University, South Africa.

Promotiecommissie

Prof. dr. G. van Noort, University of Amsterdam.
Prof. dr. ir. P.H.M. Spronck, Tilburg University.
Prof. dr. A. A. Navarro Newball, Pontificia Universidad Javeriana.
Dr. D. Pimentel, University of Oregon.

"Don't wish it was easier, wish you were better.
Don't wish for less problems, wish for more skills.
Don't wish for less challenge, wish for more wisdom."
Jim Rohn

"No desees que sea más fácil las cosas, desea que fueras mejor.
No desees menos problemas, desea más habilidades.
No desees menos desafíos, desea más sabiduría".
Jim Rohn

Abstract

Despite significant technological innovations and breakthroughs, one of the main obstacles to modern education is that many students and teachers continue to find it uninspiring. The temporary changes introduced by the global COVID-19 pandemic brought about greater uptake of online education, as well as numerous new approaches and further research. Still, teachers reported that many students failed to diligently attend online classes. Students were said to be distracted and to lack focus, and their learning attainment dropped dramatically compared to traditional attendance in a physical classroom (Guo, 2020; Serhan, 2020). According to Weldon et al. (2021), the major problems of online education were accessibility to technology and the quality of educational material in existence for such technology. Papanastasiou et al. (2019) considered that immersive technologies could improve long-term memory retention, content comprehension, collaboration skills, individual differences among learners, and unsuccessful classroom integration. Immersive technologies can quickly foster students' 21st Century Learning Skills (e.g., communication, collaboration, critical thinking, and creativity) due to their potential to allow the manifestation, creation, manipulation, navigation and interaction with virtual objects and virtual environments in a free, flexible and immersive way. They are able to interact with other people worldwide, as well as with non-player characters (NPCs), which are artificial intelligence agents that are able to simulate human-like intelligence within a virtual learning environment, and also stimulate their communicative, collaborative and social skills. Inspired by the insight that virtual characters can potentially be empathized with and readily accepted by students, this dissertation's general research question seeks to explain the means by which virtual characters are effectively visually designed in order to support communication and collaborative skills within immersive technology. This thesis starts by presenting the research question, central concepts, and thesis overview in **Chapter 1**.

We analyze the instruments deemed valid for the measurement of 21st Century Learning Skills (e.g., collaboration, communication, creativity, and critical thinking) in **Chapter 2** to solve our **RQ1** *"How can we measure 21st Century Learning Skills (creativity, collaboration, communication, and critical thinking skills) with existing tests and what is the reliability and validity of these tests?"*. The results showed that no tests existed by which to evaluate attainments in collaborative skills, unlike personality and aptitude tests (in the form of psychological evaluative tests). In the articles that formed the basis

for this review, no tests were found on collaborative skills and critical thinking tests that had been used at kindergarten and primary school levels, and communicative skills tests had not been developed or used for university or kindergarten levels. We assumed that a possible reason behind this is the difficulty of assessing individual academic performance in terms of collaboration and communication; abilities that involve an interactive experience between multiple persons. However, a degree of uncertainty remains as to whether collaboration and communication skills can be stimulated by a virtual character, in primary and secondary education and whether there are differences across cultures.

Virtual animals have been used in both entertainment films for children's audiences (e.g., Finding Nemo, Ice Age, Ratatouille, A Bug's Life, and Duck Tales) and video games (e.g., Pokémon Crash Bandicoot, and Nintendo's Little Friends: Dogs and Cats) and also for educational purposes (e.g., 4D+Utopia 360 Animal Zoo, Digoo 3D AR Flashcards, and Avatar Zoo) and in advertising (e.g., Animal-based logos such as Lacoste, Puma, ING, Swarovski, and Penguin Books). Human-Computer Interaction studies by Krekhov et al. (2019b) explained that simulating a virtual body with non-human characteristics is not an easy task. One of the main obstacles is that three traits differ significantly between animal and human bodies: shape, skeleton, and posture. For instance, a spider differs in terms of its skeletal characteristics, body shape and the number of limbs compared to a human body. Another example is a lion or dog, which has a similar skeleton to a human, but differs in its posture when walking or moving. However, the impact of virtual animals on human-computer interaction has remained understudied. To date, most major studies have instead focused on virtual humans (Schwind et al., 2018a). This dissertation offers a deeper insight into the influence of the design of virtual characters applied to Human-Computer Interaction Studies.

Next, to answer **RQ2** *"Does a virtual animal (e.g., panda) also adhere to the uncanny valley effect?"*, we investigate possible ways to design the appearance of a virtual animal, taking into account the effects described by Uncanny Valley Theory through images and videos in **Chapter 3**. The results indicated the manifestation of the uncanny valley effect for virtual animals for robot and zombie animal appearances (morbid images), especially concerning familiarity and commonality, for both still and moving images. The uncanny valley effect was not found in the participant-based ranking for naturalness and attractiveness, except in the expert-based ranking. No uncanny valley effect was observed for animateness and interestingness. Following this research, the concept of naturalness and artificiality in visual animal design took on a significant role in this thesis. However, no controlled studies have compared differences in the natural appearance and facial expression of animal characters in video games.

Subsequently, in **Chapter 4,** we employed a computer game to investigate the possible effects of the visual appearance of the virtual animal on empathy and immersion, in order to answer **RQ3** *"Can a virtual animal's visual appearance influence the level of empathy of users?"* and **RQ4** *"Can a virtual animal's visual appearance influence the level of immersion of users?"*. The result of this study helped clarify the role of interaction between the expressiveness and artificial/natural appearance of a virtual character on self-reported situational empathy and immersion in players. We found that research has so far failed to determine whether a player being able to embody an animal yields the same or comparable results to those that we found for our video game. A virtual reality simulation was used to analyze the effects of a virtual animal's appearance on pain perception, empathy, immersion, embodiment, and animal conservation, and this is presented in **Chapter 5** in order to support our already achieved answers to **RQ3** and **RQ4**, and to answer **RQ5** *"Can a virtual animal's visual appearance influence the level of perceived pain of users?"* and **RQ6** *"Can a virtual animal's visual appearance influence the level of embodiment of users?"*. The single most striking observation to emerge was that natural and artificial appearance has an effect on immersion and perceived pain in players but lacks an effect on empathy, avatar embodiment and animal conservation tendencies in virtual reality.

Following these studies, in **Chapter 6**, we explored the effect of the virtual animals' different natural and artificial appearances in their roles as virtual instructors on students' knowledge recall and perception of characteristics, to solve our **RQ7** *"How can different versions of virtual animals be used as virtual tutors in video instruction, which may have different effects on affective and cognitive outcomes, depending on their visual appearance?"*. This research was developed through an experiment conducted in two secondary public schools with children between 11 and 17 years of age in Bogota, Colombia. The results showed that a virtual instructor with an artificial appearance (e.g., robotic appearance and animal appearance) elicited negative reactions compared to human instructors. The uncanny valley effect of the virtual animals was confirmed for virtual instructors, and it had a vital effect on the knowledge recall of students. The instruments used or found in the experiments of this dissertation are presented in **Appendices A, B, C, E**, and **F**. Further, a future research framework that could be developed for a virtual animal that could possess intelligence attributes that would stimulate 21st Century Learning Skills in students is described in **Appendix D. Appendix E** is described SIKS dissertations, and **Appendix H** explains the professional biography of the author of this dissertation.

Finally, a discussion and conclusions are presented in **Chapter 7**, which addresses our statement of the problem and subsequent research questions in detail. In the conclusion, we elaborate on our discovery of the fact that the visual design of the virtual character can have a significant effect on users' cognition, empathy and immersion perception. However, the visual appearance does not impact users' perception of virtual embodiment and situational empathy in certain circumstances. A key element is the effect the naturalness and artificiality of the visual design of the virtual character has upon students, especially when it has non-human form, as is the case with virtual animals. The contribution of this dissertation can be applied to reduce rejection or apathy reactions and increase a high level of immersion in non-human virtual characters. Likewise, we are looking to foster natural conservation toward animals and nature from virtual worlds. In the robotics field, new design possibilities for zoomorphic robots can be explored in online education. A natural progression of this work would be to analyze whether, and to what extent, utilizing artificial intelligence to develop authentic and believable conversational virtual characters can foster 21st Century Learning Skills in virtual learning environments.

Contents

Listing of Figures and Tables

Acknowledgments

** English

To my dear God, to myself for writing this dissertation (I thought it would be easier to undertake this doctorate, but I was wrong), to my promotor and supervisor Marie Postma who admired her intelligence, from whom I learned immensely, and had the opportunity to share the richness of our Colombian jungles. My co-promotor and supervisor Menno van Zaanen who taught me great values such as loyalty, organization, and unconditionally. Also, I admired him for their great love for music and gave me with his lovely family great support with my son in the Scouts for many years. Marie and Menno have a great integrity toward science, patience with my learning process, and their role as natural leaders in the Humanities field. To Max Louwerse (who has always inspired me with his virtual reality projects), to collaborative teacher Pieter Spronk (who has always supported me with the necessary resources for this research), to Maria Jose, Eva, and Karin for fascinating conversations, to my beautiful parents Rosalba and Alejandro, to my patient and wise sister Angie Geraldine, and my supportive boyfriend Peter, to my incredible friends Mariana (a colleague in our adventures and a great friend), Javiera, Sofia, Gabo, Victor, Laura, Keyttin, Anaika, Ivar, Lane, Yueqiao, Yaser, Julija, Thiago, Andres, Yaser, Gabriela, Irena, Cindy Carolina, Sindey, Agnes, Sabrina, Miruna, Laura, Charlie, Nitin, Gulbike, Nathalia, and Karo. Thanks also to Chrissy Cook and Andrew Lawler for help me with my academic writing. I greatly appreciate the efforts of the designers and programmers who have helped me to perform these experimental simulations, Esteban and Johnny. Finally, I want to say thanks to my wise and lovely son Jorge Alejandro for having patience and sometimes sacrificing our time together over these six years.

I dedicate this book to them, and I hope it will inspire Latin American children who want to be scientists or PhD students in the future at Tilburg University, and also students who consider abandoning their doctoral studies due to the challenge they represent. I was able to achieve this, despite being born in an impoverished neighborhood in Colombia, not learning English until the age of 27, living without my family in a place where the language (Dutch) is very different from Spanish, raising my son alone in the Netherlands throughout my six years of study, and surviving these six years with a Colombian Scholarship (Colciencias) to support me during my doctoral studies.

During my PhD thesis, I realized that it is my calling to create awareness and foster values in the virtual worlds for future generations. I therefore pursued this PhD to the

end as I could no longer consider the option of abandoning it; in such circumstances as mine, educating yourself is an excellent way to progress. I also discovered how beautiful science and research could be when I undertook my first experiment in the laboratory, and felt the emotions of being a real scientist in the field of humanities and technology. Moreover, my passion for this doctoral topic has sustained throughout, and I feel that I could continue researching it for many more years to come.

At the end of this process, this PhD dissertation was done with love and commitment. I hope that you enjoy reading this thesis.

With love,
Alexandra Sierra

** Español

A mi amado Dios, a mí misma por escribir esta disertación (yo pensaba que era más fácil hacer este doctorado, pero me equivoque), a mi promotora y supervisora Marie Postma quien admiró su inteligencia, de quien aprendí muchísimo y tuve la oportunidad de compartir la riqueza de nuestras selvas colombianas. Mi co-promotor y supervisor Menno van Zaanen quien me enseño grandes valores como la lealtad, la organización y la incondicionalidad. Además, lo admiré por su gran amor por la música y me brindó junto a su encantadora familia un gran apoyo con mi hijo en los Scouts durante muchos años. Marie y Menno tienen una gran integridad hacia la ciencia, paciencia con mi proceso de aprendizaje y su papel como líderes naturales en el campo de las humanidades. A Max Louwerse (quien siempre me ha inspirado con sus proyectos de realidad virtual), al profesor colaborador Pieter Spronk (él siempre me apoyó con los recursos necesarios para esta investigación), grandes conversaciones con María José, Eva, y Karin, a mis hermosos padres Rosalba y Alejandro, a mi paciente y sabia hermana Angie Geraldine, al gran apoyo de mi novio Peter, a mis increíbles amigos Mariana (Colega de aventuras y una gran amiga), Javiera, Sofia, Gabo, Victor, Laura, Keyttin, Anaika, Ivar, Lane, Yueqiao, Yaser, Julija, Thiago, Andres, Cindy Carolina, Sindey, Agnes, Sabrina, Miruna, Laura, Charlie, Nitin, Gulbike, Nathalia, Yaser, Gabriela, Irena, and Karo. Gracias a Chrissy Cook y Andrew Lawler por ayudarme con mi redacción académica. Agradezco mucho a los disenãdores y programadores que me ayudaron a hacer estas simulaciones experimentales, Esteban y Johnny. Finalmente, quiero agradecerle en especial a mi sabio y querido hijo Jorge Alejandro, gracias por tenerme paciencia y a veces sacrificar nuestro tiempo juntos en estos seis anõs.

A ellos les dedico este libro, y espero que sirva de inspiración a los niños latinoamericanos que quieren ser científicos o estudiantes de doctorado en el futuro en la Universidad de Tilburgo y a los estudiantes que desean abandonar sus estudios debido a que es un reto terminar un doctorado. Yo pude lograrlo a pesar de que: nací en un barrio pobre de Colombia, aprendí inglés a los 27 anõs, viví sin mi familia en un lugar donde el idioma (Holandes) es muy diferente al espanõl, crié sola a mi hijo en Holanda durante mis seis anõs de estudio y sobreviví estos seis años con una Beca Colombiana que me apoyó durante mis estudios de doctorado (Colciencias).

Durante mi tesis doctoral, descubrí que parte de mi misión es crear conciencia y fomentar valores en los mundos virtuales para las generaciones futuras. Entonces, obtuve este doctorado ya que no tuve opción de abandonarlo y educarse es una excelente manera de progresar en mis circunstancias. También descubrí lo hermosa que era la

ciencia y la investigación cuando hice mi primer experimento en el laboratorio. Me sentí como una verdadera científica en el campo de las humanidades y la technología. Además, me apasionaba el tema doctoral de esta tesis y podría seguir investigandola durante más anõs.

Al final de este proceso, esta tesis doctoral fue hecha con amor y compromiso. Espero que disfrutes leyendo esta tesis.

Con amor,
Alexandra Sierra

CHAPTER 1

Introduction

Introduction

During my master's studies, I discovered that teaching educational robotics through a traditional approach was useless in the classroom. By this traditional approach, in general, students received low grades in this course, and the class was perceived by students as extremely complex (i.e., robotics demands interdisciplinary knowledge and techniques such as electronics, programming, design, maths, physics, mechanical, electricity, and others), and their learning process required an extended period. I used a constructivist learning approach for my classes and showed the importance of educational robotics to foster 21st Century Skills in K-12 education in Colombia (Sierra, 2018). The 4Cs were stimulated in this research as: (1) Critical thinking skills were observed in the capacity of our students to solve a problem (i.e., problems in human body) actively and reflexively by applying the knowledge learned in class and online sources, (2) Creativity skills were detected when students produced new ideas in an inventive way through artistic and mechanical prototypes, (3) Communication skills were fostered during the writing and sharing process of students' ideas and reflections in their digital journals, and (4) Collaborative skills was observed in the student's responsibilities, decision-making processes, and social interaction in the group. Previous studies recognized the critical role played by 21st Century Learning Skills in educational robotics. For instance, Eguchi (2014) considered that during the programming of LEGO Mindstorms, students can develop collaboration and communication skills. Likewise, in the international competition called RoboCupJunior, students improved their 21st Century Skills by participating in advanced competitions and solving complex robotics projects (Eguchi, 2016). This previous master study contributed to improving our understanding of educational robotics and its potentiality in 21st Century Learning Skills in the classroom.

However, most studies in the field of educational robotics and their effect on 21st Century Learning Skills have only focused on qualitative or anecdotal data. The main weakness of this study was the absence of quantitative tests to evaluate the 21st Century Learning Skills during this research. The results were predominantly qualitative. The study was limited by the lack of information on available tests to measure 21st Century Skills and their reliability in secondary education. Other essential aspects were the technological tools' localization, accessibility, and interaction. The localization issue was determined when several students lost the small items of their robotic prototypes, and others could no longer use the robots. Likewise, the accessibility problem was observed by the lack of resources in the educational institution to make enough robotics materials available in the classroom, which could influence the results of this study. Notwithstanding these two limitations, the study suggests using virtual robots or virtual environments for teaching robotics. A natural progression of this master's study into doctoral research was to expand the diversity of robots in education using virtual animal robots, which could be perceived as friendly, funny and attractive user interfaces for children compared with human robots.

During my doctoral studies, the global pandemic in 2019 and 2020 accelerated the digitalization of education. Existing virtual learning environments and new communication tools were explored by teachers, students and parents alike, as a means by which to transplant traditional teaching methods into an online approach. Weldon et al. (2021) reported that significant benefits of this included increasing students' time by allowing them to learn at home, and increased openness and awareness of teachers to new software that could be implemented in their classes. However, students' performance, motivation and attendance were observed to drop significantly during this period (Guo, 2020; Serhan, 2020). The new challenges faced by virtual learning environments request significant improvements in their design in terms of quality and functionality. Recent trends in education have led to a proliferation of studies of digital environments, and recent evidence suggests that education ought to explore other ranges of technology (e.g., virtual reality, augmented reality, or mixed reality) that could prove more engaging and immersive for their students and teachers. Unfortunately, recent research in this domain has consistently shown that, in some instances, immersive technology is unable to efficiently promote knowledge gains. For example, studies developed on environmental topics have compared the effects of virtual reality simulations with those of traditional computer simulations, with an experiment conducted by Makransky et al. (2019) being an example of this. Their results showed that students had a higher presence sensation in the Virtual Reality condition but lower learning than in the traditional computer condition. Likewise, Parong and Mayer (2018, p. 1) explored college students who watched the biology lesson ("The Body VR: Journey Inside a Cell") in a virtual reality lesson compared with a desktop-based lesson had any effects on learning and motivation. The results showed that the Virtual Reality lesson did not have a significant impact on learning gain outcomes (post-test knowledge recall) compared with the desktop-based lesson. However, the Virtual Reality lesson did achieve higher scores for students' motivation engagement and interest. Surprisingly, an inefficient effect of virtual reality on learning outcomes in environmental education was noted compared with standard digital settings. However, there was no close examination of whether or not results could be influenced by levels of distraction inherent to being introduced to a Virtual Reality environment or the lack of quality of the specific simulation for the undertaking of the experiment. Whether or not the development of 21st Century Learning Skills in the classroom can be fostered by this immersive technology remains unclear. Data about the efficacy of immersive technology for the fostering of 21st Century Learning Skills are limited, but there is a widespread awareness of the fact that such technology can potentially stimulate these skills. Preliminary work by Papanastasiou et al. (2019) considered that Virtual Reality and Augmented Reality can affect students' critical thinking skills by allowing students to control their own learning through navigation and digital manipulation within a virtual learning space free from the barriers that time, location and distance impose upon both students and instructors. These virtual world experiences can positively impact the experience of learning due to their ability to promote an analysis of abstract and deep concepts and self-questioning with regard to them, which is very difficult in traditional education. Likewise, Virtual Reality and Augmented Reality can affect students' emotional skills (social empathy) and creativity by allowing them to create and develop their own objects in the virtual world. It can foster decision-making skills, novelty, flexibility, and personalization based on their personality, performance and personal interests. Finally, Papanastasiou et al. (2019, p. 431) explained that VR/AR can affect students' communication, collaboration and social skills through an avatar: students are able to communicate and collaborate with other people in a "realistic ambience" in their virtual learning environment. A particular concern with using this type of immersive technology in education is the current imperfections of visual displays, the quality of environmental simulation, avatar embodiments' movement, and type(s) of social interaction (Gerschutz et al., 2019; Papanastasiou et al., 2019; Porssut et al., 2022). Aesthetic and interaction problems with immersive technology and virtual learning environments can negatively impact or reduce their potential application in fostering 21st Century Learning Skills within students. This dissertation investigates what factors are necessary to design a virtual character that allows students to successfully communicate and collaborate (21st Century Learning Skills) with other people within a virtual learning environment. The following sections introduce the main concepts used in this thesis. Section 1.1 defines "21st Century Learning Skills" and their relationship with virtual learning environments. In Section 1.2, the concept of design for conventional and immersive technologies is described. In Section 1.3, which focuses on the design of virtual characters, the concept of the Uncanny Valley Effect is outlined, alongside its application to virtual animals, and an understanding of the definition of naturalism versus artificiality will be discussed in this research. In Section 1.3.1, the concept of empathy, immersion, perceived pain, and embodiment are introduced. The research questions are proposed in Section 1.5. Section 1.6 provides a full overview of the structure of this thesis. Section 1.7 describes the list of the publications that form the basis of this thesis, with Section 1.7.3 and Section 1.8 are presented additional information such as data sets created during the research described in this thesis.

1.1 21st Century Learning Skills and virtual learning environments

Recently, global societal changes have surprised us with new challenges relating to teaching and learning, such as collaboration, communication, informed decision-making, adaptability, and creative problem-solving skills that should be developed for this unpredictable world. Previously, steps had already been taken to teach using digital environments for learning outside of the classroom. However, the need was no more urgent than when the lockdown(s) produced by the COVID-19 pandemic started. The lockdown was part of large-scale social restrictions and corresponding isolation that were imposed upon teachers and students, forcing them to work (and study) from home. It exposed the fact that the digital tools for teaching and learning did not provide sufficient quality virtual learning environments to cover students' educational processes around the world. (Kurbakova et al., 2020; Torres Martín et al., 2021; Mundiri et al., 2021). Indeed, this situation showed us the importance of studying current technologies and the academic concepts that should be used to improve teaching processes, making them more effective and consistent with the social challenges of the 21st Century. For many years, education has worked on finding and testing more effective and simpler automated and digitized ways to teach skills and also to evaluate the applications in virtual learning environments (Schleicher, 2012). At the same time, according to Niemi et al. (2014), the relevance of developing 21st Century Skills, for instance, creativity, collaboration, critical thinking, problem-solving, decision making, communication and other skills related to ICT technologies, is growing. However, these authors suggest developing these skills is much harder. They consider that the advantages of technology-mediated learning environments include the fact that they are more flexible, faster, synchronic with digital content in various information sources, and that their level of interactivity is higher than traditional content. However, they also claim that the major challenge is promoting 21st Century Skills through virtual learning environments that can be used in informal and formal learning situations. Although extensive research has been carried out on 21st Century Skills, an insufficient number of studies exists to define what 21st Century Skills are, how they can be assessed, and how they can be applied to virtual learning environments. Despite the phrase's common usage, "21st Century Skills" is employed in different disciplines to mean different things. For instance, there is a degree of confusion as to what precisely 21st Century Skills and 21st Century Learning Skills consist of and what differences lie between the two. According to Trilling and Fadel (2009), 21st Century Skills encompass three different taxonomies of skills (pp. 45–84): Digital Literacy Skills (i.e., information literacy, media literacy and ICT literacy), Career and Life Skills (i.e., flexibility, adaptability, initiative, self-direction, social skills, cross-cultural interaction, productivity, accountability, leadership and responsibility), and Learning and Innovation Skills (i.e., the "4Cs": critical thinking, creativity, communication and collaboration). Although Trilling and Fadel (2009) collectively termed such skills "Learning and Innovation Skills", much of the subsequent literature adapted the nomenclature used, terming them simply "Learning Skills" (Pardede, 2020; Sanguna, 2021; Selvaratnam, 2021). Following this distinction, we consider only 21st Century Learning Skills in this dissertation. In this dissertation, we will use "21st Century Learning Skills" as a term through which we predominantly focus on the above-mentioned "4C" skills: creativity, collaboration, communication and critical thinking. Previous studies have demonstrated that this group of learning skills already gained importance during the last century (Brown and Campione, 2002; Glăveanu, 2011; McCroskey, 2009; Wagner-Döbler, 2001), but also due to their general applicability, they are considered yet more pertinent skills in the present and more recent past (Häkkinen et al., 2017). Their primary importance is inherently linked to the fact that having good learning skills allows for more efficient learning of other types of knowledge and practical skills. Teaching these skills, however, is complex due to their meta-cognitive properties. It is precisely due to their impact on the ability to learn other (21st Century) skills that we will concentrate on these particular learning skills (i.e., the 4Cs in this thesis). Following an intensive study on 21st Century Learning Skills in this thesis, we considered the possibility of developing a virtual character with artificial intelligence properties that can foster these abilities in the users. The main objective of this intelligent virtual character would be to foster users' cognitive abilities for educational purposes. On the one hand, we could create a virtual character that could communicate with the users, in turn, possibly desiring to collaborate with the virtual agent. On the other hand, we could create a virtual character that could foster creativity, communication, collaboration, or critical thinking skills in users through interaction with another virtual character. However, before designing a virtual character that looked more intelligent or possessed artificial intelligence properties, further information about what "Leaning Skills" actually are, and how they are evaluated, was needed in this dissertation. From this knowledge, we can define the minimal characteristics that a virtual character must have to foster these skills, and apply these to future research. For this reason, we considered it crucial to study what are learning skills, how currently valid instruments assess these learning skills, and which tests are already available. These instruments can give us clues as to the fundamental elements that our intelligent virtual character must have in order to stimulate learning skills in the users. We discuss these learning skills in more detail in **Chapter 2**.

1.2 Design for conventional and immersive technology

Advances in the 21st Century have provided us with numerous tools through conventional technology such as videos, smart blackboards, games, online questionnaires, and other elements that can be accessed employing digital media. Recently, advances have been promoted in a type of technology termed "immersive technology". Such immersive technology is manifested in augmented reality, virtual reality, and mixed reality, as well as in (future) advances in artificial intelligence. According to Beck et al. (2020), the potentials of immersive technology are:

1. It can simulate or imitate the physical world.
2. It can provide visual and textual information of an object or location in an innovative way.
3. It can improve learning and soft skills.
4. It has inputs and outputs of multimodal interaction (e.g., facial expression recognition, touch screen, and eye gaze training).
5. It facilitates possibilities for collaborative work (e.g., remote, synchronous and asynchronous), a characteristic recognized as highly pertinent to the global pandemic situation.

6. It allows a high degree of combination of conventional educational technologies with immersive elements.
7. It can generate visual elements in the virtual world that are invisible in the physical world, allowing the exploration of the visualization of abstract concepts in learning environments.
8. It can change human perceptions and social responses through the embodiment of other humans, animals or objects in the virtual world, whereby it can create emotional and cultural responses, interactive exploration and foster engagement and motivation in individuals.
9. It can generate the possibility to switch perspective roles and viewpoints, allowing data collection from the users.
10. It can help students who lack access or availability to specific immersive resources to learn about the real world, far away from risky environments, or be used by persons with disabilities. All potentialities of these immersive technologies pose challenges for instructional designers to create and develop quality content and simulations to help foster the learning processes of users in virtual worlds.

In recent years, there has been a dramatic increase in the interest to improve the quality of virtual environments (Delgado-Mata et al., 2007). Particularly, in terms of how to create virtual characters similar to the biological world, and how users can have a degree of acceptability toward virtual characters and elements in the digital world. However, the degree of influence of types of design on the virtual character to allow it to be considered highly realistic and empathetic remains unclear. In this dissertation, the design of the main virtual character in conventional technology (digital images, videos, and computer games) and immersive technology (virtual reality) is examined. The elements deemed necessary for the design of virtual characters are discussed in greater detail in Section 1.3.

1.3 Design of virtual characters

In this dissertation, before setting out on the design of our virtual character, we ask ourselves: What should a virtual character look like? The answer to such a question is not obvious. We first need to determine what forms virtual characters have taken throughout history. The first virtual character was arguably the facial animation called the Phonoscope in 1892, whereby photographs of particular expressions were projected sequentially in order to resemble the movement and emotions of a human face (Waters and Levergood, 1995). This allowed the mixing of the facial expressions of the animated images, essentially creating a "life" for the character inside the technological device. A century later, this concept has evolved to allow the generation of facial expressions upon virtual faces through a software called "DECface". This software was able to manipulate two primary muscles of the face to design universally recognized facial expressions (Waters and Levergood, 1995; Wagner et al., 2006). With the most recent technological advances, the challenge today is to give more life to this type of virtual character through the integration of precise animated facial expressions (Rapuano et al., 2021), gestures (Rebol et al., 2021), natural speech (Thézé et al., 2020), conversation (Garcıa-Carbajal et al., 2020), realistic movement (Niay et al., 2020; Thaler et al., 2020), and similar facets of biological beings in 3D models (Fribourg et al., 2020; Kocur et al., 2020). However, much debate has surrounded the issue of how realistic a virtual character should be, because many such characters can evoke a mismatch sensation between the natural and the artificial, sometimes leading users to reject them (Wagner et al., 2006; Kätsyri et al., 2015; Thézé et al., 2020). This possible mismatch or incongruent sensation related to the design of the virtual character has been attributed to a famous hypothesis, termed "Uncanny Valley Effect".

1.3.1 Uncanny valley effect

In 1970, Masahiro Mori proposed the Uncanny Valley Effect, developed in robotics and automation (Mori, 1970; Mori et al., 2012). He described humans' emotional responses toward robots and other artificial entities with life-like appearances. He suggested that humans have different levels of emotional response to artificial entities depending on their appearance. The expectations are that the higher the level of similarity between the artificial entity and a real-life human, the higher its level of familiarity or affinity of user perception. However, he found that certain types of appearance within an artificial entity closely replicating their biological equivalent can foster a feeling of mismatch or revulsion towards it. This phenomenon is called Uncanny Valley Effect. This effect has been supported by a variety of studies demonstrating its validation in robotics (Kim et al., 2020; Laakasuo et al., 2021; Thepsoonthorn et al., 2021) and virtual characters (Mousas et al., 2021; Weisman and Peña, 2021). However, other studies have determined that such a phenomenon is nonexistent or have been critical of the theory (Brenton et al., 2005; Geller, 2008). Despite the controversies surrounding this effect, there are several important areas where this study makes an original contribution to robotics, human-computer studies, video games, animation, neuroscience and psychology. What is not yet clear is whether the uncanny valley effect can be observed in non-human characters such as virtual animals.

1.3.2 Uncanny valley effect when applied to virtual animals

Previous research has shown that the uncanny valley effect on human likeness is well-studied compared with animal-likeness studies. Schwind et al. (2018a) identified that:

"The question whether realistic artificial depictions of animals can fall into an uncanny valley is important because it would have a significant impact on research investigating that phenomenon. A hypothetical "uncanny valley of animals" would either mean that Mori's dimension of human likeness is not only related to humans and has to be enhanced or that the phenomenon appears in a different shape (or not at all). However, there is currently no empirical investigation of Mori's hypothesis, which explicitly considers an uncanny valley of virtual animals or discusses how and whether animals should be incorporated into the uncanny valley hypothesis." (p. 50).

This question occurs due to the presence of a stuffed animal among the figures plotted on Mori's original graphs. However, this stuffed animal was evaluated in terms of its human-like attributes. The issue, therefore, arises as to whether animals should also be evaluated on their corresponding animal-like attributes. If so, then it is vital to study if design choices for virtual animals lead to the most appropriate degree of user acceptance or whether such virtual animals can be faced with the rejection resulting from the uncanny valley effect, in greater depth. In summary, two crucial issues are addressed in this dissertation: the possibility of a presence of an uncanny valley effect in virtual animals, and the types of properties that should be evaluated in a virtual animal (**Chapter 3**).

1.3.3 Naturalism versus artificiality

The degree of realism to render an animal is one of the most challenging decisions to make when designing a virtual animal (Schwind et al., 2018a), especially considering that animals have very varied bodies, and different types of skin, hair, and other properties, which vary from species to species. However, perhaps a digital model of a virtual animal with more artificial properties might receive the same level of user acceptance. In order to study this, we decided to look at the realism of virtual animals from the perspective of their naturalistic or artificial appearance in a digital model. Coeckelbergh (2011, p. 199) describes that a robotic animal is considered "natural" when it looks like its "biological" equivalent, while it is considered "artificial" when it looks like its "technological" representation. He considered that a robot being visually imperceptible from an animal, with no discernible physical differences, could be somewhat problematic. Hence there is a need to study differences in the field of animal appearance. Likewise, in the field of virtual worlds or virtual environments, this type of gap arises in the design of virtual animals, as highlighted by Schwind et al. (2018a). For this dissertation, we designed the animal with a body and color similar to its biological traits, and we called this design "natural". In a virtual world, the concept of "natural" does not have the same meaning as for organic objects in the real world. Therefore, when we refer to "natural", we refer to the digital model design that appears natural or biological, preserving its original properties in comparison to the original (i.e., biological) animal's body. Contrarily, we refer to artificial as a digital model with mechanical and metallized parts and attributes, with a color dissimilar to that of the original animal. This dissertation addresses the definition of naturalness and artificiality from the perspective of uncanny valley effect in **Chapters 3** and **4**. Subsequently, virtual animal designs with natural and artificial appearances were introduced to the settings of a video game and a virtual simulation, where the effects of natural and artificial appearances of the virtual animals on empathy, immersive, perceived pain and embodiment are studied.

1.4 Impact on empathy, immersion, perceived pain, and embodiment

This dissertation sets out from the belief that the elements of the visual appearance of virtual characters can give us the keys for users to desire to cooperate with these virtual agents, creating emotional ties and acceptability in human-computer interaction.

However, while this may be a valid assumption, it may also be the case that visual appearance is not a significant factor in creating effects such as empathy, immersion, perception of pain or embodiment amongst users. To allow a greater understanding of this, it is important to define what we mean by each of these concepts in further detail.

1.4.1 Empathy

Historically, the term "empathy" has been used to describe the ability to share the feelings and beliefs of other people, including their psychological states of pain and distress, or experiences of excitement (Keen, 2006). Empathy can be related to both positive and negative cognitive and affective reactions. Patient and Skarlicki (2010) describe two types of empathy: (1) cognitive empathy, the ability to predict and recognize the feelings of others; and (2) affective empathy, the ability to distinguish affective responses in a distress situation. As empathy facilitates the process of social interactions and can repress antisocial behaviors, bullying behavior, and aggression towards others, it is a useful measure in educational, research and design environments (Paiva et al., 2005; Garandeau et al., 2021). Existing research identifies the critical role played by empathy upon human behavior. For this dissertation, empathy was deemed a crucial facet to explore and evaluate, in terms of whether it is possible for a virtual character to evoke it. Studies by Ochs et al., 2008 described that a empathetic virtual character can be found in two situations: (1) Virtual characters or agents can display empathetic and emotions reactions (e.g., emotional facial expressions, automatic emotional response) during their interaction with users, and (2) users can have empathetic reactions toward the virtual characters during their interaction with them (e.g., due to their appearance, personality, movement, story). In this dissertation, we were not assumed that our virtual character is directly empathetic by their emotional design as in situation (1) without an evaluation that confirms it. On the contrary, we explored if the visual appearance of the virtual characters cantake affect the empathy of the user in a situation (2) by measuring before and after their interaction with them. Previous research has established that there are two types of measure for empathy: dispositional and situational (Eisenberg et al., 1994). Dispositional empathy indicates a person's tendency to react toward the experiences of other people, animals or virtual stimuli in general (Konrath et al., 2011; Pallavicini et al., 2020; Paul, 2000). For this dissertation, we used the most well-known questionnaire for measuring dispositional empathy, Davis's Interpersonal Reactivity Index (IRI) (Choi et al., 2021; Davis, 1983; Garcia-Barrera et al., 2017; Hojat, 2016; Otterbacher et al., 2017). This questionnaire is deemed to possess outstanding reliability and validity, having been validated, for example, in numerous Chinese studies (Siu and Shek, 2005), Dutch studies (De Corte et al., 2007), South American studies (Fernández et al., 2011), and French studies (Gilet et al., 2013). This questionnaire has four subscales: perspective taking, personal distress, empathic concern, and fantasy. These measurement scales seem appropriated for experiments regarding virtual characters. Dispositional empathy can explain the current level of an individual's empathy but cannot define whether (or not) a situation or stimulus has modified — or is currently modifying — this. Situational empathy can be assessed during or after exposure to a situation or stimulus. Situational empathy can evaluate whether specific stimuli, conditions or situations can provoke empathic reactions. The principal technique for evaluating situational empathy is with a self-reporting or empathy-related responses (Eisenberg et al., 1994; Holmgren et al., 1998). Another technique for measuring situational empathy is through users' facial reactions. The advantage of the use of self-reporting in this dissertation is that this method has previously been applied to game and virtual world interaction studies. The self-reporting method can give information as to whether a virtual character induces an empathetic response in the users (McQuiggan et al., 2008; Paiva et al., 2005). Both measures used in this dissertation can provide a more straightforward answer as to whether changes in the design of the virtual animal's appearance can affect situational empathy, and we can further explore whether dispositional empathy has an influence on the results of our experiments.

1.4.2 Immersion

Another vital aspect in creating virtual worlds and an attractive context for users in the context of their virtual animal is immersion. The term "immersion" is generally understood to relate to the degree of an individual's disposition toward the realism of a virtual world from an experiential perspective as analogous to the real world (Hou et al., 2012). People's immersion can be influenced by various elements of a virtual world's design (e.g., a character's aesthetics, screen size, music, viewing angle, light effects and storyline). Self-reporting questionnaires are currently the most popular technique by which to examine immersion. The immersion questionnaire formulated by Jennett et al. (2008) is frequently used to measure immersion in video games or computer interfaces. This questionnaire was analytically based upon Brown and Cairns (2004b) work and is based around two subscales: engagement and engrossment. Engagement can be assessed through factors such as attention, time, and energy expended upon the virtual world. Engrossment can be associated with the appearance and alteration of emotions attributable to the virtual simulation's atmosphere (e.g., graphics, vibrations, plot and sounds). This dissertation addresses the question whether a virtual experience and virtual character's appearance designed for these studies embodied a strong enough level of immersion for users was critically important.

1.4.3 Perceived pain

In order to analyze the users' empathy, we designed a distressing event at the end of the simulation where another virtual character hunts the virtual animal. We aimed to determine whether or not it was possible to measure whether users could directly connect with the pain of the virtual character. Virtual reality simulation allows people to have a virtual body where we assume that users can experience pain reactions in distress situations. In broad terms, pain can be defined as any negative stimulus or adverse reaction that gives rise to a physical, emotional, or cognitive sensation of discomfort. In virtual reality, the simulation evaluated the perception of pain among users in an attempt to determine what decreased such sensations (Matamala-Gomez et al., 2019; Niharika et al., 2018; Patel et al., 2020). Previous studies have measured perceived pain in dental practices (Niharika et al., 2018). These studies used a questionnaire whereby respondents rated the intensity of pain through a series of illustrations of faces expressing different emotions. This questionnaire is called the "Wong-Baker Face Pain Rating Scale". This study concluded that virtual reality acted as an efficient tool for decreasing the level of perceived pain for its analgesic effects on users (Niharika et al., 2018; Sullivan et al., 2000; Wiederhold et al., 2014). This questionnaire was used in the research presented in this dissertation, to observe whether the stressful situation and the appearance of the virtual character could generate the perception of pain in the users.

1.4.4 Embodiment

Embodiment is considered to provoke the illusion of perceiving or feeling that this virtual body is their biological body in virtual reality. According to Matamala-Gomez et al. (2019), the embodiment is a mixture of sensations involving vision, touch, perception or awareness, the body's internal state, motor control, and the inner ears. In this dissertation, an experiment in virtual reality was developed from the users' first-person perspective. This allows users to be "placed" in a virtual body representing a virtual animal in a process known as embodiment. Moreover, we used a haptic sensation feedback vest to induce an elevated level of realism in the virtual body representation. To measure the avatar embodiment, we used a questionnaire developed by Gonzalez-

Franco and Peck (2018). This questionnaire measures tactile sensations, body ownership, location of the body, agency and motor control, response to external stimuli, and external appearance. We expected that the greater the natural appearance of the virtual character and embodiment experience, the higher the intensity of the empathy, immersion, perceived pain and avatar embodiment would be for the participants in this virtual reality simulation.

1.5 Problem statement and research questions

An analysis of the appearance of virtual animals may give rise to benefits in terms of alleviation of the manifestation of the uncanny valley effect, thereby helping to foster empathy, immersion, perceived pain and embodiment concerning an avatar, thus giving rise to improvements in terms of future application to 21st Century Learning Skills. For the purposes of this research, it has been determined that only the natural and artificial visual appearances of the virtual animals were used, so that the results can be applied widely. Consequently, our problem statement reads as follows:

Problem statement

Which is the more effective visual design in the virtual characters that can invite users to foster communication and collaborative skills within virtual environments?

Research questions

To answer the problem statement, we study various aspects of uncanny valley theory, empathy, immersion, perceived pain, embodiment, and 21st Century Skills in relation to our goals. The following Research Questions (RQs) have therefore been formulated:

1. How can we measure 21st Century Learning Skills (creativity, collaboration, communication, and critical thinking skills) with existing tests and what is the reliability and validity of these tests?
2. Does a virtual animal (e.g., panda) also adhere to the uncanny valley effect?
3. Can a virtual animal's visual appearance influence the level of empathy of users?
4. Can a virtual animal's visual appearance influence the level of immersion of users?
5. Can a virtual animal's visual appearance influence the level of perceived pain of users?

6. Can a virtual animal's visual appearance influence the level of embodiment of users?
7. How can different versions of virtual animals be used as virtual tutors in video instruction, which may have different effects on affective and cognitive outcomes, depending on their visual appearance?

1.6 Thesis overview

In **Chapter 2**, a study of 21st Century Learning Skills and testing applied to them is performed in order to address RQ1. A total of 195 articles, published between 1943 and 2022, and relating to the current instruments concerning the testing of the defined learning skills (i.e., the 4Cs), as well as to different aspects of measurement, such as type, educational level, reliability and validity, are reviewed. In **Chapter 3**, we applied a study from Uncanny Valley Theory to answer RQ2. In the preliminary study presented in Section 3.1, we described the impact of the appearance of robotic and natural characters by means of an example of six images of a virtual panda, and propose future applications for the use of virtual animals in human-computer interaction, education, and robotics studies. In Section 3.2, the study elaborated in Section 3.1 is extended in order to answer RQ2. We improved upon the survey of Section 3.1, and formulated a survey of six different still and moving panda images to be assessed upon conditions relating to the following properties: familiarity, commonality, naturalness, attractiveness, interestingness, and animateness. We explore these using two types of animal-likeness scales. In **Chapter 4**, the design of a computer game simulation created with the aim of answering RQ3 and RQ4 is described and presented. To answer these research questions, we analyzed the effect of the visual appearance of a virtual animal character in a third person view on empathy and immersion. We designed an experimental game called "Justin Beaver". The computer game was designed so that the user takes care of the virtual animal (a beaver) by providing it with energy, break time and fun. In **Chapter 5**, we extended the study of **Chapter 4** in order to answer RQ3, RQ4, RQ5 and RQ6, converting this PC version to an immersive VR technology environment. We analyzed the effect of the visual appearance of a virtual animal character on empathy, immersion, perceived pain and avatar embodiment. This new version of the experimental simulation in virtual reality was called "Justin Beaver VR". The virtual reality simulation was designed to explore the virtual animal's life and natural habitat. Participants explored their virtual animal appearance from the first-person perspective and took on the virtual body. In **Chapter 6**, we analyzed the virtual animal appearance and its impact when such virtual animals are used as virtual instructors in an attempt to answer RQ7. The results showed that the virtual instructor's appearance had an important effect on knowledge recall during a class about robotics and the culture of the Netherlands. This work extends **Chapter 3** on the uncanny valley effect and explores these virtual animal characters in their role as virtual instructors in secondary education in Colombia. In **Chapter 7**, a general discussion is provided, and the most critical findings are put forward and a general discussion is provided, and the most critical findings are put forward for RQ1 to RQ7. Conclusions and proposals as to the applicability for future research are also given in robotics, education of the natural conservation and artificial intelligence in education. Moreover, we explain some possible improvements of our research for the generalized the findings of this dissertation (see also Figure 1.6.1).

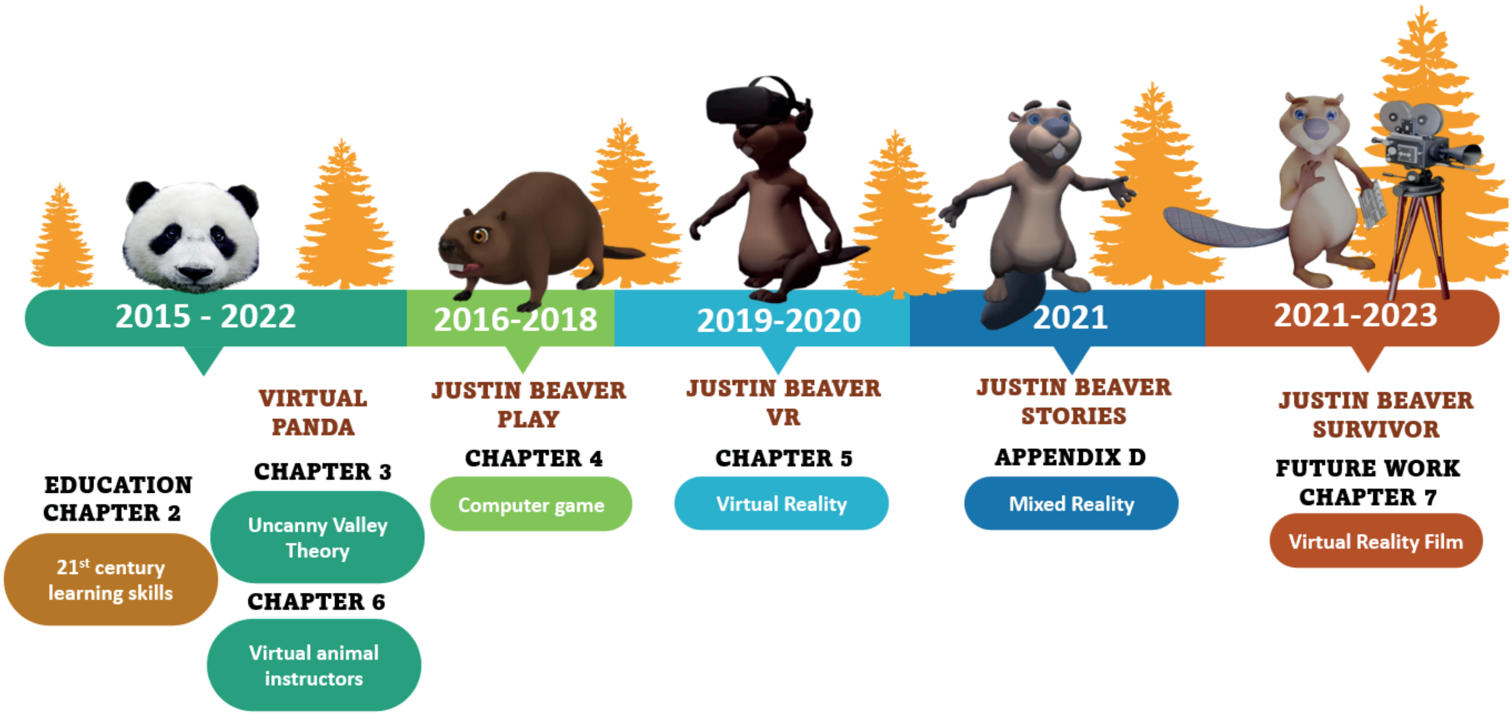


Figure 1.6.1: Infographic of the thesis overview.

1.7 Published work

For all co-authored chapters within this dissertation, I am the principal author. I am responsible for the complete process of this dissertation. As my principal supervisors, Marie Postma and Menno van Zaanen, supported me with conceptualization and methodology and reviewed my writing in the academic articles.

1.7.1 Structure of the chapters

The following chapters have been previously published or submitted to journals or conferences. The modifications made here are limited to realignment to conform to the format of a thesis dissertation and the resizing of some figures and tables.

Chapter 1: Introduction.

Chapter 2: Sierra Rativa, A., Postma, M., & van Zaanen, M. (2021). 21st Century Learning Skills: A systematic review of the literature on critical thinking, creativity, communication, and collaboration skills tests. Manuscript submitted to *Sage Open* for publication.

Chapter 3: The uncanny valley of the virtual animals

Section 3.1: Sierra Rativa A., Postma M., & van Zaanen M. (2020). The Uncanny Valley of the Virtual (Animal) Robot. In Merdan M., Lepuschitz W., Koppensteiner G., Balogh R., & Obdržálek D. (Eds.), *RiE 2019: Robotics in Education*, Vol 1023. Advances in Intelligent Systems and Computing, 419-–427. Springer. https://doi.org/10. 1007/978-3-030-26945-6_38

Section 3.2: Sierra Rativa A., Postma M., & van Zaanen M. (2022). The Uncanny valley of the virtual animals. *Computer Animation and Virtual Worlds Journal*, 33(2), 1–21. https://doi.org/10.1002/cav. 2043

Chapter 4: Sierra Rativa, A., Postma, M., & Van Zaanen, M. (2020). The influence of game character appearance on empathy and immersion: Virtual non-robotic versus robotic animals. *Simulation & Gaming*, 51(5), 685–711. https://doi.org/10.1177/1046878120926694

Chapter 5: Sierra Rativa, A. S., Postma, M., & Van Zaanen, M. (2022). Try walking in my paws: Is it possible to increase empathy, immersion, and perceived pain in virtual reality environments by manipulating animal character appearance?. Manuscript submitted to *Virtual Reality* for a journal publication.

Chapter 6: Sierra Rativa A., Vasquez C.C., Martinez F., Orejuela Ramirez W., Postma M., & van Zaanen M. (2021). The Effectiveness of a Robot Animal as a Virtual Instructor. In

Lepuschitz W., Merdan M., Koppensteiner G., Balogh R., & Obdržálek D. (Eds.), *RiE 2020: Robotics in Education*, Vol 1316. Advances in Intelligent Systems and Computing, 329–338.Springer. https://link.springer.com/chapter/10.1007/978-3-030-67411-3_30

Chapter 7: Discussion and conclusion.

1.7.2 Publications finished during the PhD

These conference publications were finalized during my PhD process, but they were not included in this dissertation:

1. Sierra Rativa A. (2019). How can we Teach Educational Robotics to Foster 21st Learning Skills through PBL, Arduino and S4A?. In Lepuschitz W., Merdan M., Koppensteiner G., Balogh R., & Obdržálek D. (Eds.), *RiE 2018: Robotics in Education*, vol 1316. Advances in Intelligent Systems and Computing, 149-–161. Springer. https://doi.org/10.1007/978-3-319-97085-1_15

2. Sierra Rativa, A. S., Postma, M., & Van Zaanen, M. (2020). Can virtual reality act as an affective machine? The wild animal embodiment experience and the importance of appearance. In Urrea C. (Eds.). *Proceedings of the MIT LINC 2019 Conference*, 26 Vol 3. EPiC Series in Education Science, 214–223. EasyChair. https://doi.org/10.29007/dc7s

3. Sierra Rativa, A., Bakker, G. A., & Sierra Rativa, A. G. (2022). Animal Embodiment: Embodying a beaver in immersive virtual environments to create empathy and teach about the impact of global warming in a playful way. In Richir S. (Eds.). *Proceedings of Virtual Reality International Conference (VRIC)*, 33–36. [Poster presentation], Laval, France. Laval Virtual.

1.7.3 Award recognition

The best revolution research in virtual reality technology and immersive technology (2021) called "Justin Beaver Stories". Laval Virtual. France (See **Appendix D**).

1.8 Databases

For this dissertation, we publicly share through Dataverse the data collected in the doctoral process:

1. Sierra Rativa, A. S., Postma, M., & Van Zaanen, M. (2019). Virtual animals versus virtual robotic animals: The influence of appearance of game characters on empathy and immersion [Data set]. *Dataverse NL*. https://doi.org/10.34894/3DX4VN
2. Sierra Rativa, A. S., Postma, M., & Van Zaanen, M. (2022). Study I: FaceReader Data (Emotional Facial Expressions) after watching a Virtual Instructor [Data set]. *Dataverse NL*. https://doi.org/10.34894/O1S3N9
3. Sierra Rativa, A. S., Postma, M., & Van Zaanen, M. (2022). Study II: FaceReader Data (Emotional Facial Expressions) after watching a Game Character [Data set]. *Dataverse NL*. https://doi.org/10.34894/BXLX8E
4. Sierra Rativa, A. S., Postma, M., & Van Zaanen, M. (2022). Virtual reality experimental data and Justin Beaver VR Simulation [Data set]. *Dataverse NL*. https://doi.org/10.34894/YDVRDC
5. Sierra Rativa, A. S., Postma, M., & Van Zaanen, M. (2022). Data of uncanny valley of a virtual animal. [Data set]. *Dataverse N*L. https://doi.org/10.34894/JIBXBU

CHAPTER 2

Measuring 21st Century Learning Skills: A systematic review of the literature on critical thinking, creativity, communication, and collaboration skills tests

This Chapter 2 tackles research question RQ1. It is founded on journal paper publication.

Research Question:

1. How can we measure 21st Century Learning Skills (creativity, collaboration, communication, and critical thinking skills) with existing tests and what is the reliability and validity of these tests?

This chapter is based on:

1. Sierra Rativa, A., Postma, M., & van Zaanen, M. (2021). Measuring 21st Century Learning Skills: A systematic review of the literature on critical thinking, creativity,communication, and collaboration skills tests. Manuscript submitted to *Sage Open* for publication.

2.1 Abstract

In this chapter, we present a systematic review of the literature on the assessment of 21st Century Learning Skills. The 21st Century Learning Skills can be subdivided into the 4Cs: critical thinking, creativity, communication, and collaboration skills. Due to the differences between the 4C learning skills, each of the skills requires specialized forms of testing. In total, 195 articles published between 1943 and 2022 form the basis for this review. We investigate how the current instruments relate to the different learning skills, as well as different aspects of the measurements such as type, educational level, and the reliability and validity test.

2.2 Related work

As the name 21st Century Skills indicates, educational researchers and practitioners alike agree that the new century requires a specific set of skills (Rios et al., 2020; Van Laar et al., 2020). As an umbrella term, the term 21st Century Skills refers to sets of skills, including learning skills (e.g., critical thinking, creativity, communication, and collaboration skills, also called the 4Cs), literacy skills (e.g., information literacy, media literacy, and technology literacy skills, also called IMT), and life skills (e.g., flexibility, leadership, productivity, and social skills, also called FLIPS) (Pardede, 2020; Stone et al., 2017). These skills appear to be particularly useful in our fast-paced daily environment where the widespread use of online communication has led to new forms of interaction and collaboration. For the purposes of this article, we focus on tests designed to evaluate learning skills, i.e., the 4Cs. The set of learning skills contains skills that are specifically targeting the improvement of learning or, in other words, learning about learning. Many authors have agreed that the following are all considered 21st Century Learning Skills and Innovation Skills: critical thinking and problem solving, communication, collaboration, creativity, and innovation (Dede, 2010; Fadel and Trilling, 2010; Germaine et al., 2016; Häkkinen et al., 2017; Kivunja, 2014; Voogt and Roblin, 2010). Without a doubt, learning skills play a crucial role in the acquisition of other 21st Century Skills and are, therefore, of paramount importance. Evaluating these skills, however, is a nontrivial task due to the fact that they concern general learning strategies rather than specific knowledge and understanding. As teachers try to improve the 21st Century Learning Skills in students, it is essential for them to be able to compare the starting levels to the levels achieved by training. Additionally, measuring skills levels is also valuable for understanding the effectiveness of different types of instruction. For instance, when technological innovations such as video games or simulations (e.g., computer, virtual reality, augmented reality, mixed reality, and other versions), educational software, and other online tools are developed to support these learning skills (for example, on the basis of self-directed education), it is necessary to assess their effectiveness. For this purpose, objective means of measuring the level of learning skills are indispensable. In this article, we collect information about several tests that are available for the evaluation of 21st Century Learning Skills which allow for obtaining objective performance metrics. Typically, the level of these abilities is measured through the use of psychological tests. We address two specific objectives: 1. To provide a broad overview of existing tests designed to measure the learning skills and 2. to examine the reliability and validity of these tests. Below, we first discuss a general definition of the learning skills (the 4Cs), i.e., critical thinking, creativity, communication, and collaboration.

2.2.1 Critical thinking

Several definitions of critical thinking can be identified in the literature. For example, John (1997) defines this term as a thought process focused on the discovery and verification of the information received. According to Facione et al. (2011), it is a set of cognitive skills and attitudes (dispo sitions towards critical thinking) for studying a problem in a thoughtful way, acknowledging these skills, and applying them (Mwalongo, 2014).Ennis (2015), on the other hand, understands this concept as a type of reflective reasoning which corresponds to the ability to make a decision about one's beliefs. Fisher (2011) argues that critical thinking can also be subsumed under meta-cognition, which means that critical thinking is a high-level cognitive skill (Dumitru, 2019) and more than one ability could be involved. Thonney and Montgomery (2019) mentioned that critical thinking is the ability to apply "knowledge to new situations, considering different viewpoints, evaluating options, facts and suggestions across different disciplines" (p. 174). Summarizing, critical thinking can be defined as an active, thoughtful, and continuous process of the act of independent thinking, which is vital for decision-making and problem-solving.

Critical thinking thus appears to be an umbrella term that can be subdivided into lower-level skills. In fact, Facione et al. (2011) defines six cognitive skills within critical thinking: interpretation, analysis, evaluation, inference, explanation, and self-regulation. Additionally, a set of dispositions (attitudes) is introduced, consisting of the following terms: inquisitive, judicious, truth-seeking, systematic, analytical, open-minded, and confident in reasoning. A full characterization of the ideal critical thinker is a combination of cognitive skills and dispositions.

2.2.2 Creativity

Creativity is challenging to define because it is assigned somewhat dif ferent meanings in different disciplines (Gomez, 2007). For example, Reid and Petocz (2004) describe creativity for education as a way of solving problems, for business it is entrepreneurship, and in the arts (e.g., music) it may be a way to create compositions. Götz (1981) uses the term to refer to its Greek etymological origin "ktidzo" which means "to create". There is thus a creative process that is evaluated by its originality (e.g., new product or insight), results, effects (e.g., impact in the society), antecedents (e.g., previous products or services), and capacities (e.g., the ability to create). In line with this definition, Sternberg and Lubart (1999) state that creativity can be defined as a skill to develop original and appropriate work or tasks (e.g., new ways of problem-solving, new discoveries in research, new products). Likewise, Kim (2019) expressed that in order to define creativity, it is important to eliminate misconceptions of what we believe is creativity, such as that creativity is unique to the arts, is fostered by mental illnesses, is rapidly recognizable by the majority of people, is a supernatural inspiration, and that creative people always work alone, or that novelty is a prerequisite to be creative. Summarizing, creativity is an evolving concept that can change over time, but in general is associated with the development of ideas, which can result in novel and useful solutions to different problems, new findings, new inventions, and new products or services. As such, it is a crucial ability in the 21st Century, because of its direct effects on development of new levels of research, and collaboration.

2.2.3 Communication

With the term "communication", in this article, we specifically refer to human social interaction expressed via language (Argyle et al., 1970; Burns, 1991). Traditionally, communication can broadly be defined as an act to transfer, or send information and response to it (Stevens, 1950). Although differences in opinion still exist, there appears to be some consensus that communication refers to social connections through language, which requires at least two interlocutors (sender-receiver) in a specific medium (verbal and non-verbal channels). In the context of the 21st Century Learning Skills, Bellanca (2010) argues that engaging in complex communication requires specific skills in order to manage the large amounts of verbal and non-verbal information. For example, one such communication skill is the ability to accept, articulate, and transmit feedback about ideas or information in technological media (Germaine et al., 2016). Coffelt et al. (2019) describe that in addition to commonly speaking and writing communication skills, visual communication skills (e.g., data visualization, producing visual aids, or non-verbal communication), and electronic communication skills (e.g., writing effective text messages) are considered important in a modern context. Nevertheless, Liberman et al. (2017) noted that developing communication skills is no simple task. This is because of the requirement that people need to have the compatible linguistic and pragmatic cues in order to have effective communication. Indeed, a primary goal of the 21st Century Learning Skills has been to develop efficient communication skills in people despite this inherent complexity. Therefore, an effective communicator can engage in major interactions with multiple audiences, while facilitating problem-solving (Bellanca, 2010).

2.2.4 Collaboration

The term "collaboration" is generally understood as the ability to work together to solve a mutual goal or problem (Mattessich and Monsey, 1992; Winn and Blanton, 2005). For John-Steiner et al. (1998), it refers to dialogues with the goal of sharing knowledge and workload in a team setting. For Wood and Gray (1991), collaboration relates to an interac tive decision-making process regarding a common problem or objective in which all participants have same interest, autonomy, and organization (e.g., norms, rules, and structures). Likewise, Katz and Martin (1997) describe collaboration as a social process which is motivated by participation and contribution (sharing ideas and data). In contrast, Henneman et al. (1995) find that a generally accepted definition of col laboration is lacking. Instead, they refer to its Latin etymological origin "collaborate" which means "to labor together". Arguably, working in online environments brings about new challenges, such as the complexity of the technological environment, as well as the lack of an emotional connection, and problems in intercultural understanding while working with distant collaborators in a project (Hur et al., 2020; Moore, 2016). To sum up, collaboration is associated with a social interactive process through which people share knowledge and responsibility, make decisions, and experience autonomy in order to solve complex problems or challenges.

The broad use of the term "collaboration" is sometimes equated with "cooperation". Both have been used to describe the ability to work in a group on a common issue (Bruffee, 1987). However, the terms have also been viewed as conceptually distinct. Cooperation insinuates an informal relationship in which the participants' responsibilities (e.g., teach ers assign the task and students solve it) and resources (e.g., human, technology) can be individually distributed. In terms of task performance, the product is the only importance. In contrast, collaboration implies a formal relationship and interactive planning between members, in which participants' responsibility (e.g., teachers help students with a project) and resources can be employed jointly for long-term results. Hence, process and product are the focus for task performance (Mattessich and Monsey, 1992). In the context of 21st Century Learn ing Skills, researchers are focusing on collaboration skills rather than cooperation skills (Binkley et al., 2012; Blumenfeld et al., 1996; Dede, 2010).

2.3 Method

The study presented here is a systematic literature review. Criteria for selecting sources for this literature review are based on the work by Chalkiadaki (2018), as shown in Table 2.3.1.

Table 2.3.1: Inclusion and exclusion criteria for the literature review.

Type of criterion	Criteria	Inclusion	Exclusion
Type of publication	Journal articles	x	
	Conference papers	x	
	Reports	x	
	Dissertations Books	x	
		x	
Access	Online	x	
Publication period	1943–2022	x	
Place of publication	Worldwide	x	
Type of study	Empirical investigation	x	
	Theoretical studies	x	
Research methods	Qualitative	x	
	Quantitative	x	
Language	English	x	

We conducted this systematic literature search in two phases. In the first phase, we searched databases for studies associated with the four learning skills for further analysis and apply these inclusion and exclusion criteria. In the second phase, we classified the studies according to their measurement properties such as type, applicability, and validity.

2.3.1 Literature search: Phase 1

To identify the tests that can be used to assess the 21st Century Learning Skills, we first searched for overview articles that discuss relevant tests. In particular, we searched in the Google scholar database (using the terms "critical thinking" and "test" and "list", similarly for the other 21st Century Learning Skills) with a publication period from 1943 to 2022. This search led to the following articles described in Table 2.3.2:

The complete list of the tests found in these articles is provided in the **Appendix A**. As may be clear, there are too many tests to be practically analyzed in full detail.

In our analysis, we first and foremost prioritized tests that were reported in the following review studies. For critical thinking, we partly relied on the review of Ennis (1993), for creativity, on the review provided in Venable (1994), and for collaboration on Thannhauser et al. (2010); for communication skills, we were not able to find a relevant overview article to identify a list of communication tests. Next to that, we gave preference to tests for which validity and reliability have been reported in the literature. Finally, we made sure that for each learning skill, the same number of tests was available, in order to make an equal comparison on various criteria, such as type, applicability and others. From the resulting list, we randomly selected ten tests for each type of 21st Century Learning Skill for an in-depth analysis.

To identify articles that describe properties of each of the selected tests, we searched for relevant articles using two databases: Google Scholar and ERIC (see Table 2.3.3). List of the articles identified on the Google Scholar and ERIC, also they were on SCOPUS and PsycINFO datasets. The search action consisted of the name of the test. This resulted in the numbers as described in Table 2.3.3: 427 publications with critical thinking tests, 343 with creativity tests, 299 with communication tests, and 396 with collaboration tests.

Table 2.3.2: Literature search of the list of 21st Century Learning Skills: Phase 1.

#	Publication	#	Publication
1	Acar and Runco (2012)	35	Klebig et al. (2016)
2	Afzali et al. (2011)	36	Ku (2009)
3	Arden et al. (2010)	37	Kuncel et al. (2005)
4	Baer and Kaufman (2008)	38	Lee and Kim (2011)
5	Bar et al. (2018)	39	Liu (1998)
6	Batey and Furnham (2008)	40	Lu and Xie (2019)
7	Bataineh and Zghoul (2006)	41	Maloa and Bux (2015)
8	Behar-Horenstein and Niu (2011)	42	Marsh et al. (1996)
9	Bensley et al. (2016)	43	Matsumoto et al. (2000)
10	Broadleaf (2020)	44	McCrae et al. (1993)
11	Brown et al. (2020)	45	Morrison and Shriberg (1992)
12	Castleberry and Shepherd (1993)	46	Sańchez and Olivares (2011)
13	Charmello (1993)	47	Sfetsos et al. (2006)
14	Charyton et al. (2008)	48	Smith and Gregg (2020)
15	Chen et al. (2006)	49	Soni and Bakhru (2019)
16	Cheung et al. (2003)	50	Soto and John (2009)
17	Cropley (2000)	51	Stone et al. (2001)
18	Dollinger et al. (2004)	52	Spell and Frank (2000)
19	Dougherty and Larson (2005)	53	Sustekova et al. (2019)
20	Ellis et al. (2016)	54	Pastor and David (2017)
21	Ennis (1993)	55	Pease and Colton (2011)
22	Fee and Gray (2012)	56	Rear (2019)
23	Furnham and Bachtiar (2008)	57	Ritter and Mostert (2017)
24	Furnham et al. (2011)	58	Rosip and Hall (2004)
25	Han (2003)	59	Runco et al. (2011)
26	Holton (2001)	60	Taube (1997)
27	Hu and Adey (2002)	61	Thannhauser et al. (2010)
28	Hu et al. (2013)	62	Ting et al. (2006)
29	Huang et al. (2017)	63	Tiruneh et al. (2017)
30	Huang and Wang (2019)	64	Vaida (2019)
31	Huhn et al. (2011)	65	Venable (1994)
32	Hullman et al. (2010)	66	Wagner and Harvey (2006)
33	Karpova et al. (2011)	67	Wolfradt and Pretz (2001)
34	Kaufman et al. (2016)	68	Ziv and Keydar (2009)

Table 2.3.3: Overview of the tests selected for the systematic literature review.

Skills	Name of the test	Identification	Selection
Critical Thinking	California Critical Thinking Disposition Inventory	143	22
	California Critical Thinking Skills Test	103	7
	Cornell Critical Thinking Test	2	8
	Graduate Management Admission Test	2	2
	Halpern Critical Thinking Assessment	8	3
	HEIghten Critical Thinking Assessment	6	3
	International Critical Thinking Reading and Writing Test	33	3
	Critical Reasoning Test Battery	66	3
	Management and Graduate Item Bank	41	4
	Watson-Glaser Critical Thinking Appraisal	23	6
	Total	427	61
Creativity	Creative Scientific Ability Test	4	3
	California Psychological Inventory	11	5
	Creative Achievement Questionnaire	27	9
	Creative Product Semantic Scale	7	7
	Omnibus Personality Inventory	1	3
	Scientific Creativity Test	22	6
	Schaefer's Biographical Inventory Creativity	5	3
	Creative Attitude Survey	71	3
	Gough's Creative Personality Scale	23	3
	Torrance Test of Creative Thinking	172	12
	Total	343	54
Communication	Affective Communication Test	28	3
	Behavioral Indicants of Immediacy Scale	54	3
	Diagnostic Analysis of Nonverbal Accuracy	20	4
	Gates-MacGinitie Reading Tests	80	8
	Generalized Immediacy Scale	29	3
	Interpersonal Communication Competence Scale	8	3
	Kentucky Comprehensive Listening Test	29	3
	Templin-Darley Test of Articulation	8	3
	Templin Test of Auditory Discrimination	40	4
	Watson-Barker Listening Test	3	3
	Total	299	37

Collaboration	Big Five Personality Traits	289	18
	Gallup Clifton Strengths Test	4	4
	Collaboration and Satisfaction about Care Decisions	1	2
	Index of Interdisciplinary Collaboration	9	4
	Interdisciplinary Education Perceptions Scale	32	2
	Keirsey Temperament Sorter	8	4
	Multidisciplinary Collaboration Instrument	14	2
	Riso-Hudson Enneagram Type Indicator Test	3	3
	Role Perceptions Questionnaire	33	2
	Readiness for Interprofessional Learning Scale	3	2
	Total	396	43

After the exclusion of duplicates and texts that did not satisfy the inclusion criteria (as described in Table 2.3.1), 61 texts remained for analysis on critical thinking, 54 publications related to creativity tests, 37 on communication tests, and 43 on communication tests. A total of 195 articles were used to analyze the current status of the tests in detail in this literature review.

2.3.2 Literature search: Phase 2

We provide an analysis of the selected tests that may be used to evaluate each of the learning skills. For each of the tests, we describe several properties:

1. Classification of the type of psychological test (aptitude, achievement, or personality test);
2. Educational level for which the test was applied (kindergarten, primary school, secondary school, university, or unspecified);
3. Validity and reliability.

To collect information on the tests, we consider publications between 1943 and 2022 to identify the different tests (as some tests have been in use for a longer time) and their validity and reliability investigations. We focus on more recent use (2016–2022) of the tests for the identification of other properties (e.g., educational level).

As the description of the types of psychological tests used in this article requires more explanation, we will discuss these first. Next, we describe how validity and reliability are operationalized for the purposes of this review.

Classification of the type of psychological test

Psychometricians have developed many types of psychological tests that can be used to evaluate a wide range of abilities. Here, we explore ways in which psychological tests can be used specifically for the assessment of 21st Century Learning Skills. Depending on their primary use, we classify available tests into three test categories: achievement, aptitude, and personality. Each of these will be discussed here briefly.

Achievement: This type of test can be defined as an instrument that shows the degree of acquisition of abilities through training or experience. Goldstein et al. (2019) note that an achievement test is designed to evaluate the level of learning in different fields. In particular, in education settings, achievement tests are used to evaluate the skill deficit or requirements of learning in students.

Aptitude: This type of test can be defined as an instrument that examines the potential of a person to acquire a specific set of abilities, as opposed to their current knowledge about a topic or their level of ability in a skill (Kline, 2000). The focus of this type of test is on measuring the possibility of acquiring knowledge or an ability.

Personality: This type of test can be defined as an instrument that examines an individual's behavioral characteristics. It does not measure knowledge or skill nor does it measure the possibilities of acquiring knowledge or skill. It concentrates completely on how people behave. This type is typically used for educational selection, in clinical settings, and in the field of psychology. The principal characteristics of personality tests are that questions refer to specific properties of behaviors, attitudes, beliefs, reactions, wishes, interests, and judgments in specific situations. The questions are typically a combination of yes-no, true-false, like-dislike, trichotomous items, and rating scales (Kline, 2000).

2.3.3 Literature search: Phase 3

Validity and reliability

There is a growing body of literature that deals with investigating the quality of psychological testing. There are two interesting aspects that define the quality of tests: validity and reliability. Validity refers to the degree of the accuracy of the variable which is assessed, whereas reliability refers to the degree of consistency of an assessment instrument in diverse situations (Field, 2013; Goodwin and Goodwin, 2016; Kelley, 1999; Thissen and Wainer, 2001).

In this article, we conducted a qualitative content analysis of 46 empirical studies or publications about the validity and reliability of the tests using the software Atlas.ti (Version 22). From these empirical studies or publications, we extracted quotations which described the validity and reliability of the forty 21st Century Learning Skills Tests. The complete quotations found in these publications are provided in the **Appendix A**. For analysis of these quotations, we conducted a content analysis of validity and reliability found in publications by authors, a positive mark (+) is given when the publication indicates positive information regarding the test's validity and reliability, a negative mark (-) when the test's validity or reliability is considered negative, and a question mark (?) when either the results of validity or reliability studies are inconclusive or if we could not find any studies by authors that identify the test's validity and reliability. Table 2.3.4 provides an overview of the different situations that may result from the validity and reliability evaluations.

Table 2.3.4: Criteria of evaluation or reliability and validity tests.

Criteria of evaluation	Reliability	Validity
Reliable and valid	+	+
Reliable, but not valid	+	-
Valid, but not reliable	-	+
Neither reliable nor valid	-	-
Valid, but unknown reliability	?	+
Not valid and unknown reliability	?	-
Reliable, and unknown validity	+	?
Not reliable and unknown validity	-	?
Unknown reliability and validity	?	?

Note: Positive (+): Validity or reliability; Negative (-): No validity or no reliability; Question mark (?): No results.

2.4 Results

The first aim of this study is to ascertain which tests are available to assess 21st Century Learning Skills. To answer this question, we will provide a general overview of the tests below. We identified 195 articles on the four learning skills, which allows for the analysis of ten tests for each learning skill (critical thinking, creativity, communication, and collaboration). These tests are organized depending on their psychological test type, and educational level. The educational level indicates the (educational) context in which the test has been applied. It may be the case that a specific test is grouped in more than one educational level.

Next, we will discuss the tests based on the different property groups (psychological test type and educational level) in more detail. In particular, we will take a look at the distribution of the tests within these property groups.

2.4.1 Classification by learning skill

Critical thinking skills

A National Panel of forty-six Experts who participated in the Delphi research project for the American Philosophical Association, created a theoretical construct to recognize and evaluate the critical thinking skills (Facione, 1990). As a result of this study, two tests have become available to evaluate critical thinking: the "California Critical Thinking Disposition Inventory" and the "California Critical Thinking Skills Test". The California Critical Thinking Skills Test is a standardized test, which is used in educational settings, in particular to measure in college level of the critical thinking. The duration of the test is close to 45 minutes. It has 34 standardized multiple-choice questions and is only available in English. This test has been used since 1990 in the fields of education and medicine.

Other tests that assess critical thinking are available as well. For example, the Cornell Critical Thinking Test is used in educational settings. In particular, it has two versions: one for children and another for adults. This test assesses the skills of deduction, semantics, credibility, induction-judging conclusions, induction-planning experiments, definition, and assumption identification. Another test is called the Halpern Critical Thinking Assessment; it evaluates abilities of decision making and problem-solving, argument analysis, likelihood, and uncertainty analysis, hypothesis testing, and verbal reasoning. Similarly, the International Critical Thinking Reading and Writing Test assesses the ability to use reading and writing for gathering information on analysis and evaluation.

There are other critical thinking tests which are usually used in commercial enterprises, in particular for professional and administrative recruitment. For example, the Watson-Glasser Critical Thinking Test evaluates the ability to perform comprehension, analysis, and evaluation for separating facts from suppositions and viewpoints. Other tests are the Critical Reasoning Test Battery and the Management and Graduate Item Bank, which evaluate numerical reasoning (e.g., the ability to evaluate numerical information in logical way) and verbal critical reasoning (e.g., the ability to assess critical verbal arguments). The Graduate Management Admissions Test assesses the ability to analyze logical arguments and opinions. This test is similar to the HEIghten Critical Thinking Assessment which assesses the ability of analysis (e.g., analyzing arguments) and synthesis (e.g., understanding inferences and consequences). These tests are used in a business context and in particular in the staff selection process. In conclusion, there are several critical thinking tests that are used depending on the field or purposes of application.

Table 2.4.1 provides an overview of the different tests investigated in this article for critical thinking.

Table 2.4.1: Critical thinking tests.

#	Test	Type of test	Educational level	V	R
1	California Critical Thinking Disposition Inventory	Personality	University/Secondary	+	+
	(Facione et al., 1995, 2011)		(Akgun and Duruk, 2016; Barin, 2019; Bayram et al., 2019; Bulgurcuoglu, 2016; Demiral, 2019; Demirbag et al., 2016; Kabeel and Eisa, 2016; Kilic et al., 2017; Kizilet, 2017; Fettahlıo˘glu and Kaleci, 2018; Gul and Akcay, 2020; Maltepe, 2016; Orhan, 2022; Ordem, 2017; Pepe, 2018; Saglam et al., 2017; Unlu and Dokme, 2017; Uslu, 2020; Y¨ukselir, 2020)	(Yeh, 2002)	
2	California Critical Thinking Skills Test	Aptitude	University	+	+
	(Facione et al., 1995)		(Aghajani and Gholamrezapour, 2019; Alkharusi et al., 2019; Ng et al., 2022)	Khalili and Hossein, 2003; Khalili and Soleymani, 2003; Khoda et al., 2007)	
3	Cornell Critical Thinking Test	Aptitude	University/Secondary	+	-
	(Ennis, 2015)		(Baki et al., 2016; Demirci and Ozy¨urek, 2017; Erdogan, ¨ 2019; Fadhlullah and Ahmad, 2017; Paulsen and Kolstø, 2022)	(Leach et al., 2020; Verburgh et al., 2013)	
4	Graduate Management Admission T	Aptitude	No articles (2016–2022)	+	?
	(Coetzee et al., 2015)			(Kuncel et al., 2007)	
5	Halpern Critical Thinking Assessment	Achievement	No articles (2016–2022)	+	-
	(Halpern, 2010)			(Butler, 2012; Verburgh et al., 2013)	
6	HEIghten Critical Thinking Assessment	Aptitude	No articles (2016–2022)	+	+
	(ZlatkinTroitschanskaia et al., 2016)			(Liu et al., 2016, 2018)	
7	International Critical Thinking Reading and Writing Test	Achievement	University	+	+
	(Paul and Elder, 2006)		(Lu and Xie, 2019)	(Hollis et al., 2020)	
8	Critical Reasoning Test Battery	Aptitude	No articles (2016–2022)	+	+
	(Bawtree et al., 1991; Moutafi et al., 2005)			(Moutafi et al., 2005)	

9	Management and Graduate Item Bank	Aptitude	No articles (2016–2022)	?	+
	(Howard and Choi, 2000; Saville and Ltd, 1991; Van der Merwe, 2002)			(Kotze and Griessel, 2008)	
10	Watson-Glaser Critical Thinking Appraisal	Aptitude	No articles (2016–2022)	+	+
	(Neimark, 1984; Watson, 1980; Yeh, 2009)			(Gadzella and Baloğlu, 2003; Gadzella et al., 2006; Sternod and French, 2016)	

Note: In this table, the first column contains the name of the test including a reference to the original publication of this test or studies that use this instrument. In the columns on reliability and validity, references to articles that contain information on reliability and validity have been added. The last two columns represent Validity (V) and Reliability (R) and quotations are provided in the **Appendix A**.

Creativity skills

Focusing on techniques for measuring creativity, Lubart (1994) describes eight methods:

1. Cognitive abilities tests (e.g., Torrance Test of Creative Thinking, Creative scientific ability test, Scientific Structure Creativity Model);
2. Personality inventories (e.g., Omnibus Personality Inventory, California Psychological Inventory, or Gough's Creative Personality Scale);
3. Biographical inventories (e.g., Schaefer's Biographical Inventory);
4. Attitude and interest surveys (e.g., Schaeffer's Creative Attitude Surveys);
5. Person-centered ratings;
6. Eminence;
7. Self-reports of achievements (e.g., Creative Achievement Questionnaire); and
8. Judgments of work samples (e.g., Creative Product Semantic Scale).

Even though several techniques to measure creativity exist, Gomez (2007) claims that creativity cannot be perfectly measured because the existing literature fails to describe the strong relationship between creativity and intelligence. Some tests assess intelligence traits, but these do not necessary correspond to creativity traits. Also, creative people do not always come up with the same unique answer, which leads to invalid results on some creativity tests because they allow only one possible correct answer or lack an alternative right answer. However, some studies have validated the Torrance Tests of Creative Thinking (Davis, 1989; Kim, 2006), although other studies have also analyzed some aspects to improve this test (Clapham, 2004; Hee Kim, 2006). The test has been used and supported by many researchers, positioning it as the most used international form of evaluation of creative thinking, addressing various aspects of creativity as fluency, flexibility, originality, and elaboration. For instance, in a study on the validation of this test, Kim (2006) concludes that it is a useful method to evaluate creativity. It is easy to administer, and it is useful to identify talented students. In this sense, Lubart (1994) concludes that each method for measuring creativity has positive aspects, but also negative aspects that can be improved. Table 2.4.2 provides an overview of the different tests for creativity investigated in this article.

Table 2.4.2: Creativity tests.

#	Test	Type of test	Educational level	V	R
1	Creative Scientific Ability Test (Sak and Ayas, 2013)	Achievement	Secondary (Huang and Wang, 2019)	+ (Ayas and Sak, 2014)	+
2	California Psychological Inventory (Gough, 1984; Helson, 1965; Hocevar, 1981; McClure and Mears, 1984)	Personality	No articles (2016–2022)	+ (Rosen and Schalling, 1974)	?
3	Creative Achievement Questionnaire (Carson et al., 2003)	Achievement	University/Unspecified (Carter et al., 2019; Dostál et al., 2017; Lennartsson et al., 2017; McKibben and Silvia, 2017; Perach and Wisman, 2019; Wesseldijk et al., 2019; Zabelina et al., 2021)	+ (Wang et al., 2014)	+
4	Creative Product Semantic Scale (Jackson and Games, 2015; O'Quin and Besemer, 2006; Tsai, 2016)	Achievement	University/Secondary (Tsai, 2016; Wandari et al., 2018)	+ (Besemer and O'Quin, 1999; O'Quin and Besemer, 1989)	?
5	Omnibus Personality Inventory (Heist et al., 1968; Hocevar, 1981)	Personality	No articles (2016–2022)	+ (Griffin and Hopson, 1978)	-

6	Scientific Creativity Test (Hu and Adey, 2002)	Achievement	Primary/Secondary /University (Akcanca and Cerrah Ozsevgec, 2018; Astutik and Prahani, 2018; Eroglu and Bektas, 2022; Huang and Wang, 2019; Yang et al., 2016)	+ (Ayas and Sak, 2014)	?
7	Schaefer's Biographical Inventory Creativity (Cropley, 2000; Glover et al., 2013; Hocevar,1981; Schaefer and Anastasi,1968)	Personality	No articles (2016–2022)	+ (Schaefer, 1972)	?
8	Creative Attitude Survey (Bonk, 1988; Tesser and Shaffer, 1990)	Aptitude	No articles (2016–2022)	- (Bonk, 1988)	?
9	Gough's Creative Personality Scale (Gough, 1979)	Personality	Primary/University (Zampetakis, 2010)	+ (Perry, 2020)	+
10	Torrance Test of Creative Thinking (Almeida et al., 2008; Bailey et al., 1970; Hamlen, 2013; Torrance, 1972; Walton, 2003)	Aptitude	Primary/University (Chi et al., 2016; Derman and Yavuz, 2022; Kralova et al., 2018; Oncu, 2016; Shabahang and Amani, 2016)	+ (Kim, 2006; Wechsler, 2006)	+

Note: In this table, the first column contains the name of the test including a reference to the original publication of this test or studies that use this instrument. In the columns on reliability and validity, references to articles that contain information on reliability and validity have been added. The last two columns represent Validity (V) and Reliability (R) and quotations are provided in the **Appendix A**.

Communication skills

Regarding techniques for measuring basic verbal and non-verbal communication competencies, several tests have been proposed. The following have been used to evaluate verbal communication:

1.Templin-Darley Test of Articulation, which assesses the ability to pronounce words through picture stimuli; 2. Templin Test of Auditory Discrimination, which assesses the ability of auditory discrimination; 3. Kentucky Comprehensive Listening Tests, or Watson-Barker Listening Test, which assess short-term listening, rehearsal, long-term listening, interpretive listening, and listening with distraction; and 4. Gates-MacGinitie Reading Tests which assess the ability to read and assigns corresponding levels.

Likewise, some non-verbal communication tests are:

1. Behavioral indicants of immediacy scale or the generalized immediacy scale, which assess the ability to perceive the subjective gestalt perceptions, immediacy behaviors, and objective counting individual non-verbal behaviors; 2. Diagnostic Analysis of Nonverbal Accuracy, which assesses the ability to effectively recognize and send emotional information through non-verbal communication; 3. Affective Communication Test, which assesses expressiveness; and 4. Interpersonal Communication Competence Scale, which assesses empathy support, flexibility, social relaxation, and general competence.

Each test is designed to assess a specific ability or set of abilities. However, there currently appears to be no test available that can mea sure all verbal and non-verbal communication skills simultaneously. Depending on the situation, teachers, trainers, and researchers may select one or more of these tests (Kearney and Beatty, 2020).

Table 2.4.3 shows an overview of the different tests for communication skills investigated in this article.

Table 2.4.3: Communication tests.

#	Test	Type of test	Educational level	V	R
1	Affective Communication Test (Friedman et al., 1980; Schutte et al., 1998)	Personality	No articles (2016–2022)	?	- (Hensley, 1986)
2	Behavioral Indicants of Immediacy Scale (Andersen et al., 1979; Richmond et al., 1987)	Aptitude	No articles (2016–2022)	?	+ (Ellis et al., 2016)
3	Diagnostic Analysis of Nonverbal Accuracy (Nowicki and Duke, 1994)	Achievement	Primary (Booth et al., 2019; Lamer et al., 2022)	+	+ (Pitterman and Nowicki Jr, 2004)
4	GatesMacGinitie Reading Tests (Hoge and Butcher, 1984; MacGinitie et al.,1989)	Achievement	Primary/Secondary (Choi, 2017; Coggins et al., 2017; Giusto and Ehri, 2019; Sanabria et al., 2022; Young et al., 2019)	+	? (Jenkins and Jewell, 1993)
5	Generalized Immediacy Scale (Andersen et al.,1979; Plax et al.,1986)	Aptitude	No articles (2016–2022)	?	- (Ellis et al., 2016)
6	Interpersonal Communication Competence Scale (Cupach and Spitzberg, 1983; Rubin and Martin, 1994)	Aptitude	No articles (2016–2022)	+	- (Rubin et al., 1993)
7	Kentucky Comprehensive Listening Test (Bostrom and Waldhart, 1983; Davenport Sypher et al., 1989)	Achievement	No articles (2016–2022)	-	? (Fitch-Hauser and Hughes, 1992)
8	TemplinDarley Test of Articulation (Sherman and Geith, 1967; Templin and Darley, 1960)	Achievement	No articles (2016–2022)	+	+ (Blake, 1978)
9	Templin Test of Auditory Discrimination (Shearer, 1976; Sherman and Geith, 1967; Templin, 1943)	Achievement	Achievement (Shelton et al., 1977)	-	?
10	Watson-Barker Listening Test (Watson and Barker, 1988; Worthington et al., 2014)	Achievement	No articles (2016–2022)	+	? (Roberts, 1986)

Note: In this table, the first column contains the name of the test including a reference to the original publication of this test or studies that use this instrument. In the columns on reliability and validity, references to articles that contain information on reliability and validity have been added. The last two columns represent Validity (V) and Reliability (R) and quotations are provided in the **Appendix A**.

Collaboration skills

There are several commonly used questionnaires focusing on the measurement of collaboration:

1. Big Five personality traits, which shows tendencies towards openness to experience (e.g., imaginative or independent personality), conscientiousness (e.g., dependable and self-discipline personality), extraversion (e.g., sociability and talkativeness personality), agreeableness (e.g., compassionate and cooperative personality), and neuroticism (e.g., depression and vulnerability personality)1; 2. The Gallup Clifton Strengths-Finder Assessment, which assesses talents or strengths in different themes such as strategic thinking, executing, influencing, and relationship building; 3. Collaboration and Satisfaction about Care Decisions, which assesses the satisfaction, shared decision-making, joint planning, communication, and cooperation within a group; 4. Index of Interdisciplinary Collaboration, which assesses interdependence, professional activities, collective ownership of goals, flexibility, and reflection in a group context; 5. Interdisciplinary Education Perceptions Scale, which assesses professional competence, autonomy, cooperation, resource sharing, and contributions with professional values from group members; 6. Keirsey Temperament Sorter II, which assigns four temperaments through roles such as artisan or tactic temperament (e.g., operators and entertainers), guardian or logistics (e.g., administrators and conservator), idealist or diplomacy (e.g., mentor or advocates), and rational or strategy (e.g., coordinators or engineers); 7. The Multidisciplinary Collaboration instrument, which assesses the ability to work together, patient care process, communication, and teamwork; 8. Riso-Hudson Enneagram Type Indicator, which investigates nine personalities types such as helper, motivator, ambitious, sensitive, investigator, loyalist, enthusiast, spontaneous, leader, peacemaker, and reformer; 9. Role Perceptions Questionnaire, which assesses the ability to work in other professions different to group members' primary studies, communication, caring, and dedication; and 10. Readiness for Interprofessional Learning Scale, which assesses the ability to perceive teamwork and collaboration, professional identity and roles.

There are a number of instruments available for measuring collaboration skills, but the majority are developed in the health field. In other fields, different researchers have tried to develop their own questionnaires or surveys to measure collaboration skills (Thannhauser et al.,2010).

1 This test is sometimes used to measure creativity as well.

Table 2.4.4 shows an overview of the different tests for collaboration skills investigated in this article.

Table 2.4.4: Collaboration tests.

#	Test	Type of test	Educational level	V	R
1	Big Five Personality Traits (Barrick and Mount, 1991; Gosling et al., 2003)	Personality	Secondary / University/ Unspecified (Atabek, 2019; Bozgeyikli, 2017; Dal, 2018; Ercan, 2017; Fortis, 2019; Karduz and Sar, 2019; Katrimpouza et al., 2019; Kirkagac and Oz, 2017; ¨ K¨oseoglu, 2016; Kulig et al., 2019; Linvill, 2019; Rieger et al., 2017; Syed et al., 2020; Tabak et al., 2022; Wild and Alvarez, 2020)	+ (Mount et al., 1994)	-
2	Gallup Clifton Strengths Test (Asplund et al., 2007)	Personality	Unspecified (Busch and Davis, 2018; Rosson and Weeks, 2018)	+ (Schreiner, 2006)	+
3	Collaboration and Satisfaction about Care Decisions (Gedney, 1994)	Aptitude	No articles (2016–2022)	- (Aaberg et al., 2019)	+
4	Index of Interdisciplinary Collaboration (Bronstein, 2002; Parker-Oliver et al., 2005; Wittenberg-Lyles et al., 2010)	Aptitude	No articles (2016–2022)	+ (Thannhauser et al., 2010)	+
5	Interdisciplinary Education Perceptions Scale (Luecht et al., 1990)	Aptitude	No articles (2016–2022)	+ (Thannhauser et al., 2010)	+
6	Keirsey Temperament Sorter (Kiersey and Bates, 1984)	Personality	Secondary (West, 2016)	+ (Dodd and Bayne, 2007; Kelly and Jugovic, 2001)	+
7	Multidisciplinary Collaboration Instrument (Carroll, 1999)	Aptitude	No articles (2016–2022)	+ (Thannhauser et al., 2010)	+
8	Riso-Hudson Enneagram Type Indicator Test (Riso and Hudson, 1996; Vaida and Pop, 2014)	Personality	Personality	+ (Newgent et al., 2004)	+
9	Role Perceptions Questionnairie (Mackay, 2004)	Aptitude	No articles (2016–2022)	+ (Thannhauser et al., 2010)	+

Table 2.4.4: Continued.

10	Readiness for Interprofessional Learning Scale (Mattick et al., 2009)	Aptitude	No articles (2016–2022)	+ (Thannhauser et al., 2010)	+

Note: In this table, the first column contains the name of the test including a reference to the original publication of this test or studies that use this instrument. In the columns on reliability and validity, references to articles that contain information on reliability and validity have been added. The last two columns represent Validity (V) and Reliability (R) and quotations are provided in the **Appendix A**.

2.4.2 Distribution by type of psychological test

If we consider the distribution of the tests over the psychological test types, we can group the tests according to the skill that is being measured and its corresponding psychological test type. The results in Figure 2.4.1 show the proportions of psychological tests that are available for each learning skill. For example, the critical thinking skills are measured mainly by aptitude tests (70%). In contrast, both achievement tests (20%) and the personality tests (10%) show low percentages. When comparing critical thinking skills with creativity skills, we see that the creativity tests have a different distribution. The differ ent types of psychological tests available to measure creativity skills two type of tests have similar percentages. For the creativity skills, personality tests (40%) show same percentage of achievement (40%). In the case of communication skills, these are measured mainly using achievement tests (60%), whereas the other percentages are 30% for the aptitude tests and 10% for the personality test. Finally, collaboration skills are measured mainly by aptitude tests (60%) with personality tests (40%) on the second place. Interestingly, we were not able to find any achievement tests for collaboration skills.

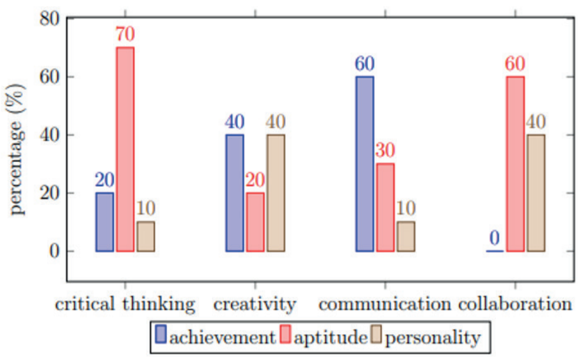


Figure 2.4.1: The distribution by type of psychological test.

2.4.3 Distribution by educational level

Figure 2.4.2 provides information on the educational level to which the 40 learning skills tests have been applied. These educational levels were reported in academic articles from 2016 to 2022. We reviewed 195 studies in this systematic literature review, focusing on the educational level on which the tests have been applied. We identified 68 studies between 2016 and 2022 that detailed the level of education of the participants. We found 26 studies for the critical thinking skills test, 20 studies for the creativity skill tests, 5 studies for the communication skill tests, and 17 studies for the collaboration skill tests. The results show that the kindergarten, primary, secondary school, and university levels each received different amount of attention from researchers depending on the type of the test. For critical thinking skills tests, we found 33.3% of the tests in secondary school, and 66.7% of the tests at university level. For creativity skill tests, we found the tests mostly at university level (41.6%), 25% of the tests at primary and secondary school level, and for 8.4% the educational level was not specified. Communication skill tests were found mostly at primary (66.6%) and secondary (33.4%) school level. Finally, collaboration skill tests were found mostly to be unspecified educational level (50%), followed by secondary (33.4%) and university (16.6%) level. We have not found any studies applied to kindergarten level. For primary school level, however, no critical thinking and collaboration skill tests were found. In contrast, communication skill tests were absent at university level.

2.4.4 Distribution by validity and reliability

Table 2.4.5 groups information regarding validity and reliability information for the 40 tests. For the 40 learning skills tests, 42.5% are found to be both reliable and valid. Another 20% of the tests are found to be valid, but their reliability is unknown. For 37.5% of the tests, we did not find any publications indicating validity or reliability.

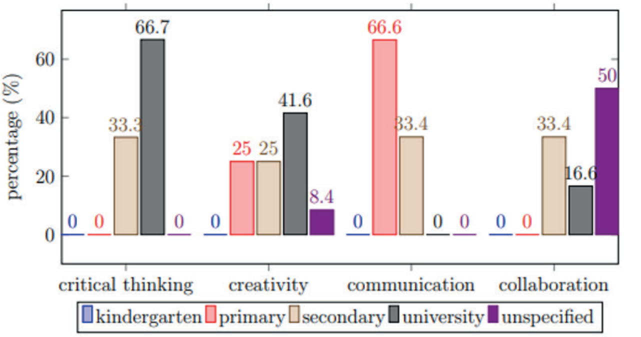


Figure 2.4.2: The distribution of tests by educational level.

Comparatively, more critical thinking, creativity, and collaboration skill tests have been shown to be valid and reliable compared to communication skill tests. However, the differences are small. In particular, communication skill tests have few publications that investigate their validity and reliability.

Table 2.4.5: Distribution of information on validity and reliability references by external author publications for the different types of tests.

Validity Reliability	+ +	+ –	– +	– –	? +	? –	+ ?	– ?	? ?
Critical thinking	6	2	0	0	1	0	1	0	0
Creativity Skills	4	1	0	0	0	0	4	1	0
Communication skills	2	1	0	0	1	2	2	2	0
Collaboration skills	8	1	1	0	0	0	0	0	0
Total	20	5	1	0	2	2	7	3	0
Percentage	50%	12.5%	2.5%	0%	5%	5%	17.5 %	7.5%	0 %

2.5 Discussion

2.5.1 Type of test

This article investigates the evaluation of 21st Century Learning Skills, which are a subset of the 21st Century Skills. In particular, we consider how the 21st Century Learning Skills can be tested. For the different tests, we look at how the tests are applied (with respect to the educational level) and the nature of the tests (achievement, aptitude, and personality). Additionally, we investigated their reported reliability and validity.

We found that although the 21st Century Skills (including the learning skills) are currently relevant and essential in the education and economy domains (Larson and Miller, 2011; Silva, 2009), in particular, teachers are interested in these type of abilities as they try to improve the level of these skills in their students. In order to measure the level, they require the availability of tests, and these tests should show high quality (providing reliable and valid results). Some authors have described different evaluation models for 21st Century Skills (Binkley et al., 2012; Griffin et al., 2012; Kyllonen, 2012; Silva, 2009), but the impact of the tests used to evaluate them is understudied. The current study concentrates on the characteristics of currently available learning skill tests. The primary aim of this study is to determine which tests are available to assess 21st Century Learning Skills depending on psychological type, and educational level, and their reliability and validity.

The first aspect of the tests deals with the three psychological types of tests: achievement, aptitude, and personality. Achievement tests evaluate the abilities and knowledge learned through instruction or other experiences. When considering the distribution of the available achievement tests for the different learning skills, some of the findings were surprising. To measure communication skills several achievement tests are available. A possible explanation for this might be that communication skills are closely related to the sub-skill of language (e.g., reading, writ ing, speaking, and listening). Katz and Slomka (2000) describe that in humanities, sciences, and social sciences, standard achievement tests are frequently employed. However, even though critical thinking and creativity can also be measured using achievement tests, fewer tests are available compared to the communication skills. Surprisingly, in the sample analyzed for the purposes of this review, no achievement tests were found to evaluate collaboration skills. One reason for this may be that collaboration skills cannot be measured based on individ ual academic performance due to the fact that collaboration implies an interactive experience between more than one person.

Aptitude tests show a different distribution compared to achievement tests. Using this type of psychological test, we can measure the natural capacity of a person for a particular skill, such as understanding, analysis and interpretation of information, logical reasoning, or manual skills. This results in information about a potential personal performance in a specific place or environment. For instance, critical thinking and collaboration skills are mainly measured using aptitude tests. A possible explanation for this may be that aptitude tests are designed to analyze the cognitive or dispositional ability related to task performance. Apti tude tests are mostly used for "task-oriented job analysis", for example, in employee selection procedures (Schmidt et al., 1981). Information on critical thinking and collaboration may form an important aspect in job interviews and admissions to educational institutes (Mackay, 2004;Watson, 1980). In contrast, other skills such as creativity and communication abilities have fewer aptitude tests available. However, the results show that for all learning skills aptitude tests are available. In sum, all 21st learning skills have an aptitude test which can project a future performance.

Personality tests are mainly used to measure the psychological profile or character of a person. In contrast to aptitude tests, this study shows that fewer personality tests are available to measure the learning skills when compared to achievement and aptitude tests. However, creativity and collaboration skills are mainly measured through personality tests, whereas critical thinking and communication skills are typically not measured through personality tests. In the case of collaboration skills, personality tests are the second most available type (after aptitude). A possible explanation for this is proposed by Morgeson et al. (2005), who discuss that collaboration and social skills are associated to specific personality characteristics (e.g., emotional stability, conscientiousness, extraversion, teamwork knowledge, and agreeableness). Moreover, they show that it is important to assess social skills and teamwork through personality tests because they can predict task performance. Reid and Petocz (2004) claim that personality tests are frequently used to measure creativity due to the fact that creativity skills are associated with personality traits (e.g., reflective, unconventional, sensitive). According to Kim (2006), personality studies can analyze the characteristics of creative people. These relationships may partly explain why creativity is mostly measured using personality tests. In conclusion, personality tests are available tools used to test learning skills that are associated with personality characteristics.

2.5.2 Educational level

To investigate the educational level (kindergarten, primary, secondary school levels, and university) in which the tests are applied, we reviewed 195 studies related to each available learning skills test. All of these were selected from 2016 to 2022. The overview showed that critical thinking tests are mostly applied at university level compared to kindergarten, primary, and secondary school level. We found that the universities specifically applied these critical thinking tests for the evaluation of language skills in their students (Aghajani and Gholamrezapour, 2019; Demirbag et al., 2016), to measure the ability to negotiate in business programs (Fadhlullah and Ahmad, 2017), and to assess analytical skills for the body performance of sports (Demiral, 2019). Interestingly, the studies that were used at university and secondary level were used to assess the critical thinking instruction skills of teachers at these institutions (Barin, 2019; Bayram et al., 2019).

Surprisingly, communication skills tests were not found to be used at the university level. The communication skills tests are mostly applied on the primary and secondary school levels. Some of these tests were used in primary and secondary level to analyze reading or writing disabilities in their students (Booth et al., 2019; Giusto and Ehri, 2019).

Creativity tests were found to be applicable to all educational levels. At university level, studies reported on application in psychology or art classes (Carter et al., 2019; McKibben and Silvia, 2017), in the secondary and primary levels they are mainly measuring scientific creativity (Huang and Wang, 2019; Wandari et al., 2018; Yang et al., 2016). Some of the studies analyzed did not specify the level of education. These studies were mostly related to the fields of creative arts and music (Lennartsson et al., 2017; Wesseldijk et al., 2019).

Most collaboration skills tests did not provide a specific educational context, followed by publications applied to secondary school and university level. We did not find any kindergarten, and primary school level studies in our sample. One reason for this may be that collaboration skills are too complex to measure at kindergarten and primary school level with formal tests. Studies that do not specify the educational level of participants were used for various topics such as manager personality, social media, and victims and vulnerable population (Atabek, 2019; Fortis, 2019; Kulig et al., 2019). At a university level, these tests have been used for analyzing relationships between student and instructors, social media, resiliency, collaboration abilities for working with sporting teams (Dal, 2018; Linvill, 2019). The secondary school level used these tests to determine social cognitive constructs with school students, school counselors and ethical decisions (West, 2016; Riegeret al., 2017; Katrimpouza et al., 2019; Syed et al., 2020).

In summary, the primary school level frequently described studies dealing with communication and creativity skills tests. The secondary school level can be found in publications on all four learning skills. University level is the context for studies on critical thinking, creativity, and collaboration skills, with a notable absence of communication skills tests. Creativity and collaboration skills tests can mostly be found in publications not specifying educational level. We did not find any stud ies in our sample that measured all learning skills at the kindergarten level.

2.5.3 QUALITY OF TESTS

Another aim of this study was to determine the reliability and validity of the available 21st Century Learning Skills tests. According to Silva (2009), critics of 21st Century Skills claim that these skills cannot be measured properly. Learning skills, such as collaboration, creativity, communication, and critical thinking, need a long period to develop in a rich environment to reach their peak. To measure these skills, proper assessment methods are required that are reliable, flexible, adaptable, and accessible. The assessment of this type of skills must guarantee that its evaluation is done with good standards of quality and reliability. This is not, however, a simple task. Each test must be rigorous in its selection of methods, procedures, appropriate scales, type of groups on which they are to be evaluated, the language to be used, and a wide range of other factors.

Currently, there is a large amount of research available regarding the assessment of skills. Binkley et al. (2012), who previously wrote a literature review on 21st Century Skills, already concluded that it is important to guarantee the validity of the forms of evaluation of these skills. Here, we analyze the reliability and validity of 40 tests designed to measure either critical thinking, creativity, communication, or collaboration skills. We also report the results of existing research regarding the reliability and validity of these tests. The analysis of the quality of the tests is based on results published in academic articles that perform qualitative analysis on the validity and reliability of the tests. The results show that only for 50% of the tests positive information on reliability and validity was available. For 17.5% of the tests, positive information on validity but no reference to reliability is given, and negative reference to reliability is a 12.5%. For 5% of the test, no information on validity but positive or negative reference to reliability is given and 2,5% of the tests, the negative information on validity but positive reference to reliability is found. Surprisingly, negative reports regarding the reliability and validity of these tests are not found in the existing literature. Some publications note that the tests need to be improved on specific aspects of their design without explicitly stating that the test is invalid or unreliable.

One unanticipated finding is that collaboration skill tests have higher levels of known validity and reliability compared to communication, creativity, and critical thinking skill tests. Although there is a wide range of tests available to assess communication skills (e.g., reading, writing, speaking, listening tests), it is very difficult to find publications that analyze their validity and reliability. In conclusion, this study identifies fourteen tests to measure 21st Century Learning Skills that lead to valid and reliable results. However, it also shows that there is still room for improvement by determining the reliability and validity of the other available tests reporting 21st Century Learning Skills.

2.6 Conclusions

There are many ways to evaluate the 21st Century Learning Skills and this study provides a systematic literature review of 195 studies analyzing the current instruments. The results are grouped by psychological type, educational level and the tests' reliability and validity.

Firstly, we found that the aptitude, achievement, and personality tests used to assess learning skills are distributed unevenly across the different skills. Achievement tests are mainly used to evaluate communication skills, aptitude tests are mostly used to assess collaboration skills, and personality tests are used to measure creativity and collaboration skills.

Secondly, studies on critical thinking, collaboration, and creativity skills test can be found at university level. The secondary school level has been the context for all four learning skills. The primary level is the setting, mostly of application studies related to communication and creativity skill tests. There is an absence of studies that focus on the kindergarten educational level for all four learning skills. Creativity and collaboration skills tests dominate studies that do not specify the educational level. Finally, we found that half of the tests are valid and reliable, but for the other tests, reliability and validity are unknown or negative. This study serves as a basis for future research designed to stimulate creativity, collaboration, critical thinking, and communication skills on educational technology projects and to assess their effects on learning.

CHAPTER 3

The uncanny valley of virtual animals

This first and second sections of this Chapter 3 called "**the Uncanny valley effect of virtual animals**" tackles research question RQ2. The first section is published on the conference paper and the second section is founded on journal paper publication.

Research Questions:

1. Does a virtual animal (e.g. panda) also adhere to the uncanny valley effect?

Published Work:

1. Sierra Rativa A., Postma M., & van Zaanen M. (2020). The Uncanny Valley of the Virtual (Animal) Robot. In Merdan M., Lepuschitz W., Koppensteiner G., Balogh R., & Obdrzalek D. (Eds*.), RiE 2019: Robotics in Education*, Vol 1023. Advances in Intelligent Systems and Computing, 419-–4279.Springer. https://doi.org/10.1007/978-3-030-26945-6_38
2. Sierra Rativa A., Postma M.,& van Zaanen M. (2022). The Uncanny valley of the virtual animals. *Computer Animation and Virtual Worlds Journal*, 33(2), 1–21. https://doi.org/10.1002/cav.2043

Database: Sierra Rativa, Alexandra; Marie Postma; Menno van Zaanen, 2022, "Data of uncanny valley of a virtual animal", https://doi.org/10.34894/JIBXBU, *Dataverse NL*

3.1 The uncanny valley of the virtual (animal) robot

3.2 Abstract

We explored whether the uncanny valley effect, which is found for human-like appearances, can also be found for animal-like virtual characters such as virtual robots and other types of virtual animals. In contrast to studies that investigate human-like appearance, there is much less information about the effects concerning how a virtual character's animal-likeness influences their users' perception. In total, 162 participants evaluated six different virtual panda designs in an online questionnaire. Participants were asked to rate different panda faces in terms of their familiarity, commonality, naturalness, attractiveness, interestingness, and animateness. The results show that a robot animal is perceived as less familiar, common, attractive, and natural. The robot animal is interesting and animateness to users, but no big differences with the other images are found. We propose future applications for the human-(animal) robot interaction as tutorial agents in videogames, virtual reality, simulation robot labs using real-time facial animation.

3.2.1 Related work

There is a growing body of literature that recognizes the importance of human-robot interaction in educational settings. A recent study on the diversity of robots in education conducted by Belpaeme et al. (2018) provides an overview of which robots are used most in pedagogical studies. They found that the NAO robot (with human-like appearance) was the most used often (in 48% of the studies). Other types of robots that were used (in a range between 6–4% of the studies), were Keepon (with animal appearance), Wakamaru (human appearance), Robovie (human appearance), Dragonbot (animal appearance), iCat (animal appearance), and Bandit (android appearance). Interestingly, although there was a predominance of robots with human characteristics, robots with an animal appearance were also found. Additionally, data from several studies with HRI reported more affective effects (66%) than cognitive effects (34%).

Studies on human-(animal) robot interaction show positive emotional effects on users. For instance, the seal robot "Paro" has demonstrated to help people to reduce stress, anxiety, depression, dementia, and other behavioral and psychological symptoms (Takayanagi et al., 2014). Previous research with seal and dog companion robots found that loneliness of the users can be reduced as robot animals have an effect comparable to a live animal (Banks et al., 2008; Robinson et al., 2013). Interestingly, some physical robot animals also have a virtual model prototype. For instance, (1) the bear robot called "Keio U Robot-phone" has a digital version called "Keio U Robot-phone animation" (Li and Chignell, 2011); (2) the dinosaur robot called "Pleo " has a digital version called "Pleo animation" (Fernaeus et al., 2010; Rosenthal-von der Pütten et al., 2013; Ryokai et al., 2009); (3) the dog robot called "Sony Aibo" has a digital version called "Simulated Aibo" (Carpin et al., 2007; Coghlan et al., 2018; Kertész and Turunen, 2019); the cat robot called Simulated Aibo "Philips iCat" has a digital version called "Philips iCat animation" (Heerink et al., 2009; Leite et al., 2008; Looije et al., 2010; Pereira et al., 2008); and the rabbit robot called Philips iCat animation "NTT Cyber Solution Lab robot" has a digital version called "NTT Cyber Solution Lab animation" (Shinozawa et al., 2005).

One of the reasons that animal robots may result in a friendly user interface is due to their funny and attractive behavior as a possible robot pet (Goris et al., 2009). Other reasons may be that virtual robots or agents can offer abilities similar to a physical robot with lower prices, fewer problems with installation and additional technical requirements, personalization (also in curricula), and can be used by a large number of students into a classroom concurrently. However, previous research which compared physical robots

3

with virtual models have found that virtual robots lead to less engagement, empathy, and smaller effects in educational outcomes (Belpaeme et al., 2018). This can be associated with problems with the robot's appearance or behavior in virtual environment. For this reason, we investigate whether the uncanny valley can give answers to these problems.

3.2.2 The uncanny valley in virtual animals

The effects of the uncanny valley theory related to virtual humans (which are characters with anthropomorphic features) are currently well-studied. Virtual humans are found in the form of virtual android robots, bots, chatbots, teaching avatars, and related agents (Ciechanowski et al., 2019; Lugrin et al., 2018). However, one may wonder whether the uncanny valley theory can also be applied to non-human characters like "stuffed animals" with human-like features. Unfortunately, it is unclear why the research in the uncanny valley area has been mostly focused on virtual humans only, when virtual non-human characters are also regularly used in the industry, but not yet widely analyzed in the academic field. In particular, there is a lack of research regarding the uncanny valley theory on whether virtual animals and their resemblance can spark a feeling of familiarity. Specifically, we are interested in the questions that are related to the existence of the uncanny valley theory in the area of virtual animals. Additionally, we would like to know more about the possible effect of the uncanny valley theory on the design of the appearance of the virtual animal and its relationship with the perception of users. This study sets out to investigate whether the uncanny valley theory can be applied to virtual animals. The main research question addressed in this paper is: *Can we identify the effect of the uncanny valley theory in the familiarity, commonality, naturalness, attractiveness, interestingness, and animateness perception towards a virtual animal?*

3.2.3 Methods

Several researchers have investigated the uncanny valley effect and its relation to their familiarity (MacDorman, 2006; Schwind, 2018; Schwind et al., 2018a; Tinwell et al., 2011). According to Schwind et al. (2018a), one may expect the uncanny valley effect to occur when the agent is familiar (similarly to human agents). The reason for this is that animals are very familiar to humans. As such, we selected a panda due to it being known worldwide. We expected the panda to inspire a sense of familiarity in participants of the study. The panda is considered a charismatic species (Ducarme et al., 2013; Kandel et al., 2015). Moreover, people perceive pandas as attractive, cute, and charming. This means we can analyze the panda not only on its familiarity aspect but also on other features such as attractiveness, commonality, naturalness, interestingness, and animateness in a virtual setting. For this study, an image of a real panda and five different virtual pandas were used as stimuli.

Survey procedure

Participants accessed the online survey using a hyperlink in Qualtrics. The survey took less than 5 min to complete. Participants were asked for their age, gender, current location, and how frequently they played videogames. After these demographic questions, all participants saw six different images of pandas and for each image they answered questions using semantic differential scales for each of the properties: familiarity, commonality, naturalness, attractiveness, interestingness, and animateness. The participants indicated their answers on a 9-point scale where the extremes were labeled (e.g., 1=Very strange and 5=Very familiar).

Participants

We recruited participants via Facebook and directly from the student population of Tilburg University. A total of 162 participants came from age groups of 10 to 20 (15.43%), 21 to 30 (74.07%), 31 to 40 (8.03%), and 41 to 50 (2.47%) years old.

3.2.4 Stimulus

As this study focuses on virtual animals, we propose a scale similar to the human-likeness scale: the animal-likeness scale. We placed six different faces of a panda (animal) on the animal-likeness scale, similar to that developed by Mori (Figure 3.2.1) (Mori, 1970; Mori et al., 2012; Tobe and Greenberg, 2004). A real panda face (World, 2013) is located on the high side of animal-likeness scale.

Subsequently, a face used with a photorealistic design (MotionCow, 2018) is placed lower. Next, models of a zombie panda, robot panda, stuffed panda, and mechanical panda are placed on the scale, gradually being less animal-like. The mechanical panda is considered to have the lowest number of animal-like features in this range.

Figure 3.2.1: Images of pandas adapted from Mori's uncanny valley graph on an animal-likeness scale.

3.2.5 Results

The main question in this study explores whether there is an uncanny valley effect in the perception of familiarity, commonality, naturalness, attractiveness, interestingness, and animateness perception towards a virtual animal (Figure 3.2.2). We found that the uncanny valley effect is present for measures of familiarity, commonality, naturalness, and attractiveness perception. However, no uncanny valley effect was found for perception of interestingness, and animateness. The robot animal is found in the uncanny valley for almost all the variables along with the zombie, except for perception of interestingness, and animateness.

3.2.6 Discussion

Unexpectedly, the robot panda (virtual robot) is found in the uncanny valley for the properties familiarity, commonality, attractiveness, and naturalness. Contrary to our expectations, these results show that robot animals are perceived less familiar, common, attractive, natural, animate, and interesting when compared to photorealistic and real animals. The robot animal is interesting and animate to users, but no big differences with the other images are found. This result may be explained by the fact that although many researchers are investigating how to increase the empathy and emotional connection of social interaction with robots for long term interaction, this relationship is not resolved yet (Yang et al., 2018). The robots still lack the social interaction skills of humans because machines cannot operate perfectly natural under the different real-world conditions. Additionally, these robots require more natural or biological appearances due to the uncanny valley effect. This finding, while preliminary, suggests that is better to use a virtual robot animal with natural appearance than an artificial appearance

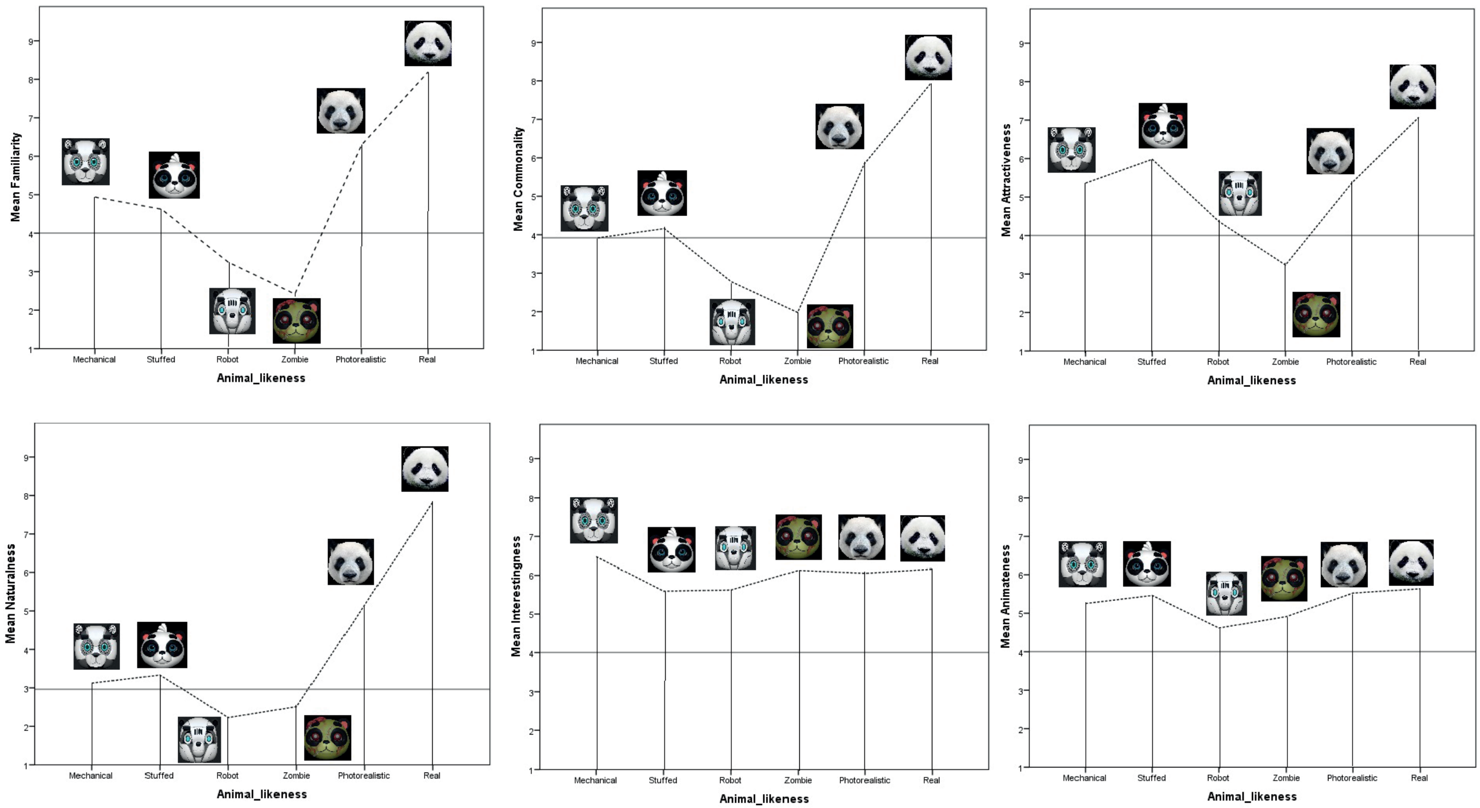


Figure 3.2.2: Mean perception of the six stimuli of the virtual panda organized on the x-axis on different properties y-axis.

3.2.7 Future work

Applications for the human-(animal) robot interaction as tutorial agents

Previous research developed by Li et al. (2016) found that students who watched video lectures with an animated robot (virtual NAO robot) had a better knowledge recall compared with a recording with a real robot (NAO robot). They concluded that video instructions can be affected by the appearance of the tutor (human or virtual appearance). For this reason, in future investigations, we are interested in using dynamic images (e.g., animations) compared to static images of the virtual animals to analyze whether the movement of virtual robot has similar uncanny valley effects for the user. Moreover, we are curious to understand how this research can be applied practically on robotics in education. For instance, we can use different versions of virtual animals (in particular robot animals) as virtual tutors in video instruction (Chen et al., 2007), where they may lead to different effects on affective and cognitive outcomes (see Figure 3.2.3) depending of their visual appearance. In these future investigations, the animal robot can have other types of effect to that we found in this study of uncanny valley of virtual animals.

Virtual animal robots in serious games and virtual reality

The results of the uncanny valley appearance of the robot animal extend our knowledge of the empathic reactions of users towards robots (Mori et al., 2012). Previous studies have explored the relationships between empathy and robots, especially in uncanny valley studies (Misselhorn, 2009; Riek et al., 2009; Zlotowski et al., 2018). These investigations are limited to physical prototypes of social robots, however, less is known of digital prototypes. Serious games and virtual reality simulations can be an attractive alternative as it provides interactive experiences that are similar to the real world but with robots animals. Further research should focus on determining whether users can have empathic reactions towards robots in virtual environments. This may have a number of important implications for future practice in new technology environments for teaching with robots in the classroom.

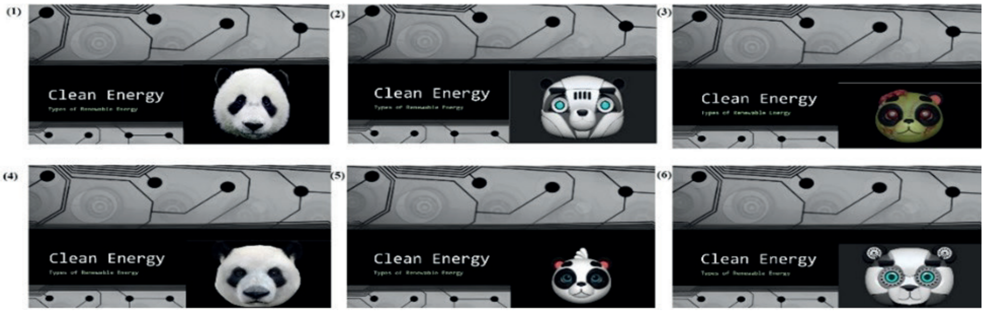


Figure 3.2.3: Example of a future application of the images into animal-likeness scale as virtual tutors.

Simulation robot labs: virtual animals

Simulation labs are fundamental tools in robotics education. In these simulation labs, students perform laboratory exercises, including mathematical models, electrical, mechanical, hydraulic systems, robotic models, physics, automated systems, and others. The android and mechanical robots are currently the major area of interest within the field of simulation lab (Balamuralithara and Woods, 2009). Up to now, too little attention has been paid to investigate the benefits of employing robot animals. As such, further investigation and experimentation into simulation robot labs with animal models is strongly recommended.

Interactive robotic avatars with tion real-time facial animal

One of the latest advances in technology is the "modeling for real-time facial animation" (Bouaziz et al., 2013). This software allows the natural use of avatars. Facial recognition software tracks the eyes, mouth and face posture of a human and maps this onto the interface (an image of humans, robots, animals, or objects). This software can be used by students and teachers, where they can create robotic avatars to make class presentations or interactive activities more engaging with more motivating effects (see Figure 3.2.4). These interactive robotic avatars can have a great potential in human-robot interaction. However, users should experience these avatars as human or animal like.

3

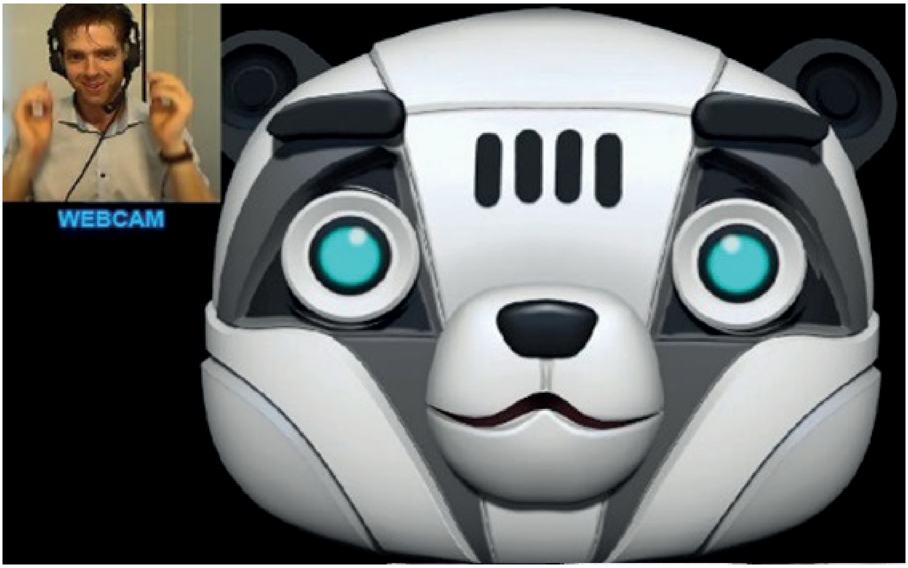


Figure 3.2.4: Example of a future application robotic panda as a real-time facial interactive animation.

3.2.8 Conclusion

In this investigation, the aim was to explore the presence of the uncanny valley effect associated to virtual animals, especially in robot animals. The results of this study indicate that robot animals fall into uncanny valley on the properties of perception of familiarity, commonality, naturalness, and attractiveness. One unanticipated finding was that for the robot animal no uncanny valley effect is found for the properties of perception of interestingness, and animateness. The generalizability of these results is subject to certain limitations. For instance, in our experiment, we analyzed the perception of a panda as a virtual animal, but we are not sure whether similar uncanny valley effects are found when other types of animals are used. Keeping the knowledge of the uncanny valley effects into account, further research should explore the effects of virtual animal robot interaction when used as tutorial agents, in video games, simulation labs, and used for real-time facial animation.

3.3 The uncanny valley of a virtual animal

3.4 Abstract

Virtual robots, including virtual animals, are expected to play a major role within affective and aesthetic interfaces, serious games, video instruction, and the personalization of educational instruction. Their actual impact, however, will very much depend on user perception of virtual characters as the uncanny valley theory has shown that the design of virtual characters determines user experiences. In this article, we investigated whether the uncanny valley effect, which has already been found for the human-like appearance of virtual characters, can also be found for animal-like appearances. We conducted an online study (N = 163) in which six different animal designs were evaluated in terms of the following properties: familiarity, commonality, naturalness, attractiveness, interestingness, and animateness. The study participants differed in age (under 10-60 years) and origin (Europe, Asia, North America, and South America). For the evaluation of the results, we ranked the animal-likeness of the character using both expert opinion and participant judgments. Next to that, we investigated the effect of movement and morbidity. The results confirm the existence of the uncanny valley effect for virtual animals, especially with respect to familiarity and commonality, for both still and moving images. The effect was particularly pronounced for morbid images. For naturalness and attractiveness, the effect was only present in the expert-based ranking, but not in the participant-based ranking. No uncanny valley effect was detected for interestingness and animateness. This investigation revealed that the appearance of virtual animals directly affects user perception and thus, presumably, impacts user experience when used in applied settings.

3.4.1 Related work

Technological advances in animation over the past decade have brought upon new challenges regarding the design of virtual characters (in robotic, animal, or human forms). In particular, a high quality, realistic appearance can negatively impact user acceptability (Schwind et al., 2018a). This negative effect has been described by the widely known uncanny valley theory (Mori, 1970). According to the theory, human-like appearance of artificial objects (e.g., robots) can, in some cases, have a detrimental impact on their perceived familiarity and affinity towards them. The uncanny valley theory provides a useful account of which features of the appearance give rise to a low user acceptance. This theory has been used in robotics, video games, virtual reality, and computer-animated movies in which virtual character appearance serves to establish emotional connection with the audience.

Past studies of the uncanny valley theory clearly demonstrate the importance of human-like properties. In this context, researchers have evaluated the impact of virtual humans used as pedagogical agents, bots, chatbots, virtual android robots, and avatars (Ciechanowski et al., 2019; Lugrin et al., 2018). Recent work by Schwind et al. (2018a) established the possibility to analyze the uncanny valley theory for virtual animals as well. However, in their work, a human-likeness scale was employed rather than an animal-likeness scale. Arguably, when studying the perception of virtual animals, an animal-likeness scale would be more appropriate.

To investigate the uncanny valley effect for virtual animals using an animal-likeness scale, we designed a virtual panda with different features, building on previous work including the scale originally developed by Mori (1970). Next to that, we explored the effect of using either still or moving (animated) images and of morbid appearance features. Below, we first summarize existing research findings and subsequently report on the outcomes of our online experiment in which characters were ranked for animal likeness both by experts and by naive participants.

3.4.2 The uncanny valley theory

The uncanny valley model, as visualized in the well-known graph presented by Mori (1970) and Mori et al. (2012), captures the relation between the appearance of robots and the sense of "shinwakan" a person feels towards them. "Shinwakan" is a Japanese word that can be translated with terms such as "familiarity", "affinity", and "comfort level" (Ho and MacDorman, 2010). The uncanny valley theory describes how the feeling of familiarity increases steadily when an artificial object obtains more human-like aspects. However, at some point, the familiarity suddenly drops, creating what is known as the uncanny valley effect. After this dip, the familiarity rises again sharply when the character's appearance approaches the resemblance of a the uncanny valley effect (the "dip") is most likely to occur when it is difficult to determine whether an object is alive/animate or dead/inanimate, for example, in the case of a zombie or an extremely realistic human-robot (Tinwell et al., 2011).

Mori et al. (2012), in a translation of Mori's original article, described the shape of the uncanny valley as follows:

"*The mathematical term monotonically increasing function describes a relation in which the function y=f(x) increases continuously with the variable x. [. . .] An example of a function that does not increase continuously is climbing a mountain—the relation between the distance (x) traveled by a hiker toward the summit and the hiker's altitude (y)—owing to the intervening hills and valleys. I have noticed that, in climbing toward the goal of making robots appear like a human, our affinity for them increases until we come to a valley [. . .] which I call the uncanny valley*" (See Figure 3.4.1, pp. 98).

Despite some inconclusive evidence, the effect has been demonstrated both for robots and virtual characters. While some studies in the past did not find an uncanny valley effect for familiarity, strangeness, and eeriness (Bartneck et al., 2009; MacDorman, 2006), a number of recent articles describe the effect in likability and eeriness for the appearance of android robots (Li, 2015; Zlotowski et al., 2018) and familiarity for computer-generated characters (Dill et al., 2012).

3.4.3 The uncanny valley in virtual animals

Previously published studies on the uncanny valley theory have been limited to robots or virtual characters resembling humans, such as androids or virtual humans. Virtual animals, however, are currently quite popular in the entertainment industry and have only received limited attention in the area of human-computer interaction studies. To our knowledge, there are only two studies that explore the uncanny valley effect for virtual animals. The first study, by Schneider et al. (2007), reported that non-human characters (e.g., animals and robots) are perceived favorably if they display human-like features. The second study, developed by Schwind et al. (2018a), noted that it is possible to avoid the uncanny valley for a virtual animal when it is completely natural or when it has a stylized appearance. Their study concentrated on how realism has an effect on a virtual animal (e.g., a cat). As far as we know, no studies analyzed in depth whether animal-like appearance leads to an uncanny valley effect. In this sense, a systematic understanding of how the uncanny valley theory contributes to the perception of virtual animals and their possible effect on the users' experience is still lacking.

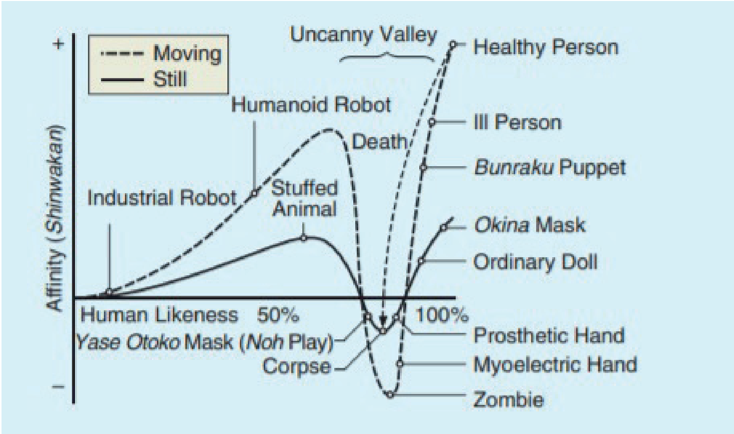


Figure 3.4.1: Original graph of the uncanny valley (reproduced from Mori et al. (2012)).

3.4.4 Current study

In our previous study (Sierra Rativa et al., 2019) we investigated whether the uncanny valley effect observed for human-like objects can be found for virtual animals. In particular, we determined that there is an uncanny valley effect for the user's perception towards still images of avirtual animal with respect to familiarity, commonality, naturalness, attractiveness, interestingness, and animateness. The current study first sets out to replicate this result and confirm that the uncanny valley theory can be applied to virtual animals using the original scales. Second, we tested the so-called 'movement hypothesis' according to which the movement of the object can influence affinity because bodily activity is important for the perception of live entities (Kätsyri et al., 2015). For this aim, we included dynamic stimuli in our experimental material. Finally, we explored the effect of morbid virtual character features in combination with other character traits Kätsyri et al., 2015).

Figure 3.4.2 shows a simulation graph of the expected uncanny valley effect for animal-likeness when directly related to the original human likeness uncanny valley graph presented by Mori (1970). To determine the existence of the uncanny valley for the animal-likeness scale, we propose that four conditions need to be met: (1) the graph decreases when x is between points (a) and (c); (2) the uncanny valley (point c) must be found near the maximum value of x; (3) there are significant differences on the y-axis when varying on the x-axis; and (4) the real animal (at the right side of the graph) compared to uncanny valley images of the animal at point c, are on opposite extremes in the plot.

According to Kätsyri et al. (2015, pp. 4–7), five main hypotheses can influence the uncanny valley effect. Of these hypotheses, we are interested in the movement hypothesis, which indicates that movement in the stimuli can amplify the uncanny valley curve (both in positive and negative direction), and the morbidity hypothesis, which states that morbid characters lead to a more negative perception compared to other characters. Based on these hypotheses, we expect to find an uncanny valley graph that can be affected by ranking the animal likeness scale, the movement of stimuli and the morbidity of images. To our knowledge, these hypotheses have not previously been tested on virtual animals.

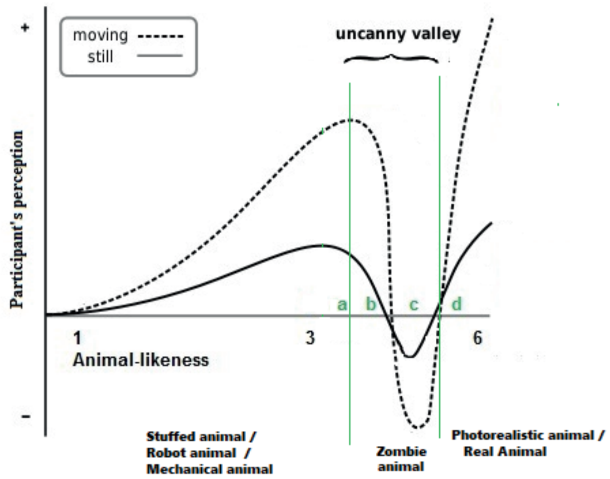


Figure 3.4.2: Expected uncanny valley graph relating participant's perception to the animal-likeness scale. The uncanny valley occurs between points a and c.

The overarching research question addressed in this article is as follows:

Which conditions influence the existence of an uncanny valley effect for virtual animals?

In order to answer this question, in our experiment, we focused on three design features likely to influence the uncanny valley effect, two of which movement and morbidity are based on the findings of Kätsyri et al. (2015). The following research sub-questions were investigated:

1. **Ranking**. Does the expert-based ranking of virtual animals lead to different uncanny valley effects than the participant-based ranking of virtual animals?
2. **Movement**. Does movement of the virtual animals amplify the affinity responses (changes in familiarity, commonality, naturalness, attractiveness, interestingness, and animateness) compared to still images of the virtual animals?
3. **Morbidity**. Does a morbid virtual animal (e.g., one with zombie features) elicit more negative familiarity, commonality, naturalness, attractiveness, interestingness, and animateness than other characters?

In the sections below, we first describe our experimental methods followed by the quantitative and qualitative results of the study and their discussion.

3.4.5 Methods

This study aimed to collect information on the perception of familiarity, attractiveness, commonality, naturalness, interestingness, and animateness using six still images as well as six moving videos of virtual pandas. Participants ranked the pandas on their level of animal-likeness using a scale adapted from Mori's original human-likeness scale. The participant-based ranking was compared to expert-based ranking.

Participants

In total, 163 participants were recruited via the University of Gran Colombia (Colombia), High School "I.E.D Quiroga Alianza" (Colombia), High School "I.E.D Almirante Padilla" (Colombia), and directly from the student population of Tilburg University (the Netherlands). We provide demographic information of the participants in Table 3.4.1.

Survey procedure

The survey was developed in Qualtrics (a web-based platform for distributing surveys). Participants accessed the online survey using a hyperlink. For the participants younger than 18 years old, informed consent was obtained from their parents and caretakers. The survey, which contained 79 questions, took approximately 10 minutes to complete. We collected participants' information about their age, gender, the country where they lived longest, the highest educational degree they received, and how frequently they played video games.

Table 3.4.1: Demographic data of the participants.

Variable	Description	n=163	%
Age	under 10	2	1.2
	10-19	65	39.9
	20-29	78	47.9
	30-39	16	9.8
	40-50	1	0.6
	over 50	1	0.6
Gender	Female	81	49.7
	Male	82	50.3
Geographic location	Asia	8	4.9
	Asia/Europe	2	1.2
	Europe	73	44.7
	North America	1	0.7
	South America	79	48.5
Frequency of playing video games	Daily	27	16.6
	Several times a week	43	26.4
	Several times a month	18	11
	Several times a year	35	21.5
	Never	40	24.5
Highlest grade or school level	Primary school	21	12.9
	Highschool	60	36.8
	MBO	7	4.3
	HBO/University of Applied Sciences	8	4.9
	University Bachelor	32	19.6
	University Master	32	19.6
	PhD	3	1.8

Each participant was presented all six (different) still images and movies of the pandas in a random order. For each image and movie, they answered questions using semantic differential scales anchored at the following poles: familiar/strange, common/unusual, attractive/ugly, interesting/boring, natural/artificial, and animate/inanimate. Data preprocessing and analysis were performed using SPSS version 27.0 (a software package used to perform statistical analysis on quantitative data).

Stimuli

The panda is considered a charismatic species (Ducarme et al., 2013; Kandel et al., 2015) and is perceived as attractive, cute, and charming by most respondents. As such, we expected the panda to stimulate a sense of familiarity in participants of the study. Researchers have explored the relationship between the uncanny valley effect on familiarity (MacDorman, 2006; Schwind et al., 2018a; Tinwell et al., 2011). However, here we extend this work and use the panda to investigate not only its familiarity aspect, but also attractiveness, commonality, naturalness, interestingness, and animateness traits in a virtual setting.

In this study, we use six versions of the virtual panda: 1. mechanical panda (own design), 2. stuffed toy panda (own design), 3. robot panda (own design), 4. zombie panda (own design), 5. photo-realistic panda (MotionCow, 2018), and 6. real panda (World, 2013). The original images were 1024 x 768 pixels, and Qualtrics automatically resized the images to 551 x 301 pixels.

The moving panda images used in this study are based on the still images and have been animated using software called "Crazy animator 8" (https://www.reallusion.com/crazytalk/download.html).

Figure 3.4.3 shows the fitting face editor used in Crazy Animator 8 to control the facial points of the animal's face that are used to animate it. All pandas have a neutral expression. Each animation has a 10 second duration. In total this means that six still images and six animated versions of the virtual pandas are used as stimulus material in this study. The videos of the virtual pandas can be found via the following links:

1. Mechanical panda (https://youtu.be/6SrCiUxszuY)
2. Robot panda (https://youtu.be/W-D2fLekMd8)
3. Stuffed toy panda (https://youtu.be/2ULKk4-CX98)
4. Zombie panda (https://youtu.be/O3pzsEVu-Ys)
5. Photorealistic panda (https://youtu.be/hpSN9l2fsZ0)
6. Real panda (https://youtu.be/tL2WJTO272E)

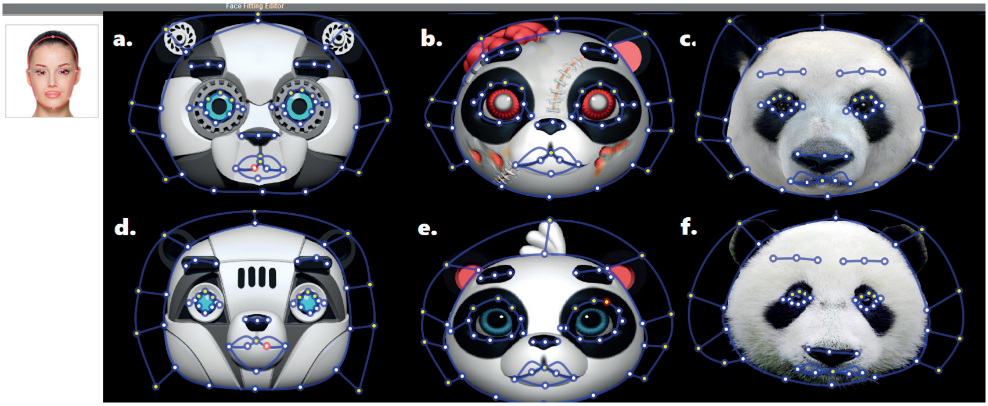


Figure 3.4.3: Face fitting of the still images of the pandas to create animated pandas.

3

Measurement of animal-likeness scale

Expert-based ranking

We ranked the six different faces of a panda according to the animallikeness scale, which is based on the human-likeness scale developed by Mori (1970) (see Figure 3.4.4, left column). Masahiro Mori, a Japanese roboticist and the author of the uncanny valley theory proposed a human-likeness scale on which he placed thirteen types of stimuli in the original graph (see Figure 3.4.1): an industrial robot, a humanoid robot, a stuffed toy animal, a zombie, a bunraku puppet, and a healthy human. Following the same expert-based ranking, we ordered the panda characters as follows:

1. Mechanical panda (i.e., a simulation of the industrial robot in Mori's studies);
2. Stuffed toy panda (i.e., a simulation of the stuffed animal in Mori's studies);
3. Robot panda (i.e., a simulation of the humanoid robot in Mori's studies);
4. Zombie panda (i.e., a simulation of the zombie in Mori's studies);
5. Photo-realistic panda (i.e., a simulation of the bunraku puppet in Mori's studies) (MotionCow 2018);
6. Real panda (i.e., a simulation of the healthy person in Mori's studies) (World 2013).

Here, panda 1 represents the least animal-like panda, whereas panda 6 represents the most animal-like panda. We call this ranking the expert-based ranking in this article as it follows the ranking proposed in Mori's research.

Participant-based ranking

Since it is possible that non-experts perceive the animal-likeness of the pandas differently from the expert-based ranking, participants were instructed to rank the six panda images based on animal-likeness with the value of 1 representing the most animal-like and 6 the least animal-like character. The order of the presentation of the images in the questionnaire was randomized. To align the participant-based ranking with the expert-based ranking (x-axis in the figures), the participant scale was subsequently inverted in the analysis.

Questionnaire of perception regarding virtual characters

Following the ranking, each participant was presented with six still images of the pandas (randomized) as shown in Figure 3.4.4a and for each image answered the questions provided in Table 3.4.2. Next, the participants were presented with six animations of the pandas (randomized) as shown in Figure 3.4.4b and answered the same questions as for the still images. A reliability analysis using Cronbach's alpha revealed that the questions were answered consistently (α=0.948).

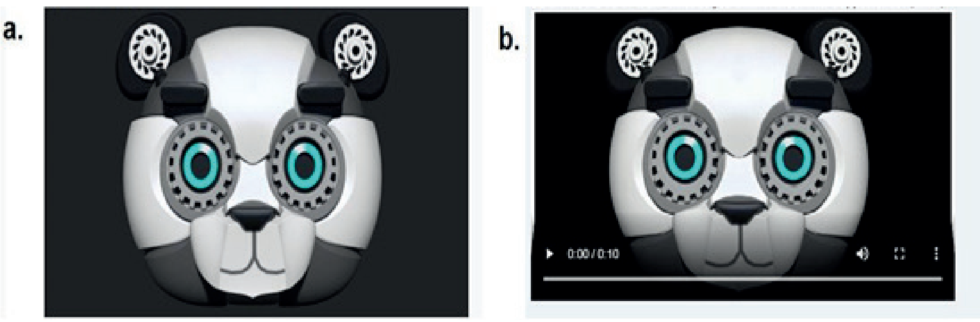


Figure 3.4.4: A still image of the mechanical panda (a) and the interface of the mechanical panda as a moving image (b).

Ethics Approval

Ethics approval was received from the Research Ethics Committee of the Tilburg School of Humanities with the reference REDC#2019/89. The data that support the findings of this study are openly available in the Dataverse repository name "Data of uncanny valley of a virtual animal" at https://doi.org/10.34894/JIBXBU.

3.4.6 Results

In the analysis below, we first compared the expert-based ranking to the participant-based ranking. Next, we investigated the relation between the judgments of familiarity, commonality, naturalness, attractiveness, interestingness, and animateness, as well as the effect of movement and morbidity.

Table 3.4.2: Perception of virtual characters questionnaire.

What do you think of the animal's appearance?											
Familiarity	Strange	1	2	3	4	5	6	7	8	9	Very familiar
Commonality	Very unusual	1	2	3	4	5	6	7	8	9	Very common
Attractiveness	Very ugly	1	2	3	4	5	6	7	8	9	Very attractive
Interestingness	Very boring	1	2	3	4	5	6	7	8	9	Very interesting
Naturalness	Very artificial	1	2	3	4	5	6	7	8	9	Very natural
Animateness	Very inanimate	1	2	3	4	5	6	7	8	9	Very animate

Raking

Figure 3.4.5 provides an overview of the participant-based ranking in terms of animal-likeness. The real panda was ranked to be most animal-like by 134 (82%) respondents. The photorealistic panda was ranked second by 127 respondents (77%). The stuffed toy panda was ranked next on the scale by 110 respondents (67%), followed by the zombie panda 76 respondents (46%) and the mechanical panda 67 respondents (41%). Finally, the robot panda was ranked as least animal-like by 65 respondents (40%). Note that participants agreed mostly on the ranking of the real, photorealistic, and stuffed toy panda, but were less consistent in their ranking of the other characters. Overall, this resulted in the following ranking from most to least animal-like:
real —photorealistic —stuffed toy —zombie —mechanical —robot.

Comparing the expert-based ranking to the participant-based ranking (as illustrated in Figure 3.4.6), we see that the rankings of the real and photo-realistic pandas are the same, but the other pandas are ranked differently. In particular, participants ranked the stuffed toy panda as more animal-like compared to the expert-based ranking and the robot and mechanical pandas swapped places in the ranking. This means that the expert-based ranking is different from the participant-based ranking.

	Ranking according to animal likeness by participants					
	1	2	3	4	5	6
Robot	39.63%	32.32%	14.02%	9.76%	3.05%	1.22%
Mechanical	18.90%	40.85%	21.95%	12.80%	3.05%	2.44%
Zombie	33.54%	9.15%	46.34%	5.49%	4.27%	1.22%
Stuffed	3.66%	9.15%	10.98%	67.07%	3.66%	5.49%
Photo realistic	3.05%	6.10%	3.05%	2.44%	77.44%	7.93%
Real panda	1.22%	2.44%	3.66%	2.44%	8.54%	81.71%

Figure 3.4.5: Result of the animal-likeness ranking of the pandas chosen by participants (1 is least animal-like and 6 is most animal-like).

Uncanny valley

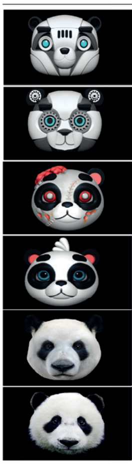
Below, we present the analysis of the uncanny valley scales for perceived familiarity, commonality, naturalness, attractiveness, interestingness, and animateness.

Figure 3.4.7a plots the mean values of the different properties assigned to the different still panda images according to the expert-based ranking. The familiarity perception starts with the mechanical panda (M=3.62, SD=2.311) and moves up toward the stuffed toy panda (M=5.54, SD=2.335). After this, familiarity drops toward the robot panda (M=3.49, SD=2.330) and zombie panda (M=2.67, SD=2.000).

a.

Type of animal	Expert-based ranking of virtual animals
Mechanical	1
Stuffed	2
Robot	3
Zombie	4
Photo realistic	5
Real Animal	6

b.



Type of animal	Participant-based ranking of virtual animals
Robot	
Mechanical	
Zombie	1 to 6
Stuffed	
Photo realistic	
Real Animal	

Figure 3.4.6: Expert-based ranking and participant-based ranking (1 is least animal-like and 6 is most animal-like).

Next, the familiarity rises gradually from the photorealistic panda (M=6.89, SD=2.316) to the real panda (M=7.88, SD=2.044).

Likewise, the perception of commonality starts from the mechanical panda (M=3.23, SD=2.157) and rises toward the stuffed toy panda (M=4.98, SD=2.252). After the stuffed toy panda, the commonality value decreases for the robot panda (M=3.09, SD=1.823) and zombie panda (M=2.70, SD=1.995). After that, the commonality increases progressively from the photorealistic panda (M=6.61, SD=2.247) to the real panda (M=7.71, SD=1.922).

The attractiveness perception shows similar behavior to familiarity and commonality. There is a lower value in attractiveness perception for the mechanical panda (M=4.65, SD=2.121). Next, it increases toward the stuffed toy panda (M=5.31, SD=2.320). Then attractiveness decreases from the robot panda (M=4.32, SD=2.190) to the zombie panda (M=3.70, SD=2.348). Following these low values, the attractiveness increases gradually from the photorealistic panda (M=5.98, SD=2.206) to the real panda (M=7.18, SD=2.197).

The perception of naturalness follows the same behaviour as the previous perceptions. The perception of naturalness begins with the mechanical panda (M=2.93, SD=2.390), and increases to the stuffed toy panda (M=4.88, SD=2.332). After that, it decreases to the robot panda (M=2.51, SD=1.974) and zombie panda (M=2.94, SD=2.239). Next, the naturalness rises gradually from the photorealistic panda (M=6.62, SD=2.382) to the real panda (M=7.74, SD=2.215).

In contrast, interestingness and animateness do not show the similar trends as familiarity, commonality, attractiveness and naturalness. The interestingness values show only moderate changes from the mechanical panda (M=5.76, SD=2.212) to the real panda (M=6.46, SD=2.353). The animateness perception also shows moderate changes between the mechanical panda (M=5.42, SD=2.818) and the real panda (M=5.32, SD=3.107). Summarizing, we find an uncanny valley effect for the measures of familiarity, commonality, attractiveness, and naturalness toward still images when organized according to the expert-based ranking, but no such effects are found for the measures of interestingness and animateness.

Figure 3.4.7b plots the mean familiarity assigned to the different still panda images where the pandas are organized by the participant-based ranking. The familiarity perception starts with the robot panda (M=3.49, SD=2.330) and moves toward the mechanical panda (M=3.62, SD=2.311). After this, the familiarity decreases toward the zombie panda (M=2.67, SD=2.83), after which the familiarity increases again gradually from the stuffed toy panda (M=5.54, SD=2.335) to the real panda (M=7.88, SD=2.044).

Similarly, the perception of commonality starts with the robot panda (M=3.09, SD=1.823), then decreases from the mechanical panda (M=3.23, SD=2.157) to the zombie panda (M=2.70, SD=1.995). After this, the commonality increases gradually from the stuffed toy panda (M=4.98, SD=2.252) to the real panda (M=7.71, SD=1.922).

The attractiveness perception follows similar behavior of familiarity and commonality, where there is a decrease in attractiveness starting with the robot panda (M=4.32, SD=2.190) and decreasing from the mechanical panda (M=4.65, SD=2.121) to the zombie panda (M=3.70, SD=2.348).

Next, the attractiveness increases gradually from the stuffed toy panda (M=5.31, SD=2.320) to the real panda (M=7.18, SD=2.197). In contrast, naturalness, interestingness, and animateness show different behavior. The naturalness perception shows a steady rise from the robot panda (M=2.51, SD=1.974) to the real panda (M=7.74, SD=2.215). The interestingness perception shows moderate changes between the robot panda (M=5.30, SD=2.303) to the real panda (M=6.46, SD=2.323) and the animateness perception also shows moderate changes between the robot panda (M=4.90, SD=2.846) to the real panda (M=5.32, SD=3.107). Summarizing, we find an uncanny valley effect for the measures of familiarity, commonality, and attractiveness toward still images organized on the participant-based ranking, but no such effects are found for the measures of naturalness, interestingness, and animateness.

When considering the perception of the moving images according to the expert-based ranking (as shown in Figure 3.4.8a), we see that the familiarity perception shows the lowest value for the mechanical panda (M=3.66, SD=2.279), increases to the stuffed toy panda (M=4.80, SD=2.188). After this, familiarity drops toward the robot panda (M=3.76, SD=2.297) and zombie panda (M=2.83, SD=2.050). Next, the familiarity rises steadily from the photorealistic panda (M=5.90, SD=2.525) to the real panda (M=6.49, SD=2.551).

The commonality scores reveal similar behavior to familiarity with the mechanical panda (M=3.51, SD=2.074), then rising to the stuffed toy panda (M=4.38, SD=2.055). After this, the score decreases to the robot panda (M=3.54, SD=2.068) and zombie panda (M=2.80, SD=1.922). The commonality score then increases gradually from the photorealistic panda (M=5.77, SD=2.441) to the real panda (M=6.04, SD=2.429).

The perception of familiarity and commonality of moving images had the same behavior as that of the still images by expert-based ranking. The attractiveness perception values follow similar behavior to that of familiarity and commonality, where we find a decrease in attractiveness starting with the mechanical panda (M=4.79, SD=2.236) and increasing to the stuffed toy panda (M=5.84, SD=2.079). After this the scores decrease to the robot panda (M=4.72, SD=2.296) and zombie panda (M=3.46, SD=2.197). Next, the attractiveness increases gradually from the photorealistic panda (M=5.63, SD=2.241) to the real panda (M=6.52, SD=2.077).

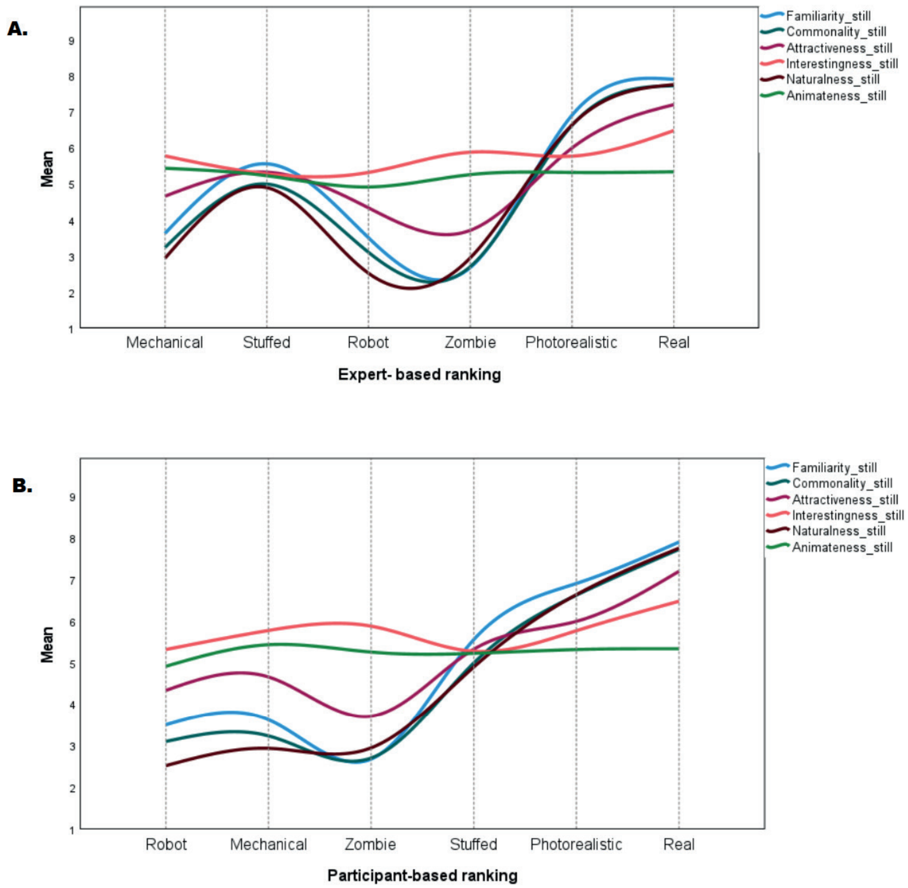


Figure 3.4.7: Graphs of the perceived familiarity, commonality, naturalness, attractiveness, interestingness, and animateness for the six moving stimuli organized on the x-axis according to (a) expert-based ranking and (b) participant-based ranking of animal-likeness.

Likewise, the naturalness perception begins with the mechanical panda (M=3.25, SD=2.327) and increases with the stuffed toy panda (M=4.05, SD=1.981). After dropping to the robot panda (M=3.01, SD=2.193) and zombie panda (M=2.84, SD=1.915), the naturalness score rises gradually from the photorealistic panda (M=5.93, SD=2.420) to the real panda (M=6.43, SD=2.420).

The perception of interestingness only shows slight changes between the mechanical panda (M=5.50, SD=2.239) and the real panda (M=6.24, SD=2.199). Similarly, the animateness perception values show moderate variation between the mechanical panda (M=5.80, SD=2.531) and the real panda (M=5.68, SD=2.491). To summarize, uncanny valley effects can be found for familiarity, commonality, attractiveness, and naturalness, but no such effects are found for interestingness and animateness for the moving images when considering the expert-based ranking.

Looking at the perception values for the moving images ordered according to the participant-based ranking, we see that the familiarity perception shows the lowest value for the robot panda (M=3.76, SD=2.297), moving toward the mechanical panda (M=3.66, SD=2.279). After this, familiarity decreases toward the zombie panda (M=2.80, SD=1.922). Next, the familiarity increases progressively from the stuffed toy panda (M=4.38, SD=2.055) to the real panda (M=6.04, SD=2.429).

The commonality perception shows similar behavior to that of familiarity. As can be seen from Figure 3.4.8b, commonality starts with the robot panda (M=3.54, SD=2.068), showing a decrease from the mechanical panda (M=3.51, SD=2.074) to the zombie panda (M=2.80, SD=1.922). After this, commonality increases gradually from the stuffed toy panda (M=4.38, SD=2.055) to the real panda (M=6.04, SD=2.429). The perception of familiarity and commonality of moving images has similar behavior to that of the still images.

Likewise, the naturalness perception starts with the robot panda (M=3.01, SD=2.193), then showing a decrease from the mechanical panda (M=3.25, SD=2.327) to the zombie panda (M=2.84, SD=1.915). Next, the naturalness increases gradually from the stuffed toy panda (M=4.05, SD=1.981) to the real panda (M=6.43, SD=2.420). In contrast, attractiveness shows significant variation between the stuffed toy (M=5.84, SD=2.079) and photorealistic pandas (M=5.63, SD=2.241) and we expected the photorealistic panda to have a higher score than the stuffed toy panda.

The perception of interestingness toward the moving images has moderate variation between the robot panda (M=5.21, SD=2.275) and the real panda (M=6.24, SD=2.199). Also, the animateness perception has moderate changes between the robot panda (M=5.75, SD=2.541) and the real panda (M=5.68, SD=2.491). To summarize, uncanny valley effects are found for the perception of familiarity, commonality, and naturalness, but no such effects are found for attractiveness, interestingness, and animateness for the moving images when considering the participant-based ranking.

We can observe the different orders of animal-likeness (participant-based ranking versus expert-based ranking) in relation to the uncanny valley effects by comparing the graphs of Figures 3.4.7a and 3.4.8b as well as 3.4.8a and 3.4.9b. Overall, we see consistent patterns: perception of familiarity, commonality, attractiveness, and naturalness display an uncanny valley effect. However, for naturalness in the participant-based ranking, the uncanny valley effect is unclear and the shape of the graph for attractiveness for the still images is slightly different in the participant-based ranking. In contrast, measures of interestingness and animateness do not show such effects. This holds for both still as well as moving images. The differences between the participant-based and expert-based ranking are consistent as well. Due to the differences in ranking, the uncanny valley effects are already found with less animal-like pandas (to the left of the graph) for the participant-based ranking. To sum up, the ranking of animal-likeness does not seem to affect the main characteristics of the uncanny valley line graphs. The only difference is in the naturalness case, which cannot be found for the participant-based ranking. Similar patterns are also found comparing still and moving images.

Movement

The second research question focuses on whether the movement of the virtual animals can amplify the affinity responses (familiarity, commonality, naturalness, attractiveness, interestingness, and animateness) compared to still images of the virtual animal. Figure 3.4.9 provides graphs of both still and moving images for the perception of familiarity, commonality, naturalness, attractiveness, interestingness, and animateness. This representation shows small changes in the participant's perceptions when the virtual animal is still versus when the virtual animal is moving. We can observe that the values for the moving pandas are slightly amplified for familiarity and commonality. Similarly, for the expert-based ranking, we see an amplification for naturalness, but this is less clear for the participant-based ranking. In contrast, for attractiveness, we see that the still images have more extreme values. Given that interestingness and animateness do not have uncanny valleys, no clear effect on the affinity responses can be established.

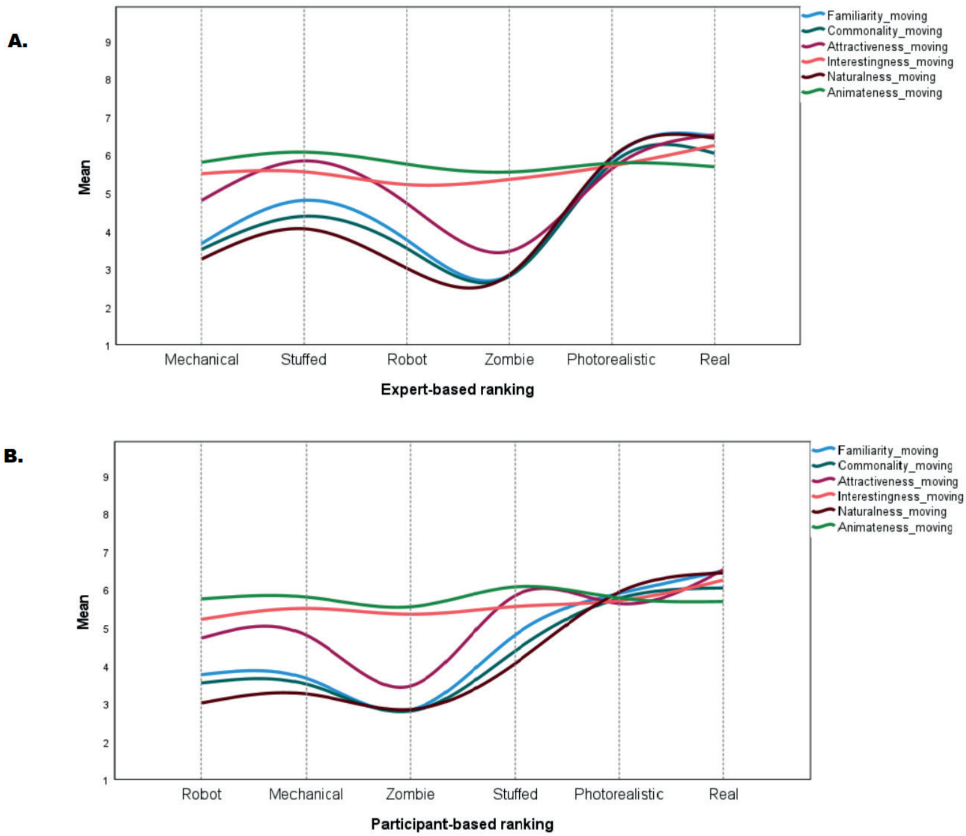


Figure 3.4.8: Graphs of the perceived familiarity, commonality, naturalness, attractiveness, interestingness, and animateness for the six moving stimuli organized on the x-axis according to (a) expert-based ranking and (b) participant-based ranking of animal-likeness.

Morbidity

In line with previous literature, we hypothesized that a morbid virtual animal (e.g., one with zombie features) will elicit more negative familiarity, commonality, naturalness, attractiveness, interestingness, and animateness than other characters.

The Kruskal-Wallis tests[2] (data not normally distributed) show that there were significant differences for the familiarity of still images ($\chi^2(5) = 412.310$, $p \leq 0.001$), familiarity of moving images ($\chi^2(5)=227.446$, $p \leq 0.001$), commonality of still images ($\chi^2(5)=434.438$, $p \leq 0.001$), commonality of moving images ($\chi^2(5)=222.322$, $p \leq 0.001$),

2 Additional Median information about statistics analysis can be found in Appendix E.

attractiveness of still images ($\chi^2(5)=210.907, p \leq 0.001$), attractiveness of moving images ($\chi^2(5)=163.988$, $p \leq 0.001$), naturalness of still images ($\chi^2(5)=419.731$, $p \leq 0.001$), naturalness of moving images ($\chi^2(5)=275.973, p \leq 0.001$), interestingness of still images ($\chi^2(5)=30.473$, $p \leq 0.001$), and interestingness of moving images ($\chi^2(5)=21.448$, $p \leq 0.001$). However, no significant differences were found for animateness of still images ($\chi^2(5)=3.362$, $p \leq 0.644$) and animateness of moving images ($\chi^2(5)=2.316$, $p \leq 0.804$).

The differences between the panda images were further explored using post-hoc tests for familiarity, commonality, attractiveness, naturalness, and interestingness (see Figure 3.4.10). We will first consider the still images.

Most of the combinations between images for the familiarity perception were significantly different ($p \leq 0.001$), except for the robot and mechanical versions of the panda ($p = 0.712$). The majority of the combinations between images for the perception of commonality were significantly different ($p < 0.001$), except for the zombie and robot versions of the panda ($p = 0.125$), and the photorealistic and real versions ($p = 0.699$). Similarly, most of the combinations between images in the attractiveness perception were significantly different ($p \leq 0.001$), except for the robot and mechanical versions of the panda ($p = 0.291$).

For naturalness, we found that the majority of the combinations were significantly different, except of the robot and mechanical versions of the panda ($p = 0.228$), the robot and zombie versions of the panda ($p = 0.194$), and the mechanical and zombie versions of the panda ($p = 0.925$). In contrast, for interestingness many of the combinations were not significantly different and the only combinations that were significantly different were the robot, mechanical, zombie, and stuffed toy pandas with respect to the real panda ($p \leq 0.001$).

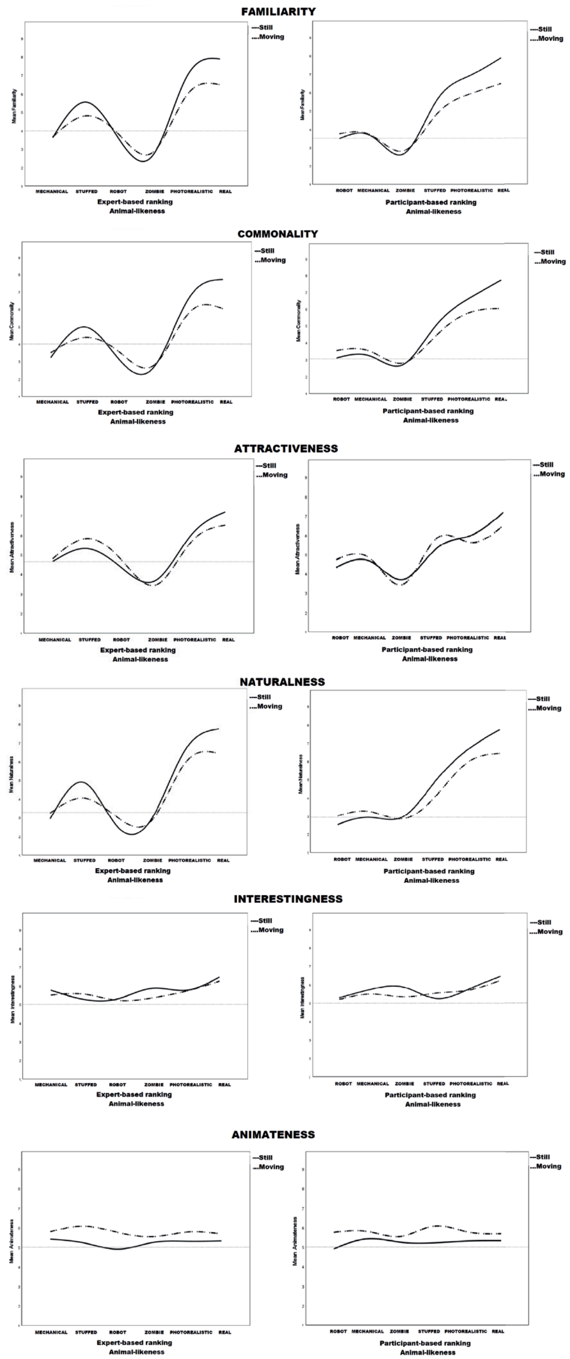


Figure 3.4.9: Graphs of the perceived familiarity, commonality, attractiveness, naturalness, interestingness, and animateness of the virtual pandas for still and moving images according to the expert-based and participant-based rankings.

We now consider the differences between the values of the moving images. The majority of the combinations between the images for the familiarity perception were significantly different ($p \leq 0.001$), except for the robot and mechanical versions of the panda (p=0.730) and the photorealistic and real versions (p=0.068). Most of the combinations between images for the commonality perception were significantly different $p \leq 0.001$, except for the mechanical and robot versions of the panda (p=0.905), and the photorealistic and real versions (p=0.382). Similarly, the majority of the combinations between images for the attractiveness perception were significantly different ($p \leq 0.001$), except for the robot and mechanical versions of the panda (p=0.886), and the photorealistic and stuffed toy versions (p=0.488). For naturalness, we found that the majority of the combinations were significantly different, except for the zombie and robot versions of the panda (p=0.650), the zombie and mechanical versions of the panda (p=0.196), the mechanical and robot versions of the panda (p=0.402), and the photorealistic and real version (p=0.132). In contrast, for interestingness many of the combinations were not significantly different. The only combinations that were significantly different were the robot, zombie, and stuffed toy versions with respect to the real version ($p \leq 0.001$).

We compared the results of the current study to those of our previous study (Sierra Rativa et al., 2020) where we used the identical still images with the exception of the color of the zombie panda (green as opposed to white in the current study). To investigate the effect of color, we performed a series of independent-samples t-tests (Figure 3.4.11). These demonstrated significant differences for commonality, $t(324) = -3.983$, $p \leq 0.001$ and attractiveness, $t(324) = -1.973$, p=0.049. No significant differences for familiarity, $t(324) = -1.249$, p=0.212; naturalness, $t(324) = -1.850$, p=0.065; interestingness, $t(324) = 1.090$, p=0.276, and animateness, $t(324) = -1.042$, p=0.298 were found.

3.4.7 Discussion

This study investigated the existence of the uncanny valley effect for virtual animals. The original publication on the theory of the uncanny valley by Mori (1970) and Mori et al. (2012) does not mention explicitly how the ordering on the human-likeness scale was established. However, Mori (1970) and Mori et al. (2012) notes that characters higher on the human-likeness scale have more anthropomorphic characteristics and characters lower on the scale have more object-like characteristics. In another study conducted by (Burleigh et al., 2013) participants were instructed to rate a face on a

Likert scale of human-likeness. In general, the human-likeness scale is associated with the perception of realism by participants (Schwind et al., 2018b; Burleigh et al., 2013). According to Kätsyri et al. (2015), human likeness perception can be influenced by aspects of the appearance or aesthetics of the character (e.g., a healthy or morbid appearance). Our study assumed that the human likeness scale can be converted into animal-likeness. We first compared two rankings of animal-likeness, one based on expert judgment and the other one participant-based. Building on our previous study, we explored the concept of animal-likeness in more detail using six subscales representing the properties of familiarity, commonality, naturalness, attractiveness, interestingness, and animateness. In order to test the effect of movement, participant responses were collected for both still images and moving images of virtual pandas. Finally, we examined the impact of morbid features given that morbidity has been assumed to influence the uncanny valley effect in the literature.

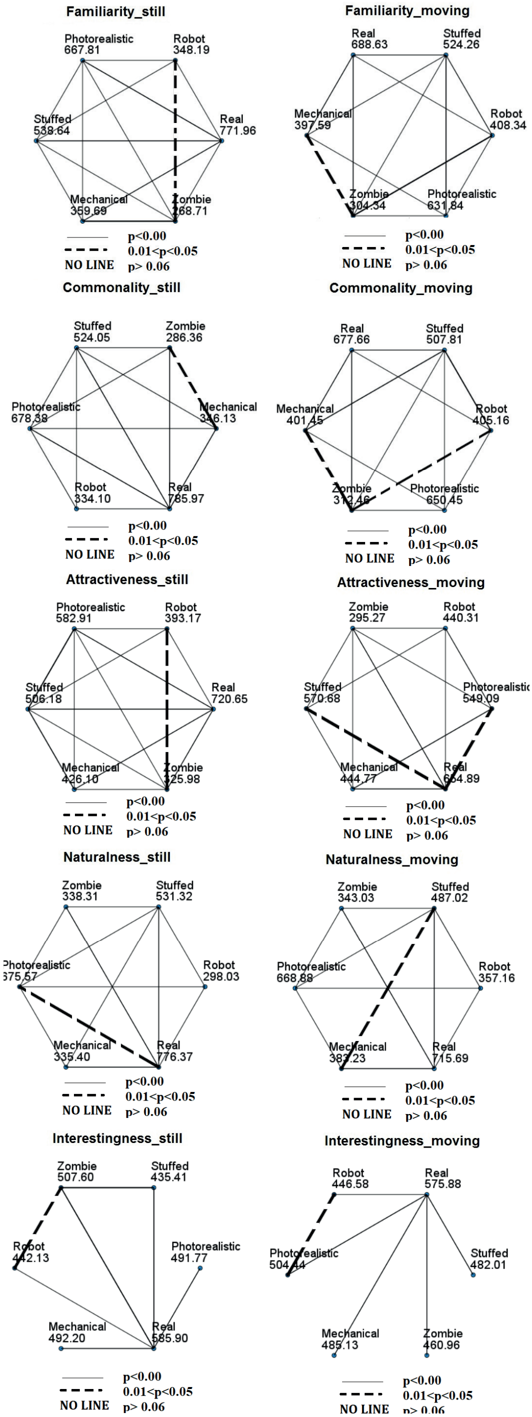


Figure 3.4.10: Summary graphs of familiarity, commonality, naturalness, attractiveness, interestingness, and animateness perception of the six stimuli of the virtual panda. Dashed lines indicate significant differences at $p<0.05$ and solid lines indicate significant differences at $p \leq 0.001$. The absence of lines indicates no significant differences.

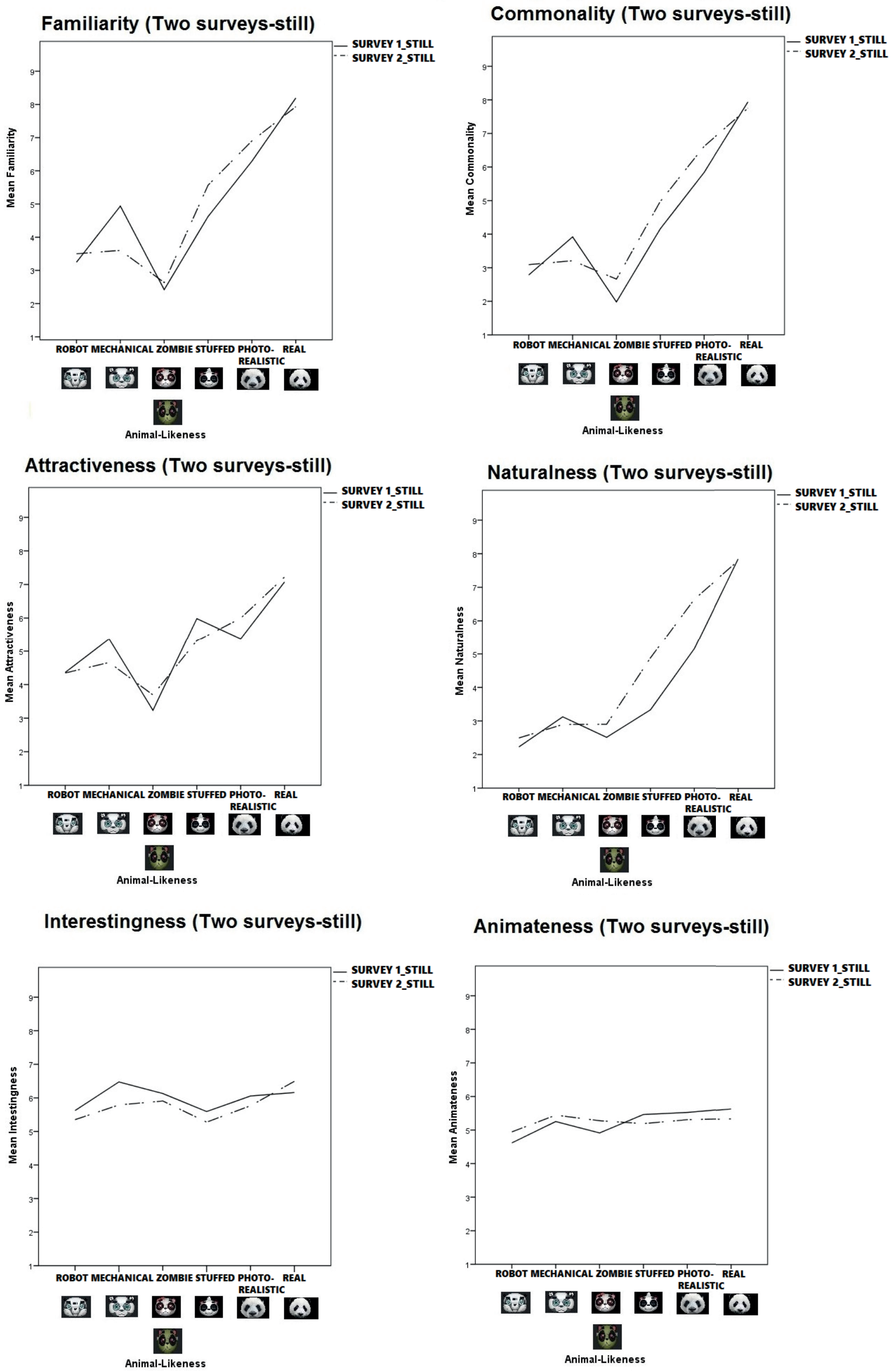


Figure 3.4.11: Graphs comparing the results of our earlier study (study I) and the current study (study II) for still images, where all results are ordered on participant-ranked animal likeness.

Ranking

The first research question of this study investigated the shape of an animal-likeness scale. We compared an expert-based ranking (which was based on Mori's ranking) with the ranking provided by the participants. Both groups of respondents considered the photorealistic and real animals the most animal-like. One unanticipated finding is that the mechanical, stuffed toy, robot, and zombie pandas were ordered differently by participants compared to the expert-based ranking. For the experts, the mechanical panda was the lowest on the animal-likeness scale, whereas this was the robot panda for the participants.

The results show that the different rankings do not affect the properties of familiarity and commonality. However, participant-based ranking eliminates the uncanny valley effect of naturalness for still and moving images when compared to the expert-based ranking. Also, the uncanny valley effect of attractiveness in participant-based ranking for moving images has a slightly different shape. Interestingness and animateness show no uncanny valley, no matter which ranking. Overall, this shows that ranking of the virtual animals has an effect on uncanny valley studies. This means that the way the animals are ranked should be taken into account in the analysis for the future studies of the uncanny valley effect.

Uncanny valley effect on familiarity

The results of the study show that an uncanny valley effect can be found for familiarity, both for still and moving images and for both rankings. This demonstrates that the results reported by Mori (1970) and Mori et al. (2012), which focus on human-likeness, can also be applied to virtual animals, specifically in the area of familiarity perception. The results also show that the participants feel less familiar with a zombie and a robot panda when using the expert-based ranking and with a zombie panda for participant-based ranking.

It is possible that familiarity can be related to certain aesthetic characteristics that are conventionally associated with a given entity. However, in the results on the perception of familiarity, we did not find a significant difference between the zombie rendered in two different colors (green as opposed to white).

Uncanny valley effect on commonality

The commonality dimension describes how much the viewer recognizes characteristics of an object or living being frequently or rarely occurring in nature (Intons-Peterson, 1981).

In virtual reality technology, some experiments reported that when the virtual human image is more unusual (meaning its features are less anthropomorphic), this affects the presence, engagement, and interest of the participants (Lanier, 1992; Nowak and Biocca, 2003). This means that the sense of unusualness relates to the anthropomorphic features of a virtual character. Another example of models of strange living creatures (e.g., reptile, beast, and alien) applied to uncanny valley theory can be found in the study by Burleigh and Schoenherr (2015). This study confirmed an uncanny valley phenomenon in non-human characters for both still and moving images. When we relate that result to our study, we see that the perception of commonality can be associated with the animal-like features. Interestingly, our results also confirm an association between familiarity and commonality.

Uncanny valley effect on attractiveness

A possible uncanny valley effect on attractiveness of the virtual character can form a crucial factor in the desirability for a large number of users. According to Kim et al. (2012), users preferred to play with an attractive virtual character rather than an ugly one. Likewise, Royse et al. (2007) suggested that users, especially women, do not want to play with ugly characters, which is likely a reaction to the players' desire to look attractive in the virtual world. Schneider et al. (2007) showed there is a second highest mean attraction towards non-human virtual characters (e.g., animals or robots) when they present mild human-like features. The results on attractiveness in this study are interesting in that we observed an uncanny valley effect for still but not for moving images, independently of whether the ranking on animal-likeness was done by experts or by non-experts.

Uncanny valley effect on naturalness

The appearance of naturalness is an essential component for representing biological systems. Participants experience a lack of naturalness when, for example, the image of the animal has some missing traits or there is a perceptual mismatch with the real animal (Schwind et al., 2018a). Moreover, Schwind et al. (2018a) described that three factors can affect humans' reactions toward virtual animals: "the violations of the naturalness of the virtual animal, the facial expression, and body pose, as well as how the animal fits into the scene" (pp. 58). Likewise, Löffler et al. (2020) identified the animal-likeness scale from extreme values as machine-like to animal-like, artificial to natural, and inanimate to living. They found an uncanny valley effect for zoomorphic robots, where the robot animals located in the middle level of animal-likeness were perceived as less attractive or likeable. The authors conclude that mismatching between the realistic and natural appearance of the animal at the same time can foster an unfavourable reaction in the users.

We found that in our data, there was an uncanny valley effect on the perception of naturalness for still and moving images with expertbased ranking, but this was not the case for still and moving images with participant-based ranking.

Uncanny valley effect on interestingness, and animateness

Surprisingly, no uncanny valley effect was found for aspects of interestingness and animateness. This holds for the expert-based ranking and the participant-based ranking alike with both still and moving images. Put differently, participants considered all of the six animal images to be similarly interesting and animate. This finding suggests that despite there being a clear conceptual relation between all the uncanny valley subscales, the subscales for interestingness and animateness may behave differently. To our knowledge, there has not yet been any detailed investigation of the relationship between the perception of interestingness and animateness in the context of the uncanny valley in regard to virtual humans or virtual animals.

Movement

According to the original hypothesis of Mori and the hypothesis proposed by Kätsyri et al. (2015), the uncanny valley effect can be amplified by movement of the virtual character. In particular, movement can provide an artificial object with an unnatural quality and thus magnify its creepiness. Therefore, in our stimulus material, we included stimuli depicting a subtle movement of the virtual figure. In line with the expectation formulated in previous studies, the movement slightly amplified the perception responses for the less animal-like virtual characters (i.e., the robot, zombie, mechanical, and stuffed toy panda).

Morbidity

On the one hand, we might expect a zombie virtual character to be attractive to participants, as zombies are a common staple in popular culture. On the other hand, according to the morbidity hypothesis of Kätsyri et al. (2015), a morbid virtual character should give rise to negative affinity, in line with a human "revulsion to death" (Criscuolo,

3

2019). Indeed, in our study, we found that the zombie panda was considered to be unattractive, as well as less familiar, common, and natural.

3.4.8 Conclusion

The current study investigated the existence of the uncanny valley effect for virtual animal characters. Considering an expert-based ranking of animal-likeness, we found uncanny valley effects for familiarity, commonality, naturalness, and attractiveness. For participant-based ranking, we found uncanny valley effects for familiarity and commonality for both still and moving images, attractiveness for still images and naturalness for moving images. In both rankings, no uncanny valley effect was found for the perception of interestingness and animateness.

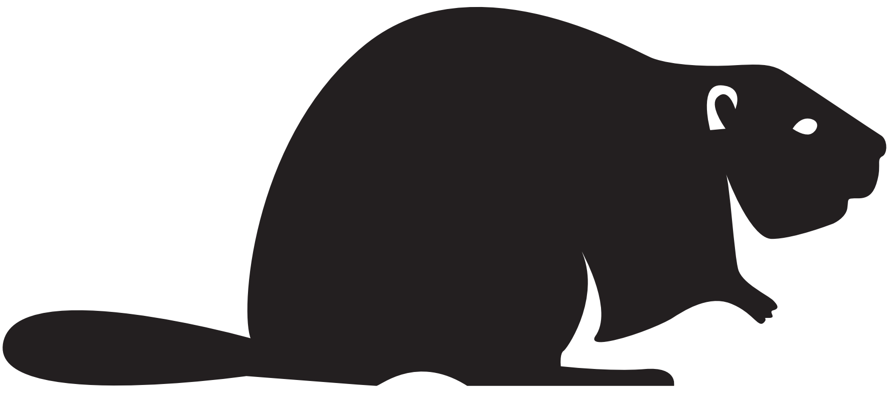
The results of this investigation show that the uncanny valley effect can generally be observed not just for human likeness but also for animal likeness (Mori, 1970). The findings contribute to our understanding of uncanny valley theory of non-human virtual characters and provide a basis for further animal-likeness studies. Taken together, they suggest that design features of virtual animals can have a direct impact on the user experience.

The generalized of these results is subject to certain limitations. In particular, we observed that the expert-based ranking of animal-likeness is different from that of the participants. More research is needed to fully understand these discrepancies. Also, it seems to be the case that movement has an effect on the animal-likeness, possibly leading to a new ranking. The color of the animals (in a way a property of the visualization of the animal) can influence the uncanny valley effect as well and more research into the visual appearance of the virtual animals is thus needed. Finally, it remains to be seen to what extent the results concerning virtual pandas as stimuli carry over to other virtual animals (e.g., fish, birds, worms, beavers, insects, spiders, etc) (Cenydd and Teahan, 2013; Sierra Rativa et al., 2020). Future studies should concentrate on testing the uncanny valley effect on other species and their virtual visualizations.

3

CHAPTER 4

The influence of game character appearance on empathy and immersion: virtual non-robotic versus robotic animals



This chapter tackles research question RQ3 and RQ4.

Research Question

3. Can a virtual animal's visual appearance influence the level of empathy of users?
4. Can a virtual animal's visual appearance influence the level of immersion of users?

Published Work: Sierra Rativa, A., Postma, M., & Van Zaanen, M. (2020). The influence of game character appearance on empathy and immersion: Virtual nonrobotic versus robotic animals. *Simulation & Gaming*, 51(5), 685–711. https://doi.org/10.1177/1046878120926694

4.1 Abstract

In this Chapter 4, we examine empathetic interactions with animated game characters can help improve user experience, increase immersion, and achieve better affective outcomes related to the use of the game.We used a 2x2 between-participant design and a control condition to analyze the impact of the visual appearance of a virtual game character on empathy and immersion. The four experimental conditions of the game character appearance were: Natural (virtual animal) with expressiveness (emotional facial expressions), natural (virtual animal) with non- expressiveness (without emotional facial expressions), artificial (virtual robotic animal) with expressiveness (emotional facial expressions), and artificial (virtual robotic animal) with non-expressiveness (without emotional facial expressions). The control condition contained a baseline amorphous game character. 100 participants between 18 to 29 years old (M=22.47) were randomly assigned to one of five experimental groups. Participants originated from several countries: Aruba (1), China (1), Colombia (3), Finland (1), France (1), Germany (1), Greece (2), Iceland (1), India (1), Iran (1), Ireland (1), Italy (3), Jamaica (1), Latvia (1), Morocco (3), Netherlands (70), Poland (1), Romania (2), Spain (1), Thailand (1), Turkey (1), United States (1), and Vietnam (1). We found that congruence in appearance and facial expressions of virtual animals (artificial + non-expressive and natural + expressive) leads to higher levels of self-reported situational empathy and immersion of players in a simulated environment compared to incongruent appearance and facial expressions. The results of this investigation showed an interaction effect between artificial/natural body appearance and facial expressiveness of a virtual character's appearance. The evidence from this study suggests that the appearance of the virtual animal has an important influence on user experience.

4.2 Related work

The term empathy tends to be used to refer to the ability to share the feelings and thoughts of other people and take another person's perspective, in order to facilitate the process of social interactions (Azevedo et al., 2013; Håkansson and Montgomery, 2003; Kalisch, 1973). It is defined as "a combination of interrelated components of emotion recognition (in oneself and others), affective responsiveness (sharing the emotional experience of others), and perspective taking (cognitively assuming the perspective of others)" (Rodriguez, 2013, p. 494) or as "the capacity to (a) be affected by and share the emotional states of another, (b) assess the reasons for the other's state, and (c) identify with the other, adopting his or her perspective" (Aaltola, 2014, pp. 243-244; De Waal, 2008, p. 281). We also differentiate between cognitive empathy, as the capacity to perceive and predict the feelings of others, and affective empathy, as the capacity to recognize emotional reactions when witnessing the suffering of another person (Patient and Skarlicki, 2010).

Empathy is associated with both positive and negative feelings. On the one hand, it can be triggered by feelings of happiness, achievement, excitement, and celebration (Keen, 2006). On the other hand, empathy is often experienced by an observer of another person in a distress situation or in pain. It can also arise while observing antisocial behaviors and aggression (Batson et al., 1987a; Clark et al., 2019a; Stanger et al., 2012; Sterzer et al., 2007). There is some terminological confusion with respect to the terms of empathy and sympathy. Sympathy is interpreted as the ability to pity and understand the suffering of others, that is to care for the suffering of others, where it can have a reaction tied to negative emotion (Eisenberg and Strayer, 1990). In contrast, empathy, as a broader concept, is the possibility of experiencing positive and negative situations or emotions as if they are one's own (Davis, 2018; Keen, 2006). Although the term sympathy is older than empathy, both terms are associated with compassion, which is the understanding of the situation of the other.

It is important to realize that some people may be more predisposed to empathic reactions than others (Adams, 2019; Eisenberg et al., 1994). Dispositional empathy indicates the manner in which a person tends to respond toward the experiences of other people in general (Konrath et al., 2011). It is a multidimensional construct that includes both cognitive and affective facets (Davis et al., 1994, p. 370).

In general, empathy is understood as the ability to "put yourself in someone else's shoes" (Faulkner, 2018, p. 218), however, "the others" is typically expected to be a fellow human being rather than a technological object (e.g., a robot) or a virtual character (Matravers, 2017, p. 86). An interesting issue that can be studied in the context of simulation games is to investigate what happens when we replace (virtual) humans by any (other) kind of character including virtual animals. Non-human characters in the form of virtual animals are mostly met with violence in computer games, where people can hunt animals as bears, deer or sharks. It has been shown that the use of violent video games can stimulate aggression, decrease pro-social behavior, increase mood changes, and impulsivity in young learners (Greitemeyer (2018, 2019); Sherry, 2001) although for a summary of an opposing view, see the discussion in Kühn et al. (2019). Arguably, they can also lead to apathy towards the fate of similar characters in the real world.

Referencing empathic virtual agents, Ochs et al. (2008) proposed that empathy in human-machine interaction can be developed in two ways:

1. Virtual agents (in this case, virtual characters) can manifest empathic emotions toward a player (also called "empathetic virtual characters" that foster immersion by McQuiggan et al. (2008, p. 1511),
2. and players express empathic emotions toward a virtual agent (which Paiva et al. (2005)) named "empathic synthetic characters").

In our study, we focus on the second type of relation. In line with Paiva et al. (2005, p. 265), we make use of "distress-inducing stimuli" (Belman, 2016, p, 64; Eisenberg et al., 1989, p. 42) to create conditions favorable to experiencing possible empathic reactions toward virtual characters. We vary the game character appearance (expressiveness and artificiality) to investigate its influence.

Past research showed that simulation games can be used as powerful tools to develop learning and social skills through virtual characters (Hofstede et al., 2010; Ke and Moon, 2018). Moreover, simulation games are typically considered well-suited to support educational programs fostering empathy, because they allow players to adopt new perspectives in an immersive way (Bachen et al., 2012; Belman and Flanagan, 2010). Research on human-computer interaction (HCI) (Pan and Hamilton, 2018; Scassellati et al., 2018), human robot interaction (HRI) (Ishiguro and Nishio, 2018), and video games (Wu et al., 2018) has led to a proliferation of studies using human characters. In contrast, recent research, such as that conducted by Schwind et al. (2018a), has shown a lack of systematic studies on the effect of different design features for non-human artificial characters, such as animal-like characters or virtual robots, despite their growing popularity in simulated environments including learning applications.

A few studies analyzed empathy toward non-human characters like virtual pets. Virtual pets are defined by Tsai (2008) as the simulations of "life-like agents" and could possibly be used to build relationships with players (pp. 49). Tsai and Kaufman (2014, p. 149) used a "real-time pet simulation videogame" called Nintendogs and reported a positive effect of interacting with virtual pets on users' relationship to pets in real life, suggesting that children playing the game developed an empathic disposition toward the virtual characters. This result supports the conclusion that empathic feelings toward a virtual animal are, in principle, possible. However, non-human characters often have human characteristics, such as movement, expressiveness, behavior, and personality (Tinwell et al., 2011; Yamane et al., 2010) and it is not clear what the effect of these properties on the users. For example, does the "uncanny valley effect" (Ho and MacDorman, 2017, p. 129; Mori, 1970) apply to the same extent to virtual animals? Which design features are necessary for the user to establish an emotional bond with a virtual animal?

According to the uncanny valley theory, an object with human-like traits can fail to elicit affinity when it is very realistic to its human counterpart (Mori, 1970; Mori et al., 2012). Aesthetic and other features of the inanimate objects can have a direct effect on their acceptation or rejection by contributing to the uncanny valley effect (Hodgins et al., 2010; Misselhorn, 2009; Riek et al., 2009). Here, a clear link to empathy can be found as it originally stems from the German word "Einfühlung", i.e., the appreciation of the aesthetics of an object as an observer projection (Eisenberg and Strayer, 1990, pp. 18–20). One of the few investigations that have been carried out on virtual animals in relation to the uncanny valley effect concerned the appearance of the character, in particular the effects of realism, stylization, and facial expressions (Schwind et al., 2018a). According to the outcomes of the study, the naturalness of the virtual animal was fundamental to create the sensation of realism with the user.

An important aspect of simulated environments involving virtual characters concerns immersion. Even though immersion is one of the most significant concepts in game design, little is known about the relationship between immersion and the empathic disposition toward a virtual character in a game environment. Arguably, empathy-centered design can improve the affective experience and emotional response of the player with respect to the virtual game character (Belman and Flanagan, 2010; Brown and Cairns, 2004a). Next to that, several studies convincingly show that the affective experience increases immersion in the game (Jennett et al., 2008; Nacke and Lindley, 2008; Nacke et al., 2011). Given the link between emotional response and empathy on the one hand, and immersion on the other hand, we might expect empathy and immersion to be strongly related to each other.

The relationship between dispositional empathy and situational empathy plays a critical role in the measurement of empathy concerning a virtual character. Eisenberg et al. (1994) describes that situational and dispositional measures can be correlated occasionally. It depends on the hypothesis of the study and their different methods of vicarious emotional responding (e.g., emotions and sensations experienced through the stories of others by watching or reading). The dispositional empathy can describe the current level of empathy of the participant, but cannot describe whether a stimulus affects the measure of empathy in the person. Likewise, the situational empathy can help to clarify if the stimulus (e.g., virtual character) can affect the empathy of the participant and possible relationships. For this reason, many studies used dispositional and situational empathy measures to identify relationships between both sub-constructs of empathy to have a better measure of empathy (Adams, 2019; Eisenberg et al., 1994; Holmgren et al., 1998; McQuiggan et al., 2008; Rosenthal-von der Pütten et al., 2013). To date, no previous studies have investigated whether the situational empathy toward a virtual character can be affected by the dispositional empathy of the users.

One of the most well-known approaches for assessing dispositional empathy (however, see Eisenberg and Strayer (1990), for a discussion of potential weaknesses of this approach) is to use a self-reported questionnaire, such as Davis's Interpersonal Reactivity Index (also known as the IRI questionnaire), which is the most frequently used (Davis, 1983; Garcia-Barrera et al., 2017; Hojat, 2016; Otterbacher et al., 2017; Rivers et al., 2016). Measurement of dispositional empathy may be affected by self-representation concerns, but the IRI questionnaire has an acceptable validity and reliability confirmed in studies with Dutch (De Corte et al., 2007), Chinese (Siu and Shek, 2005), and French population (Gilet et al., 2013). This questionnaire appears to be suitable for experiments concerning fictional characters, due to the inclusion of a fantasy measure scale.

Contrary to empathy as a trait, situational empathy (i.e., context dependent empathic reactions) can be induced through specific stimuli or situations. Many researchers have utilized situational empathy to measure empathy-related responses (Eisenberg et al., 1994; Holmgren et al., 1998). To capture the concept of empathy in its entirety, self-reported measures based on existing questionnaires should be combined with measurements of psycho-physiological responses.

Situational empathy in the context of simulated environments appears to be very closely related to the concept of engrossment and presence (or, total immersion), both sub-concepts of immersion. Immersion is the illusion of experiencing a virtual environment as if it were akin to the real world (Hou et al., 2012; Mäyrä and Ermi, 2011) and can be enhanced by realistic elements within the virtual game experience. In the same way, the term immersion can be defined as a mental process (Hou et al., 2012; Mäyrä and Ermi, 2011) or intrinsic human characteristic (Hou et al., 2012), which can be affected by different elements of game design (e.g., size screen, viewing angle, audio, character's aesthetics, or story) (Domsch, 2017; Hou et al., 2012; Mäyrä and Ermi, 2011).

One of the most well-known ways of assessing immersion within a computer interface is a questionnaire developed by Jennett et al. (2008). The questionnaire contains 31 questions on both game elements and experience. Brown and Cairns (2004a) proposed that game qualities can be described in three distinct levels of immersion: engagement, engrossment, and total immersion. Whereas engagement can be measured using factors that include time, attention, and energy required from the player, engrossment, is associated with gamers' emotions. Total immersion is associated with "presence". At this stage, the impact of player's feelings and thoughts are affected by their level of empathy and game's atmosphere (e.g., graphics, plot, and sounds). Brown and Cairns (2004a) offer anecdotal support for the idea that empathy and immersion have a strong relationship that is fostered by the game character and its interaction with the environment (character appearance and first-person player perspective).

To sum up, existing literature suggests that empathy and immersion are two strongly related concepts (engrossment and engagement) that are both affected by the game character's appearance. Likewise, the game character's appearance (artificiality/expressiveness) influences the level of self-reported situational empathy. Finally, the empathic tendency of the participants (dispositional empathy) and their possible empathetic reactions toward the virtual character (situational empathy) can influence to different levels of immersion. However, the empirical evidence supporting these conclusions is currently mainly based on anecdotal accounts.

The expected relations are shown in Figure 4.2.1, based on reports found in the literature discussed above. The purpose of the current study is to explore the effect of different design features pertaining to the virtual game character (artificiality and expressiveness) on situational empathy and immersion of the user in the simulated environment. In terms of artificiality, the body appearance of the virtual character in our experiment was manipulated using a distinction made by Coeckelbergh (2011, p. 199), who proposed that animals are "natural" and "biological" entities, a living organism in an ecosystem, while robots are "artificial" and "technological" articles and objects. Referring to a game character appearance, a natural appearance is understood to be closer to the biological animal in the real world in terms of important characteristics such as the animal's color or body shape. An artificial appearance, on the other hand, has characteristics akin to the appearance of a robot. With respect to characters' expressiveness, virtual characters can display human-like facial expressions (Beer et al., 2015; Paiva et al., 2005; Thomas et al., 1995). According to Dyck et al. (2008), it is possible to enrich virtual faces of an avatar with expressions of happiness, sadness, fear, disgust, anger, and a neutral state. Dyck et al. (2008) and colleagues found that recognition of the virtual expressions was comparable to natural facial displays of emotion. Interestingly, in their study, neutral expressions were most frequently confused with the emotion of sadness (both in human and virtual faces) but were also chosen when participants were uncertain about the emotion displayed. In contrast to Dyck et al. (2008) and Hoffmann et al. (2010), studying human expressions only, considered the neutral face to be a facial expression of its own right, but not an expression of emotion. Likewise, Tinwell et al. (2011) used basic universal emotions (happiness, anger, fear, sadness, and disgust) as described by Ekman (1992) and used the neutral expression as a kind of control state in virtual character (91.47%) and human images (89.92%), with surprisingly high recognition rates. In our study, we made use of three basic facial expressions: happiness, sadness, and neutral. These facial expressions were chosen due to their high and consistent recognition rates for virtual faces Dyck et al. (2008).

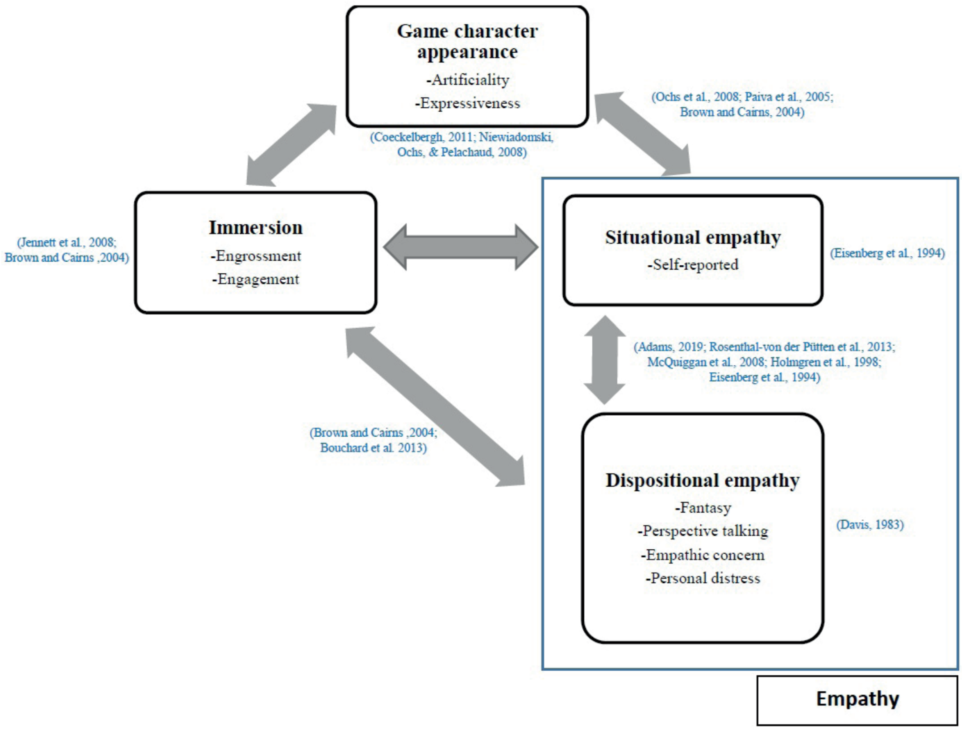


Figure 4.2.1: Expected relationships between the game character's appearance, empathy, and immersion.

Our hypothesis was that an expressive virtual character with animal-like features (i.e., low artificiality) would lead to a higher level of subjective immersion and situational empathy compared to a robotic character without emotional facial expressions. The main research questions addressed in the experimental design were as follows:

1. Does game character appearance (artificiality and expressiveness) influence the level of self-reported situational empathy?,
2. Does game character appearance (features pertaining to artificiality and expressiveness) influence the level of immersion?, and
3. Does self-reported situational empathy correlate with immersion?

4.3 Methods

We used a 2x2 between-participant design and a control condition. The four experimental conditions of the game character appearance were: Natural (virtual animal) with expressiveness (emotional facial expressions), natural (virtual animal) with non-expressiveness (without emotional facial expressions), artificial (virtual robotic animal) with expressiveness (emotional facial expressions), and artificial (virtual robotic animal) with non-expressiveness (without emotional facial expressions). The control condition contained a baseline amorphous game character. Participants were videotaped while accomplishing the experimental task. The task consisted of taking care of a virtual character by supplying it with energy, break time, and fun in a simulation game lasting 10.16 minutes (609.6 seconds). Before playing the game, the participants filled out the IRI questionnaire (Davis, 1983). After playing the game the participants filled out an "Immersion Questionnaire" (Jennett et al., 2008, p. 644), and a situational empathy question about the experiment.

4.3.1 Participants

We recruited 100 participants between 18 to 29 years old (M=22.47, SD=2.914). The participant sample was balanced for gender, with 50 females (50%) and 50 males (50%), drawn from a university student population in the Netherlands. The participants were randomly assigned to one of five experimental groups. Each condition had 20 participants (10 female and 10 male). The participants received a course credit for their participation and their participation was voluntary. They originated from several countries: Aruba (1), China (1), Colombia (3), Finland (1), France (1), Germany (1), Greece (2), Iceland (1), India (1), Iran (1), Ireland (1), Italy (3), Jamaica (1), Latvia (1), Morocco (3), Netherlands (70), Poland (1), Romania (2), Spain, Thailand (1), Turkey (1), United States (1), and Vietnam (1). The participants reported playing videogames: daily (10%), several times per week (19%), several times per month (19%), several times per year (31%), and never (18%).

4.3.2 Stimulus

The beaver is not considered a worldwide charismatic wild species, compared to other species such as the panda, polar bear, wolve, tiger, dolphin, whale, or ape (Albert et al., 2018; Ducarme et al., 2013). Since the beaver is not perceived as very attractive, cute, and charming by default, we expected that it would not directly stimulate a sense of empathy in the participants of the study. The beaver was depicted in a virtual setting with two different body design features; a natural body (virtual animal) and an artificial body (virtual robotic animal). Both designs are shown in Figure 4.3.1. We chose an amorphous figure as a control character, since it does not contain a body easily recognized by the user, such as any geometric figure used in other studies (Heider and Simmel, 1944).

Figure 4.3.1: Modeling in 3D design of the natural beaver (left side) and artificial robot beaver (right side).

Artificiality and expressivity were combined into four experimental conditions and one control condition (as shown in Figure 4.3.2):

1. Natural body with expressiveness [experimental]: This condition showed a beaver with a natural body in 3D. The character had three facial expressions: sadness, happiness, and neutral.
2. Natural body with non-expressiveness [experimental]: This condition showed a beaver with a natural body in 3D. However, it had no human-like facial expressions.
3. Artificial body with expressiveness [experimental]: This condition showed a beaver with an artificial robot body without any biological traits. The character had three facial expressions: sadness, happiness, and neutral.
4. Artificial body without non-expressiveness [experimental]: This condition showed a beaver with an artificial robot body without any biological traits. However, it had no human-like facial expressions.

5. Amorphous figure as a virtual character [control]: This condition had no clearly defined kind of character. Its shape was indeterminate, and lacking a definite form (marshmallow-like).

4.3.3 Procedure

For the purposes of our experiment, we designed an experimental game called "Justin Beaver" (as shown in Figure 4.3.3). This game environment was inspired by the virtual pet game Tamagotchi, which also had an autonomous virtual character, but in a 2D environment (Higuchi and Troutt, 2004). Our game has a virtual simulation environment in 3D graphics, where the virtual character (animal or robot-animal beaver) explores a natural habitat. The virtual character moves randomly in the game environment. The player is instructed to take care of the character by supplying it with energy, break time, and fun during 10.16 minutes (609.6 seconds) through a "drag & drop" system. The players' performance was indicated by increasing or decreasing level bars on the screen.

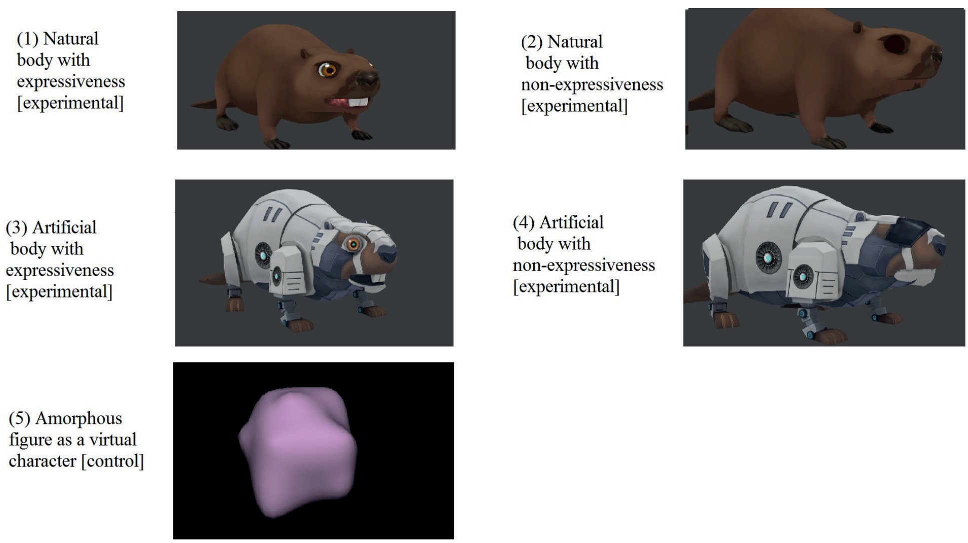


Figure 4.3.2: Five conditions of the game character.

Figure 4.3.3: Screenshot of the game environment used in this study.

On the initial screen of the game, five virtual characters were displayed which allowed the researcher to choose with which character the participant would play. Then, to start playing the game, the participant had to press the "F9" key which saved an initial timestamp on the computer used later to synchronize different streams of data. The virtual characters in the experimental conditions (1) and (3) (the beaver or robot beaver) displayed facial expressions in the following order: 3 minutes of sadness, 3 minutes of neutral, 3 minutes of happiness, and 1 minute and 16 seconds of neutral (as shown in Figures 4.3.4 and 4.3.5). As the virtual characters in the other two experimental conditions (2) and (4) did not show emotional facial expressions, their appearance did not change during the game.

Figure 4.3.4: Timeline of emotional facial expressions of the natural beaver (virtual animal).

Figure 4.3.5: Timeline of facial expressions for the artificial beaver (virtual robotic animal).

After 10 minutes, a "distress-inducing stimuli" was evoked in the game when another character (a hunter) came out from behind a bush and shot the beaver. After the distress situation, there was a delay of 16 seconds and then a screen appeared with the message: "Game Over".

Our control group (5) played with an amorphous figure as their virtual character. This virtual character had no emotional facial expressions, artificial or natural properties, or any particular kind of character. Its shape was indeterminate, and lacking a definite form (somewhat comparable to a marshmallow). The player took care of the character by supplying it with energy, break time, and fun during 10.16 minutes and they watched the same distress event as in the experimental situations and a message "Game Over".

4.3.4 Ethics approval

Ethics approval was obtained from the Research Ethics Committee of the Tilburg School of Humanities with the reference REC#2017/01. We found that this experimental game did no harm to the participants, other than feeling perhaps a little uncomfortable when the virtual character found itself in the distress situation. However, this situation was necessary for the successful induction of a distressful event in the experiment. After the experiment, the researcher explained to the participant that it was a virtual character and that no harm was actually done to any characters in the real world. Data and experimental game are publicly available on the Dataverse platform at Tilburg University (see https://dataverse.nl/dataset.xhtml?persistentId= hdl:10411/NJJATW).

4.3.5 Instrumentation

The participants took part in the task individually in front of a computer screen. Before of the game, IRI questionnaire was administrated online. After playing the game, the participants filled out an online Immersion Questionnaire and answered the Self-reported situational empathy about the character. Finally, the researcher conducted a semistructured individual interview with the participant, taking between 3 to 7 minutes. In total, the experiment took approximately 1 hour.

Self-reported situational empathy

We measured self-report situational empathy with a single item: "To what extent did you really empathize with the character (animal/robotanimal/amorphous figure)? The participants indicated the answer on a 5-point scale where the extremes were labeled (1=Not at all and 5=Very much).

Dispositional empathy

To measure dispositional empathy, the participants filled out the IRI questionnaire (Davis, 1983, p 1) which measures empathy on four subscales: perspective taking ("the tendency to spontaneously adopt the psychological point of view of others"), fantasy ("taps respondents' tendencies to transpose themselves imaginatively into the feelings and actions of fictitious characters in books, movies, and plays"), empathic concern ("assesses 'other-oriented' feelings of sympathy and concern for unfortunate others"), and personal distress ("measures 'self-oriented' feelings of personal anxiety and unease in tense interpersonal settings"). The participants indicated answers on a 5-point scale where the extremes were labeled (A=Does not describe me well and E=Describes very well). The test has 28 items (see supplemental material #3). The items (1–2, 4–6, 8–11, 16–17, and 20–28) were scored: A=0, B=1, C=2, D=3, E=4. The reversed-scored items of this test were: 3, 7, 12, 13, 14, 15, and 19 (see **Appendix B**). The Cronbach's alpha for the IRI questionnaire was reliable with α=0.809. The principal component analysis (PCA) of the IRI questionnaire indicates a possible reduction from a larger number of variables of the IRI questionnaire to a smaller number of factors. The Keiser-Meyer-Olkin measure of sampling adequacy was 0.706 and the Bartlett test of Sphericity was highly significant, suggesting the data were suitable for PCA. The scree plot (as shown in Figure 4.3.6) of the PCA shows that four groups (factors) appear stacked and separated from the rest. Factor 1 was labeled "fantasy", factor 2 "personal distress", factor 3 "perspective taking", and factor 4 "empathy concern"(see supplemental material #1). The outcomes of the PCA were in line with the sub-scales of the IRI questionnaire as they have been used in the past. In the data analysis below, we, therefore, treated dispositional empathy both as a single complex concept (validated by the Cronbach's alpha value) and in terms of the four sub-scales (validated by the PCA).

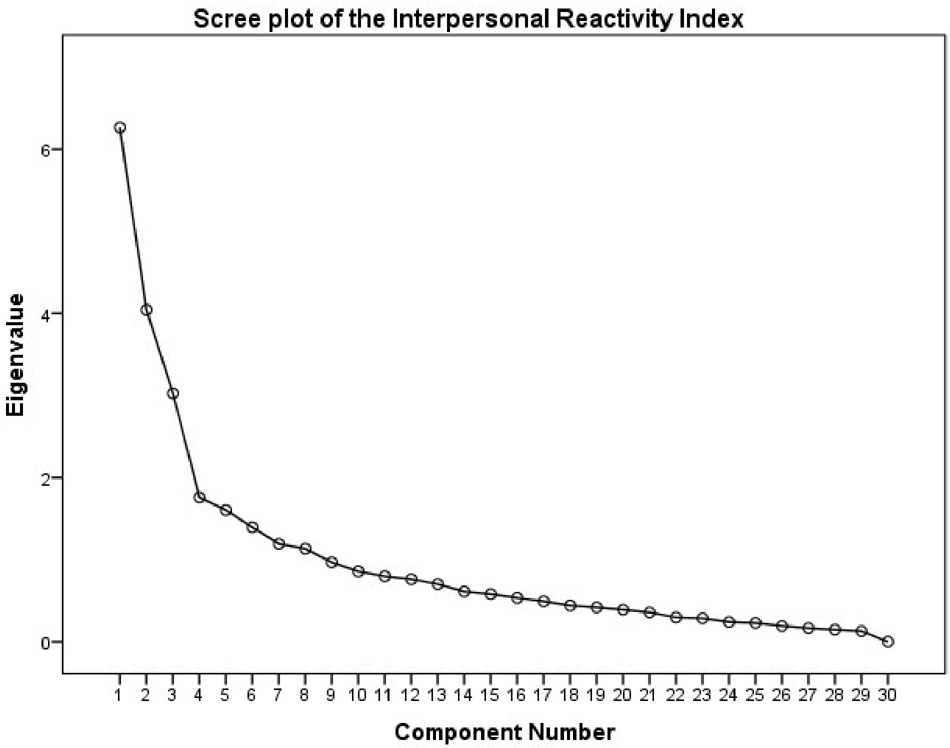


Figure 4.3.6: Scree plot showing four groups (factors) of the IRI test.

Immersion

We measured immersion using a questionnaire that was originally developed by Jennett et al. (2008) (see **Appendix B**). Participants indicated their answers on a 5-point scale where the extremes were labeled (see supplemental material #2). The test had 31 items. The items (1–5, 7, 11–17, 19, and 21–31) was scored: 1 to 5. The reversed-scored items of this test were: 6, 8, 9, 10, 18, and 20. The Cronbach's alpha for the Immersion questionnaire was reliable with α=0.906.

The principal component analysis (PCA) of the Immersion questionnaire indicates a possible reduction from a larger number of variables of the Immersion questionnaire to a smaller number of factors. The Keiser-Meyer-Olkin measure of sampling adequacy was 0.787 and the Bartlett test of Sphericity was highly significant, suggesting the data were suitable for PCA. The scree plot of the PCA (illustrated in Figure 4.3.7) shows that two groups (factors) appear stacked and separated from the rest. Factor 1 was labeled "engrossment", and factor 2 "engagement" (Brown and Cairns (2004), pp. 1297–1298; see supplemental material #2).

The outcomes of the PCA of the results of the Immersion Questionnaire was associated with two labels, which were used on the construct of immersion developed by Brown and Cairns (2004a). Engagement was considered the first phase of immersion in which the players show their "effort, invest time, and attention" (p. 1298). Engrossment was considered the second phase of immersion in which the players show their "emotional investment in the game" (p. 1299). Total immersion was considered the total score of the immersion which the players are totally involved in the game.

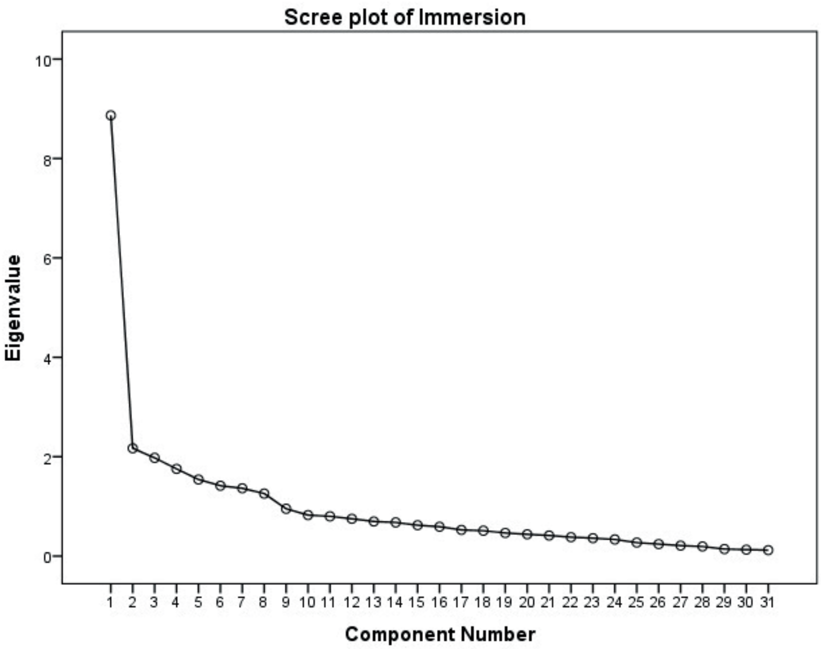


Figure 4.3.7: Scree plot showing two groups (factors) of the Immersion test.

Semi-structured interview

During the debriefing the participants answered three open question in the form of a short semi-structured interview. This was used to collect additional information about experiment. The questions were:

1. Did you feel any emotional attachment to the character (animal/ robotanimal/amorphous figure)? Why?,
2. How did you feel when the character was in a distress situation?, and
3. What aspect could we improve in the video game to improve the emotional link to the character?

We analyzed whether the level of immersion of participants in the game correlated with their level of situational empathy answer. In addition, we compared the results of the IRI questionnaire with the Immersion Questionnaire and self-reported situational empathy. To identify whether the virtual character's artificiality and expressiveness influenced the level of immersion and situational empathy of a participant, we compared the average results of the Immersion Questionnaire and self-reported situational empathy of the participants in the experimental conditions using two-way ANOVA. Moreover, we compared the results with the control group consisting of 20 participants. Data collected in the semi-structured interviews with open questions was used to gain qualitative insight and to acquire additional information about the experiment. Data preprocessing and analysis was performed using SPSS 24.0.

4.4 Results

4.4.1 Interaction effect between artificiality and expressiveness toward self-report situational empathy

The first question of this study aimed to explore if the game character appearance (artificiality/expressiveness) influenced the self-reported situational empathy of the participants. A two-way ANOVA showed that there was no statistically significant interaction between the effect of artificiality and self-reported situational empathy $F(1, 95)=.698$, $p=.405$, $\eta p^2=.007$. Likewise, there was no statistically significant interaction between the effect of expressiveness and self-reported situational empathy $F(1, 95)=.393$, $p=.532$ $\eta p^2=.004$. There was, however, a statistically significant interaction between the effects of game character appearance (artificiality and expressiveness) and self-reported situational empathy, $F(1,95)=11.171$, $p=.001$, $\eta p^2=.105$, as shown in Figure 4.4.1.

Interaction Effect Between Artificiality and Expressiveness Toward Immersion

The second question of this study ex-plored whether game character appearance (artificiality/expressiveness) influenced the immersion level of the participants. A two-way ANOVA showed no statistically significant effect of artificiality: $F(1, 95) = .116$, $p = .734$, $\eta p^2=.001$. Likewise, there was no statistically significant effect of expressiveness: $F(1, 95)= .055$, $p= .815$ $\eta p^2=.001$. However, there was a statistically significant interaction between the game character appearance and immersion, $F(1, 95) = 5.965$, $p=.016$, $\eta p^2=0.059$, as shown in Figure 4.4.2.

Relationship Between Dispositional Empathy, Self-Reported Situational Empathy and Immersion

The third question of this study was to explore whether dispositional empathy and self-reported situational empathy were correlated with immersion. A Pearson product moment correlation coefficient was computed to assess the relationship between the dispositional empathy, self-reported situational empathy, and immersion. There was a positive correlation between the dispositional empathy and immersion level, r=0.249, n=100, p=0.012. There was also a positive correlation between the dispositional empathy and self-reported situational empathy variables, r = 0.384, n = 100, $p < 0.001$. Finally, there was a positive correlation between the self-reported situational empathy and immersion level, r=0.593, n=100, $p < 0.001$. Table 4.4.1 summarizes the results of the correlation analysis. For example, other significant relationships were found: There was also a positive correlation between the dispositional empathy and their four sub-scales: empathic concern (r=0.736, n=100, $p < 0.001$), fantasy (r=0.692, n=100, $p < 0.001$), perspective taking (r=0.588, n=100, $p < 0.001$), and personal distress (r=0.438, n=100, $p < 0.001$). There was a positive correlation between the dispositional empathy and a sub-scale of immersion level: engrossment (r=0.270, n=100, p=0.007). There was a positive correlation between fantasy with empathy concern (r=0.316, n=100, p=0.001), self-reported situational empathy (r=0.293, n=100, p=0.003), immersion engrossment (r=0.240, n=100, p=0.016), perspective taking (r=0.220, n=100, p=0.028), and immersion (r=0.213, n=100, p=0.033). There was a positive correlation between perspective taking with empathy concern (r=0.395, n=100, $p < 0.001$), self-reported situational empathy (r=0.271, n=100, p=0.006), immersion (r=0.238, n=100, p=0.017), immersion engagement (r=0.217, n=100, p=0.030), and immersion engrossment (r=0.216, n=100, p=0.032). There was a positive correlation between empathy concern and self-reported situational empathy (r=0.293, n=100, p=0.003). There was a positive correlation between self-reported situational empathy with immersion engrossment (r=0.599, n=100, $p < 0.001$) and immersion engagement (r=0.398, n=100, $p < 0.001$). There was a positive correlation between immersion with immersion engrossment (r=0.962, n=100, $p < 0.001$) and immersion engagement (r=0.782, n=100, $p < 0.001$). Finally, there was a positive correlation between immersion engrossment and immersion engagement (r=0.582, n=100, $p < 0.001$). Overall, increases in dispositional empathy and self-reported situational empathy were positively correlated with increases in ratings of immersion.

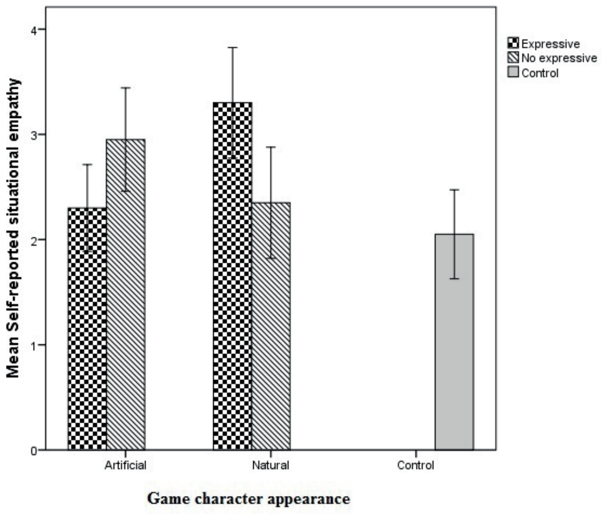


Figure 4.4.1: Differences between mean values of self-reported situational empathy for artificiality/naturality and expressiveness/non-expressiveness (game character appearance). An interaction effect was found between game character appearance and self-reported situational empathy. Standard errors are represented in the figure by error bars attached to each column.

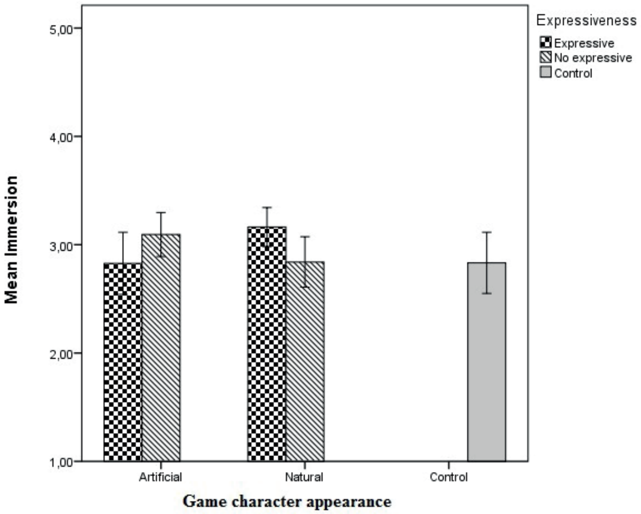


Figure 4.4.2: Mean difference values representing immersion for artificiality/naturality and expressiveness/non-expressiveness (game character appearance). Standard errors are represented in the figure by error bars attached to each column.

Table 4.4.1: Bivariate correlations between the measured variables.

Variables	M	SD	1	2	3	4	5	6	7	8	9
(1) Dispositional empathy -IRI	2.44	0.39	-								
(2) IRI Fantasy	2.58	0.76	0.692**	-							
(3) IRI Personal Distress	1.68	0.72	0.438**	0.013	-						
(4) IRI Perspective Taking	2.74	0.56	0.588**	0.220*	-0.064	-					
(5) IRI Empathic Concern	2.63	0.52	0.736**	0.316**	0.15	0.395**	-				
(6) Self-reported situational empathy	2.59	1.14	0.384**	0.293**	0.084	0.271**	0.293**	-			
(7) Immersion	2.95	0.54	0.249*	0.213*	0.047	0.238*	0.108	0.593**	-		
(8) Immersion Engrossment	2.73	0.59	0.270**	0.240*	0.076	0.215*	0.12	0.599**	0.962**	-	
(9) Immersion Engagement	3.48	0.63	0.126	0.087	-0.032	0.217*	0.049	0.398**	0.782**	0.582**	-

Note. A Spearman correlation analysis was conducted. It did not have a significant variation in its values compared to the outcomes of the Pearson correlation.
*p<.05.**p<.01.

4.5 Discussion

The general aim of this study was to investigate if the appearance of the virtual animal character can be adapted to foster empathy and immersion. The game character appearance was systematically manipulated in terms of (1) artificiality (robotic or natural), and (2) expressiveness (with or without emotional facial expressions). Next to that, we measured dispositional empathy, situational empathy, and immersion in the game.

The first aim of this study was to explore whether game character appearance (manipulated in terms of its artificiality and expressiveness) can influence the level of self-reported situational empathy toward the character. Referring just to expressiveness, the present study set out with the aim of assessing whether emotional facial expressions of the virtual character can foster empathy in the user. In reviewing the literature, Ochs et al. (2008) described that a virtual character or agent can be called empathic when there are two situations: (1) users can feel empathy toward the virtual character/agent, or (2) the virtual character/agent shows empathic emotions concerning the users. For situation (1), previous studies have used facial expressions onto virtual characters to develop empathetic towards virtual agents or characters (Niewiadomski et al., 2008; Prendinger et al., 2005; Vinayagamoorthy et al., 2006). For instance, Ochs et al. (2008) used facial expressions in the virtual human characters to simulate the perception of emotions in agents. This means that a virtual character was perceived as being more empathetic when it had positive and negative emotions than when it was non-expressive. Can an emotionally expressive virtual character also foster the user's empathy towards itself? The aim of the present study was to examine if a user would display more empathetic reactions towards virtual agents with facial expressions. We found that emotional facial expressions of the virtual character by itself cannot foster the empathy of users, as there was no statistically significant effect of expressiveness on self-reported situational empathy.

Another important aspect of the appearance of the game character is artificiality and naturality, with the theoretical construct of artificiality being based on the work of Coeckelbergh (2011), who defined a robot as an "artificial" and "technological objects" object (p. 199). In robotics, artificiality plays a very important role in realism and its effects with the emotional connection with the users. Andrews (2013) has suggested some concerns about artificiality of the robots when he described that modern robots look more like machines than live animals in natural environments such as a zoo. He noted that due to advances in technology in a virtual environment, it is easier to design realistic robot animals that simulate fine motor movement, appearance, and unpredictable behavior of a live animal. Previous studies investigating the uncanny valley have shown how artificiality (more or less human-like) of robots can affect the familiarity or affinity of the users (Mori, 1970; Mori et al., 2012). Riek et al. (2009) showed that appearance of a robot can affect the empathy of the users. They found that humanlike robots foster more empathy in the users compared to mechanicallooking robots. However, they did not associate the mechanical looks of robots with artificiality but with the absence of anthropomorphic (human-like) traits. In reference to zoomorphic (animal-like) traits, there is a study that shows how a video where a dinosaur robot was tortured can affect self-reported empathy (Rosenthal-von der Pütten et al., 2013). This study showed that participants had a significant main effect on self-reported empathy related to "Pity for robot/angry at torturer", but not "empathy with the robot" (pp. 24–25). This dinosaur robot had a more biological appearance than mechanical, which was called naturality in this study. For this study, the "natural" condition was designed through a virtual animal with a "natural" and "biological" body appearance (Coeckelbergh, 2011, p. 199). Contrary to expectations, this study did not find that the virtual character was perceived with more empathy when it was manipulated only in terms of its artificiality or naturality as it did not increase situational empathy in users. The most interesting finding of the current study was that the impact of artificiality/naturalness depended on the expressiveness of the character in that congruent appearances gave rise to the highest levels of situational empathy. We found that an artificial (virtual robotic animal) body appearance with absence of expressiveness and a natural (virtual animal) body appearance with expressiveness appeared to generate most empathetic reactions. We assume that these effects were due to participants preferring congruent virtual characters. In fact, similar choices can be found outside of the research domain. For example, in the movie WALL·E (film produced by Pixar Animation Studios), the animators designed a robot character called "Eve" with harmonic characteristics where the eyes of a robot were more mechanical (two blue moving lights) than human (e.g., an iris and a pupil).

The second aim of this study was to determine how the game character appearance (features pertaining to artificiality and expressiveness) influenced the level of immersion. The results for immersion were comparable to those for situational empathy: Immersion levels were higher for expressiveness when the game character's appearance was natural (virtual animal). However, immersion was higher for non-expressiveness when game character's appearance was artificial (virtual robotic animal).

The third aim of this study was to identify the relationship between empathy and immersion produced by the game character's appearance. Our findings were consistent with the expected relationship between game character appearance, empathy, and immersion presented in the literature review. We found a moderately positive correlation (see Figure 4.5.1) between self-reported situational empathy and immersion, in line with the conclusions of Brown and Cairns (2004a) who observed that players who were not completely immersed, experienced a lack of empathy with respect to some game design features (such as the appearance of the virtual character). Self-reported situational empathy was significantly correlated with both sub-concepts of immersion: engrossment and engagement. Interestingly, engrossment had a greater correlation with situational empathy (r=.599) compared to engagement (r=.398). This result may be explained by the fact that engrossment is associated with the emotional investment of the player in the game (Brown and Cairns, 2004a; Jennett et al., 2008).

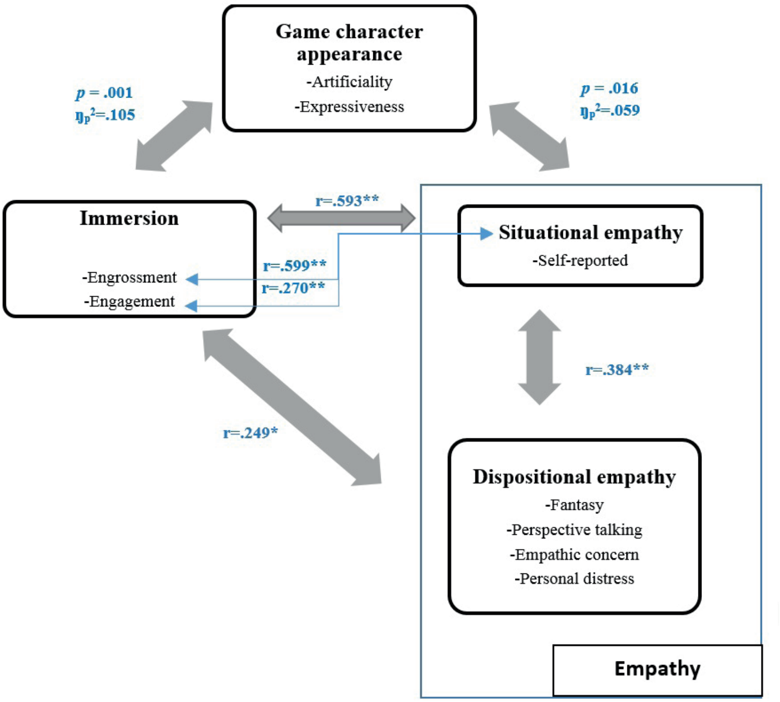


Figure 4.5.1: Results of the expected relationship between game character appearance, empathy, and immersion.

Finally, we examined if dispositional empathy had any relation to situational empathy and immersion. The results of our study further support the idea that dispositional empathy is likely related to situational empathy towards a virtual character in simulated environments. This finding is in line with the proposal of Eisenberg et al. (1994). Participants with a high level of dispositional empathy are more likely to empathically respond to the virtual stimuli, though the response may differ depending on the stimulus or circumstances suitable for it (Adams, 2019). With respect to immersion, the results of our study showed a low significant correlation between dispositional empathy and immersion. This finding, while preliminary, suggests that there is a relationship between dispositional empathy, situational empathy and immersion, which is an important aspect for future research.

4.6 Limitations and recommendations

In our research, we found that the appearance of a virtual animal character can be manipulated to foster empathic and immersion reactions in users of simulated environments. However, there are other aspects of these environments than need to be examined in future research, such as possible negative effects of empathy on players. An interesting example was discussed by Happ et al. (2013) in relation to a game character who was considered to be a victim of circumstances, generating empathy and justifying a high level of acceptance towards violence perpetrated by the character. One should thus be cautious when employing empathy in simulated environments as there is something inherently dangerous to people's emotions being manipulated through game characters.

From a methodological point of view, our measure of situational empathy relied on a single-item instrument. A single item can give us a general idea about participants' empathic feelings towards the character in an exploratory manner, however only one item is not sufficient to check the reliability and validity of the concept. For future studies, we intend to develop a multi-item measurement of situational empathy in simulated environments and to test its construct validity. Next to that, we aim to include socio-demographic variables such as the participants' age, gender, and cross-cultural background, in the experimental design. Finally, in future research, we will examine more closely the links between self-reports of situational empathy and psychophysiological measurements (e.g., heart rate and facial expressions) to externally validate the construct.

4.7 Conclusion

The purpose this study was to determine how a virtual character's visual appearance affects players' experiences of empathy and immersion in virtual environments. In our experiment, the body appearance and facial expressions of the virtual character were manipulated in terms of (1) artificiality (robotic or natural), and (2) expressiveness (with or without facial expressions). The major finding of this study was the discovery that the interaction between the artificial/natural appearance and the expressiveness of a virtual character — particularly, their congruence — affects self-reported situational empathy of a player, as well as the level of immersion experienced in a simulation game. We found a positive correlation between dispositional empathy, self-reported situational empathy, and immersion, thereby empirically confirming the link between empathy and immersion previously proposed on conceptual grounds. The findings shed new light on studies of non-human characters (e.g., virtual animal/virtual robotic animals) and their effect on user experience. Further research might explore other types of situational empathy reactions. These may include but are not limited to measuring the emotions of participants using facial recognition software or psychophysiological measures during the same virtual interactive experience developed in this study.

CHAPTER 5

Try walking in my paws: empathy, immersion, and perceived pain in virtual reality environments by manipulating animal character appearance?

This chapter tackles research questions RQ3, RQ4, RQ5, and RQ6.

Research Question

3. Can a virtual animal's visual appearance influence the level of empathy of users?
4. Can a virtual animal's visual appearance influence the level of immersion of users?
5. Can a virtual animal's visual appearance influence the level of perceived pain of users?
6. Can a virtual animal's visual appearance influence the level of embodiment of users?

This chapter is based on:

1. Sierra Rativa, A. S., Postma, M., & Van Zaanen, M. (2020). Can virtual reality act as an affective machine? The wild animal embodiment experience and the importance of appearance. In Urrea C. (Eds.). *Proceedings of the MIT LINC 2019 Conference*, Vol 3. EPiC Series in Education Science https://doi.org/10.29007/dc7s
2. Sierra Rativa, A. S., Postma, M., & Van Zaanen, M. (2022). Try walking in my paws: Is it possible to increase empathy, immersion, and perceived pain in virtual reality environments by manipulating animal character appearance?. Manuscript submitted to *Virtual Reality* for a journal publication.

5.1 Abstract

Empathy towards a virtual character is an essential component of digital simulations due to that an empathetic virtual character can stimulate user immersion in a virtual environment and the sharing of another perspective of a non-human character. Past research indicates that the appearance of a character plays an important role to foster emotional reactions in users. To explore the effect of virtual animal appearance on induced perceived pain, empathy, embodiment, immersion, and con- servation tendencies, we created a simulation in virtual reality called "Justin Beaver VR". In the virtual world, the participants experienced the natural environment of a beaver from a first-person perspective. To facilitate immersion, users received physical feedback through a haptic vest with vibration motors, which conveyed the sensations of touching trees, swimming in a lake past virtual fish, and being shot at by a hunter. The appearance of the virtual character was manipulated to be either natural or robot-like. The results indicate that visual appearance has an important effect on perceived pain and immersion during the simulated experience. Participants reported that the sensation of pain through pressure in the haptic vest was stronger in the natural body appearance condition than in the robot-like body and control conditions. Similar effects were associated with immersion during the game. No effects on situational empathy, avatar embodiment or animal conservation tendencies were found.

5.2 Related work

As humans, we can experience feelings of pain that are comparable to direct pain when observing another person hurting (Singer et al., 2004). The ability to empathize with others comes about by means of reasoning about the perspective of the other (cognitive empathy) or by a direct emotional experience akin to the emotion felt by the other (affective empathy) (Shen, 2010).[3] Interestingly, empathetic reactions can be evoked not just by witnessing within species suffering (Kelly et al., 2020) but also the suffering of members of other species (Akerman, 2019; Angantyr et al., 2016) and even inanimate objects (Paiva et al., 2017). Previous research has established that the degree of empathy an observer experiences in a particular situation is largely determined by an individual disposition for empathy (Bujic et al., 2020; Davis et al., 1980). When evaluating the impact of a particular intervention, it is thus important to measure both situational and dispositional empathy in order to take into account individual differences. The effects can be measured by means of self-reports using validated questionnaires in combination with biofeedback (Spreng et al., 2009), which is considered to be an objective measure of situational empathy (Alimardani et al., 2020).

Empathy has been the subject of many classical studies in humancomputer interaction and empathetic characters or agents (Becker-Asano et al., 2005; Kober and Neuper, 2013; Kors et al., 2020; Fisher, 2017; Hoorn et al., 2004; Paiva et al., 2017; Parmar et al., 2022; Milcent et al., 2022; Moroz and Krol, 2018; Salminen et al., 2019). One of the reasons to include empathy in HCI design is the assumption that using characters or agents with cognitive and emotional traits increases their believability from the point of view of the users (Becker-Asano et al., 2005). Milcent et al. (2022) and Paiva et al. (2017) discussed five main factors that can foster empathy in HCI simulations: (1) *context & situation* where the user becomes a witness of a simulated situation with an affective component that is perceived by the user; (2) *empathy mechanisms* when the user interacts and possibly imitates an agent that has the ability to display emotions, for example, by means of facial expressions (see also Wang and Ruiz (2021)); (3) user characteristics such as personal history, age, gender, and other relevant factors; (4) *empathy modulation*, i.e., lowering or heightening the empathic reaction of the user due to mechanisms such as perceived similarity with the agent, the affective relationship between the user and the agent, or mood and personality of the user; and (5) *virtual*

3 While empathy and sympathy are closely related, they likely refer to two distinct processes. With respect to sympathy, the observer does not experience the emotion observed in another person but rather the feeling of caring for or loving the other person (Ramirez, 2017, p. 510).

agent characteristics such as appearance, experience, behavior, and intellectual ability (Van Vugt et al., 2006; van Vugt et al., 2009a). With respect to the last factor, it is important to note that relatively little is known about empathy towards virtual agents that produce the Uncanny Valley effect which is likely to have a negative impact (Diel and MacDorman, 2021).

The goal of this study was to examine the possibility of eliciting empathetic reactions towards animals in distress in a simulated environment. The advantage of using a simulation rather than original material is that the user can be provided with a first-person perspective, the features of the environment can be manipulated to achieve the desired impact, and the experience can be controlled in terms of its emotional intensity. Virtual reality environments that can be used to evoke cross-species empathy have a high potential to be employed for educational purposes in the domain of nature conservation and environmentalism.

5.3 Background

5.3.1 Empathy and virtual reality

Virtual reality technology has been referred to as an "empathy machine" (Bujic et al., 2020) due to its potential for enabling users to embody another character in a highly immersive setting, thus shutting out the influence of the outside world and perceptions of self that are associated with it. According to Fairclough et al. (2020), virtual reality can even produce an analgesic effect supported by the illusion of spatial presence, top-down attention, as well as the engagement of auditory and visual senses. Different kinds of VR experiences can result in distinct levels of immersion, with head-mounted display, (such as Oculus Rift or Quest 2) applications typically being considered as high-immersive, and 360-degree videos as less or non-immersive (Kim and Lee, 2022). Additional features can support the degree of embodiment and immersion, including, but not limited to, the format of the simulation, the virtual character's appearance, the content of the narrative presented in the environment, the degree of user control, design quality, and any accompanying music.

Three different main formats of virtual reality simulations have been used in the past to explore empathy: (1) realistic 360-degree videos, (2) narrative or storytelling, and (3) animation. In the first format, 360-degree videos are used to show dramatic or sensitive situations to examine whether participants can connect with virtually presented characters in these situations (Schutte and Stilinović, 2017). The second format refers to environments that represent VR renderings of written stories (Pianzola et al., 2019). This allows for an additional type of empathy to be measured, namely associative empathy, defined in terms of identification of the character with the message of the story (Shen, 2010). Finally, the third format refers to simulations of animated characters. Examples include a meditation scenario, where participants were represented by two illuminated statue-like avatars (Salminen et al., 2019).

5.3.2 Perceived pain and empathy

Most studies of empathy focus on situations that involve personal distress, including pain, anxiety, concern, and confusion, since these lead to clearly identifiable and measurable reactions in the observer (Ickes and Decety, 2009; Nickerson et al., 2009). Pain, in particular, is a powerful reaction to discomforting sensations, which might be of a physical, emotional, or cognitive nature (Niharika et al., 2018). This experience appears to be translatable into virtual environments, as demonstrated by Cheetham et al. (2009), who analyzed the effect of an avatar in relation to pain (Cheetham et al., 2009) and by Das et al. (2005), Matamala-Gomez et al. (2019), Niharika et al. (2018), Patel et al. (2020), Sullivan et al. (2000), Wiederhold et al. (2014) who focused on how virtual reality can help alleviate pain in patients. To assess perceived pain, scales involving emotional expressions, which are associated with a specific score, have been utilized, comparable to the Wong-Baker Face Pain Rating Scale (Niharika et al., 2018) or a pain assessment scale for children developed by Das et al., (2005), who applied this to virtual reality games.

Eliciting the illusion of pain in a virtual world without actually affecting the physical body requires that the boundary between the user's biological body and virtual body is blurred. Past studies convincingly demonstrate that certain conditions can lead to an illusory ownership of a virtual body or a body part (e.g., a hand or a leg). This has been repeatedly demonstrated in the well-known Rubber Hand illusion experiments (Botvinick and Cohen, 1998; Slater et al., 2008) and the outcomes are presumably generalizable to other limbs (Kokkinara and Slater, 2014). In another study, painful, neutral and pleasant stimulations of a virtual hand were performed, in order to examine behavioral and psychological reactions in the participants (Fusaro et al., 2016). Participants who recognized the painful stimuli of the virtual hand reported a more significant illusory ownership compared with pleasure and neutral stimulations. To our knowledge, the illusion has not been tested with nonhumanoid avatars.

Virtual reality allows for a first-person perspective embodied experience, i.e., the ability to experience a virtual body as one's own biological body. With the studies of Vinayagamoorthy et al. (2004) and Kilteni et al. (2012), we can identify two key elements that must be considered for the naturalness of the virtual experience, namely, the appearance of the avatar, and the sensory feedback that the user usually observes in their biological body (Kilteni et al., 2012; Vinayagamoorthy et al., 2004). A visual and auditory simulation combined with a virtual haptic experience can offer internal body sensations such as temperature, vibration, pressure, and texture (Gibbs et al., 2022). These elements can even be used with virtual bodies of non-human life forms, as in the study of Liu et al. (2017) who designed a virtual tree storytelling, where users were able to achieve a first-person perspective embodied experience of being a tree. Arguably, the illusion of acceptance of the virtual body as a person's own is more likely to occur with robotic and cartoon bodies compared to virtual human bodies (Lugrin et al., 2015) due to the Uncanny Valley effect (Mori et al., 2012), particularly when an "imperfection of the recreation becomes highly disturbing, or even repulsive" (Vinayagamoorthy et al., 2004, pp. 87). Also, similarity, realism, and idealistic characteristics of the embodied characters can affect the engagement reactions of users (Van Vugt et al., 2006; van Vugt et al., 2009a,b; Vugt et al., 2008).

The effects of imperfection of the representation of the natural world in virtual designs can also be detected when users interact with virtual characters or avatars that depict animals (Schwind et al., 2017). Our previous studies on the character appearance of a virtual animal roughly distinguish between two kinds: (1) the natural appearance, where the virtual design representation is similar to the biological animal in reality (e.g., similarities to the biological color), and (2) the artificial appearance, where the virtual design was easily differentiated from the real-life appearance of the biological animal (e.g., significantly different from the biological color). However, it is not clear if users consider the naturalness or artificiality of an animal character to be of importance in virtual environments.

One of the greatest challenges is developing avatar control for animal characters in virtual reality is due to their shapes, skeletons, and postures which differ from those of a human agent (Krekhov et al., 2019b; Škola and Liarokapis, 2021). For example, as demonstrated by Pimentel and Kalyanaraman (2021), there are technological limitations to visuomotor synchrony in virtual reality when the user attempts to control a virtual tail of a turtle. A possible solution lies in providing virtual animals with anthropomorphic properties that differ from their original biological characteristics (Krekhov et al., 2019a), such as an erect posture which may also help to decrease the occurrence of cybersickness.

The main purpose of this study is to investigate how the embodiment of a life form, namely a virtual wild animal, can affect the user's own perception of pain, empathy, embodiment, and immersion, as well as their animal conservation tendencies, thus providing a basis on which to improve our relationship with nature with the help of immersive technologies. Furthermore, this research can contribute to our understanding of virtual reality avatar technology, which is primarily designed as a mirror of the human body, and mostly disregards the possibilities of other types of physiognomies.

5.3.3 Research questions

The main research questions addressed in this article are:

1. Can the visual body appearance (artificiality/naturality) of an animal or robot animal influence perceived pain, situational empathy, avatar embodiment, immersive experience, and animal conservation tendency in virtual reality?
2. What are the relationships between dispositional empathy, situational empathy, perceived pain, immersion, avatar embodiment, immersive experience, and animal conservation tendency in this virtual reality simulation?
3. Which emotions are felt by participants when their virtual character dies during such a virtual reality simulation?

5

5.4 Methods

The principal objective of this study is to explore the effect of virtual appearance on immersion in a virtual world, i.e., the degree to which users would feel personally affected by virtual events experienced from a first-person perspective. The study followed the recommendations and procedures of the institutional Ethics Review Board and received an ethical approval (reference Addendum RECD#2017/01).

5.4.1 Participants

Ninety undergraduate students (38 female, 18–35 years old, M_{age}=22.63, *SD*=3.682) from Tilburg University. They participated in the experiment in exchange for course credits. All participants had a good command of English. The participants reported their frequency of playing (34.5%), occasionally or several times a month (21.1%), and rarely or never (44.4%).

5.4.2 Material

Virtual environment

We adapted the virtual reality environment of the game on the basis of previous research conducted into the appearance of virtual characters in a computer game environment in a game called "Justin Beaver". The simulation for this study in virtual reality is called "Justin Beaver VR" (see Figure 5.4.1). For the new version of the game, we made a number of modifications in relation to the previous version concerning the appearance of the animal, the game perspective (first person view), we reduced the time duration of the game, we added the possibility for a participant to see themselves in a mirror at the beginning of the game to enhance the embodiment of the character, and, finally, we added sensory feedback through an external vest.

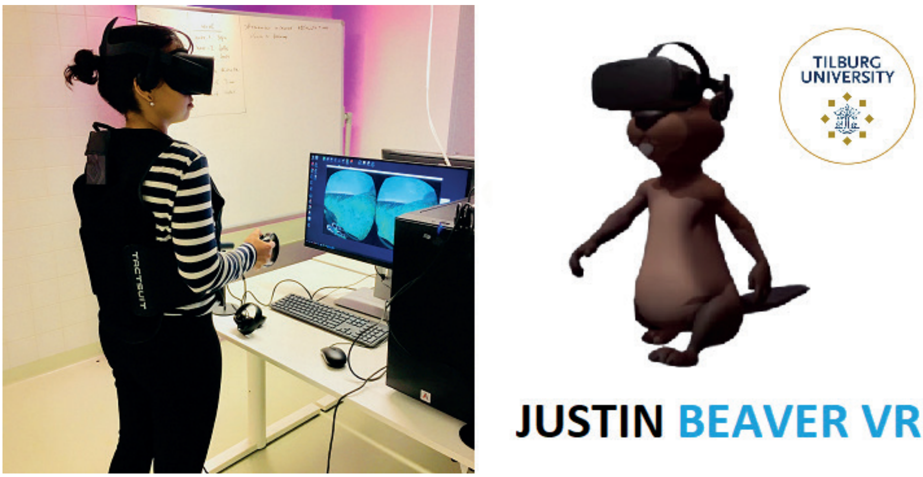


Figure 5.4.1: The virtual reality simulation used in this study is called "Justin Beaver VR". In this simulation, the participant wears VR glasses, touch controllers and a VR vest.

VR apparatus

Participants were invited to sit individually in front of a computer and were accompanied by a researcher throughout the simulation. They wore a Head-Mounted Display device (Oculus Rift S), two touch controllers, and a haptic virtual reality vest (TactSuit bHaptics; see Figure 5.4.1).

Haptic vest

The vest makes it possible to create the illusion of touch in a virtual reality environment through a total of 40 Haptic Feedback Points (vibration motors). It is adjustable to the size of the wearer, and supports a wireless connection. In Figure 5.4.2, we show how the vest's sensory feedback was automatically activated in the following situations:

1. The sensation of eating or transforming trees into pieces of wood was created by the activation of four frontal motors with a slight intensity (Figure 5.4.2a);

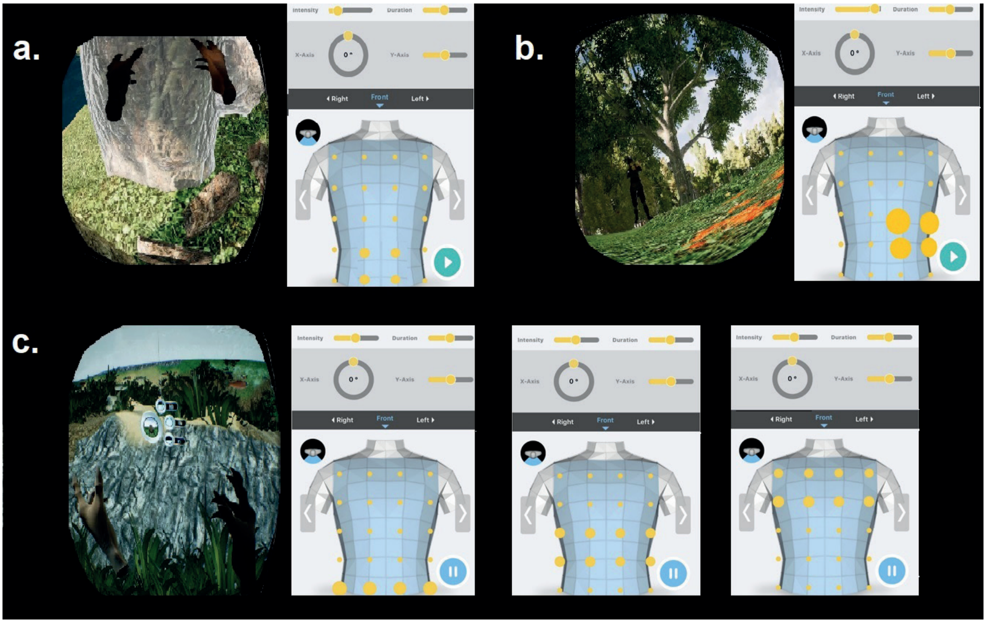


Figure 5.4.2: Activation of feedback sensations of the beaver: (a) sensation of eating; (b) sensation of receiving a shot; (c) sensation of swimming with the vest.

2. The sensation of being shot was created with the activation of four frontal motors with high intensity (Figure 5.4.2b);
3. The sensation of swimming was created with the horizontal linear sequential activation of four frontal motors with moderate intensity (Figure 5.4.2c).

Appearance of the virtual character

In this study, the primary virtual character (a virtual beaver) was presented with three different body appearances to different groups of participants, with the first two conditions presenting a natural or an artificial appearance and the third being a control condition:

1. Natural appearance: The body of the beaver had a brown color with a humanoid cartoon manifestation of the biological beaver (Figure 5.4.3a). In the scenes where only the beaver's 'hands' were observed, they were designed with a brown color and claws resulting in a more naturalistic appearance (Figure 5.4.3d).
2. Artificial appearance: The body of the beaver had a gray color with a humanoid cartoon manifestation of the robotic beaver (Figure 5.4.3b). In scenes where only the robotic beaver's 'hands' were seen, they were rendered with a metallic color and claws (Figure 5.4.3e).
3. Amorphous figure: The character had no particular shape, and a white color (Figure 5.4.3c). In scenes where only the 'hands' are observed, they appeared similar to the amorphous figure itself (Figure 5.4.3f).

After the pilot tests of the experiment, the participants reported feeling dizzy when they had walked on four legs. For this reason, we decided to provide the virtual character with a more humanoid body. In the first minute, the participant could see the complete body of the virtual character in front of the mirror. After that, the participants only saw their virtual hands, in order to reduce potential symptoms of sickness when they walked in the virtual environment, and to make the simulation more pleasant.

5.4.3 Procedure

Prior to the experiment, the participants were informed that they would be placed in a virtual reality environment with sensory feedback. We excluded students with a history of epilepsy, migraines, and the possibility of being pregnant. Once in the lab, the experimental leader informed participants that they could possibly observe some stressful events in the game, and only if they agreed, were they presented with further instructions and the consent form. Participants could stop the game at any time without providing an explanation. After signing the consent form, the participants filled out an online questionnaire called the Interpersonal Reactivity Index (IRI) and a pretest questionnaire of self-reported animal conservation tendencies, which took about 10 minutes to complete. The experimental leader explained that the participant's task was to care for their virtual character and keep it alive until the end of the game. The experimental leader also informed the participant that they would not be left alone during the entire simulation, which would last 5 minutes and 16 seconds. After that, the participant was randomly assigned to one of the three conditions (natural appearance, virtual appearance, or control). The participant received instructions about how to move their virtual character with the touch controllers, and was equipped with the VR headset, two controllers, and a virtual reality vest.

The game consisted of three main parts. The first part involved presentation of the character in a virtual mirror (see Figures 5.4.3a, 5.4.3b, and 5.4.3c), and had a duration of 1 minute. In this part, the participants could visually explore their virtual body (natural, artificial or control) in mirror. A countdown timer was visible to the participant, showing when the game would begin. The second part was contained interaction with the virtual habitat of the character (see Figures 5.4.3d, 5.4.3e), and had a duration of 4 minutes. In this part participants explored a virtual forest, provided the character with energy by "eating" wood from the ground and fish from the river, had a break time while napping in a wooden house, and experienced fun activities such as swimming in the river, walking in the habitat, and transforming trees into logs. Participants received visual and haptic feedback when their game character was eating, playing with trees and swimming in the river. The third part of the simulation involved the actual distress situation (see Figures 5.4.3g, 5.4.3h, and 5.4.3i), and had a duration of 16 seconds. The purpose of this scene was to create a context where a virtual character could perceive pain and to induce empathy towards the character. During the scene, the participants observed another virtual character, a hunter, heard the sound of a gunshot, and saw "blood" in their virtual abdomen. At the same time, the vibration motors were activated in that particular area of the haptic vest. Their virtual body fell to the ground and participants could conclude that another virtual character caused pain to their avatar. After the distress situation, an on-screen message appeared stating that the game was over.

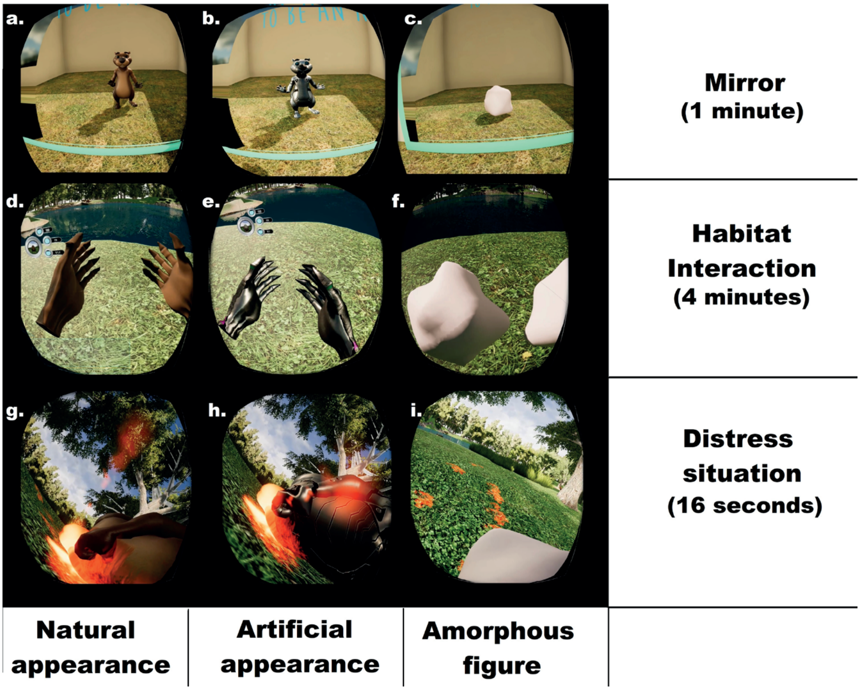


Figure 5.4.3: Scenes of the experiment with the three different virtual appearances: (a) virtual beaver in front of the mirror, (b) virtual robot beaver in front of the mirror, (c) amorphous figure in front of the mirror, (d) virtual beaver's hands during the experimental task, (e) virtual robot beaver's hands during the experimental task, (f) amorphous figure's hands during the experimental task, (g) virtual beaver's scene in the distress situation, (h) virtual robot beaver's scene in the distress situation, and (i) amorphous figure's scene in the distress situation.

Once the virtual reality simulation was completed, the participants filled out an online post-test of self-reported animal conservation tendencies, two situational empathy questionnaires, an immersion questionnaire, avatar embodiment questionnaire, a questionnaire concerning self-reported emotions about the distressing event, perceived pain questionnaire, and virtual reality sickness symptoms questionnaire (see below for more information about the instruments). The participants generally took 20 minutes to complete these. Finally, the experimenter conducted a semi-structured verbal interview about the level of emotional discomfort that the participant experienced due to the distress event. The interview took 5 minutes on average and provided the participants with the opportunity to report potential emotional distress and to receive help if needed.

Before leaving the lab, the participants obtained a letter that explained that their virtual character was in good condition, and that they could contact the university psychologist at any time if they suffered any discomfort as a result of this experiment. In total, the experiment took 40 minutes on average.

5.4.4 Instruments

Perceived pain questionnaire

In order to measure the sensation of pain during the virtual distress event, we used a perceived pain rating scale based on the face scale proposed by Das et al. (2005). As is illustrated in Figure 5.4.4, the questionnaire employs a 5-point scale for participants to assess the level of pain they felt (see **Appendix C**). The simple question used was: Q1. 'Which picture best depicts how the virtual shot made you feel?'.

Dispositional empathy

Dispositional empathy was measured using the Interpersonal Reactivity Index (IRI) questionnaire (Davis et al., 1980). This questionnaire consists of four subcomponents: perspective taking, fantasy, empathic concern, and personal distress. It contains 28 items with responses based on a 5-point scale (A='Does not describe me well' to E='Describes me very well'). The twenty items (Q1–Q2, Q4–Q6, Q8–Q11, Q16–Q17, and Q20–Q28) were scored: A=0, B=1, C=2, D=3, E=4. The seven reverse-scored items of this questionnaire were: Q3, Q7, Q12, Q13, Q14, Q15 and Q19 (See Table 5.4.1 and **Appendix C**).

The Keiser-Meyer-Olkin was 0.700 and the Bartlett test of Sphericity was highly significant, suggesting the data were suitable for PCA. The analysis and the accompanying scree plot (as illustrated in Figure 5.4.5) of the PCA revealed four major factors corresponding to the original IRI test. Factor 1 was categorized as 'Fantasy', factor 2 was 'Personal Distress', factor 3 stood for 'Perspective Taking', and factor 4 for 'Empathic Concern' (see **Appendix C**).

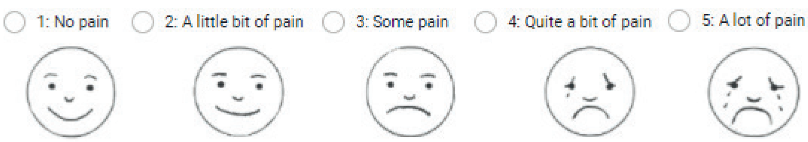


Figure 5.4.4: Faces of Perceived-Pain Questionnaire.

Table 5.4.1: Interpersonal Reactivity Index (IRI).

Question	Description
Q1	I daydream and fantasize, with some regularity, about things that might happen to me.(FANTASY)
Q2	I often have tender, concerned feelings for people less fortunate than me. (EMPATHIC CONCERN)
Q3	I sometimes find it difficult to see things from the "other guy's" point of view. (PERSPECTIVE TAKING)
Q4	I sometimes I don't feel very sorry for other people when they are having problems. (EMPATHIC CONCERN)
Q5	I really get involved with the feelings of the characters in a novel. (FANTASY)
Q6	In emergency situations, I feel apprehensive and ill-at-ease. (PERSONAL DISTRESS)
Q7	I am usually objective when I watch a movie or play, and I don't often get completely caught up in it. (FANTASY)
Q8	I try to look at everybody's side of a disagreement before I make a decision. (PERSPECTIVE TAKING)
Q9	When I see someone being taken advantage of, I feel kind of protective towards them. (EMPATHIC CONCERN)
Q10	I sometimes feel helpless when I am in the middle of a very emotional situation. (PERSONAL DISTRESS)
Q11	I sometimes try to understand my friends better by imagining how things look from their perspective. (PERSPECTIVE TAKING)
Q12	Becoming extremely involved in a good book or movie is somewhat rare for me. (FANTASY)
Q13	When I see someone get hurt, I tend to remain calm. (PERSONAL DISTRESS)
Q14	Other people's misfortunes do not usually disturb me a great deal. (EMPATHIC CONCERN) (-)
Q15	If I'm sure I'm right about something, I don't waste much time listening to other people's arguments. (PERSPECTIVE TAKING)
Q16	After seeing a play or movie, I have felt as though I were one of the characters. (FANTASY)
Q17	Being in a tense emotional situation scares me. (PERSONAL DISTRESS)
Q18	When I see someone being treated unfairly, I sometimes don't feel very much pity for them. (EMPATHIC CONCERN)
Q19	I am usually pretty effective in dealing with emergencies. (PERSONAL DISTRESS)
Q20	I am often quite touched by things that I see happen. (EMPATHIC CONCERN)
Q21	I believe that there are two sides to every question and try to look at them both. (PERSPECTIVE TAKING)
Q22	I would describe myself as a pretty soft-hearted person. (EMPATHY CONCERN)
Q23	When I watch a good movie, I can very easily put myself in the place of a leading character. (FANTASY)
Q24	I tend to lose control during emergencies. (PERSONAL DISTRESS)
Q25	When I'm upset at someone, I usually try to "put myself in his shoes" for a while.(PERSPECTIVE TAKING)
Q26	When I am reading an interesting story or novel, I imagine how I would feel if the events in the story were happening to me. (FANTASY)
Q27	When I see someone who badly needs help in an emergency, I go to pieces. (PERSONAL DISTRESS)
Q28	Before criticizing somebody, I try to imagine how I would feel if I were in their place. (PERSPECTIVE TAKING)

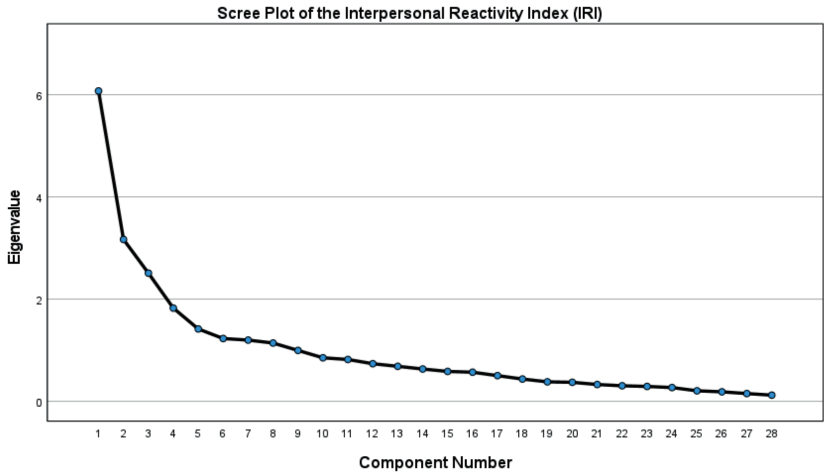


Figure 5.4.5: Scree plot showing four groups (factors) of the IRI test.

The scales of the IRI questionnaire were calculated after of a factor analysis (rotated component matrix) as:

1. Perspective taking = Q8+Q9+Q15+Q21+Q25+Q28
2. Fantasy = Q1+Q5+Q7+Q12+Q16+Q23+Q26
3. Empathic concern = Q3+Q4+Q10+Q14+Q18+Q20
4. Personal distress = Q2+Q6+Q11+Q13+Q17+Q19+Q22+Q24+Q27

   The Cronbach's α for the IRI questionnaire was reliable with α = 0.848.

Self-reported situational empathy

Two questions were used to measure the situational empathy toward the virtual character: Q1. 'To what extent did you empathize with the character (animal/robot-animal/amorphous figure)?', and Q2. 'To what extent did you empathize with the game character?'. We used a 5-point scale where the extremes were scored (1='Not at all' and 5='Very much'). The Self-reported situational empathy was calculated as the mean score of the two questions (see **Appendix C**).There was a positive correlation between the two questions, r = 0.713, n = 90, p = 0.001. The Cronbach's α for the Self-reported situational empathy was reliable with α = 0.832.

5

Self-reported emotions in the distressing event

We evaluated self-reported emotions when the virtual animal was placed in the distressing event. The single item asked participants to answer the question 'What did you feel when the game character died?' The list of the emotions was as follows: joy or happiness, surprise, anger, sadness, fear or scared, disgust, nothing, and other (see **Appendix C**). The participants had the possibility to select more than one answer.

Immersion

Immersion was measured using a questionnaire constructed by Jennett et al. (2008). This questionnaire contains 31 items, with responses based on a 5-point scale (1='Not at all' to 5='Very much'). The twentyfive items Q1-Q5, Q7, Q11-Q17, Q19 and Q21-Q31 were scored from 1 to 5. The six reverse-scored items of this test were: Q6, Q8, Q9, Q10, Q18 and Q20 (See Table 5.4.2 and **Appendix C**). The Cronbach's alpha for the Immersion questionnaire was reliable with $\alpha=0.883$

Avatar embodiment questionnaire

The avatar embodiment from a first-person perspective was measured using a questionnaire based on Gonzalez-Franco and Peck (2018) with some questions adapted to the purpose of this study. The questionnaire is based on six scales: body ownership, agency and motor control, tactile sensations, location of the body, external appearance, and response to external stimuli. It contains 25 items, with responses based on a 7-point scale (1 (strongly disagree), 2 (disagree), 3 (somewhat disagree), 4 (neither agree nor disagree), 5 (somewhat agree), 6 (agree), 7 (strongly agree)). Subsequently, the items were recoded in line with the previous use of the questionnaire. The twenty-three items (Q1, Q3 to Q4, Q6 to Q25) were recoded as: 7=3, 6=2, 5=1, 4=0, 3=−1, 2=−2, and 1=−3.

The seven reverse-scored items were: Q2 and Q5: 7=−3, 6=−2, 5=−1, 4=0, 3=1, 2=2, and 1=3 (See Table 5.4.3 and **Appendix C**).

The Keiser-Meyer-Olkin was 0.860 and the Bartlett test of Sphericity was highly significant, suggesting the data were suitable for PCA. The scree plot (as illustrated in Figure 5.4.6) and the accompanying analysis revealed six factors. Factor 1 was categorized as 'Response to external stimuli', factor 2 as 'Body ownership', factor 3 as 'Tactile Sensations', factor 4 as 'Location of the body', factor 5 as 'Agency and motor control', and factor 6 as 'External appearance' (see **Appendix C**).

Table 5.4.2: Immersion questionnaire.

Question	Description
Q1	To what extent did the game hold your attention? (ENGAGEMENT)
Q2	To what extent did you feel you were focused on the game? (ENGAGEMENT)
Q3	How much effort did you put into playing the game? (ENGAGEMENT)
Q4	To Did you feel that you were trying you best? (ENGAGEMENT)
Q5	To what extent did you lose track of time? (ENGROSSEMENT)
Q6	what extent did you feel consciously aware of being in the real world whilst playing? (ENGROSSEMENT) (-)
Q7	To what extent did you forget about your everyday concerns? (ENGROSSEMENT)
Q8	To what extent were you aware of yourself in your surroundings? (ENGROSSEMENT) (-)
Q9	To what extent did you notice events taking place around you? (ENGROSSEMENT) (-)
Q10	Did you feel the urge at any point to stop playing and see what was happening around you? (ENGROSSEMENT) (-)
Q11	To what extent did you feel that you were interacting with the game environment? (ENGROSSEMENT)
Q12	To what extent did you feel as though you were separated from your real-world environment? (ENGROSSEMENT)
Q13	To what extent did you feel that the game was something you were experiencing, rather than something you were just doing? (ENGROSSEMENT)
Q14	To what extent was your sense of being in the game environment stronger than your sense of being in the real world? (ENGROSSEMENT)
Q15	At any point did you find yourself become so involved that you were unaware you were even using controls? (ENGROSSEMENT)
Q16	To what extent did you feel as though you were moving through the game according to you own will? (ENGAGEMENT)
Q17	To what extent did you find the game challenging? (ENGROSSEMENT)
Q18	Were there any times during the game in which you just wanted to give up? (ENGAGEMENT) (-)
Q19	To what extent did you feel motivated while playing? (ENGAGEMENT)
Q20	To what extent did you find the game easy? (ENGROSSEMENT) (-)
Q21	To what extent did you feel like you were making progress towards the end of the game? (ENGAGEMENT)
Q22	How well do you think you performed in the game? (ENGAGEMENT)
Q23	To what extent did you feel emotionally attached to the game? (ENGROSSEMENT)
Q24	To what extent were you interested in seeing how the game's events would progress? (ENGROSSEMENT)
Q25	How much did you want to "win" the game? (ENGROSSEMENT)
Q26	Were you in suspense about whether or not you would win or lose the game? (ENGROSSEMENT)
Q27	At any point did you find yourself become so involved that you wanted to speak to the game directly?(ENGROSSEMENT)
Q28	To what extent did you enjoy the graphics and the imagery? (ENGROSSEMENT)
Q29	How much would you say you enjoyed playing the game? (ENGROSSEMENT)
Q30	When interrupted, were you disappointed that the game was over? (ENGROSSEMENT)
Q31	Would you like to play the game again? (ENGROSSEMENT)

Table 5.4.3: Avatar embodiment questionnaire.

Question	Description
Q1	I felt as if the virtual body I saw when I looked down was my body (BODY OWNERSHIP)
Q2	It felt as if the virtual body I saw was someone else (-) (BODY OWNERSHIP)
Q3	It seemed as if I might have more than one body (BODY OWNERSHIP)
Q4	I felt as if the virtual body I saw when looking in the mirror was my own body (BODY OWNERSHIP)
Q5	I felt as if the virtual body I saw when the game character looked in the mirror was another person (-) (BODY OWNERSHIP)
Q6	It felt like I could control the virtual body as if it was my own body (AGENCY AND MOTOR CONTROL)
Q7	The movements of the virtual body were caused by my movements (AGENCY AND MOTOR CONTROL)
Q8	I felt as if the movements of the virtual body were influencing my own movements (AGENCY AND MOTOR CONTROL)
Q9	I felt as if the virtual body was moving by itself (AGENCY AND MOTOR CONTROL)
Q10	It seemed as if I felt the touch of water when the game character was swimming (TACTILE SENSATIONS)
Q11	It seemed as if I felt the trees when the game character touched them (TACTILE SENSATIONS)
Q12	It seemed as if I felt the virtual shot when I was felt the vibrations in the vest (TACTILE SENSATIONS)
Q13	It seemed as if I felt the virtual shot when I was felt the vibrations in the vest (TACTILE SENSATIONS)
Q14	I felt as if my body was located where I saw the virtual body (LOCATION OF THE BODY)
Q15	I felt out of my body (LOCATION OF THE BODY)
Q16	I felt as if my (real) body was drifting towards the virtual body or as if the virtual body was drifting towards my (real) body" (LOCATION OF THE BODY)
Q17	I felt as if my (real) body was turning into an 'avatar' body (EXTERNAL APPEARANCE)
Q18	At some point it felt as if my real body was starting to take on the posture or shape of the virtual body that I saw (EXTERNAL APPEARANCE)
Q19	At some point it felt that the virtual body resembled my own (real) body, in terms of shape, skin tone or other visual features (EXTERNAL APPEARANCE)
Q20	I felt like I was wearing different skin from when I came to the laboratory (EXTERNAL APPEARANCE)
Q21	I felt as if my own body could be affected by the virtual shot (RESPONSE TO EXTERNAL STIMULI)
Q22	I felt a distress sensation in my body when I saw the shooter (RESPONSE TO EXTERNAL STIMULI)
Q23	When the virtual shot happened, I felt the instinct to escape (BODY OWNERSHIP)
Q24	I felt as if my virtual body had vibration sensations (RESPONSE TO EXTERNAL STIMULI)
Q25	I had the feeling that I might be harmed by the virtual shot (RESPONSE TO EXTERNAL STIMULI)

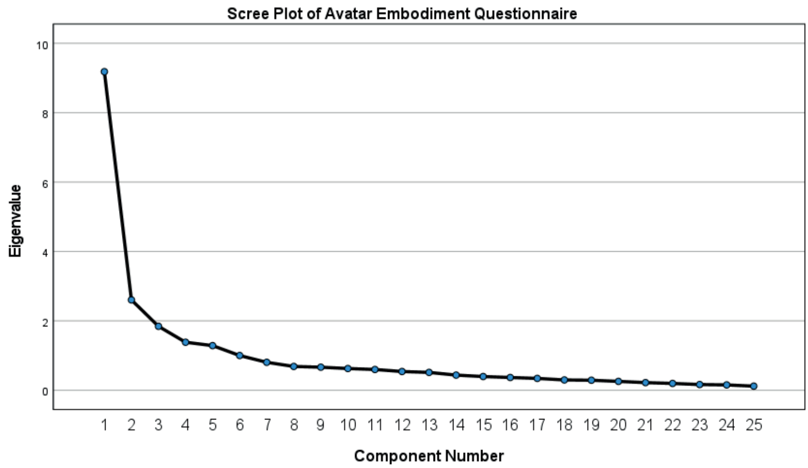


Figure 5.4.6: Scree plot showing six groups (factors) of the Avatar Embodiment Questionnaire.

The scales of the avatar embodiment questionnaire were subsequently calculated as follows:

1. Body ownership = (Q1–Q2)+(Q4–Q5)
2. Agency and motor control = Q6+Q7
3. Tactile sensations = (Q10–Q11)+Q12+Q13+Q18
4. Location of the body = Q8+Q14–Q15+Q16+Q17
5. External appearance = Q3+Q9
6. Response to external stimuli = Q19+Q20+Q21+Q22+Q23+Q24+Q25
7. Total Avatar Embodiment = ((Ownership/4) * 2 + (Agency/2) * 2 + Tactile Sensation/5 + (Location/5) *2 + Appearance/2 + Response/7)) / 9

The Cronbach's alpha for the Avatar Embodiment Questionnaire was reliable with $\alpha = 0.922$.

Animal conservation tendency

Participants' animal conservation tendencies were measured using four items: I1. 'I am really fond of (like) animals', I2. 'I am interested in protecting endangered species', I3. 'I like the animal in the picture', and I4. 'I am interested in protecting the animal in the picture'. The responses were measured on a 7-point scale (1='Strongly disagree' to 7='Strongly agree'). We employed this questionnaire before (pre-test) and after (post-test) the game. The animal conservation tendency score was calculated with the difference of the average: Post-test((Q1+Q2+Q3+Q4)/4) – Pre-test((Q1+Q2+Q3+Q4)/4) (see **Appendix C**). The Cronbach's alpha for the self-reported animal conservation tendency was reliable with $\alpha = 0.872$.

Virtual reality sickness symptoms

We measured the virtual reality sickness symptoms on the basis of an original questionnaire developed by Lu (2016), which used six items: (1). General discomfort during the game, (2). Boredom, (3). Nausea (dizziness), (4). Headache, (5). Disorientation, and (6). Stomach awareness. The responses were measured on a 5-point scale (1='None' to 5='Severe') (see **Appendix C**). The Cronbach's alpha for the Virtual Reality Sickness Symptoms was reliable with $\alpha = 0.802$.

5.5 Results

5.5.1 Dispositional empathy

The first series of ANOVA analyses examined a potential difference with respect to dispositional empathy tendencies in the three experimental conditions. As illustrated in Table 5.5.1, the results showed no significant difference concerning dispositional empathy (IRI) and its sub-scales (perspective taking, fantasy, empathic concern, and personal distress) in the two experimental and one control conditions. Therefore, dispositional empathy was not included as a factor in the subsequent analyses of the effects of the experimental conditions on situational empathy and other variables.

Table 5.5.1: One-way ANOVA on the effects of character appearance on empathy.

Source	Sum of Squares	Df	F	η^2	p
Dispositional Empathy	0.119	2	0.286	0.007	0.752
Perspective Taking	1.119	2	1.537	0.034	0.221
Fantasy	1.030	2	1.053	0.024	0.353
Empathic Concern	0.406	2	0.546	0.012	0.581
Personal Distress	0.267	2	0.363	0.008	0.696

5.5.2 Virtual reality sickness symptoms

Prior to the testing of experimental hypotheses, we ran a one-way MANOVA and found no significant difference in effect between character appearances on Virtual Reality Sickness Symptoms $F(2, 87) = 0.823$, $p = 0.443$, $\eta^2 = 0.019$.

5.5.3 Effect of virtual appearance

The first question of this study aimed to explore if the visual bodily appearance (artificiality/naturality) of an animal or robot animal can influence perceived pain, situational empathy, avatar embodiment, immersion experience, and animal conservation tendencies in virtual reality. A multivariate analysis of variance (MANOVA) revealed a statistically significant difference in character appearance with respect to these variables; $F(10, 166) = 2.113$, $p = 0.003$; Wilks $\wedge$=0.787, $\eta^2 = 0.170$.

We found a statistically significant effect of appearance on perceived pain, F (2, 87) = 8.904, $p = 0.001$, $\eta^2 = 0.170$. Follow-up comparisons (Tukey's HSD) indicated a significant difference between the condition with a natural appearance ($M = 2.73$, $SD = 1.202$) and the condition with the amorphous figure ($M = 1.60$, $SD = 0.855$, $p = 0.001$), but no significant difference between the natural appearance and the artificial appearance ($M = 2.37$, $SD = 1.098$, $p = 0.378$), see Figure 5.5.1). Moreover, we found a significant difference between the condition with an artificial appearance ($M = 2.37$, $SD = 1.098$) and the condition with the amorphous figure ($M = 1.60$, $SD = 0.855$), $p = 0.017$.

As shown in Figure 5.5.2, we found a statistically significant effect of character appearance on immersion, F (2, 87) = 3.284, $p = 0.042$, $\eta 2 = 0.070$. Follow-up comparisons (Tukey's HSD) indicated a significant difference between the condition with the natural appearance ($M = 3.67$, $SD = 0.47$) and the amorphous figure ($M = 3.35$, $SD = 0.43$), $p = 0.050$.

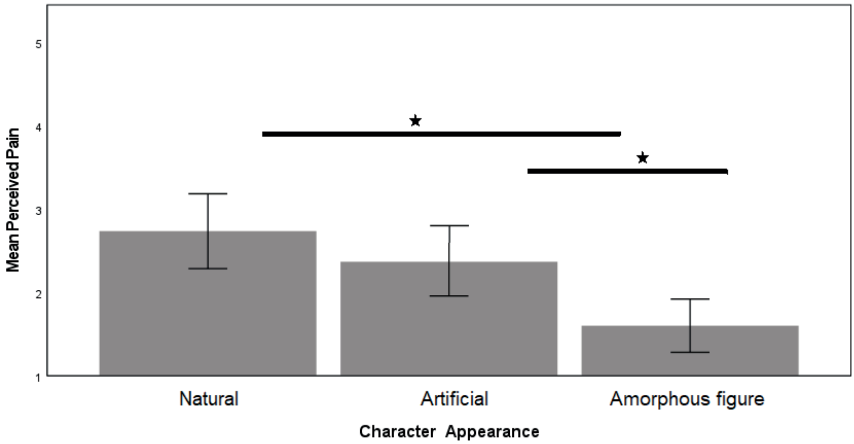


Figure 5.5.1: A bar chart of the mean perceived pain in relation to character appearance. Asterisks (*) represent statistical significance in this graph (* = p < 0.05).

There was no significant difference between the artificial appearance (M = 3.40, SD = 0.62) and the amorphous figure (M = 3.35, SD = 0.43), p = 0.950, and between the artificial appearance (M = 3.40, SD = 0.62) and the natural appearance (M = 3.67, SD = 0.47), p = 0.106.

No significant effect of character appearance was found for self-reported Situational Empathy $F(2, 87) = 2.057$, $p = 0.134$, $\eta^2 = 0.045$, or avatar embodiment $F(2, 87) = 2.277$, $p = 0.252$, $\eta^2 = 0.031$. For animal conservation tendency, we calculated a new variable as the difference between the postand pre-test of the conservation animal. Also for this variable, we found no significant effect of the character appearance $F(2, 87) = 0.299$, $p = 0.742$, $\eta^2 = 0.007$.

5.5.4 Correlations with embodiment

The second question of this study aimed to explore the relationships between dispositional empathy, situational empathy, perceived pain, immersion, avatar embodiment, immersion experience, and animal conservation tendency to embodiment in virtual reality.

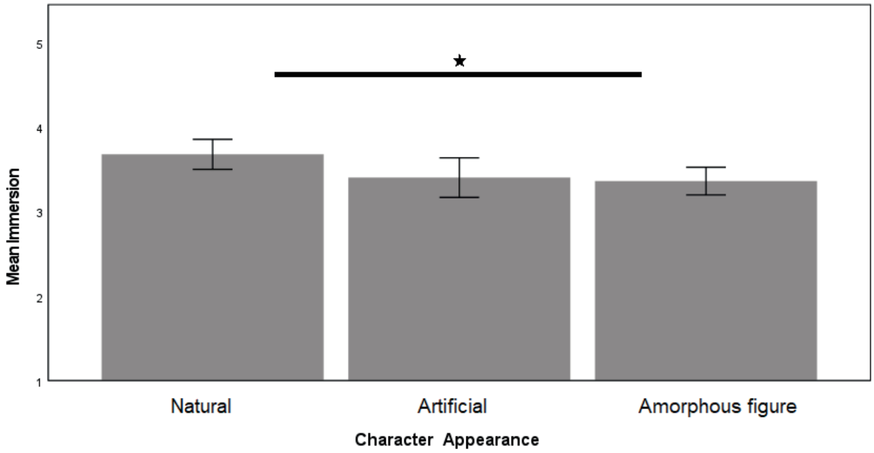


Figure 5.5.2: A bar chart of the mean immersion in relation to character appearance. Asterisk (*) represents statistical significance on this graph (* = $p < 0.05$).

A Pearson product-moment correlation coefficient was computed to assess the relationship between perceived pain and visual body appearance (see Table 5.5.2). There was a significant positive association between perceived pain and dispositional empathy ($r(90) = 0.318$, $p = 0.002$), and with three sub-scales of dispositional empathy: perspective- taking ($r(90) = 0.307$, $p = 0.003$), personal distress ($r(90) = 0.217$, $p = 0.040$), and empathic concern ($r(90) = 0.240$, $p = 0.023$). The association with fantasy was not significant ($r(90) = 0.169$, $p = 0.110$). Likewise, there was a significant correlation between perceived pain and situational empathy ($r(90) = 0.467$, $p = 0.001$), and immersion ($r(90) = 0.434$, $p = 0.001$). There was also a positive correlation between perceived pain and avatar embodiment ($r(90) = 0.346$, $p = 0.001$). Perceived pain had a positive correlation with four sub-scales of avatar embodiment: Response to external stimuli ($r(90) = 0.494$, $p = 0.001$), Tactile sensations ($r(90) = 0.271$, $p = 0.010$), Location of the body ($r(90) = 0.248$, $p = 0.019$), and Agency and motor control ($r(90) = 0.240$, $p = 0.023$). The associations with body ownership ($r(90) = 0.163$, $p = 0.123$) and external appearance ($r(90) = -0.008$, $p = 0.940$) were not significant.

As is to be expected, situational empathy toward a character had a significant positive correlation with dispositional empathy ($r(90) = 0.259$, $p = 0.014$), and with one sub-scale of dispositional empathy, namely personal distress ($r(90) = 0.202$, $p = 0.005$). There was a significant correlation between situational empathy and perceived pain ($r(90) = 0.467$, $p = 0.001$), and also between situational empathy and immersion

($r(90) = 0.567$, $p = 0.001$). Situational empathy also had a positive significant correlation with avatar embodiment ($r(90) = 0.399$, $p = 0.001$) and three sub-scales of avatar embodiment: location of the body ($r(90) = 0.446$, $p = 0.001$), response to external stimuli ($r(90) = 0.426$, $p = 0.001$), and tactile sensations ($r(90) = 0.422$, $p = 0.001$). No significant correlation was found for body ownership ($r(90) = 0.114$, $p = 0.286$), agency and motor control ($r(90) = 0.174$, $p = 0.100$), and external appearance of the avatar ($r(90) = 0.106$, $p = 0.322$).

Immersion had a significant positive correlation with avatar embodiment ($r(90) = 0.522$, $p = 0.001$), and the five sub-scales of avatar embodiment: Body ownership ($r(90) = 0.281$, $p = 0.007$), Agency and motor control ($r(90) = 0.471$, $p = 0.001$), Tactile sensations ($r(90) = 0.283$, $p = 0.007$), Location of the body ($r(90) = 0.495$, $p = 0.001$), and Response to external stimuli ($r(90) = 0.330$, $p = 0.001$). No correlation was found for External appearance of the avatar ($r(90) = 0.082$, $p = 0.441$).

Finally, animal conservation did not correlate with any of the following variables: Perceived pain, self-reported situational empathy, animal conservation tendency, immersion, and avatar embodiment questionnaire). There was a small positive correlation with tactile sensations ($r(90) = 0.240$, $p = 0.023$) and location of the body ($r(90) = 0.252$, $p = 0.017$).

Table 5.5.2: Bivariate Correlations between the measured variables.

Variable	M	SD	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
(1) Perceived Pain	2.233	1.152	-															
(2) Self-reported Situational Empathy	2.911	1.146	0.467**	-														
(3) Immersion	3.478	0.533	0.434**	0.567**	-													
(4) Animal conservation tendency	0.664	0.694	0.153	0.241*	0.272**	-												
(5) Dispositional Empathy (IRI)	2.406	0.452	0.318**	0.259*	0.097	0.017	-											
(6) Fantasy (IRI)	2.510	0.700	0.169	0.106	0.049	0.038	0.693**	-										
(7) Personal Distress (IRI)	2.030	0.602	0.217*	0.292**	0.106	0.001	0.765**	0.344**	-									
(8) Perspective Taking (IRI)	2.633	0.626	0.307**	0.141	0.048	0.040	0.622**	0.144	0.295**	-								
(9) Empathic Concern (IRI)	2.622	0.606	0.240*	0.179	0.065	0.066	0.766**	0.406**	0.405**	0.499**	-							
(10) Avatar Embodiment Questionnaire	0.116	0.730	0.346**	0.399**	0.522**	0.207	0.223*	0.096	0.270*	0.187	0.051	-						
(11) Body ownership (AEQ)	0.944	3.146	0.163	0.114	0.281**	0.072	0.038	0.082	0.041	0.148	0.027	0.635**	-					
(12) Agency and motor control (AEQ)	1.700	2.537	0.240*	0.174	0.471**	0.074	0.142	0.034	0.188	0.212*	0.040	0.724**	0.417**	-				
(13) Tactile sensations (AEQ)	1.522	5.037	0.271**	0.422**	0.283**	0.240*	0.212*	0.087	0.354**	0.067	0.024	0.705**	0.226*	0.325**	-			
(14) Location of the body (AEQ)	0.444	4.818	0.248*	0.446**	0.495**	0.252*	0.142	0.126	0.253*	0.011	0.063	0.765**	0.260*	0.363**	0.687**	-		
(15) External appearance (AEQ)	2.233	2.566	0.008	0.106	0.082	0.114	0.121	0.110	0.019	0.128	0.114	0.433**	0.404**	0.002	0.195	0.211*	-	
(16) Response to external stimuli (AEQ)	4.689	9.516	0.494**	0.426**	0.330**	0.135	0.299**	0.277**	0.269*	0.153	0.108	0.733**	0.217*	0.400**	0.639**	0.596**	0.243*	-

Note:*p<.05,**p<.01

5.5.5 Self-reported emotions during the distressing event

The third aim of this study was to explore which emotions were experienced by participants when the virtual character died. The bar chart in Figure 5.5.3 represents the degree to which each emotion was reported by participants, split into conditions by the appearance of the character (natural appearance, virtual appearance, or amorphous figure). Overall, it can be seen that the emotions most frequently reported during the distressing event were surprise (48 times) and sadness (31 times). Following these were neutral emotion (14 times), fear (12 times), happiness (9 times), disgust (8 times), anger (7 times), and other emotions (3 times). In terms of conditions, sadness, surprise, and neutral emotion were reported with similar frequency in the two natural and artificial appearance conditions. Interestingly, happiness and fear were reported more frequently in the artificial appearance condition compared to the amorphous figure and natural appearance. The participants reported significantly more anger in the natural appearance condition compared to the artificial appearance condition, while the anger emotion was absent for the amorphous figure. In contrast, the participants reported significantly more frequently disgust in the artificial appearance condition compared with the amorphous figure and natural appearance.

5.6 Discussion

The general aim of our study was to explore whether the visual appearance of a virtual animal can be used to foster empathy by eliciting the perception of pain in a simulated environment providing first-person view. We assumed that it is possible to elicit virtual pain since pain has a considerable psychological component (Panda, 2017; Wismeijer and Vingerhoets, 2005, p. 268, p. 280). In our experimental design, we included the impact of character appearance on situational empathy, avatar embodiment, immersion experience, and animal conservation tendency, and we also took into account potentially confounding variables such as dispositional empathy and VR sickness symptoms. The virtual character appearance was systematically manipulated in terms of naturalness versus artificiality, with an amorphous figure included as a control condition.

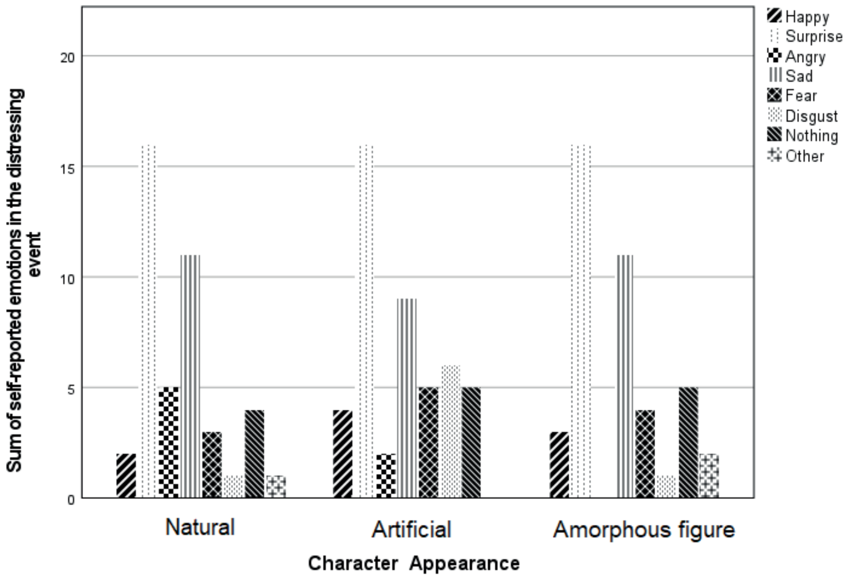


Figure 5.5.3: Bar chart of the frequency of each emotion reported by participants during a distress situation.

The most important finding was that the naturalistic appearance of the virtual character resulted in a higher degree of perceived pain compared to the artificial appearance and the control condition. A possible explanation of this result is that at the beginning of the game, the users could more easily identify with a character that was more familiar to them. This explanation is supported by the overall positive correlation between perceived pain and avatar embodiment.

According to Bujic et al. (2020), a first-person view of tragic situations promotes empathy. Contrary to previous findings of this research (Sierra Rativa et al., 2020), the appearance of the virtual character did not affect situational empathy in our study. We did, however, find a relatively strong positive relation between perceived pain and situational empathy, as well as a correlation with dispositional empathy. It is worth noting that the questionnaire used to assess situational empathy only consisted of two items. Despite a high internal consistency, it is possible that including additional items could improve the assessment of empathy users experience in a particular context. Next to that, the self-reported situational empathy measure could be enriched by psychophysiological and EEG data. It would also be interesting to compare the effects of first-person versus third-person view on situational empathy in relation to the distinction between 'imagine-self' and 'imagine-other' that is sometimes made in the literature (See Figure 5.6.1).

Figure 5.6.1: Third-person view versus first-person with a virtual animal (e.g., beaver).

Virtual reality simulations typically lead to high immersion due to their visual, auditory and other sensory stimulation (Bowman and McMahan, 2007). The outcomes of this study revealed that a naturalistic appearance results in a higher degree of immersion, compared to artificial appearance and the control condition. Immersion was, in turn, strongly correlated with situational empathy, which raises the question if there is a potential causal relation between the two constructs. This result is in line with a previous observation of (Kober and Neuper, 2013), who suggested that there is a strong relationship between the degree of immersion, presence, and empathy.

According to Banakou et al. (2013), users in virtual reality environments can adopt a body different to their biological one and this may affect their perceptual processes. For instance, Banakou et al. (2013) found that if a user embodies a child body, they can develop an overestimation of the sizes of objects in virtual reality (Banakou et al., 2013). Another study developed by Schloss et al. (2021). Schloss showed that children prefer to embody a child body compared with animal and anthropomorphise fictional Muppet virtual body in virtual reality. However, children wanted to touch an animal compared with other features. The interesting results of this study are that children showed great curiosity and engagement regardless of body type in virtual reality. We analyzed avatar embodiment from a first person perspective using a questionnaire based on one developed by Gonzalez-Franco and Peck (2018). While our study seems to confirm that adopting the body of a virtual animal is possible, as demonstrated by the outcome that users can feel virtual pain when the animal is attacked, surprisingly, the appearance of the virtual animal did not seem to have an effect on the avatar embodiment. Perhaps additional sensory feedback could further enhance the adoption of a virtual body as one's own and reveal potential impact of character appearance.

We did not find any significant impact on participants' animal conservation tendency. Even though the questionnaire appeared to have a high internal consistency and we did observe a small positive correlation between situational empathy, immersion, and conservation we think that further work is required to develop and validate a questionnaire that could be used in experimental settings. Next to that, the game developed for the purposes of this study was relatively short and as such the experience may not have been sufficient to bring about an attitude change.

Finally, we explored which emotions were experienced by the participants in relation to the virtual distressing event. The participants most frequently reported feeling surprised and sad. In a follow-up study, it would be interesting to combine the VR headsets with a measurement of muscle movements in the participants facial area in order to combine the self-reports with behavioral data.

5.7 Conclusion

The current study aimed to determine if the visual bodily appearance (artificiality/naturality) of an animal virtual character can lead to empathy towards the character including a perception of virtual pain. The results showed that a naturalistic visual bodily appearance had the strongest effect on perceived pain and immersion during the virtual reality experience. We found no statistically significant effect for situational empathy, avatar embodiment, and animal conservation tendencies. However, there was a positive correlation between perceived pain and dispositional empathy, situational empathy, immersion and avatar embodiment. Finally, the emotions of sadness and surprise were heightened within participants when their virtual character died during the virtual reality simulation. These results have vital implications for understanding how the visual design of the virtual animals influences users' reactions during a virtual reality simulation. In future research, it would be interesting to explore if empathetic responses can be affected by the user perspective (first or third-person view) within the virtual reality environment. Moreover, other concepts such as compassion, sympathy, or character identification could provide additional insights into the effect of virtual appearance.

CHAPTER 6

The effectiveness of a robot animal as a virtual instructor

This chapter tackles research question RQ7.
Research Question

7. How can different versions of virtual animals be used as virtual tutors in video instruction, which may have different effects on affective and cognitive outcomes, depending on their visual appearance?

Published Work: Sierra Rativa A., Vasquez C.C., Martinez F., Orejuela Ramirez W., Postma M., & van Zaanen M. (2021).*The Effectiveness of a Robot Animal as a Virtual Instructor.* In Lepus- chitz W., Merdan M., Koppensteiner G., Balogh R., & Obdržálek D. (Eds.), *RiE 2020: Robotics in Education*. Advances in Intelligent Systems and Computing,,1316, 329-–338.Springer. https: //doi.org/10.1007/978-3-030-67411-3_30

6.1 Abstract

In this Chapter 6, we study the use of virtual robot animals (VRAs) can have a potential impact on applications with affective and aesthetic interfaces. In particular, VRAs can be used in instructional videos in order to develop new ways to engage young learners and to foster personalization of educational instruction. In this paper, we explore the perception of the virtual instructor appearance and its effect on knowledge recall outcomes for young learners. We conducted an experiment with three different virtual instructor appearances: (1) robot animal, (2) animal, and (3) human. The content of the video instruction had two themes: (A) A topic related to robotics (e.g., introductory concepts about robotics), and (B) a topic unrelated to robotics (e.g., Dutch culture). A total of 131 students participated in this study. They originated from two secondary public schools in Bogota, Colombia. Our results showed that the robot animal as a virtual instructor was perceived as the least familiar, common, attractive, interesting, and natural compared with the virtual instructors with the animal and human appearance. Moreover, learners in the condition with the virtual robot animal scored significantly lower on knowledge recall for both topics. A follow-up study can focus on ways to increase positive reactions toward robotic animals as virtual instructors. Video about this research: (https://youtu.be/PY1CN0DoKF4).

6.2 Related work

New trends, challenges, and developments in robotics have led to the question whether a robot could become an effective teacher. In past studies, social robots have been used, for example, to foster language acquisition, and in support of science education, and technology education (Mubin et al., 2013). In this pedagogical process, the robots were used as a tool, a peer, a tutor or teacher (Causo et al., 2016; Mubin et al., 2013). While actual robots can be employed in direct physical interactions with students, there is also a possibility of developing virtual agents with robotic appearance to be used in simulated environments. For instance, Li et al. (2016) showed that it is possible to use an embodied pedagogical character with a robotic appearance as a virtual instructor. Li et al. (2016) suggest that the use of a virtual robot could have almost the same effect on students' recall compared to a human teacher in instructional video content. To test this claim, we examined whether a virtual robot with non-human traits could replace a human instructor in pre-recorded instructional video's. The result of our investigation significantly contributes to research on pedagogical agents that can be used in the area of educational robotics.

Robot instructors

One of the important goals in the field of social robotics is to design robots to be used as education companions and tutors (Causo et al., 2016). The decision on whether a robot design is useful for educational purposes is highly dependent on its impact on learning outcomes. For example, Belpaeme et al. (2018) showed than the Nao (human appearance with arms and legs) and Keepon (yellow snowman appearance without arms and legs) robots had a medium-sized effect on cognitive learning gains in students. A virtual prototype of the Nao robot was used in an instructional video by Li et al. (2016), who found that a Nao as a virtual robot agent can have a positive effect on students' knowledge recall compared to a non-virtual robot, and a similar effect to a non-virtual human teacher. This study offers some important insights into the effect of robotic appearances as a virtual instructor and their wide possibilities in the pedagogical context.

6.2.1 Robotic appearance and the uncanny valley effect

Research in robotics on the effect of robot appearances has primarily been based on the "uncanny valley theory" (Mori et al., 2012). This theory explores the effect of human-like appearance on the affinity and/or familiarity of the users towards the character. An interesting aspect of the theory is that when the robot comes closer to having human traits, it may be perceived with an increasing revulsion. This effect is called the "uncanny valley" and it is found when the robot is close to but fails to attain a realistic appearance Mori et al. (2012). While anthropomorphic features have been the subject of many classic studies in robotic appearance, recently, the focus has begun to shift toward animal appearances. For example, Schneider et al. (2007) suggested that a safe strategy preventing the uncanny valley effect is to design virtual animals with non-anthropomorphized appearance, in such a way that they could emote and communicate as a human. However, Schwind et al. (2018a) note that, similarly to virtual/robotic humans, the level of realism of the virtual animal-like characters could generate negative reactions.

6.2.2 Summary

To sum up, the literature shows that the theory of the uncanny valley can be applied to robots both with a human and non-human appearance, such as animals. However, currently very little is known on whether the familiarity's perception described in uncanny valley theory can also affect a robot with animal appearance when it acts as a virtual instructor. In our experiment, we tested different appearances of a virtual animal instructor, compared to a human instructor, on students' perception of the character and its possible effect on cognitive learning outcomes.

6.3 Method

In our previous research on the uncanny valley effect toward virtual animals (Sierra Rativa et al., 2019), we designed virtual pandas both with a robotic and a natural appearance. This study used the same virtual character appearances in order to explore if the robotized and natural version of the panda, when used as a virtual instructor have (1) a different effect on users' perception of the appearance of these characters, and (2) if the appearance of the virtual character has an effect on the cognitive learning outcomes, particularly, knowledge recall.

6.3.1 Design and material

We conducted an experiment with three different renditions of the instructor appearance, (1) virtual robot animal, (2) virtual animal, and (3) (video recorded) human, each presented in an instructional video. The content of each video instruction focused on one two themes: (A) A topic related to robotics (e.g., introduction concepts about robotics), and (B) a topic unrelated to robotics (e.g., the Netherlands and Dutch culture). This resulted in a 3 × 2 design with six experimental conditions (three virtual instructors and two topics), see Figure 6.3.1. As the native language of the participants was Spanish, all experimental material was in Spanish.

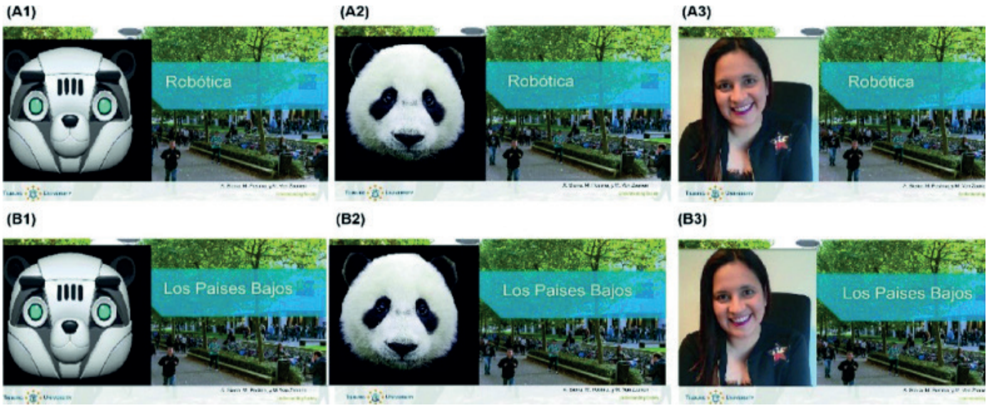


Figure 6.3.1: Participants viewed a video featuring: (A1) Robotics with robot panda instructor, (A2) Robotics with panda instructor, (A3) Robotics with a human instructor, (B1) The Netherlands with robot panda instructor, (B2) The Netherlands with panda instructor, and (B3) The Netherlands with human. Video about this research: https://youtu.be/PY1CN0DoKF4.

6.3.2 Procedure

Prior to the experimental sessions, participants were asked to fill out a pre-test called "domain knowledge". The pre-test contained questions related to robotics and Dutch culture, to ensure that participants were not already knowledgeable on the content of the instructional videos.

At the start of the experimental session, participants were asked about their age, gender, and nationality. Each participant watched one of the six (different) video instructions for 10 min. After the video, participants answered questions about: (1) knowledge recall of the presentation, (2) social presence, (3) interpersonal attraction, (4) presentation skills, (5) enthusiasm, (6) overall experience, (7) concentration, and (8) a questionnaire on the perception of virtual instructors. This questionnaire was administrated online and its completion took approximately 10 min. In total, the experiment took less than 30 min to complete by each participant.

6.3.3 Participants

Ethics approval was obtained from the Research Ethics and Data Management Committee of the Tilburg School of Humanities and Digital Sciences with the reference REC2019/89. The experiment was conducted in two secondary public schools in Bogota, Colombia (see Figure 6.3.2). A total of 131 students participated in this study. All students have either given their written consent or received consent by their legal representatives for participants younger than 18 years. The directors of the schools I.E.D Almirante Padilla and I.E.D Prado Veraniego gave permission to the teacher and researcher to conduct the study on the school premises. We recruited a total of 68 students from the Almirante Padilla School, and a total of 63 students from the School Prado Veraniego. Participants ranged in age from 11–17 years (M=13.6). The sample included 56 females (42.7%) and 75 male (57.3%) participants. The participants' courses were distributed between the seventh (59.5%) and eighth (40.5%) grade. The participants were more or less evenly distributed between the six experimental conditions.

Figure 6.3.2: Student participants (a. Prado Veraniego School and b. Almirante Padilla School).

6.3.4 Instrumentation

To measure domain knowledge and knowledge recall, we designed a pre-test and post-test for this study. The pre-test 'domain knowledge' consisted of 10 items with four multiple-choice answers on topics related to the Netherlands and to robotics. In order to test if the pre-tests questions are not easy or intuitive to answer by participants, we analyzed the pre-test results before to do the experiment. We found that the majority of participants had a low percentage of assertiveness in their answers on the pretest for all questions except question (3) "Which of these words are associated with robotics?" on the topic of robotics. For this reason, question (3) was modified in the post-test to have a higher level of difficulty in this question compared to the pre-test to ensure that the post-test depended on the video instruction stimulus and not for other reasons. The post-test, which we referred to as 'knowledge recall', consisted of the pre-test questions apart from question on the topic of robotics (which was replaced by another question). Domain knowledge and knowledge recall of the same topic were measured as the number of correct responses where the maximum possible score was 5 (see Table 6.3.1 below).

Table 6.3.1: Pre-test and post-test questionnaires.

Robotics	The Netherlands
1. DATA: How many robots are there in the world? **a. 1.2 million (Correct answer** b. 1.8 million c. 5 million d. 5.8 million	1. DATA: How man people live in the Netherlands? **a. 17 million (Correct answer)** b. 10 million c. 40 million d. 47 millions
2. CURIOUS DATA: The first robot was: a. Vacuum cleaner b. A toothbrush **c. A pigeon (Correct answer)** **d.** A mouse	2.CURIOUS DATA: The Netherlands is famous for having many: a.Houses b. Building **c. Bikes (Correct answer)** d. Cars
3. CURIOUS WORDS: Which of these words are associated with robotics and these are correctly written? (Post-test) a. Nanotenologic, zomorphic, mecatronic **b. Nanotechnology, zoomorphic, mechatronics (Correct answer)** c. Nanotecnology, zoomorphic, mechatronics Nanotechnology, zomorphic, mecathronics	
4. LIST: What robot names do you remember? a. Nao, Asimo, Mariana, Aibo, Walker **b.Nao, Asimo, Paro, Aibo, Walker (Correct answer)** c. John, Victor, Caxi, Porto, Mili d. John, Fox, Caxi, Porto, Mili	5. LIST: What names of the Netherlands do you remember? a. Tilburg, Rotterdam, Barcelona, Haya **b. Tilburg, Rotterdam, Amsterdam, Haya (Correct answer)** c. Berlin, Hamburg, Munich, Cologne d. Berlin, Bremen, Munich, Cologne
5. INTERESTING PEOPLE: What is the name of the engineer who created a robot equal to himself? a. Tanmay Bakshi b. Tanmay Baksshyn c. Hiroshi Hoshiguros **d. Hiroshi Ishiguro (Correct answer)**	5. INTERESTING PEOPLE: What are the names of the King and Queen of the Netherlands? a. Harry and Meghan b. Harry II and Meghan c. William II and Maxima **d. Willem-Alexander and Maxima (Correct answer)**

6.3.5 Liking

Li et al. (2016) developed a set of questions that measure participants' affinity towards the virtual instructor. The questionnaire consisted of three sub-components: social presence, interpersonal attraction, and presentation skills. Social presence was measured using five items. The reliability for social presence was good (standardized Cronbach's alpha α=0.707). Interpersonal attraction was measured using four items with good reliability (standardized Cronbach's alpha α=0.787). Presentation skills were measured using three questions, with good reliability (standardized Cronbach's alpha α=0.748).

Finally, a question about concentration of the participant was assessed using a single-item question with a scale from 1 to 5 (with 1: Not at all, and 5: Absolutely).

6.3.6 Appearance perception questionnaire

We used a questionnaire from our previous study with virtual animals (Sierra Rativa et al., 2019). It contains semantic differential questions that are designed to measure participant perception of the familiarity, commonality, naturalness, attractiveness, interestingness, and animateness of the virtual instructor. All measures were assessed using fivepoint Likert scales. This scale corresponds with the evaluation used in the scholar system of Colombia (see Table 6.3.2 below). The appearance perception was measured using six items with good reliability (standardized Cronbach's alpha α=0.771).

6.4 Results

Data preprocessing and analysis was performed using SPSS 25. In the first set of analyses, we examined the impact of the virtual instructor appearance on the participant's perception. Since the variables obtained in the data collection phase all showed non-normal distributions, we made use of non-parametric tests. A Kruskal-Wallis H test[4] showed that there was a statistically significant effect of the virtual instructor appearance on participants' perception of familiarity: ($\chi^2(2) = 9.109$, $p \leq 0.011$), attractiveness: ($\chi^2(2) = 8.576$, $p \leq 0.014$), interestingness: ($\chi^2(2) = 6.322$, $p \leq 0.042$), naturalness: ($\chi^2(2) = 20.478$, $p \leq 0.001$), and animateness: ($\chi^2(2) = 6.942$, $p \leq 0.031$) (see Figure 6.4.1). The mean rank of the familiarity score was 53.70 for the virtual robot panda, 66.59 for the virtual panda and 77.38 for the human instructor. The mean rank of the attractiveness score was 57.58 for the virtual robot panda, 61.57 for the virtual panda and 78.97 for the human instructor. The mean rank of the interestingness score was 58.96 for the virtual robot panda, 62.23 for the virtual panda and 76.91 for the human instructor. The mean rank of the naturalness score was 51.44 for the virtual robot panda, 60.52 for the virtual panda and 86.08 for the human instructor. The mean rank of the animateness score was 62.89 for the virtual robot panda, 58.11 for the virtual panda and 77.48 for the human instructor. There was no significant effect of appearance on commonality ($\chi^2(2) = 2.190$, $p \leq 0.335$).

4. Additional Median information about statistics analysis can be found in **Appendix F**.

Table 6.3.2: The appearance perception toward virtual instructor questionnaire.

1. What do you think of the virtual instructor's appearance?							
Familiarity	Very strange	1	2	3	4	5	Very familiar
Commonality	Very unusual	1	2	3	4	5	Very common
Attractiveness	Very ugly	1	2	3	4	5	Very attractive
Interestingness	Very boring	1	2	3	4	5	Very interesting
Naturalness	Very artificial	1	2	3	4	5	Very natural
Animateness	Very inanimate	1	2	3	4	5	Very animate

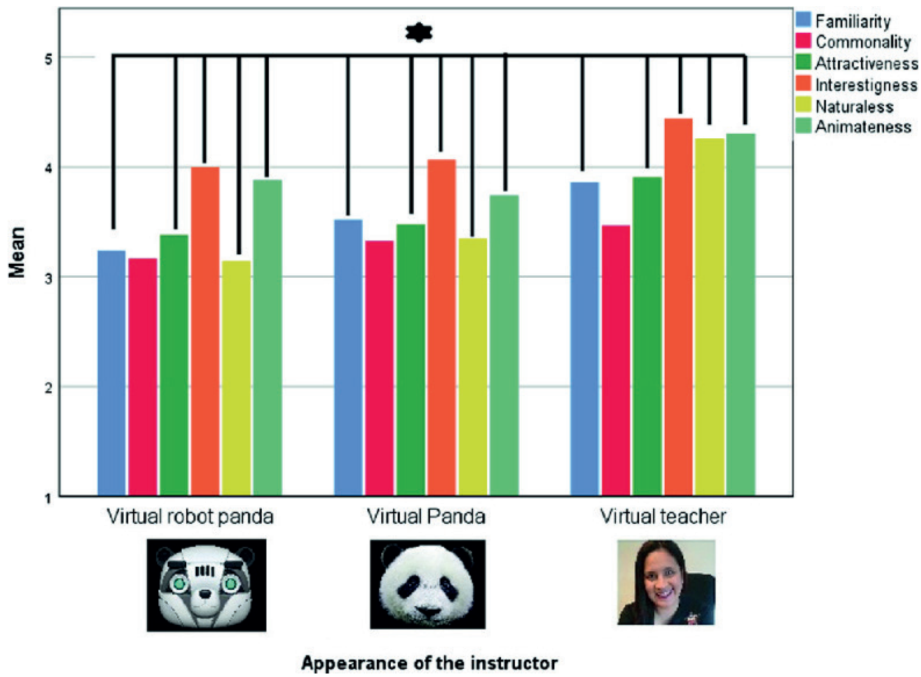


Figure 6.4.1: Mean scores of participants' perception of familiarity, commonality, attractiveness, interestingness, naturalness, and animateness in relation to the appearance of the instructors.

The second set of analyses examined the impact of the appearance knowledge recall and concentration[5]. There was no significant difference in domain knowledge (pre-test) between the experimental conditions with different virtual instructors: ($\chi^2(2) = 0.006, p \leq 0.997$). A Kruskal-Wallis H test was conducted to measure the potential effect of on domain knowledge (pre-test), knowledge recall (post-test), and concentration (see Figure 6.4.2). There was a statistically significant effect of appearance on knowledge recall (post-test): ($\chi^2(2) = 6.533, p \leq 0.038$), with a mean rank of the knowledge recall score of 55.39 for the virtual robot panda, 67.83 for the virtual panda and 72.81 for the human instructor. Finally, there was no effect of appearance on concentration:($\chi^2(2) = 0.382, p \leq 0.826$).

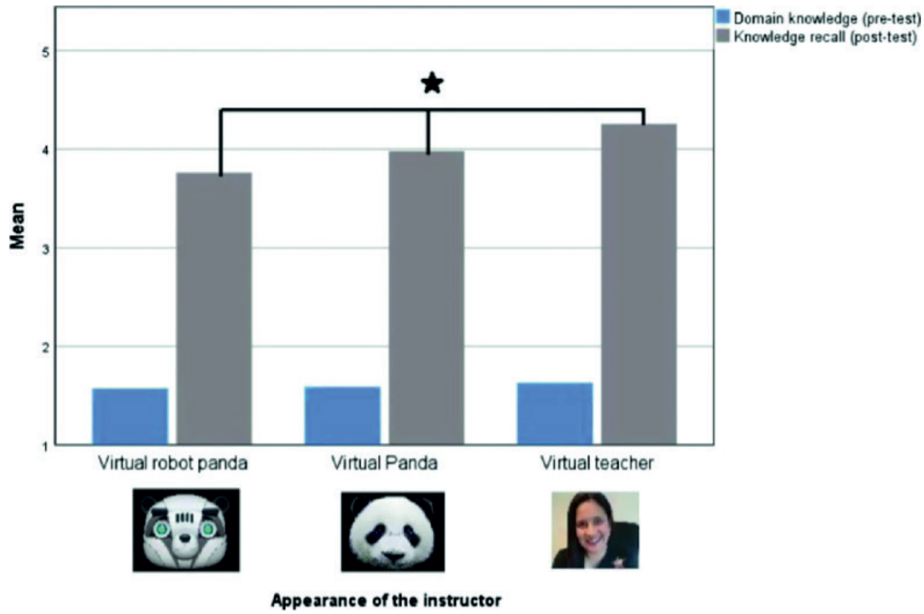


Figure 6.4.2: Participants' domain knowledge (pre-test) and knowledge recall (post-test) in relation to the appearance of the instructors.

6

5. No learning gain was calculated based on pre and post-test because according with (Boud, 2018, pp. 56–57), *"For the purposes of identification of learning gains this implies that...assessments need to be directly related to explicitly articulated course/program learning outcomes, not unit or module outcomes."*

6.5 Discussion

The first goal of the current study was to explore the relationship between the appearance of the virtual instructor and participants' perception. The results showed a significant effect of the appearance on familiarity, attractiveness, interestingness, naturalness, and animateness, with no significant effect on commonality. Contrary to expectations, the virtual robot animal instructor was considered to be the least familiar, attractive, interesting, and natural virtual character, compared to the natural version of the animal instructor and the human instructor. These results corroborate the findings of much of the previous work on the uncanny valley of virtual animals, where the robot animal led to a less pronounced 'uncanny' feeling compared with the natural version (Sierra Rativa et al., 2019). This finding, however, is contrary to previous studies which have suggested that using a virtual animal that has more animal-likeness but emotes and communicates like a human is a good strategy (Schneider et al., 2007). In our study, the virtual character retained its animal appearance, while it had the human role of being an instructor. However, students considered a human teacher more familiar, interesting and natural. This result may be explained by the fact that the learners who participated in the study are not in contact with robots during a normal school day. Another possible explanation for this is that humans tend to feel more attracted to objects with anthropomorphic properties in their appearance, such as androids or humanoid robots that have more typically been explored in the theory of the uncanny valley (Mori et al., 2012).

The second goal of this study was to determine if the appearance of the virtual instructor influences knowledge recall and another learning outcomes. The results indicate that the appearance of the virtual instructor indeed had a significant effect on knowledge recall, also when taking into account their previous knowledge about the topics. We found no effect of the appearance of the virtual instructor on concentration. On the one hand, these results are consistent with the outcomes of Li et al. (2016), who showed that the appearance of a virtual character had a recall effect on participants. On the other hand, the results showed that the virtual robot had less effect on knowledge recall compared to the animal and human instructors. In the results of Li et al. (2016), the virtual robot had a similar effect on knowledge recall compared to the human instructor, although the android appearance of their instructor could have affected the results. This study supports evidence from previous observations (e.g., Belpaeme et al. (2018)) that robotic appearances can affect learning outcomes, also in a virtual environment. One of the issues that emerge from these findings is that a virtual robot animal does not appear to be a suitable virtual instructor (yet) in instructional video content.

This study was performed in Colombia. We do not know if the type of the animal (e.g., panda) would be made a difference with an animal that the pupils have more familiarity within this country. However, the panda animal was used because we wanted to make continuity with the study of "uncanny valley of the virtual animals" where the panda animal was the main character in this previous study (Sierra Rativa et al., 2019). Current findings are consistent with the previous study where the panda robot had a low score in familiarity (which fell into the uncanny valley) compared with real panda. Likewise, these findings are consistent when it comes to an instructional video.

In future investigations, we may want to focus on virtual robotic animals or humans with different designs of virtual instructors in order to increase positive reactions by the users.

6.6 Conclusion

The purpose of the current study was to determine if a virtual robot animal can be used effectively as a virtual instructor. Our results show that a virtual robot animal does not elicit particularly positive reactions in terms of familiarity, interestingness, and naturalness compared with virtual animal and human instructors. These findings confirms our previous findings on the uncanny valley of the virtual animals, where the users had a more positive reaction toward the animal with natural traits compared to the robotic version. One of the more significant findings to emerge from this study is that the appearance of the virtual instructor had a significant effect on knowledge recall. This finding offers some insight into the design of other agents to be used as virtual instructors. To further investigate these relationships, the study should be repeated using other designs of virtual robot instructors (e.g., using other animals or humans, improving the realisticness of the animation, adding the full body, manipulating emotional facial expressions, etc.) to establish whether it will have the same effects as in this study.[6]

6. An example of a smart virtual tutor can be observed in **Appendix D**

CHAPTER 7

General discussion and conclusion

7.1 Main findings

In this dissertation, we primarily researched whether virtual animals' visual appearances can affect users' perception. This thesis has provided a deeper insight into eight research questions, which can help answer the general problem statement.

RQ1: *How can we measure 21st Century Learning Skills (creativity, collaboration, communication, and critical thinking skills) with existing tests and what is the reliability and validity of these tests?*

The answer to this question lies in the exploration of a variety of tests available to measure 21st Century Learning Skills in **Chapter 2**. We conducted a systematic review of the literature on the assessment of 21st Century Learning Skills to analyze their characteristics in terms of assessment, such as type, educational level, and test reliability and validity. The results have shown that personality, aptitude, and achievement tests are unequally used to assess learning. While personality tests are primarily used to measure collaboration and creativity skills, aptitude tests are mostly used to assess critical thinking and collaboration skills, and achievement tests are mainly used to evaluate communication skills. Critical thinking, creativity, and collaboration tests are primarily used at the university level, showing evidence of the distribution of tests by educational level. The secondary level uses all four learning skill tests, primary level uses communication and creativity skills tests, and the kindergarten educational level lacks tests to measure all four learning skills. Some studies do not specify the educational level for creativity and collaboration skills tests. Likewise, we discovered that 42.5% of the tests are valid and reliable, while for the other tests validity and reliability remain undetermined. We collected a list of possible tests to measure all four learning skills to be used for future research. However, we found an evident lack of tests focusing on collaborative and critical thinking skills. Likewise, some tests measure the simultaneity of multiple learning skills or measure subscales of the learning skills individually.

These findings contribute to our understanding of concepts such as collaboration, communication, critical thinking and creativity skills, and provide a basis for future advances of intelligent virtual agents, which can stimulate these skills in the users in conventional or immersive technologies. However, before considering the idea of giving properties of artificial intelligence to a virtual agent that stimulates these 21st Century Learning Skills, and especially collaboration skills, in the virtual world, it is essential to address the design, in order to ensure that it is effective and motivational for users when they engage in social interactions with it.

RQ2: *Does a virtual animal (e.g. panda) also adhere to the uncanny valley effect?*

The answer to this question is, put simply, yes. In **Chapter 3**, we explored whether the visual design of virtual animals can influence the uncanny valley effect in some properties such as familiarity, commonality, naturalness, attractiveness, interestingness, and animateness. We observed that virtual animals can evoke an uncanny valley effect on familiarity, commonality, naturalness, and attractiveness. However, there is no effect on interestingness and animateness. Likewise, other factors can influence our results regarding the uncanny valley effect on virtual animals. For instance, ranking on animal-likeness, morbidity, and movement of the virtual animal can be crucial factors in evoking likeability or mismatch in the user's perception. We conclude that it is possible to find an uncanny valley effect on familiarity, commonality, naturalness, and attractiveness when virtual animals were ranked by an expert-based ranking. Likewise, we observed uncanny valley effects for familiarity, commonality, and attractiveness for still images for virtual animals ranked by participant-based ranking, and we discovered uncanny valley effects for familiarity, commonality, and naturalness for moving images. Moreover, the morbidity appearance of the virtual animal showed lower levels of familiarity, commonality, naturalness, and attractiveness, except for interestingness and animateness. The movement of a virtual animal does not amplify the user's perceptions of the properties analyzed in this study, except with minor variations. The study contributes to our understanding that the uncanny valley effect, already well-studied in humans and robots, can also be observed in virtual animals.

RQ3: *Can a virtual animal's visual appearance influence the level of empathy of users?*

The answer to this question is both, yes and no. In **Chapter 4**, we designed a computer game simulation to examine the effects of the visual appearance of a virtual animal character in a third person view upon empathy. The results of this investigation show a congruence interaction between the artificial and natural appearance and the expressiveness of a virtual character influence self-reported situational empathy of participants in this study. Likewise, we observed a positive correlation between dispositional and self-reported empathy toward the character. These findings suggest that, in general, the visual appearance of virtual animals has a significant effect on the level of empathy of users in computer game simulations.

On the contrary, in **Chapter 5**, we expand the study of **Chapter 4** to explore this previous effect on empathy in virtual reality simulation. One of the more significant findings to emerge from this study is that the naturality/artificiality of the visual appearance of the virtual animal had no significant effect on the situational empathy of the participants. It is difficult to explain this result, but it might be related to the perspective view, which can be a primary factor that can influence empathy. For instance, the computer game employed a third perspective view, and in the virtual reality setting a first perspective view was used. Therefore, further study with a greater focus on how empathy can be affected by perspective in virtual reality simulation on virtual animals is recommended.

RQ4: *Can a virtual animal's visual appearance influence the level of immersion of users?*

The answer to this question is yes. As is illustrated in **Chapter 4**, we discovered that virtual animals' congruence between visual appearance and expressiveness (artificial + non-expressive and natural + expressive) can significantly influence players' immersion compared to incongruent appearance and facial expressions of the virtual character (artificial + expressive and natural + non-expressive). Likewise, in the virtual reality simulation in **Chapter 5**, we saw that a virtual animal appearance with a natural body provoked the highest level of immersion in a first-person view. These findings have significant implications for understanding how engagement, engrossment, and total immersion can be affected by the visual appearance of the virtual animal.

RQ5: *Can a virtual animal's visual appearance influence the level of perceived pain of users?*

The answer to this question is yes. In **Chapter 5**, we designed a virtual reality simulation where the participants received physical feedback through haptic virtual reality suits that simulate a hunter shooting their virtual animal body. One of the more significant findings from this study is that people who had a natural virtual body appearance perceived more pain than those with artificial virtual body appearances. We concluded that a virtual animal's visual appearance can affect the level of perceived pain of users during the distressing event; where a virtual body looks more natural, the intensity of pain is greater. The research also showed that perceived pain was experienced in the two experimental conditions and the control. The findings reported here shed new light on virtual reality's impact on individuals' perceptions of uncomfortable feelings like psychological pain generated in a virtual world and stimulated by haptics systems.

RQ6: *Can a virtual animal's visual appearance influence the level of embodiment of users?*

The answer to this question is no. In the study in **Chapter 5**, we expected that the visual appearance of virtual animals would affect the embodiment of the participants due to this study's use of a first-person view character. However, the results showed a not significant effect of visual appearance upon the embodiment.

RQ7: *How can different versions of virtual animals be used as virtual tutors in video instruction, where they may lead to different effects on affective and cognitive outcomes, depending on their visual appearance?*

In **Chapter 6**, we used three different virtual animals' appearances as virtual instructors to answer this question. This study was designed to explore the effect of the visual appearance of the virtual instructor in knowledge recall and user perception among young learners. The results showed that human and virtual animal appearance had higher positive reactions in familiarity, interestingness, and naturalness than virtual robot animal appearance. This study validates the idea that those virtual animals with natural characteristics had a major positive response compared with artificial traits. One of the more significant findings in this study was that the visual appearance of the virtual instructor had a significant effect on the knowledge recall of students who participated in this study. An example of a smart virtual tutor can be observed in **Appendix D**.

7.2 Usefulness of this research

7.2.1 Robotics

Uncanny Valley Theory was a concept first identified by (Mori, 1970), which came to have a major impact on the discipline of robotics. Mori explained that a robot with human-like traits could trigger an effect of repulsion and unfamiliarity in an observer. However, different designs and characteristics would lead to different levels of reactions among people. A significant amount of research has focused on introducing anthropomorphic traits to robots, using different materials and advancing technological research to create robots with capabilities similar to humans. However, our planet has other types of living things that are also arguably worthy of study in this regard, with such research and investigation also being conducted in the field of robotics.

Robots resembling animals in the form are termed 'zoomorphic robots'. While studies of zoomorphic robots in human-robot interaction, computergraphics animation and social robots are limited to date, based on the studies of (Schwind et al., 2018a). It can be seen that there is a necessity to study the uncanny valley effect of virtual animals in greater depth. We explored if the studies of Schwind et al., (2018a) can be applied to virtual robot animals. In this research, we found that virtual robot animals with characteristics that we visually associate with artificiality have a greater possibility of 'falling into the uncanny valley' or evoking feelings of repulsion within a user when compared to a more natural representation. However, we also found that people find virtual animal robots interesting. These findings contribute in several ways to our understanding of virtual robot animals and provide a basis for future studies to reduce the uncanny valley effect.

One of the most significant current discussions in robotics is that of whether humans can have empathic reactions to these artificial machines. This dissertation demonstrated the possibility of generating empathy toward a robot virtual animal in a computer game. For instance, in **Chapter 6**, the virtual robot animal with an artificial body and lacks facial expressions can foster an empathic reaction on users. These results showed that the artificial characteristics of the body of the virtual robot animal do not combine with human emotional expressions. Likewise, in this dissertation, we prove that it is possible to generate empathy towards a virtual robot, but this is affected by its visual appearance. However, we also proved that the generation of such empathy associated with visual appearance is not possible in immersive environments, such as virtual reality.

One of the more significant findings to emerge from this study is that the participants felt a great intensity of pain with the virtual animal robot in our experiment within a virtual reality situation when their character was subjected to a distress situation. These results and this dissertation open the doors to virtual robot animals in immersive technologies such as virtual reality that could pose other challenges in HCI studies. Further, such research could take into account the great amount of variety in animals that could be transposed into robotic counterparts, allowing the exploration of other important facets that to date has not been considered in regard to virtual robot animals

7.2.2 Education of the natural conservation

Education about the conservation of nature is of great importance, given the urgent need to protect animal species from extinction caused by human behavior. It is essential to develop simulation games that promote empathy towards animals and stimulate the ability to respond in an ethically appropriate way in different types of distress situations (Pedwell, 2014). In this dissertation, our research aim was to design a non-human animal-like character as a virtual animal that was likeable for users, and to discover possibilities for fostering empathic reactions toward virtual animals. We hope that this research can contribute to our understanding of empathetic responses from the virtual world and their transferal to their biological manifestations in the real world. Moreover, we aimed to design an immersive virtual environment with engagement properties to improve the potential for users that wish so to explore natural avatar agents in the virtual world, become inspired, and be motivated to discover more about animals.

In the experimental context, this dissertation has identified that virtual animals with natural traits have avoided the uncanny valley effect. These results showed that users, although the characters are artificial or digital entities if they have a natural appearance (e.g., same colour, body appearance, movement, and others) and are similar to their antagonist in real life, tend to see it more familiar (e.g., such as the rabbits and turtles in picture books, (de Droog et al., 2014, 2017)). We also found that virtual animals in a computer game scenario can generate empathetic reactions in users when their appearance is more natural and also when they express human emotional traits. It can be explained in the research by Miralles et al. (2019), who showed that empathy reactions change a species to another. A possible response to empathetic and compassion reactions is determined if the animals are similar to human traits. The external appearance of the animal or virtual animal can display an essential aspect in the emotional effect to users. This research can help people feel or perceive the animal's life by the first perspective view of virtual reality simulation. For this virtual reality simulation instance, we opted to design a virtual animal that would not be perceived as an aggressor in the simulated environment but was rather intended to be perceived in the role of a victim that generates empathy in the player. We explored distress situations in **Chapters 4** and **5** to understand the possible emotional connection with such animals. We did not find such empathetic reactions of users when using virtual animal avatars in virtual reality settings. However, we considered that results on pain perception in this distress situation might help us to allow users to be able to 'step into the shoes' of nature and it can be associated with an emotional connection with the virtual animal that is not necessary empathy.

A natural progression of this work would be to analyze other aspects or combinations of the findings of the experiments developed in **Chapters 4** and **5** in a future phase of experimental research allows users the generation of immersive learning and empathetic reactions toward nature and animals. For instance, the results of this research can be used to teach about Climate Change and Global Warming from the perspective of the animals through the interactive virtual reality movie "Justin Beaver Survivor". Children and the general public can explore this virtual reality film version, which "can be a great ally in teaching children and young people and inspiring environmental awareness and love of science" (University, 2021).

7.2.3 Artificial intelligence in education

Following the principal theoretical implication of the study relating to the appearance of virtual animals and their impact on both, visual and emotional user perception, we considered the possibility of developing a virtual animal with artificial intelligence properties. One of the most relevant achievements during this dissertation was obtaining an award for revolutionary research at Laval Virtual in France. We proposed that the virtual animal studied in the chapters could become a conversational agent through immersive technology in the future. In the next step of this research, we are interested in using artificial intelligence methods and applying them to virtual animals so that users can actively interact with them. Our virtual animal could be capable of vocal interaction, creating interactive stories, becoming involved in intelligent conversations about science topics, nature, or describing objects or animals of a particular environment in the role of a guide, virtual instructor, or virtual animal. The proposal for this is outlined in **Appendix D**. We intend to continue to develop experimental studies about this intelligent virtual animal. We can use this intelligent virtual animal in education, and we will foster 21st Century Learning Skills through their interactions with children and young learners.

7.3 Improvements

The generalizability of these findings of uncanny valley effect is subject to certain limitations. Statistics of the National geographic (Geographic, 2019) estimated that there are more than a total of 8.7 million species in plant and animals on Earth, meaning a wide variety of types of animals in the world. In this research, we used both charismatic and non-charismatic animals, a panda and beaver, respectively. However, we are currently unsure how well these results would carry over to other virtual animals (e.g., fish, birds, worms, insects, spiders). Therefore, future research should also focus on applying this methodology to other types of animals, in order to provide an overview of such results.

The results relating to empathy toward virtual animals in this dissertation is also subject to certain limitations. For instance, other aspects that could potentially interfere with the experience of empathy towards virtual animals might include the human perception of the animal's abilities (e.g., fish), how threatening they are assumed to be (e.g., wolves) or the level of charisma that the animal has in real life (e.g., beaver, panda, jaguar) (Kupsala et al., 2013; Albert et al., 2018; Bautista et al., 2019).

In this dissertation, an exciting aspect was the results obtained about empathy and perception of pain in **Chapters 4** and **5**. Thanks to advances in virtual reality simulation, we were able to perform the experiment in the first-person view and explore our research questions from this perspective. However, we consider that there are other key aspects that could potentially affect the results: the person view and haptics vest. In future studies, we propose that the experiment of the **Chapter 5** can be expanded to include a third-person view and with and without haptics, in order to explore the empathy and pain perception concepts.

7.4 Conclusion

Problem Statement: *Which is the more effective visual design in the virtual characters that can invite users to foster communication and collaborative skills within virtual environments?*

Throughout several studies in this dissertation, we explored the naturalness (or biological traits) vs. artificiality in the visual appearance of a virtual character in various technologies. We discovered a major significant effect of the natural visual appearance of the virtual animals on cognition (e.g., knowledge recall) and affectation (e.g., empathy, immersion, pain perception and the uncanny valley phenomenon) outcomes on the users compared to such avatars' artificial visual appearance. However, the visual appearance does not affect the user's perception in specific situations, such as the uncanny valley effect and situational empathy. This dissertation proves that it is essential to intelligently design virtual environments, because each element of a design can have a cognitive or emotional effect upon users. For example, we focused on the main character of a virtual environment, and we explored animals as non-human virtual characters and their impact on users. We discovered and identified certain elements of virtual animals' design that impact upon human-computer interactions, and we presume that other elements exist that have not yet been studied. In this dissertation, we explored two visual appearances, i.e., those of a panda and a beaver, but more than a million species of animals that live upon this planet give rise to a wide field for future research. Further research could also be conducted to determine the effectiveness of artificial intelligence in virtual animals to develop conversational agents for the fostering of 21st Century Learning Skills.

APPENDIX

List of 21st Century Leaning Skill
Tests and Quotations

A List of 21st Century Leaning Skill Tests and Quotations

A.1 List of tests

Table A.1.1: Collaboration Tests

Name of the test	Reference
1. Belbin Team Inventory	Vaida (2019)
2. Big five personality traits	Bar et al. (2018); Soni and Bakhru (2019)
3. Collaboration and Satisfaction About Care Decisions	Dougherty and Larson (2005)
4. Collaboration Survey	Broadleaf (2020)
5. Collaborative Practice Assessment Tool	Broadleaf (2020)
6. Collective Impact	Broadleaf (2020)
7. Collaborative Practice Scale	Dougherty and Larson (2005)
8. Connor Davidson Resilience scale	Bar et al. (2018); Smith and Gregg (2020)
9. Decision about Transfer Scale	Broadleaf (2020)
10. Eudaimonic wellbeing	Soni and Bakhru (2019)
11. Gallup Clifton Strengths test	Smith and Gregg (2020)
12. IDEA Partnership Success Rating Scale	Broadleaf (2020)
13. Index of Interdisciplinary Collaboration (IIC)	Thannhauser et al. (2010)
14. Interangency Collaboration Activities Scale (IACAS)	Broadleaf (2020)
15. Interdisciplinary Education Perception Scale	Thannhauser et al. (2010); Bar et al. (2018)
16. Interprofessional Perceptions Scale (IPS)	Thannhauser et al. (2010)
17. Jefferson Scale of Attitudes Toward Physician Nurse Collaboration	Dougherty and Larson (2005)
18. Keirsey Temperament Sorter	Holton (2001); Sfetsos et al. (2006)
19. Ladder of Participation Scale	Broadleaf (2020)
20. Levels of Collaboration Scale	Broadleaf (2020)
21. Modified Belbin Group Role Questionnaire	Holton (2001)
22. Modified Index of Interdisciplinary Collaboration (MIIC)	Thannhauser et al. (2010)
23. Multidisciplinary Collaboration instrument (MDC)	Thannhauser et al. (2010)
24. MyersBrigg Type Indicator Test	Sfetsos et al. (2006); Vaida (2019)
25. National standardized science test (SIMCE)	S´anchez and Olivares (2011)
26. Nurse-Physician Questionnaire	Dougherty and Larson (2005)
27. Nurses Opinion Questionnaire	Dougherty and Larson (2005)
28. Nursing Relationship Scale (NRS)	Broadleaf (2020)
29. Organization Trust Survey	Broadleaf (2020)
30. PATNERS	Broadleaf (2020)
31. Readiness for Interprofessional Learning Scale	Thannhauser et al. (2010)
32. Riso-Hudson Enneagram Type Indicator test	Vaida (2019)
33. Role Perceptions Questionnaire (RPQ)	Thannhauser et al. (2010)
34. Scale for the perception of collaborative skills	S´anchez and Olivares (2011)
35. Scale for the perception of problem solving skills	S´anchez and Olivares (2011)
36. Strategic Alliance Formative Assessment Rubric (SAFAR)	Broadleaf (2020)
37. Team Fitness Test	Holton (2001)
38. University of Western England Interprofessional Questionnaire (UWEIQ)	Thannhauser et al. (2010)
39. Wilder Collaboration Factor Inventory	Broadleaf (2020)
40. Work-life balance	Soni and Bakhru (2019)

Table A.1.2: Communication Tests

Name of the test		Reference
1.	Active Empathetic Listening Scale (AELS)	Brown et al. (2020)
2.	Affective Communication Test	Matsumoto et al. (2000)
3.	American-Speech-Language-Hearing-Association–Functional-Assessment of Communication Skills for Adults (ASHA-FACS)	Spell and Frank (2000)
4.	Arizona Articulation Proficiency Scale	Morrison and Shriberg (1992)
5.	Assessment Link Between Phonology and Articulation	Morrison and Shriberg (1992)
6.	Assessment of Phonological Processes	Morrison and Shriberg (1992)
7.	Behavioral Indicants of Immediacy Scale	Ellis et al. (2016); Klebig et al. (2016)
8.	Big Five Inventory	Klebig et al. (2016)
9.	Big-Five personality Scale	Hullman et al. (2010)
10.	Brief Test of Head Injury (BTHI)	Spell and Frank (2000)
11.	Brigance K&1 Screen	Liu (1998)
12.	California Achievement Test	Charmello (1993)
13.	Carolina Older Adults Test of Nonverbal Communication (COAT-NC)	Spell and Frank (2000)
14.	Collaborative Strategy Reading (CSR)	Liu (1998)
15.	Communication Competency Assessment Instrument	Castleberry and Shepherd (1993)
16.	Communication and Reception of Affect Test (CARAT)	Matsumoto et al. (2000)
17.	Comprehensive Test of Basic Skills (CTSB)	Liu (1998)
18.	Compton-Hutton Phonological Assessment	Morrison and Shriberg (1992)
19.	Contextual and Affective Sensitivity test (CAST)	Matsumoto et al. (2000)
20.	Diagnostic Analysis of Nonverbal Accuracy Scale (DANVA)	Matsumoto et al. (2000); Rosip and Hall (2004)
21.	Diagnostic Assessment of Nonverbal Affect-2 (DANVA-2)	Spell and Frank (2000)
22.	English Language Testing Service (ELTS)	Afzali et al. (2011)
23.	Feldstein Affect Judgment Test	Matsumoto et al. (2000)
24.	Gates-MacGinitie Reading Comprehension Test	Charmello (1993)
25.	Gates-MacGinitie Reading Tests	Charmello (1993); Liu (1998)
26.	Generalized immediacy scale	Ellis et al. (2016)
27.	Glasgow Coma Score (GCS)	Spell and Frank (2000)
28.	Goldman-Fristoe Test of Articulation	Morrison and Shriberg (1992)
29.	Goldman-Fristoe-Woodcock Test of Auditory Discrimination	Ting et al. (2006)
30.	Incivility Twenty-three behaviors	Klebig et al. (2016)
31.	Interpersonal Communication Competence Scale (ICCS)	Hullman et al. (2010); Brown et al. (2020)
32.	Kentucky Comprehensive Listening Test	Castleberry and Shepherd (1993)
33.	Languages Assessment Scales (LAS)	Liu (1998)
34.	Listening Styles Profile	Brown et al. (2020)
35.	MIDAS questionnaire	Afzali et al. (2011)
36.	Phonological Process Analysis	Morrison and Shriberg (1992)

37.	Photo Articulation Test	Morrison and Shriberg (1992)
38.	Profile of Nonverbal Sensitivity (PONS)	Matsumoto et al. (2000)
39.	Receptive Facial Expressions (Faces) Subtest	Spell and Frank (2000)
40.	Receptive Paralanguage (Voices) Subtest	Spell and Frank (2000)
41.	Resilience at University scale (RAU)	Brown et al. (2020)
42.	Resilience Scale in Adults (RSA)	Brown et al. (2020)
43.	Self-Efficacy Questionnaire for Social Skills	Hullman et al. (2010)
44.	Self-disclosure	Klebig et al. (2016)
45.	Social Interpretations Test	Matsumoto et al. (2000)
46.	Social Skills Inventory (SSI)	Matsumoto et al. (2000)
47.	Templin Test of Auditory Discrimination	Ting et al. (2006)
48.	Templin-Darley Test of Articulation	Morrison and Shriberg (1992)
49.	Test of Emotion Styles (TES)	Matsumoto et al. (2000)
50.	Test of English as a Foreign Language (TOEFL)	Afzali et al. (2011)
51.	Test of Nonverbal Cue Knowledge (TONCK)	Rosip and Hall (2004)
52.	Trained rater's perception of immediacy scale	Ellis et al. (2016)
53.	Understanding our Feelings test	Matsumoto et al. (2000)
54.	Verbal Immediacy Scale (VIS)	Ellis et al. (2016)
55.	Washington Speech Sound Discrimination Tes	Ting et al. (2006)
56.	Watson-Barker Listening Test	Castleberry and Shepherd (1993)
57.	Weiss Comprehensive Articulation Test	Morrison and Shriberg (1992)
58.	Wepman Auditory Discrimination Test	Ting et al. (2006)

Table A.1.3: Critical Thinking Tests.

Name of the test	Reference
1. Academic Aptitude Test (AAT)	Van der Merwe (2002)
2. American Critical-Thinking Assessment Test (CAT)	Sustekova et al. (2019)
3. APIL instrument	Maloa and Bux (2015)
4. AT-20 ambiguity tolerance scale	Taube (1997)
5. California Critical Thinking Disposition Inventory (CCTDI)	Stone et al. (2001); Rear (2019)
6. California Critical Thinking Skills Test (CCTST)	Ennis (1993); Stone et al. (2001); Bataineh and Zghoul (2006); Ku (2009); Behar-Horenstein and Niu (2011); Huhn et al. (2011) Tiruneh et al. (2017); Rear (2019); Sustekova et al. (2019)
7. Career Path Appreciation (CPA)	Van der Merwe (2002)
8. Checklist of Educational Views (CLEV)	Taube (1997)
9. Computerized Adaptive Test	Maloa and Bux (2015)
10. Cornell Conditional Reasoning Test	Ennis (1993)
11. Cornell Critical Thinking Test (CCTT)	Ennis (1993); Bataineh and Zghoul (2006); Ku (2009); Behar-Horenstein and Niu (2011); Tiruneh et al. (2017); Rear (2019); Sustekova et al. (2019)
12. Critical Reasoning Test Battery	Maloa and Bux (2015)
13. Deductive Logic and Assumption Recognition	Ennis (1993)
14. Dover/Vienna Test System (DOVER)	Van der Merwe (2002)
15. Dynamic assessment APIL B (SV)	Maloa and Bux (2015)
16. Ennis-Weir Critical Thinking Essay Test	Ennis (1993); Taube (1997); Ku (2009); Tiruneh et al. (2017); Rear (2019); Sustekova et al. (2019)
17. General Aptitude Test Battery (GATB)	Kuncel et al. (2005)
18. Graduate Management Admission Test (GMAT)	Kuncel et al. (2005)
19. Graduate Record Examination (GRE)	Kuncel et al. (2005)
20. Group Interaction Exercise (GI Exerc)	Van der Merwe (2002)
21. HEIghten Critical Thinking Assessment	Rear (2019)
22. Halpern Critical Thinking Assessment (HCTA)	Ku (2009); Bensley et al. (2016); Tiruneh et al. (2017); Rear (2019); Sustekova et al. (2019)
23. Health Science Reasoning Test (HSRT)	Huhn et al. (2011)
24. High Level Battery (HL Bat)	Van der Merwe (2002)
25. High Level Figure Classification Test (HL FCT)	Van der Merwe (2002)
26. Intermediate Battery (INT Bat)	Van der Merwe (2002)
27. International Critical Thinking Reading and Writing Test (ICTRWT)	Lu and Xie (2019)
28. Law School Admission Test (LSAT)	Kuncel et al. (2005)
29. Learner Autonomy Questionnaire	Lu and Xie (2019)
30. Learning Potential Computerized Adaptive Test	Maloa and Bux (2015)
31. Logical Reasoning	Ennis (1993)

32.	Lopez Critical Thinking Test (CEU)	Sustekova et al. (2019)
33.	Managerial and Graduate Item Bank	Maloa and Bux (2015)
34.	Medical College Admission Test (MCAT)	Kuncel et al. (2005)
35.	Metacognitive Assessment Inventory (MAI)	Bensley et al. (2016)
36.	Miller Analogies Test (MAT)	Kuncel et al. (2005)
37.	Minnesota Multiphasic Personality Inventory (MMPI)	Van der Merwe (2002)
38.	Myers-Briggs Type Indicator (MBTI)	Van der Merwe (2002)
39.	Need For Cognition Scale (NCS)	Taube (1997)
40.	New Jersey Test of Reasoning Skills	Ennis (1993)
41.	Nineteen Field Interest Inventory (19 FII)	Van der Merwe (2002)
42.	Normal Battery (NORM Bat)	Van der Merwe (2002)
43.	Numerical Critical Reasoning (NMG3)	Maloa and Bux (2015)
44.	Numerical aptitude Person job score (PJS)	Maloa and Bux (2015)
45.	Occupational Personality Questionnaire	Maloa and Bux (2015)
46.	Perceptual Battery (Blox)	Van der Merwe (2002)
47.	Programmer Aptitude Battery (PAB)	Van der Merwe (2002)
48.	Psychological Critical Thinking Exam (PCTE)	Bensley et al. (2016)
49.	Raven's Progressive Matrices	Van der Merwe (2002)
50.	Ross Test of Higher Cognitive Processes	Ennis (1993)
51.	SHL Critical Reasoning Test Battery (CRTB)	Van der Merwe (2002)
52.	SHL Customer Contact Aptitude Series (CCAS)	Van der Merwe (2002)
53.	SHL Customer Contact Styles Questionnaire (CCSQ)	Van der Merwe (2002)
54.	SHL Management and Graduate Item Bank (MGIB)	Van der Merwe (2002)
55.	SHL Occupational Personality Questionnaire (OPQ)	Van der Merwe (2002)
56.	SHL Personnel Test Battery (PTB)	Van der Merwe (2002)
57.	SHL Technical Test Battery (TTB)	Van der Merwe (2002)
58.	SHL test battery	Maloa and Bux (2015)
59.	Self-Directed Search Questionnaire (SDS)	Van der Merwe (2002)
60.	Senior Aptitude Tests (SAT)	Van der Merwe (2002)
61.	Sixteen Personality Factor Questionnaire (16 PF)	Van der Merwe (2002)
62.	South African Wechsler Adult Intelligence Scale (SAWAIS)	Van der Merwe (2002)
63.	Structured-Objective Rorschach Test (SORT)	Van der Merwe (2002)
64.	Test of English as a Foreign Language (TOEFL)	Lu and Xie (2019)
65.	Test of Enquiry Skills	Ennis (1993)
66.	Test of Inference Ability in Reading Comprehension	Ennis (1993)
67.	Test on Appraising Observations	Ennis (1993)
68.	Thematic Apperception Test (TAT)	Van der Merwe (2002)
69.	Thomas Personal Profile Analysis (PPA)	Van der Merwe (2002)
70.	Verbal Critical Reasoning (VMG3)	Maloa and Bux (2015)
71.	Wagner Assessment Test (WAT)	Wagner and Harvey (2006)
72.	Wonderlic Test	Kuncel et al. (2005)
73.	Watson Glaser Critical Thinking Appraisal (WGCTA)	Ennis (1993); Taube (1997); Bataineh and Zghoul (2006); Wagner and Harvey (2006); Ku (2009); Behar-Horenstein and Niu (2011); Huhn et al. (2011); Bensley et al. (2016); Tiruneh et al. (2017); Rear (2019); Sustekova et al. (2019)

Table A.1.4: Creativity Tests.

Name of the test	Reference
1. Abbreviated Torrance Test for Adults (ATTA)	Fee and Gray (2012)
2. Abedi-Schumacher Creativity Test	Cropley (2000)
3. Adaptation-Innovation Inventory (KAI)	Cropley (2000)
4. Adjective Check List (ACL)	Cropley (2000); Soto and John (2009); Ziv and Keydar (2009)
5. Alpha Biographical Inventory (ABI)	Venable (1994); Cropley (2000)
6. Alternative Uses task (AUT)	Cheung et al. (2003); Ritter and Mostert (2017)
7. Assessment of creative products	Baer and Kaufman (2008)
8. Assessment of creativity of graduate students by their professors	Baer and Kaufman (2008)
9. Assessment of various creative products	Baer and Kaufman (2008)
10. Associational fluency	Baer and Kaufman (2008)
11. Attitude and interest inventories	Karpova et al. (2011)
12. Barron–Welsh Art Scale (BWAS)	Furnham and Bachtiar (2008); Acar and Runco (2012)
13. Basadur Preference Scale	Cropley (2000)
14. Battery of creativity tests	Cheung et al. (2003)
15. Big Five Inventory	Soto and John (2009)
16. Big Five Personality Traits	Kaufman et al. (2016)
17. Biographical Inventory of Creative Behaviours (BICB)	Batey and Furnham (2008); Furnham and Bachtiar (2008); Furnham et al. (2011)
18. Biographical inventories	Karpova et al. (2011)
19. California Psychological Inventory (CPI)	McCrae et al. (1993); Venable (1994); Soto and John (2009)
20. Circle test	Baer and Kaufman (2008)
21. Cognitive Risk Tolerance Survey	Charyton et al. (2008)
22. Composite Creativity Index	Arden et al. (2010)
23. Comprehensive Ability Battery	Acar and Runco (2012)
24. Consequences Test	Furnham et al. (2011)
25. Covington Attitudinal Inventory for Problem Solving	Venable (1994)
26. Create a story task using words presented on a screen	Arden et al. (2010)
27. Creative Achievement Questionnaire (CAQ)	Kaufman et al. (2016)
28. Creative Activities Checklist	Cropley (2000)
29. Creative Attitude Survey (CAS)	Venable (1994); Pastor and David (2017)
30. Creative Behavior Disposition Scale	Venable (1994)
31. Creative Behavior Inventory	Cropley (2000)
32. Creative Functioning Test	Arden et al. (2010)
33. Creative Personality Scale (CPS)	Cropley (2000); Wolfradt and Pretz (2001); Chen et al. (2006); Batey and Furnham (2008); Ziv and Keydar (2009)
34. Creative Problem-Solving	Ritter and Mostert (2017)
35. Creative Product Inventory	Cropley (2000)

36. Creative Product Semantic Scale	Cropley (2000)
37. Creative Reasoning Test (CRT)	Cropley (2000)
38. Creative Scientific Ability Test (C-SAT)	Huang and Wang (2019)
39. Creative Temperament Scale	Charyton et al. (2008)
40. Creative personality scales	Dollinger et al. (2004)
41. Creative personality traits list	Cheung et al. (2003)
42. Creative-thinking abilities	Fee and Gray (2012)
43. Creativity Assessment Packet	Cropley (2000)
44. Creativity Behavior Inventory	Dollinger et al. (2004)
45. Creativity Checklist (CCL)	Cropley (2000)
46. Creativity Personality Scale (CPS)	Charyton et al. (2008)
47. Creativity Styles Questionnaire (CSQ)	Cropley (2000)
48. Creativity Test	Venable (1994)
49. Creativity assessment by teachers and peers	Baer and Kaufman (2008)
50. Creatrix Inventory (C & RT)	Cropley (2000)
51. Different components of openness to experience factor	Baer and Kaufman (2008)
52. Divergent Movement Ability Test	Baer and Kaufman (2008)
53. Divergent Thinking (DT)	Furnham and Bachtiar (2008)
54. Divergent Thinking Tasks	Arden et al. (2010)
55. Divergent thinking abilities and personality traits	Baer and Kaufman (2008)
56. Divergent thinking and creativity ratings	Baer and Kaufman (2008)
57. Divergent thinking test	Baer and Kaufman (2008)
58. Divergent thinking test to predict scientific creativity	Baer and Kaufman (2008)
59. Domain-General Creativity Test (DGCT)	Huang and Wang (2019)
60. Domino Creativity Scale	Cropley (2000)
61. Eight different divergent thinking tests	Baer and Kaufman (2008)
62. Evaluative thinking	Baer and Kaufman (2008); Runco et al. (2011)
63. Eysenck Personality Questionnaire	Acar and Runco (2012)
64. Five-factor model of personality	McCrae et al. (1993); Dollinger et al. (2004)
65. Fluency	Baer and Kaufman (2008)
66. Group Inventory for Finding Creative Talent (GIFT)	Cropley (2000)
67. Group Inventory for Finding Interests (GIFFII)	Cropley (2000)
68. How Do You Really Feel About Yourself?	Cropley (2000)
69. Imagine a new design for a Pen while inside the scanner	Arden et al. (2010)
70. Independence of Judgment Scale	Venable (1994)
71. Iowa Inventiveness Inventory	Cropley (2000)
72. Khatena Torrance Creative Perception Inventory	Baer and Kaufman (2008)
73. Life Experience Inventory (LEI)	Cropley (2000)
74. Manifest needs questionnaire	Baer and Kaufman (2008)
75. Match problems	Arden et al. (2010)
76. Measure of creativity thinking in music	Baer and Kaufman (2008)
77. Medhi's verbal and figural divergent thinking tests	Baer and Kaufman (2008)
78. Mednick's RAT and study-specific divergent thinking task	Arden et al. (2010)

A

Name of the test	Reference
79. Methods of fostering classroom creativity	Baer and Kaufman (2008)
80. Mini-Markers	Soto and John (2009)
81. Minnesota Multiphasic Personality Inventory Schizophrenia scale	Acar and Runco (2012)
82. Multidimensional Personality Questionnaire (MPQ)	McCrae et al. (1993)
83. Multidimensional inventories	Karpova et al. (2011)
84. Multidimensional stimulus fluency measure	Baer and Kaufman (2008)
85. NEO Five Factor Inventory	Wolfradt and Pretz (2001); Furnham et al. (2011)
86. NEO Personality Inventory (NEO-PI)	McCrae et al. (1993); Soto and John (2009)
87. Newly Creativity Test (NCT)	Huang et al. (2017)
88. Novel Creativity Test	Arden et al. (2010)
89. Omnibus Personality Inventory	Venable (1994)
90. Openness to experience	Baer and Kaufman (2008)
91. Opinion, Attitude, and Interest Survey	Venable (1994)
92. Performance-based assessments	Han (2003)
93. Personality inventories	Karpova et al. (2011)
94. Problem-solving Scale	Wolfradt and Pretz (2001)
95. Processing novel metaphors	Arden et al. (2010)
96. Publications in creativity journals	Baer and Kaufman (2008)
97. Real World Divergent Thinking Test	Han (2003)
98. Realistic creative problem solving	Runco et al. (2011)
99. Relationship of creativity to academic achievement)	Baer and Kaufman (2008)
100. Remote Associates Test (RAT)	Venable (1994); Marsh et al. (1996); Cropley (2000); Baer and Kaufman (2008); Ritter and Mostert (2017)
101. Rorschach Ambiguous Figures	Arden et al. (2010)
102. Runner Studies of Attitudinal Patterns	Venable (1994)
103. Schaefer's Biographical Inventory	Venable (1994)
104. Schaefer's Biographical Inventory Creativity	Ziv and Keydar (2009)
105. Science Performance Test (SPT)	Huang and Wang (2019)
106. Scientific Creativity Test (SCT)	Hu and Adey (2002); Baer and Kaufman (2008); Hu et al. (2013); Huang et al. (2017); Huang and Wang (2019)
107. Scientific Structure Creativity Model (SSCM)	Hu and Adey (2002)
108. Self-Rating of Creativity (SR)	Furnham and Bachtiar (2008)
109. Self-assessments of creativity	Baer and Kaufman (2008)
110. Self-rated creativity	Batey and Furnham (2008); Furnham et al. (2011)
111. Self-report measures of creativity and self-esteem	Baer and Kaufman (2008)
112. Self-report of creative activities involving scores of quality and quantity	Baer and Kaufman (2008)
113. Self-reported creative activities and achievements	Karpova et al. (2011)
114. Self-reported creative products	Cheung et al. (2003)
115. Self-reports in different domains of creativity	Baer and Kaufman (2008)
116. Sixteen Personality Factor Questionnaire	Venable (1994)
117. Square test	Baer and Kaufman (2008)

118. Sternberg's Triarchic Abilities Test	Cropley (2000)
119. Story generation task	Arden et al. (2010)
120. Tel-Aviv Creativity Test (TACT)	Ziv and Keydar (2009)
121. Temperament and Character Inventory	Arden et al. (2010)
122. Test of Creative Thinking	Cropley (2000)
123. Test of Creative Thinking–Drawing Production (TCT-DP)	Dollinger et al. (2004)
124. Test of divergent thinking	Runco et al. (2011)
125. The Creativity Tests for Children	Cropley (2000)
126. The figural DT	Runco et al. (2011)
127. Thematic Apperception Test (TAT) story	Dollinger et al. (2004)
128. Thinking creatively with sounds and words test	Baer and Kaufman (2008)
129. Torrance Test of Creative Thinking (TTCT)	Venable (1994); Cropley (2000); Hu and Adey (2002); Chen et al. (2006); Baer and Kaufman (2008); Arden et al. (2010); Karpova et al. (2011); Lee and Kim (2011); Fee and Gray (2012); Hu et al. (2013); Kaufman et al. (2016); Huang et al. (2017)
130. Torrance repeated figures circle test	Baer and Kaufman (2008)
131. Torrence battery	Marsh et al. (1996)
132. Turing Test and Computational Creativity	Pease and Colton (2011)
133. Unusual Uses Tests of objects	Huang et al. (2017)
134. Verbal DT	Runco et al. (2011)
135. Verbal and nonverbal divergent thinking	Baer and Kaufman (2008)
136. Verbal creativity using caption writing task	Baer and Kaufman (2008)
137. Verbal creativity using poetry writing task	Baer and Kaufman (2008)
138. Verbal creativity using storytelling task	Baer and Kaufman (2008)
139. Verbal divergent production	Cheung et al. (2003)
140. Villa and Auzmendi Creativity Test	Cropley (2000)
141. Wallace-Kogan Creative Thinking Test	Chen et al. (2006)
142. Wallach and Kogan creativity battery in Hebrew	Ziv and Keydar (2009)
143. Wallach-Kogan Creativity Test (WKCT)	Han (2003); Baer and Kaufman (2008); Pastor and David (2017)
144. Wallach-Kogan ideational fluency test	Baer and Kaufman (2008)
145. Wonderlic intelligence test	Runco et al. (2011)
146. Word Association Rare Response Test (WARRT)	Acar and Runco (2012)
147. Word Association Test	Chen et al. (2006); Lee and Kim (2011)

A.2 Quotations

Table A.2.1: This table with quotations about Validity and Reliability: +R: Positive Reliability, -R: Negative Reliability, +V: Positive Validity, and -V: Negative Validity.

Name Test	Codes	Quotation Content
Affective Communication Test	+R	"Results indicated that the lack of nonverbal encoding ability is a strong correlate of communication anxiety for both males and females. Also, males and females tend to make the some distinctions of nonverbal affect; therefore, different scales are not necessary for the two genders. The shortened scale, the ACT-10, is brief enough to be administered quickly yet reliable enough to assess accurately the self-report of nonverbal encoding ability." (Hensley, 1986, p. 1)
Big Five Personality traits	+V	"Results based on a sample of 105 sales representatives supported the 2 hypotheses tested. First, supervisor, coworker, and customer ratings of the 2 job-relevant personality dimensions—conscientiousness and extraversion were valid predictors of performance ratings, and the magnitude of the validities were at least as large as for self-ratings. Second, supervisor, coworker, and customer ratings accounted for significant variance in the criterion measure beyond self-ratings alone for the relevant dimensions. Overall, the results suggest that validities of personality measures based on self-assessments alone may underestimate the true validity of personality constructs." (Mount et al., 1994, p. 272)
Big Five Personality traits	-R	"The corrected (p) and uncorrected (rxy) correlations between the Big Five personality scales for the four rating sources and the supervisor and coworker performance ratings are also shown in Table 1. There was only one supervisor rating of performance for each sales representative. Therefore, we corrected the validities (p) for unreliability in the criterion by using the average single-rater reliability of .50 obtained by Rothstein (1990), which was based on 9,975 first-line supervisors. The true validities when using coworker performance ratings as the criterion were corrected based on the correlation between two randomly selected coworkers' performance ratings for each sales representative. On the basis of 105 pairs of performance ratings, the reliability of a single coworker's ratings was .53. To avoid problems associated with common method variance for coworker ratings, we randomly selected one coworker's ratings as the predictor and used the average of the remaining coworkers' ratings as the criterion measure. Because there was an average of 1.6 coworkers for each sales representative, the Spearman- Brown prophecy formula was used to adjust the reliability upward. Consequently, we used .55 as the reliability of the composite of the coworkers' performance ratings. The validities for the two job-relevant." (Mount et al., 1994, p. 275)
Big Five personality traits	+V	"In summary, there is a distinction between the validity of personality constructs based on self-reports and the validity of the constructs based on observer ratings. Our results show that supervisor, coworker, and especially customer ratings of conscientiousness and extroversion are valid predictors of sales performance. They also show that observer ratings account for significant variance beyond that of self-ratings for the job relevant personality constructs. The substantive message here is that the validity of personality constructs may be understated through reliance on the self-report method alone." (Mount et al., 1994, p. 279)

Biographical Inventory Creativity	+V	"The inventory was found to have high concurrent validity in differentiating the cross-validation groups of creative and control students (Schaefer & Anastasi, 1968; Anastasi & Schaefer, 1969)." (Schaefer, 1972, p. 471)
California Critical Thinking	+R	"FINDINGS: The test reliable coefficient was 0.62%. Factor analysis indicated that CCTST has been formed from 5 factors (elements) namely: analysis, evaluation, inference, inductive and deductive reasoning." (Khalili and Soleymani, 2003, p. 84)
California Critical Thinking	+R and +V	"Findings showed the range of content validity between %80 and %96.87, denoting a proper amount. Coefficients of correlation for each subscale were 0.71 for analysis,0.77 for evaluation, 0.71 for inference, 0.77 for inductive reasoning, and 0.71 for deductive reasoning. These indicated a positive and significant correlation of subscales with one another as well as with the total score of the test (r=0.86). The correlation between test-retest results was 0.90 with Kappa coefficient equal to 0.81." (Khoda et al., 2007, p. 12)
California Critical Thinking	+R and +V	"Conclusion: Findings showed a proper validity and reliability of the test." (Khoda et al., 2007, p. 12)
California Critical Thinking	+R	"Results: The test coefficient for reliability was 0.62. Factor Analysis indicated that CCTST has been formed from 5 factor (element) namely: Analysis, Evaluation, inference, Inductive and Deductive Reasoning. Internal consistency method shows that All subscales have been high and positive correlation with total test score." (Khalili and Soleymani, 2003, p. 29)
California Critical Thinking	+R	"The Results revealed that the questions test is sufficiently reliable as a research tool, and all subscales measure a single construct (Critical Thinking) and are able to distinguished the persons with different level's CT." (Khalili and Soleymani, 2003, p. 29)
California Critical Thinking Disposition Inventory	+R and +V	"the results indicate that evidence for construct validity existed for truth-seeking, open-mindedness, systematicity, and maturity for the Chinese CCTDI. After allowing some error to exist and deleting three items, evidence for construct validity existed for the remaining subscales. The results of the psychometric equivalencies across Chinese and English CCTDI showed similarity for content validity and reliability for inquisitiveness. In terms of multi-sample analysis, there were equal forms across all subscales of the two versions... Consequently, although the translation adequacy of the Chinese CCTDI needs to be improved, there is evidence that it is useful for evaluating critical thinking dispositions." (Yeh, 2002, p. 123)
California Critical Thinking Disposition Inventory	+R	"Bollen's reliability estimation was the same as the original reliability of 0.68. Consequently, the error variances between items 10 and 40 and between items 18 and 49 were allowed to correlate." (Yeh, 2002, p. 129)

A

Name Test	Codes	Quotation Content
California Critical Thinking Disposition Inventory	+R	"the estimation reliability value in terms of Bollen's method decreased from 0.73 to 0.66." (Yeh, 2002, p. 129)
California Critical Thinking Disposition Inventory	+V	"In conclusion, evidence for the construct validity of the English CCTDI existed for truth-seeking, open-mindedness systematicity, inquisitiveness, self-confidence and maturity. When items 6 and 42 were deleted from analyticity, evidence for construct validity existed." (Yeh, 2002, p. 130)
California Psychological Inventory	+R	"Results support the validity of some of the subscales as indicators of the role-taking construct underlying the So scale." (Rosen and Schalling, 1974, p. 757)
Collaboration and Satisfaction about Care Decisions	+R	"The CSACD-T had a Cronbach's alpha value above the desirable 0.80 (Polit & Beck, 2017) and demonstrated a good internal consistency. Result from the previous study of the nine items CSACD showed alpha value over 0.90 (Sapnas et al., 2006). Tavakol and Dennick (2011) argue that the maximum value to be recommended is 0.90, which means that the alpha value of item 1–9 was maybe too high." (Aaberg et al., 2019, p. 647)
Collaboration and Satisfaction about Care Decisions	-V	"Regarding sample size, the recommended sample size for EFA in validation studies is disputed and no consensus exists (Polit & Yang, 2016). Some suggest a minimum of 300, but emphasize that if there is strong correlations and few distinct factors, a smaller sample is adequate (Tabachnick & Fidell, 2013). Others offer guidance on the number of respondents per items, ranging from 5–40 or 50 per item, with the most common recommendation as a minimum of 10 cases per item (Polit & Yang, 2016). The sample size of 247 in the current study was thereby considered satisfactory with 27 number of respondents per item. Multiple types of healthcare personnel from multiple types of hospital units were represented in the sample; hence, a heterogeneous study sample was obtained, as recommended for testing questionnaires (Taber,2017). " (Aaberg et al., 2019, p. 647)
Cornell Critical Thinking	-R	"This suggests that the CCTT and the HCTA are-at least partially-measuring the same constructs and that the lack of correlation is partly due to the lack of reliability." (Verburgh et al., 2013, p. 11)
Cornell Critical Thinking	+V	"Based on our findings and on the dearth of support in general for the factorial validity of the CCTT, an exploratory EFAESEM analysis was justified in consideration of the interdependency of the CT dimensions in which the Level X is based. Based on the relatively poor fit of the CFA models and lack of existing factorial validity evidence, we might have also expected the confirmatory five-factor ESEM model C1 not to meet our fit criteria; however, we conducted the analysis to demonstrate the preferred steps a researcher might take in an ESEM analysis when an a priori factor structure exists (e.g., Morin et al., 2013)." (Leach et al., 2020, p. 12)
Creative Achievement Questionnaire	+V	"This study was designed to provide multiple sources of evidence of the validity of the Creative Achievement Questionnaire (CAQ) and to clarify the hierarchy of creative achievement using Rasch analyses." (Wang et al., 2014, p. 62)

Creative Achievement Questionnaire	+V	"Content and structural validity. The Rasch model was used to examine the one-dimensionality of each of the 10 CAQ subscales. The results revealed that all items in each domain had acceptable infit MNSQs (Table 1). Data from each domain fit the Rasch model well, indicating that the items in each subscale measured a unidimensional construct. The multidimensional Rasch analysis conducted within the 10 domains of the CAQ indicated that not all items in each domain fit the Rasch model well (Table 1). the hierarchy of creative achievement using Rasch analyses." (Wang et al., 2014, p. 64)
Creative Achievement Questionnaire	+R	"The person separation reliabilities of the 10 subscales, shown in Table 2, ranged from .63 to .77. Of these subscales, the person separation reliability in the dance domain (.63) was lower than were those in the other" (Wang et al., 2014, p. 64)
Creative Attitude Survey	-V	"However, only' one of the 14 comparisons was at a significant level. No differences were found on either the critical thinking test or the creative attitude survey. A listing of packages used by each group and a table displaying statistical data are appended." (Bonk, 1988, p. 1)
Creative Attitude Survey	-V	"Critical thinking was not significantly affected by the convergent treatment. There are several possible reasons for the lack of the expected treatment effects for this group. First, the reading level of the Cornell tests might have been too difficult for these children. Almost half of the subjects were 4th-graders, the lowest applicable age group for this test. Secondly, none of the convergent packages directly addressed the credibility of sources or the assumption identification sub-tests." (Bonk, 1988, p. 22)
Creative Product Semantic Scale	+V	"Results indicated that judgments of Novelty made by naive judges were consistent with those of experts. The validity of Novelty, and to a lesser degree, that of Elaboration and Synthesis, were verified by Study 2. The validity of Resolution is yet to be established (no significant differences among products were observed). In both studies, Novelty and Resolution were independent, but Elaboration and Synthesis subscales migrated between loading with Novelty and Resolution" (O'Quin and Besemer, 1989, p. 267)
Creative Product Semantic Scale	+V	"Confirmatory factor analyses provided strong support for construct validity of the questionnaire and the three-dimensional creativity model. Participant judges were able to detect differences perceived in Novelty, Resolution, and Elaboration and Synthesis of the 4 stimulus items." (Besemer and O'Quin, 1999, p. 287)
Creative scientific ability test	+R and +V	"In this article, we discussed scientific creativity and reviewed the Creative Scientific Ability Test (C-SAT), and then we presented research carried out with 693 sixth grade students to investigate its psychometric properties. The C-SAT measures fluency, flexibility and creativity in hypothesis generation, experiment design and evidence evaluation infeed areas of science. Overall, the findings provided empirical evidence for its reliability and validity, supporting its use as a criterion variable in research and identification practices for scientifically creative students." (Ayas and Sak, 2014, p. 202)

A

Name Test	Codes	Quotation Content
Creative scientific ability test	+V	"Criterion validity Pearson product-moment correlations were calculated to examine relationships between students' C-SAT total fluency, total flexibility and total creativity scores and their grades in mathematics and science and TMT scores. Findings are presented in Table 4. Correlation coefficients ranged from .31 to .59, with the highest correlation between the total fluency and the TMT and with all the correlations being significant (p <.01)." (Ayas and Sak, 2014, p. 202)
Creative scientific ability test	+R	"The Cronbach's Alpha for internal consistency and interscorer reliability were investigated. The Cronbach's Alpha coefficient value was found to be .87. For the interscorer reliability, approximately 14% of the students' papers (n = 164) were randomly selected. Two independent scorers who were trained in the C-SAT scoring procedures scored these papers using the C-SAT standard scoring procedures. Interscorer reliability ranged from .87 to .96 for the five subtests and for the total fluency, total flexibility, and total creativity, with a mean of .92 (see Table 3). Alpha levels from .80 to .89 are considered to be good and from .90 and above are thought to be excellent (George & Mallery, 2003)." (Ayas and Sak, 2014, p. 201)
Critical Reasoning Test Battery	+R and +V	"The GRT2showed that they were all above 0.8, demonstrating a high level of reliability of the test. Furthermore, test–retest coefficients were all above 0.7. In order to test the validity of the GRT2, its sub-scales and total score were compared to the subscales and total score of the Alice Heim reasoning test (AH5)... Correlation coefficients ranged from 0.56 to 0.76 for the subscales, and for the total scores of the two tests it was 0.82(N = 81), demonstrating that the GRT2measures the same trait of reasoning ability which is assessed by the AH5, although the discriminant validity of the subscales is not very high." (Moutafi et al., 2005, p.1024)
Diagnostic Analysis of Nonverbal Accuracy	+V	"The authors report constructs validity evidence from 6 additional studies. Future applications of the DANVA2-POS test are described." (Pitterman and Nowicki Jr, 2004, p. 146)
Diagnostic Analysis of Nonverbal Accuracy	+V	"Convergent validity: Association between two tests of postures. A final piece of construct validity evidence from the initial study involved data related to convergent validity. A sample of undergraduate students (n = 30) completed both the original DANVA posture test and the DANVA2-POS scale. The correlation between the two scales was significant, r (29) = .41, p <.05, providing support for convergent validity." (Pitterman and Nowicki Jr, 2004, p. 156)
Diagnostic Analysis of Nonverbal Accuracy	+V	"Only one study has been completed with children (Pitterman, 2002) and, although it provided findings supportive of construct validity, additional work needs to be completed with a wider age range, especially younger participants. Such information would be helpful for developmental researchers in evaluating the trajectory and impact of the ability to identify emotions in postures throughout childhood." (Pitterman and Nowicki Jr, 2004, p. 159)

Diagnostic Analysis of Nonverbal Accuracy	+V	"In summary, it appears that the DANVA2-POS has met preliminary requirements of construct validity and is ready to be used with a variety of populations so that its worth can be established. Additional research will help to complete the picture of its utility in studying individual differences in the ability to read emotions in adult postures." (Pitterman and Nowicki Jr, 2004, p. 160)
Diagnostic Analysis of Nonverbal Accuracy	+R	"The coefficient alpha for college students was .92. Data regarding the test-retest reliability and validity for Version 3 have not been published" (Pitterman and Nowicki Jr, 2004, p. 152)
Gallup Clifton Strengths Test	+V	"A total of 128 (93.4%) of these predictions were confirmed by significant correlation coefficients, providing strong evidence for the construct validity of the CSF (the full table of these results can be found in Table 3)." (Schreiner, 2006, p. 7)
Gallup Clifton Strengths Test	+R	"The second way of estimating reliability is through "internal consistency," as measured by coefficient alpha. This statistic assesses the extent to which all the items on a theme are related to each other rather than to items on another theme. A perfect score of 1.00 would indicate that all of the items on a theme are related only to the other items on that theme and not to any other items, something that is statistically improbable. Since the CSF was designed so that some items intentionally appear on more than one theme, this makes a high internal consistency score unlikely. Coefficient alphas in this sample ranged from α = .42 for the Activator theme to α = .80 for the Discipline theme, with a mean alpha of .61 and a median alpha of .63." (Schreiner, 2006, p. 6)
Gates- MacGinitie Reading Tests	+V	"The technical manual of the Gates- MacGinitie (MacGinitie et aI., 1978) presents concurrent correlations for Levels D and E with corresponding subtests of the fifth edition of the MAT. These are between.79 and .92, with the higher correlations for total test scores (Stahl, 1989).In describing the validity of the 1972 edition of the Gates MacGinitie, Salvia and Ysseldyke (1978) stated, "The authors of the Gates-MacGinitie do report the results of an unpublished doctoral dissertation by Davis (1968) in which the subtests of the Gates-MacGinitie were found to correlate in the .70 to .85 range with four other standardized reading tests" (p. 154). In addition, Ryckman (1982) reported correlations between the Gray Oral Reading Test and the Gates-MacGinitie that ranged from.48 to .69. In our study, the correlations of .82 to .85 (overall) and .60 to .76 (specific grade levels) between the BASS Reading subtest and the Gates-MacGinitie Reading Tests appeared to be of a magnitude comparable to those reported in this somewhat sparse literature" (Jenkins and Jewell, 1993, p. 423)

A

Name Test	Codes	Quotation Content
Generalized immediacy Scale and Behavioral Indicants of Immediacy Scale	+R and +V	"Andersen et al combined several different non-verbal behaviors to operationalize nonverbal immediacy. Andersen et al's work resulted in the simultaneous development of three instruments: 1) generalized immediacy scale, 2) behavioral indicants of immediacy scale, and 3) trained rater's perception of immediacy scale. These scales were developed to address measurement concerns in establishing both subjective and objective measures, as well as a means to establish construct validity... The behavioral indicants of immediacy scale has been considered problematic because learners are asked to compare their teacher against another teacher and the scale has item redundancy. Building on these limitations, Gorham and Zakahi generated a 14-item instrument combining items from Andersen's early work and researcher-generated items to balance positively and negatively worded items. This revised instrument, nonverbal immediacy measure, was later reduced to ten items and was used primarily across the last decade. Use of these instruments indicated some potential reliability problems, with reliability estimates ranging from 0.67 to 0.89. As a result, the non-verbal immediacy scale was constructed. This newer scale includes a self-report and observer versions and has since 24b0een used in communication research, with reliability estimates of 0.90." (Ellis et al., 2016, p. 13)
Gough's Creative Personality Scale	+V and +R	"The participants took Gough's Creativity Personality Scale for the Adjective Check List. There is a list of 30 adjectives, 18 of them being indicative of creativity (positive) and 12 of them being indicative of a lack of creativity (negative). An example of a positive adjective is "snobbish," and an example of a negative adjective is "honest". The CPS-ACL is a unidimensional scale. The total of negative points and positive points gives a score between -12 and +18, indicating the participant's creativity level. This scale is acceptably reliable, with coefficient alphas ranging from 0.73 to 0.81. Validity is r = 0.35 with Welsh's A-48" (Perry, 2020, p. 825)
Graduate Management Admission Test	+V	"Results based on over 402 independent samples across 64,583 students indicate that the GMAT is a superior predictor to UGPA and that the two combined yield a high level of validity for predicting student" (Kuncel et al., 2007, p. 51)
Graduate Management Admission Test	+V	"The data obtained indicate that the validity of the GMAT is quite likely to generalize across situations" (Kuncel et al., 2007, p. 64)
Graduate Management Admission Test	+V	"Validity estimates for full-time students for GMAT-Verbal (N=705, k=7), GMAT- Quantitative (AT = 541, Jc = 6), and UGPA (N = 705, k = 7) were .32, .32, and .37, respectively. For part-time students, the validity estimates for GMAT-Verbal (N=735, k=10), and GMAT-Quantitative (AT = 799, Jc = 10) were .41 and .39, respectively, while the validity estimate for UGPA (N = 1,102, i: = 14) was .27. Confidence intervals did not overlap for these groups. The validity estimates for students with business undergraduate degrees and those with nonbusiness undergraduate degrees were almost identical." (Kuncel et al., 2007, p. 60)

Halpern Critical Thinking	-R	"Nine exploratory items were developed to assess the respondents' tendency to think critically or their critical thinking disposition. Overall, the scale did not achieve sufficient reliability (Cronbach's alpha .58)." (Butler, 2012, p. 724)
Halpern Critical Thinking	-R	"Overall, those with higher critical thinking scores reported fewer negative life events than those with lower critical thinking scores, r (131) =.38, p <.001." (Butler, 2012, p. 721)
Halpern Critical Thinking Halpern Critical Thinking	-V	"negative life events—avoided. The strength of the relationship between scores on the HCTA and the inventory of real-world outcomes was similar to other measures that have established the criterion validity of critical thinking assessments. Compared with past studies that used academic measures to establish the criterion validity of the HCTA (Halpern, 2010a), the strength of the relationship between HCTA scores and real-world outcomes was not as strong as some of the academic measures (e.g. SAT-Verbal r = .58, SAT-Math r = .50) but was stronger than others (GRE-Verbal r = .12, GRE- Quantitative r = .20; class grades ranged from r = .17 to .41)." (Butler, 2012, p. 725)
Halpern Critical Thinking Halpern Critical Thinking	-R	"Neither of the two tests shows a high overall reliability." (Verburgh et al., 2013)
Halpern Critical Thinking Halpern Critical Thinking	+V	"Results show a higher content validity and preference by students for the HCTA" (Verburgh et al., 2013, p. 1)
HEIghten Critical Thinking	+V	"This meta-analysis supports the validity of the GMAT for predicting grades at earlier and later stages in business school and persisting through the program across the general population of students. In addition, the combination of GMAT scales is a better predictor than prior grades with the combination of GMAT and UGPA likely yielding the best predictive validity." (Kuncel et al., 2007, p. 65)
HEIghten Critical Thinking	+R	"The test has shown adequate total score reliability at both the group and individual levels, and adequate subscale reliability at the group level." (Liu et al., 2016, p. 690)
HEIghten Critical Thinking	+V	"In summary, our analyses provided preliminary validity evidence for the HEIghten critical thinking assessment regarding adequate internal structure, positive relations to other related variables and favorable feedback from examinees." (Liu et al., 2016, p. 691)
HEIghten Critical Thinking	+R	"Individual-level reliability was not high due to the low discriminating items, but institution-level reliabilities for total and subscores were satisfactory." (Liu et al., 2018, p. 1011)
HEIghten Critical Thinking	+R	"In addition, there was a small difference in total score reliability between the two delivery modes: the reliability in Cronbach's alpha for the paper test was .60 for the experiment sample of 278 students, the reliability for the online test for the 257 students in the experiment was .64." (Liu et al., 2018, p. 1005)

A

Name Test	Codes	Quotation Content
Index of Interdisciplinary Collaboration	+R	"Modified Index of Interdisciplinary Collaboration (Oliver, Wittenberg-Lyles, & Day, 2006, 2007) Internal consistency: r=0.935; subscales: interdependence and flexibility: r=0.867; newly created activities: r=0.767; collective ownership: r=0.795; reflection on process" (Thannhauser et al., 2010, p. 349)
Index of Interdisciplinary Collaboration	+V	"Face validity" (Thannhauser et al., 2010, p. 348)
Interdisciplinary Education Perception Scale	+R	"Interdisciplinary Education Perception Scale (Luecht, Madsen, Taugher, Petterson, 1990) Internal consistency (r=0.51–0.87); (Factor 1: r=0.82; Factor 2: r=0.56; Factor 3: r =0.54; Factor 4: r=0.52)–lower reliabilities contributed to small number of items" (Thannhauser et al., 2010, p. 348)
Interdisciplinary Education Perception Scale	+V	"Content-validated by 5 faculty in nursing and allied health; factor analysis" (Thannhauser et al., 2010, p. 347)
International Critical Thinking Reading and Writing	+R	The test was also shown to have good internal reliability, and to correlate with scores on a comparison test thus demonstrating it assesses the same CT construct." (Hollis et al., 2020, p. 25)
International Critical Thinking Reading and Writing	+V	"This study showed the test to be a valid instrument for testing CT." (Hollis et al., 2020, p. 24)
Interpersonal Communication Competence Scale	+V	"We found that self-efficacy mediated the effect of past experience and situation difficulty on interpersonal communication outcomes. Interpersonal communication competence directly effected rewarding, satisfying communication. The study also provided concurrent validity information on the Interpersonal Communication Competence Scale, a skills measure of competence that taps 10 dimensions of the construct" (Rubin et al., 1993, p.210)
Interpersonal Communication Competence Scale	-R	"The second goal of this study was to examine the concurrent validity of the ICCS. The 10-item scale was moderately correlated (ranging from .26 to .55) with the 10 other measures of interpersonal constructs, indicating concurrent validity. Clearly the new interaction management scale created for this study was not reliable; it had a low coefficient alpha and it was not related to the other measures and ICC as the remaining nine were" (Rubin et al., 1993, p. 217)
Interpersonal Communication Competence Scale	+V	"The results also showed that the 10 dimensions thought to comprise ICC are closely related; all the measures were related to one another. Perhaps this is why Wiemann (1977) found one factor for his Communicative Competence Scale. The results suggest, then, that for greater content validity, ICC measures should include as many facets of ICC that can be identified. In this study, for instance, we thought it important to measure as many aspects of ICC as possible to obtain a global measure of ICC skill attainment. Previous measures have focused narrowly on only one or two dimensions (Rubin & Martin, 1992)." (Rubin et al., 1993, p. 217)
Keirsey Temperament Sorter	+R	"On this evidence, the KTSII has reasonable reliability, but the level is more appropriate for research than for individual use." (Dodd and Bayne, 2007, p. 71)

Keirsey Temperament Sorter	+V	"There is little information available to support the validity of the KTS-II. Concurrent measures of the KTS-II and Myers-Briggs Type Indicator were attained for 203 first-semester college freshmen without declared academic majors. There were strong positive correlations between the concurrent MBTI and KTS-II measures of psychological type. The relevance of these findings for career counsellors' use of this online assessment is discussed." (Kelly and Jugovic, 2001, p. 49)
Keirsey Temperament Sorter	+V	"The results have provided tentative support for the validity of the KTS II." (Kelly and Jugovic, 2001, p. 57)
Kentucky Comprehensive Listening Test	-V	"However, some degree of correlation has been reported between certain sections of the Kentucky Comprehensive Listening Test and sections of the Watson-Barker test (Applegate & Campbell, 1985). Evaluation of discriminant validity appears very infrequently in listening research reports. Notable exceptions are Kelly (1965) and Bostrom and Waldhart (1983). Kelly (1965) found high correlations between various listening scores and an IQ measure, indicating a lack of discriminate validity. Bostrom and Waldhart (1983) compared their listening test with tests of mental ability. They found low correlations between scores on their listening test and reading skills and ACT scores, which measure cognitive concepts thought to be related to the listening concept. Their findings suggest a distinction between what they termed "short-term listening" and "short-term memory." (Fitch-Hauser and Hughes, 1992, p. 10)
Management and Graduate Item Bank	+R	"In terms of the reliability of the ability tests, high alpha coefficients (between 0.82 and 0.91) have been obtained (SHL, June 2003:12; Saville et al.,1996:260-261). In terms of the reliability of the OPQ32, the alpha coefficients for the 32 scales vary between 0.65 and 0.88 (SHL, March 2001:3 November 2001:5-6; March 2002:1-2)." (Kotz´e and Griessel, 2008, p. 68)
Multidisciplinary Collaboration instrument	+R	"Multidisciplinary Collaboration (Carroll, 1999) Internal consistency across vignettes: r=0.67– 0.81; within vignettes: r=0.42–0.98; low reliabilities only for vignette #3, 7, 16." (Thannhauser et al., 2010, p. 347)
Multidisciplinary Collaboration instrument	+V	"Face validity (items drawn from the literature); construct validity (convergent & discriminant)" (Thannhauser et al., 2010, p. 345)
Omnibus Personality Inve tory	+V	"Although initial validation results indicated that certain OPI scales might be predictive of attrition in the naval aviation training program, the cross-validation study showed that the relationship was due to chance variance. The application of regression weights to a second population resulted in a significant cross-validation correlation coefficient for each subsample." (Griffin and Hopson, 1978, p. 14)

Name Test	Codes	Quotation Content
Omnibus Personality Inventory	-R	Therefore, it must be concluded that the Omnibus Personality Inventory is not sufficiently related to naval training success to be of value in the prediction of aviator motivational attrition. These results show the importance of going beyond the cross-validation correlation coefficient in the interpretation of variable significance. When numerous prediction variables are being evaluated, a significant cross-validation result may be due to a subset, rather than to all predictor variables. For example, in this study, the U. S. Naval and Marine Aviation Selection Test scores, which are known predictors, were responsible for the significant effect." (Griffin and Hopson, 1978, p. 14)
Readiness for Interprofessional Learning Scale	+R	"Revised Readiness for Interprofessional Learning Scale (McFadyen, Webster, Strachan, Figgins, Brown, & McKechnie, 2005, 2006) Internal consistency: teamwork and collaboration: r=0.79–0.88; roles & responsibilities: r=0.40–0.43; negative professional identity: r= 0.60–0.76; positive professional identity: r=0.76–0.81" (Thannhauser et al., 2010, p. 348)
Readiness for Interprofessional Learning Scale	+V	"Face, content, and construct validity based on Parsell & Bligh (1998, 1999)" (Thannhauser et al., 2010, p. 347)
Riso-Hudson Enneagram Type Indicator Test	+V	"This investigation was conducted 10 estimates the reliability and validity of scores on the Riso-Hudson Enneagram Type Indicator (D. R. Riso di R. Hudson. 1999a). Results of 2H7 participants were analyzed. Alpha suggests an adequate degree of internal consistency Evidence provides mixed support for construct validity using correlational and canonical analyses but strong support for heuristic value." (Newgent et al., 2004, p. 226)
Role Perceptions Questionnaire	+R	"Role Perception Questionnaire (generic) (MacKay, 2004) Test-retest (r ¼ 0.7)" (Thannhauser et al., 2010, p. 347)
Role Perceptions Questionnaire	+V	"Content validity verified through consultation with sample group" (Thannhauser et al., 2010, p. 347)
Scientific creativity Ability test	+R and +V	"In this article, we discussed scientific creativity and reviewed the Creative Scientific Ability Test (C-SAT), and then we presented research carried out with 693 sixth-grade students to investigate its psychometric properties. The C-SAT measures fluency, flexibility and creativity in hypothesis generation, experiment design and evidence evaluation in five areas of science... Overall, the findings provided empirical evidence for its reliability and validity, supporting its use as a criterion variable in research and identification practices for scientifically creative students." (Ayas and Sak, 2014, p. 202)
Templin Test of Auditory Discrimination	-V	"The fourth discrimination test used was the Templin Test of Auditory Discrimination (1957), a measure that involves several phonemes. The subjects were children, eight years of age or older, who made articulation errors only on /r/. The authors reported correlations of 0.93, 0.94, and 0.95 for relationships among their three discrimination tests, but correlations between each of those tests and the Templin measure ranged from 0.21 through 0.26. Correlation coefficients of 0.69, 0.59, and 0.66 were obtained for the Aungst-Frick discrimination scores and scores on McDonald's (1964) deep test for /r/. However, the Templin test scores did not correlate well with the articulation measure (r=0.03)." (Shelton et al., 1977, p. 705)

Torrance Test of Creative Thinking	+V	"In conclusion, the TTCT appears to be a good measure, not only for identifying and educating the gifted but also for discovering and encouraging everyday life creativity in the general population. When used appropriately, the TTCT is an important part of Torrance's legacy and dream: to nurture and enhance creativity among students." (Kim, 2006, p. 11)
Watson-Barker Listening Test	+V	"The significant relationships were curvilinear in nature, as expected, based on the relevant literature. It was concluded that the claims of validity for the Watson-Barker Listening Test are partially supported by this data. The study concludes with a discussion of the status of listening research and suggests directions for further research in the listening field." (Roberts, 1986, p. 115)
Watson-Glaser Critical Thinking Appraisal	+v	"The confirmatory factor analysis provided data on the construct validity of the WGCTA." (Gadzella and Baloğlu, 2003, p. 1256)
Watson-Glaser Critical Thinking Appraisal	+R	"The purposes of this study were to determine (a) the various reliabilities and validities of the Watson-Glaser Critical Thinking Appraisal, WGCTA, for students majoring in Education and (b) which variable(s) best predicted their Educational Psychology course grades. The data showed that the WGCTA is a reliable and valid test measuring critical thinking for students majoring in Education." (Gadzella and Baloğlu, 2003, p. 1256)
Watson-Glaser Critical Thinking Appraisal	+V	"The standardization sample was used in the factor structure validity analysis. Three models were compared: a unidimensional factor model, a three-factor model, and a five-factor model. The results supported a three-factor model (goodness-of-fit statistic GFI=.97, adjusted goodness of-fit statistic AGFI=.96, root mean square error approximation RMSEA= .03), providing justification for three scales." (Sternod and French, 2016, p. 609)
Watson-Glaser Critical Thinking Appraisal	+R	"The Watson–Glaser short form. In 1994 and 2006, reliability values were reported as .81 and .89, respectively. The test developers highlighted the fact that cognitive abilities, such as critical thinking, are stable trait for adults, therefore they refer to the research previously conducted. Although there may be evidence to suggest traits such as these are stable, empirical evidence is still required to show that the measurements used to assess these traits are stable as well. This is especially the case here as the changes made to the test were significant, producing a different test altogether. Therefore, updated reliability estimates are critical." (Sternod and French, 2016, p. 608)

A

B Supplemental Material of the Chapter 4

B.1 Empathy Questionnaire (IRI)

The following statements inquire about your thoughts and feelings in a variety of situations. For each item, indicate how well it describes you by choosing the appropriate SCORE. When you have decided on your answer, fill in the letter next to the item number. READ EACH ITEM CAREFULLY BEFORE RESPONDING. Answer as honestly as you can. Thank you.

1. I daydream and fantasize, with some regularity, about things that might happen to me. (FANTASY)
   Does not describe me well 0 1 2 3 4 **Describe me very well**

2. I often have tender, concerned feelings for people less fortunate than me. (EMPATHIC CONCERN)
   Does not describe me well 0 1 2 3 4 **Describe me very well**

3. I sometimes find it difficult to see things from the "other guy's" point of view. (PERSPECTIVE TAKING) (-)
   Does not describe me well 0 1 2 3 4 **Describe me very well**

4. I sometimes I don't feel very sorry for other people when they are having problems. (EMPATHIC CONCERN) (-)
   Does not describe me well 0 1 2 3 4 **Describe me very well**ll

5. I really get involved with the feelings of the characters in a novel. (FANTASY)
   Does not describe me well 0 1 2 3 4 **Describe me very well**

6. In emergency situations, I feel apprehensive and ill-at-ease. (PERSONAL DISTRESS)
   Does not describe me well 0 1 2 3 4 **Describe me very well**

7. I am usually objective when I watch a movie or play, and I don't often get completely caught up in it.
(FANTASY) (-)
Does not describe me well 0 1 2 3 4 **Describe me very well**

8. I try to look at everybody's side of a disagreement before I make a decision. (PERSPECTIVE TAKING)
Does not describe me well 0 1 2 3 4 **Describe me very well**

9. When I see someone being taken advantage of, I feel kind of protective towards them. (EMPATHIC CONCERN)
Does not describe me well 0 1 2 3 4 **Describe me very well**

10. I sometimes feel helpless when I am in the middle of a very emotional situation. (PERSONAL DISTRESS)
Does not describe me well 0 1 2 3 4 **Describe me very well**

11. I sometimes try to understand my friends better by imagining how things look from their perspective.
(PERSPECTIVE TAKING)
Does not describe me well 0 1 2 3 4 **Describe me very well**

12. Becoming extremely involved in a good book or movie is somewhat rare for me.
(FANTASY) (-)
Does not describe me well 0 1 2 3 4 **Describe me very well**

13. When I see someone get hurt, I tend to remain calm.
(PERSONAL DISTRESS) (-)
Does not describe me well 0 1 2 3 4 **Describe me very well**

14. Other people's misfortunes do not usually disturb me a great deal.
(EMPATHIC CONCERN) (-)
Does not describe me well 0 1 2 3 4 **Describe me very well**

15. If I'm sure I'm right about something, I don't waste much time listening to other people's arguments.
(PERSPECTIVE TAKING) (-)
Does not describe me well 0 1 2 3 4 **Describe me very well**

16. After seeing a play or movie, I have felt as though I were one of the characters.
(FANTASY)
Does not describe me well 0 1 2 3 4 **Describe me very well**

17. Being in a tense emotional situation scares me.
(PERSONAL DISTRESS)
Does not describe me well 0 1 2 3 4 **Describe me very well**

18. When I see someone being treated unfairly, I sometimes don't feel very much pity for them.
(EMPATHIC CONCERN) (-)
Does not describe me well 0 1 2 3 4 **Describe me very well**

19. I am usually pretty effective in dealing with emergencies.
(PERSONAL DISTRESS) (-)
Does not describe me well 0 1 2 3 4 **Describe me very well**

20. I am often quite touched by things that I see happen.
(EMPATHIC CONCERN)
Does not describe me well 0 1 2 3 4 **Describe me very well**

21. I believe that there are two sides to every question and try to look at them both.
(PERSPECTIVE TAKING)
Does not describe me well 0 1 2 3 4 **Describe me very well**

22. I would describe myself as a pretty softhearted person.
(EMPATHY CONCERN)
Does not describe me well 0 1 2 3 4 **Describe me very well**

23. When I watch a good movie, I can very easily put myself in the place of a leading character.
(FANTASY)
Does not describe me well 0 1 2 3 4 **Describe me very well**

24. I tend to lose control during emergencies.
(PERSONAL DISTRESS)
Does not describe me well 0 1 2 3 4 **Describe me very well**

25. When I'm upset at someone, I usually try to "put myself in his shoes" for a while.
(PERSPECTIVE TAKING)
Does not describe me well 0 1 2 3 4 **Describe me very well**

26. When I am reading an interesting story or novel, I imagine how I would feel if the events in the story were happening to me.
(FANTASY)
Does not describe me well 0 1 2 3 4 **Describe me very well**

27. When I see someone who badly needs help in an emergency, I go to pieces.
(PERSONAL DISTRESS)
Does not describe me well 0 1 2 3 4 **Describe me very well**

28. Before criticizing somebody, I try to imagine how I would feel if I were in their place.
(PERSPECTIVE TAKING)
Does not describe me well 0 1 2 3 4 **Describe me very well**

B.2 Immersion Questionnaire

Your experience of the game. Please answer the following questions by circling the relevant number. In particular, remember that these questions are asking you about how you felt at the end of the game.

1. To what extent did the game hold your attention?
   (ENGAGEMENT)
   Not at all 1 2 3 4 5 **A lot**

2. To what extent did you feel you were focused on the game?
   (ENGAGEMENT)
   Not at all 1 2 3 4 5 **A lot**

3. How much effort did you put into playing the game?
   (ENGAGEMENT)
   Very little 1 2 3 4 5 **A lot**

4. To Did you feel that you were trying you best?
   (ENGAGEMENT)
   Not at all 1 2 3 4 5 **Very much so**

5. To what extent did you lose track of time?
   (ENGROSSEMENT)
   Not at all 1 2 3 4 5 **A lot**

6. What extent did you feel consciously aware of being in the real world whilst playing?
   (ENGROSSEMENT) (-)
   Not at all 1 2 3 4 5 **Very much so**

7. To what extent did you forget about your everyday concerns?
   (ENGROSSEMENT)
   Not at all 1 2 3 4 5 **A lot**

8. To what extent were you aware of yourself in your surroundings?
(ENGROSSEMENT) (-)
Not at all 1 2 3 4 5 **Very aware**

9. To what extent did you notice events taking place around you?
(ENGROSSEMENT) (-)
Not at all 1 2 3 4 5 **A lot**

10. Did you feel the urge at any point to stop playing and see what was happening around you?
(ENGROSSEMENT) (-)
Not at all 1 2 3 4 5 **Very much so**

11. To what extent did you feel that you were interacting with the game environment?
(ENGROSSEMENT)
Not at all 1 2 3 4 5 **Very much so**

12. To what extent did you feel as though you were separated from your real-world environment?
(ENGROSSEMENT)
Not at all 1 2 3 4 5 **Very much so**

13. To what extent did you feel that the game was something you were experiencing, rather than something you were just doing?
(ENGROSSEMENT)
Not at all 1 2 3 4 5 **Very much so**

14. To what extent was your sense of being in the game environment stronger than your sense of being in the real world?
(ENGROSSEMENT)
Not at all 1 2 3 4 5 **Very much so**

15. At any point did you find yourself become so involved that you were unaware you were even using controls?
(ENGROSSEMENT)
Not at all 1 2 3 4 5 **Very much so**

16. To what extent did you feel as though you were moving through the game according to you own will?
(ENGAGEMENT)
Not at all 1 2 3 4 5 **Very much so**

17. To what extent did you find the game challenging?
(ENGROSSEMENT)
Not at all 1 2 3 4 5 **Very difficult**

18. Were there any times during the game in which you just wanted to give up?
(ENGAGEMENT) (-)
Not at all 1 2 3 4 5 **A lot**

19. To what extent did you feel motivated while playing?
(ENGAGEMENT)
Not at all 1 2 3 4 5 **A lot**

20. To what extent did you find the game easy?
(ENGROSSEMENT) (-)
Not at all 1 2 3 4 5 **Very much so**

21. To what extent did you feel like you were making progress towards the end of the game?
(ENGAGEMENT)
Not at all 1 2 3 4 5 **A lot**

22. How well do you think you performed in the game?
(ENGAGEMENT)
Very Poor 1 2 3 4 5 **Very well**

23. To what extent did you feel emotionally attached to the game?
(ENGROSSEMENT)
Not at all 1 2 3 4 5 **Very much so**

24. To what extent were you interested in seeing how the game's events would progress?
(ENGROSSEMENT)
Not at all 1 2 3 4 5 **A lot**

25. How much did you want to "win" the game?
(ENGROSSEMENT)
Not at all 1 2 3 4 5 **Very much so**

26. Were you in suspense about whether or not you would win or lose the game?
(ENGROSSEMENT)
Not at all 1 2 3 4 5 **Very much so**

27. At any point did you find yourself become so involved that you wanted to speak to the game directly?
(ENGROSSEMENT)
Not at all 1 2 3 4 5 **Very much so**

28. To what extent did you enjoy the graphics and the imagery?
(ENGROSSEMENT)
Not at all 1 2 3 4 5 **A lot**

29. How much would you say you enjoyed playing the game?
(ENGROSSEMENT)
Not at all 1 2 3 4 5 **A lot**

30. When interrupted, were you disappointed that the game was over?
(ENGROSSEMENT)
Not at all 1 2 3 4 5 **Very much so**

31. Would you like to play the game again?
(ENGROSSEMENT)
Definitely not 1 2 3 4 5 **Definitely yes**

B.3 Principal component analysis of the Interpersonal Reactivity Index questionnaire.

Table B.3.1: Principal component analysis of the Interpersonal Reactivity Index Questionnaire.

Factor A prior classification	Items	1	2	3	4
Factor 1: Fantasy	Q16. After seeing a play or movie, I have felt as though I were one of the characters. (FS)	0.759			
	Q23. When I watch a good movie, I can very easily put myself in the place of a leading character. (FS)	0.739			
	Q5. I really get involved with the feelings of the characters in a novel. (FS)	0.709			
	Q26. When I am reading an interesting story or novel, I imagine how I would feel if the events in the story were happening to me. (FS)	0.684			
	Q7. I am usually objective when I watch a movie or play, and I don't often get completely caught up in it. (FS)	0.665			
	Q1. I daydream and fantasize, with some regularity, about things that might happen to me. (FS)	0.662			
	Q12. Becoming extremely involved in a good book or movie is somewhat rare for me. (FS)	0.609			0.301
Factor 2: Personal distress	Q6. In emergency situations, I feel apprehensive and ill-at-ease. (PD)		0.759		
	Q10. I sometimes feel helpless when I am in the middle of a very emotional situation. (PD)	0.301	0.756		
	Q24. I tend to lose control during emergencies. (PD)		0.743		
	Q17. Being in a tense emotional situation scares me. (PD)		0.674		
	Q27. When I see someone who badly needs help in an emergency, I go to pieces. (PD)		0.605		
	Q19. I am usually pretty effective in dealing with emergencies. (PD)		0.565		
Factor 3: Perspective taking	Q8. I try to look at everybody's side of a disagreement before I make a decision. (PT)			0.732	
	Q11. I sometimes try to understand my friends better by imagining how things look from their perspective. (PT)			0.690	
	Q21. I believe that there are two sides to every question and try to look at them both. (PT)			0.670	
	Q25. When I'm upset at someone, I usually try to "put myself in his shoes" for a while. (PT)			0.627	
	Q28. Before criticizing somebody, I try to imagine how I would feel if I were in their place. (PT)			0.590	
	Q15. If I'm sure I'm right about something, I don't waste much time listening to other people's arguments. (PT)			0.464	
	Q9. When I see someone being taken advantage of, I feel kind of protective towards them. (EC)			0.313	
Factor 4: Empathic concern	Q4. Sometimes I don't feel very sorry for other people when they are having problems. (EC)	-0.337			0.659
	Q3. I sometimes find it difficult to see things from the "other guy's" point of view. (PT)		-0.401		0.570
	Q20. I am often quite touched by things that I see happen. (EC)	0.315			0.528
	Q13. When I see someone get hurt, I tend to remain calm. (PD)				0.509
	Q14. Other people's misfortunes do not usually disturb me a great deal. (EC)			0.336	0.494
	Q22. I would describe myself as a pretty soft-hearted person. (EC)				0.465
	Q2. I often have tender, concerned feelings for people less fortunate than me. (EC)			0.346	0.388
	Q18. When I see someone being treated unfairly, I sometimes don't feel very much pity for them. (EC)				0.334

Note: Extraction Method: Principal Component Analysis. Rotation Method: Oblimin with Kaiser Normalization. We supressed small coefficients =0,3

B

B.4 Principal component analysis of the immersion questionnaire.

Table B.4.1: Principal component analysis of the Immersion Questionnaire.

Factor A prior classification	Item	1	2
Factor 1: Engrossment	Q14. To what extent was your sense of being in the game environment stronger than your sense of being in the real world?	**0.781**	
	Q29. How much would you say you enjoyed playing the game?	**0.711**	0.262
	Q23. To what extent did you feel emotionally attached to the game?	**0.638**	
	Q8. To what extent were you aware of yourself in your surroundings?	**0.619**	
	Q28. To what extent did you enjoy the graphics and the imagery?	**0.616**	
	Q9. To what extent did you notice events taking place around you?	**0.582**	
	Q6. To what extent did you feel consciously aware of being in the real world whilst playing?	**0.580**	
	Q20. To what extent did you find the game easy?	**0.572**	-0.373
	Q30. When interrupted, were you disappointed that the game was over?	**0.571**	
	Q12. To what extent did you feel as though you were separated from your real-world environment?	**0.545**	0.292
	Q7. To what extent did you forget about your everyday concerns?	**0.532**	0.275
	Q27. At any point did you find yourself become so involved that you wanted to speak to the game directly?	**0.531**	-0.238
	Q24. To what extent were you interested in seeing how the game's events would progress?	**0.524**	0.228
	Q31. Would you like to play the game again?	**0.510**	0.253
	Q5. To what extent did you lose track of time?	**0.505**	
	Q15. At any point did you find yourself become so involved that you were unaware you were even using controls?	**0.455**	
	Q13. To what extent did you feel that the game was something you were experiencing, rather than something you were just doing?	**0.452**	
	Q11. To what extent did you feel that you were interacting with the game environment?	**0.418**	0.367
	Q17. To what extent did you find the game challenging?	**0.416**	
	Q25. How much did you want to "win" the game?	**0.310**	0.216
	Q26. Were you in suspense about whether or not you would win or lose the game?	**0.289**	
	Q10. Did you feel the urge at any point to stop playing and see what was happening around you?	**0.256**	0.222
Factor 2: Engagement	Q1. To what extent did the game hold your attention?	0.297	**0.629**
	Q2. To what extent did you feel you were focused on the game?	0.263	**0.620**
	Q19. To what extent did you feel motivated while playing?	0.264	**0.619**
	Q16. To what extent did you feel as though you were moving through the game according to you own will?		**0.591**
	Q21. To what extent did you feel like you were making progress towards the end of the game?		**0.496**
	Q18. Were there any times during the game in which you just wanted to give up?		**0.486**
	Q4. Did you feel that you were trying you best?	0.312	**0.458**
	Q22. How well do you think you performed in the game?		**0.392**
	Q3. How much effort did you put into playing the game?	0.325	**0.386**

Note: Extraction Method: Principal Component Analysis. Rotation Method: Oblimin with Kaiser Normalization. We supressed small coefficients =0,2

B

C Supplemental Material of the hapter 5

C.1 Dispositional Empathy Questionnaire (IRI)

(Standardized Cronbach's α = 0.848)

The following statements inquire about your thoughts and feelings in a variety of situations. For each item, indicate how well it describes you by choosing the appropriate SCORE. When you have decided on your answer, fill in the letter next to the item number. READ EACH ITEM CAREFULLY BEFORE RESPONDING. Answer as honestly as you can. Thank you.

1. I daydream and fantasize, with some regularity, about things that might happen to me.
   (FANTASY)
   Does not describe me well 0 1 2 3 4 **Describe me very well**

2. I I often have tender, concerned feelings for people less fortunate than me.
   (EMPATHIC CONCERN)
   Does not describe me well 0 1 2 3 4 **Describe me very well**

3. I I sometimes find it difficult to see things from the "other guy's" point of view.
   (PERSPECTIVE TAKING) (-)
   Does not describe me well 0 1 2 3 4 **Describe me very well**

4. I I sometimes I don't feel very sorry for other people when they are having problems.
   (EMPATHIC CONCERN) (-)
   Does not describe me well 0 1 2 3 4 **Describe me very well**

5. I I really get involved with the feelings of the characters in a novel.
   (FANTASY)
   Does not describe me well 0 1 2 3 4 **Describe me very well**

6. In emergency situations, I feel apprehensive and ill-at-ease.
   (PERSONAL DISTRESS)
   Does not describe me well 0 1 2 3 4 **Describe me very well**

7. I am usually objective when I watch a movie or play, and I don't often get completely caught up in it.
(FANTASY) (-)
Does not describe me well 0 1 2 3 4 **Describe me very well**

8. I try to look at everybody's side of a disagreement before I make a decision.
(PERSPECTIVE TAKING)
Does not describe me well 0 1 2 3 4 **Describe me very well**

9. When I see someone being taken advantage of, I feel kind of protective towards them.
(EMPATHIC CONCERN)
Does not describe me well 0 1 2 3 4 **Describe me very well**

10. I sometimes feel helpless when I am in the middle of a very emotional situation.
(PERSONAL DISTRESS)
Does not describe me well 0 1 2 3 4 **Describe me very well**

11. I sometimes try to understand my friends better by imagining how things look from their perspective.
(PERSPECTIVE TAKING)
Does not describe me well 0 1 2 3 4 **Describe me very well**

12. Becoming extremely involved in a good book or movie is somewhat rare for me.
(FANTASY) (-)
Does not describe me well 0 1 2 3 4 **Describe me very well**

13. When I see someone get hurt, I tend to remain calm.
(PERSONAL DISTRESS) (-)
Does not describe me well 0 1 2 3 4 **Describe me very well**

14. Other people's misfortunes do not usually disturb me a great deal.
(EMPATHIC CONCERN) (-)
Does not describe me well 0 1 2 3 4 **Describe me very well**

15. If I'm sure I'm right about something, I don't waste much time listening to other people's arguments.
(PERSPECTIVE TAKING) (-)
Does not describe me well 0 1 2 3 4 **Describe me very well**

16. After seeing a play or movie, I have felt as though I were one of the characters.
(FANTASY)
Does not describe me well 0 1 2 3 4 **Describe me very well**

17. Being in a tense emotional situation scares me.
(PERSONAL DISTRESS)
Does not describe me well 0 1 2 3 4 **Describe me very well**

18. When I see someone being treated unfairly, I sometimes don't feel very much pity for them.
(EMPATHIC CONCERN) (-)
Does not describe me well 0 1 2 3 4 **Describe me very well**

19. I am usually pretty effective in dealing with emergencies.
(PERSONAL DISTRESS) (-)
Does not describe me well 0 1 2 3 4 **Describe me very well**

20. I am often quite touched by things that I see happen.
(EMPATHIC CONCERN)
Does not describe me well 0 1 2 3 4 **Describe me very well**

21. I believe that there are two sides to every question and try to look at them both.
(PERSPECTIVE TAKING)
Does not describe me well 0 1 2 3 4 **Describe me very welll**

22. I would describe myself as a pretty soft-hearted person.
(EMPATHY CONCERN)
Does not describe me well 0 1 2 3 4 **Describe me very well**

23. When I watch a good movie, I can very easily put myself in the place of a leading character.
(FANTASY)
Does not describe me well 0 1 2 3 4 **Describe me very well**

24. I tend to lose control during emergencies.
(PERSONAL DISTRESS)
Does not describe me well 0 1 2 3 4 **Describe me very well**

25. When I'm upset at someone, I usually try to "put myself in his shoes" for a while.
(PERSPECTIVE TAKING)
Does not describe me well 0 1 2 3 4 **Describe me very well**

26. When I am reading an interesting story or novel, I imagine how I would feel if the events in the story were happening to me.
(FANTASY)
Does not describe me well 0 1 2 3 4 **Describe me very well**

27. When I see someone who badly needs help in an emergency, I go to pieces.
(PERSONAL DISTRESS)
Does not describe me well 0 1 2 3 4 **Describe me very well**

28. Before criticizing somebody, I try to imagine how I would feel if I were in their place.
(PERSPECTIVE TAKING)
Does not describe me well 0 1 2 3 4 **Describe me very well**

C. 2 Immersion Questionnaire

(Standardized Cronbach's α = 0.883)

Your experience of the game. Please answer the following questions by circling the relevant number. In particular, remember that these questions are asking you about how you felt at the end of the game.

1. To what extent did the game hold your attention?
   (ENGAGEMENT)
   Not at all 1 2 3 4 5 **A lot**

2. To what extent did you feel you were focused on the game?
   (ENGAGEMENT)
   Not at all 1 2 3 4 5 **A lot**

3. How much effort did you put into playing the game?
   (ENGAGEMENT)
   Very little 1 2 3 4 5 **A lot**

4. To did you feel that you were trying you best?
   (ENGAGEMENT)
   Not at all 1 2 3 4 5 **Very much so**

5. To what extent did you lose track of time?
   (ENGROSSEMENT)
   Not at all 1 2 3 4 5 **A lot**

6. What extent did you feel consciously aware of being in the real world whilst playing?
   (ENGROSSEMENT) (-)
   Not at all 1 2 3 4 5 **Very much so**

7. To what extent did you forget about your everyday concerns?
   (ENGROSSEMENT)
   Not at all 1 2 3 4 5 **A lot**

8. To what extent were you aware of yourself in your surroundings?
(ENGROSSEMENT) (-)
Not at all 1 2 3 4 5 V**ery aware**

9. To what extent did you notice events taking place around you?
(ENGROSSEMENT) (-)
Not at all 1 2 3 4 5 **A lot**

10. Did you feel the urge at any point to stop playing and see what was happening around you?
(ENGROSSEMENT) (-)
Not at all 1 2 3 4 5 **Very much so**

11. To what extent did you feel that you were interacting with the game environment?
(ENGROSSEMENT)
Not at all 1 2 3 4 5 **Very much so**

12. To what extent did you feel as though you were separated from your real-world environment?
(ENGROSSEMENT)
Not at all 1 2 3 4 5 **Very much so**

13. To what extent did you feel that the game was something you were experiencing, rather than something you were just doing?
(ENGROSSEMENT)
Not at all 1 2 3 4 5 **Very much so**

14. To what extent was your sense of being in the game environment stronger than your sense of being in the real world?
(ENGROSSEMENT)
Not at all 1 2 3 4 5 **Very much so**

15. At any point did you find yourself become so involved that you were unaware you were even using controls?
(ENGROSSEMENT)
Not at all 1 2 3 4 5 **Very much so**

16. To what extent did you feel as though you were moving through the game according to you own will?
(ENGAGEMENT)
Not at all 1 2 3 4 5 **Very much so**

17. To what extent did you find the game challenging?
(ENGROSSEMENT)
Not at all 1 2 3 4 5 **Very difficult**

18. Were there any times during the game in which you just wanted to give up?
(ENGAGEMENT) (-)
Not at all 1 2 3 4 5 **A lot**

19. To what extent did you feel motivated while playing?
(ENGAGEMENT)
Not at all 1 2 3 4 5 **A lot**

20. To what extent did you find the game easy?
(ENGROSSEMENT) (-)
Not at all 1 2 3 4 5 **Very much so**

21. To what extent did you feel like you were making progress towards the end of the game?
(ENGAGEMENT)
Not at all 1 2 3 4 5 **A lot**

22. How well do you think you performed in the game?
(ENGAGEMENT)
Very Poor 1 2 3 4 5 **Very well**

23. To what extent did you feel emotionally attached to the game?
(ENGROSSEMENT)
Not at all 1 2 3 4 5 **Very much so**

24. To what extent were you interested in seeing how the game's events would progress?
(ENGROSSEMENT)
Not at all 1 2 3 4 5 **A lot**

25. How much did you want to "win" the game?
(ENGROSSEMENT)
Not at all 1 2 3 4 5 **Very much so**

26 Were you in suspense about whether or not you would win or lose the game?
(ENGROSSEMENT)
Not at all 1 2 3 4 5 **Very much so**

27. At any point did you find yourself become so involved that you wanted to speak to the game directly?
(ENGROSSEMENT)
Not at all 1 2 3 4 5 **Very much so**

28. To what extent did you enjoy the graphics and the imagery?
(ENGROSSEMENT)
Not at all 1 2 3 4 5 **A lot**

29. How much would you say you enjoyed playing the game?
(ENGROSSEMENT)
Not at all 1 2 3 4 5 **A lot**

30. When interrupted, were you disappointed that the game was over?
(ENGROSSEMENT)
Not at all 1 2 3 4 5 **Very much so**

31. Would you like to play the game again?
(ENGROSSEMENT)
Definitely not 1 2 3 4 5 **Definitely yes**

C. 3 Avatar Embodiment Questionnaire

(Standardized Cronbach's α = 0.922)

Please select your level of agreement with the following statements: "During the experiment there were moments in which. . .

1. I felt as if the virtual body I saw when I looked down was my body
   (BODY OWNERSHIP)
   Strongly disagree 1 2 3 4 5 6 7 **Strongly agree**

2. It felt as if the virtual body I saw was someone else
   (-) (BODY OWNERSHIP)
   Strongly disagree 1 2 3 4 5 6 7 **Strongly agree**

3. It seemed as if I might have more than one body
   (BODY OWNERSHIP)
   Strongly disagree 1 2 3 4 5 6 7 **Strongly agree**

4. I felt as if the virtual body I saw when looking in the mirror was my own body
   (BODY OWNERSHIP)
   Strongly disagree 1 2 3 4 5 6 7 **Strongly agree**

5. I felt as if the virtual body I saw when the game character looked in the mirror was another person
   (-) (BODY OWNERSHIP)
   Strongly disagree 1 2 3 4 5 6 7 **Strongly agree**

6. It felt like I could control the virtual body as if it was my own body
   (AGENCY AND MOTOR CONTROL)
   Strongly disagree 1 2 3 4 5 6 7 **Strongly agree**

7. The movements of the virtual body were caused by my movements
   (AGENCY AND MOTOR CONTROL)
   Strongly disagree 1 2 3 4 5 6 7 **Strongly agree**

C

8. I felt as if the movements of the virtual body were influencing my own movements
(AGENCY AND MOTOR CONTROL)
Strongly disagree 1 2 3 4 5 6 7 **Strongly agree**

9. I felt as if the virtual body was moving by itself
(AGENCY AND MOTOR CONTROL)
Strongly disagree 1 2 3 4 5 6 7 **Strongly agree**

10. It seemed as if I felt the touch of water when the game character was swimming
(TACTILE SENSATIONS)
Strongly disagree 1 2 3 4 5 6 7 **Strongly agree**

11. It seemed as if I felt the trees when the game character touched them
(TACTILE SENSATIONS)
Strongly disagree 1 2 3 4 5 6 7 **Strongly agree**

12. It seemed as if I felt the virtual shot when I was felt the vibrations in the vest
(TACTILE SENSATIONS)
Strongly disagree 1 2 3 4 5 6 7 **Strongly agree**

13. It seemed as if I felt the virtual shot when I was felt the vibrations in the vest
(TACTILE SENSATIONS)
Strongly disagree 1 2 3 4 5 6 7 **Strongly agree**

14. I felt as if my body was located where I saw the virtual body
(LOCATION OF THE BODY)
Strongly disagree 1 2 3 4 5 6 7 **Strongly agree**

15. I felt out of my body
(LOCATION OF THE BODY)
Strongly disagree 1 2 3 4 5 6 7 **Strongly agree**

16. I felt as if my (real) body was drifting towards the virtual body or as if the virtual body was drifting towards my (real) body"
(LOCATION OF THE BODY)
Strongly disagree 1 2 3 4 5 6 7 **Strongly agree**

17. I felt as if my (real) body was turning into an 'avatar' body
(EXTERNAL APPEARANCE)
Strongly disagree 1 2 3 4 5 6 7 **Strongly agree**

18. At some point it felt as if my real body was starting to take on the posture or shape of the virtual body that I saw
(EXTERNAL APPEARANCE)
Strongly disagree 1 2 3 4 5 6 7 **Strongly agree**

19. At some point it felt that the virtual body resembled my own (real) body, in terms of shape, skin tone or other visual features
(EXTERNAL APPEARANCE)
Strongly disagree 1 2 3 4 5 6 7 **Strongly agree**

20. I felt like I was wearing different skin from when I came to the laboratory
(EXTERNAL APPEARANCE)
Strongly disagree 1 2 3 4 5 6 7 **Strongly agree**

21. I felt as if my own body could be affected by the virtual shot
(RESPONSE TO EXTERNAL STIMULI)
Strongly disagree 1 2 3 4 5 6 7 **Strongly agree**

22. I felt a distress sensation in my body when I saw the shooter
(RESPONSE TO EXTERNAL STIMULI)
Strongly disagree 1 2 3 4 5 6 7 **Strongly agree**

23. When the virtual shot happened, I felt the instinct to escape
(BODY OWNERSHIP)
Strongly disagree 1 2 3 4 5 6 7 **Strongly agree**

24. I felt as if my virtual body had vibration sensations
(RESPONSE TO EXTERNAL STIMULI)
Strongly disagree 1 2 3 4 5 6 7 **Strongly agree**

25. I had the feeling that I might be harmed by the virtual shot
(RESPONSE TO EXTERNAL STIMULI)
Strongly disagree 1 2 3 4 5 6 7 **Strongly agree**

C.4 Virtual Reality Sickness Symptoms

(Standardized Cronbach's α = 0.802)
Please, tell us about your experience with the game!:

1. General discomfort during the game
None 1 2 3 4 5 **Severe**

2. Boredom
None 1 2 3 4 5 **Severe**

3. Nausea (dizziness)
None 1 2 3 4 5 **Severe**

4. Headache
None 1 2 3 4 5 **Severe**

6. Disorientation
None 1 2 3 4 5 **Severe**

6. Stomach awareness
None 1 2 3 4 5 **Severe**

C.5 Self-reported situational empathy

(Standardized Cronbach's α = 0.832)

Please, tell us about your experience with the game!:

1. To what extent did you empathize with the character (animal/robot animal/ amorphous figure)?
   Not at all 1 2 3 4 5 **Very much**
2. To what extent did you empathize with the game character?
   Not at all 1 2 3 4 5 **Very much**

C.6 Perceived pain questionnaire

Please, tell us about your experience with the game!:

1. Which picture best depicts how the virtual shot made you feel?

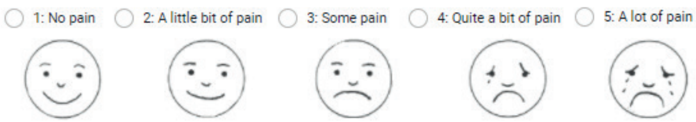


Figure C.6. 1. Perceived pain images.

C.7 Animal conservation tendency

(Standardized Cronbach's α = 0.872; Pre and Post Test) Please, tell us about your experience with the game!:

1. I am really fond of (like) animals
   Strongly disagree 1 2 3 4 5 6 7 **Strongly agree**

2. I am interested in protecting endangered species
   Strongly disagree 1 2 3 4 5 6 7 **Strongly agree**

3. I like the animal in the picture
   Strongly disagree 1 2 3 4 5 6 7 **Strongly agree**

4. I am interested in protecting the animal in the picture
   Strongly disagree 1 2 3 4 5 6 7 **Strongly agree**

C

Figure C.7.1. Beaver image.

C.8 Self-reported emotions in the distressing event

What did you feel when the game character died? (You can choose more than one)

1. Joy or happy

2. Surprise

3. Angry

4. Sad

5. Fear or scared

6. Disgust

7. Nothing

8. Other

C.9 PCA of the Questionnaire Interpersonal Reactivity Index

Table C.9.1: Principal component analysis of the Dispositional Empathy (IRI) Questionnaire.

Factor A prior classification	**Items**	**1**	**2**	**3**	**4**
Factor 1: Fantasy	Q23: When I watch a good movie, I can very easily put myself in the place of a leading character. (FS)	**0.795**			
	Q16: After seeing a play or movie, I have felt as though I were one of the characters. (FS)	**0.789**			
	Q26: When I am reading an interesting story or novel, I imagine how I would feel if the events in the story were happening to me. (FS)	**0.625**			
	Q7: I am usually objective when I watch a movie or play, and I don't often get completely caught up in it. (-) (FS)	**0.621**			0.357
	Q12: Becoming extremely involved in a good book or movie is somewhat is rare for me. (-)(FS)	**0.600**			
	Q1: I daydream and fantasize, with some regularity, about things that might happen to me. (FS)	**0.579**			-0.314
	Q5: I really get involved with the feelings of the characters in a novel. (FS)	**0.576**			
Factor 2: Personal Distress	Q24: I tend to lose control during emergencies. (PD)		**0.725**		
	Q6: In emergency situations, I feel apprehensive and ill-at-ease (awkward). (PD)		**0.630**		
	Q2: I often have tender, concerned feelings for people less fortunate than me. (EC)		**0.584**		
	Q19: I am usually pretty effective in dealing with emergencies. (-) (PD)		**0.582**	-0.464	
	Q27: When I see someone who badly needs help in an emergency, I go to pieces. (PD)		**0.563**		
	Q22: I would describe myself as a pretty soft-hearted person. (EC)	0.438	**0.492**		
	Q13: When I see someone get hurt, I tend to remain calm. (-) (PD)		**0.429**		
	Q11: I sometimes try to understand my friends better by imagining how things look from their perspective. (PT)		**0.416**		
	Q17: Being in a tense emotional situation scares me. (PD)		**0.392**		-0.495
Factor 3: Perspective Taking	Q8: I try to look at everybody's side of a disagreement before I make a decision. (PT)			**0.712**	
	Q21: I believe that there are two sides to every question and try to look at them both. (PT)			**0.697**	
	Q9: When I see someone being taken advantage of, I feel kind of protective towards them. (EC)			**0.652**	
	Q28: Before criticizing somebody, I try to imagine how I would feel if I were in their place. (PT)		0.489	**0.597**	
	Q15: If I'm sure I'm right about something, I don't waste much time listening to other people's arguments. (-) (PT)			**0.593**	
	Q25: When I'm upset at someone, I usually try to "put myself in his shoes" for a while. (PT)		0.626	**0.313**	
Factor 4: Empathic Concern	Q20: I am often quite touched by things that I see happen. (EC)	0.543	0.443		**0.374**
	Q4: Sometimes I don't feel very sorry for other people when they are having problems. (-) (EC)		0.374	0.446	**0.343**
	Q3: I sometimes find it difficult to see things from the "other guy's" point of view. (-) (PT)				**0.634**
	Q18: When I see someone being treated unfairly, I sometimes don't feel very much pity for them. (-)(EC)				**0.585**
	Q14: Other people's misfortunes do not usually disturb me a great deal. (-) (EC)	0.304	0.422		**0.513**
	Q10: I sometimes feel helpless when I am in the middle of a very emotional situation. (PD)				**-0.448**

Note. Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization. a. Rotation converged in 7 iterations. We suppressed a small coefficient (0.30).

C

C.10 PCA of the Avatar Embodiment Questionnaire

Table C.10.1: Principal component analysis of the Avatar Embodiment Questionnaire.

Factor A prior classification	**Items**	**1**	**2**	**3**	**4**	**5**	**6**
Factor 1: Response to external stimuli	Q23:When the virtual shot happened, I felt the instinct to escape (OW)	**0.786**		0.133			
	Q25:I had the feeling that I might be harmed by the virtual shot (RE)	**0.712**		0.361			
	Q22:I felt a distress sensation in my body when I saw the shooter (RE)	**0.655**			0.338		
	Q21:I felt as if my own body could be affected by the virtual shot (RE)	**0.647**					
	Q24:I felt as if my virtual body had vibration sensations (RE)	**0.624**			0.464		
	Q19:At some point it felt that the virtual body resembled my own (real) body, in terms of shape, skin tone or other visual features (AP)	**0.559**	0.403	0.428			
	Q20:I felt like I was wearing different skin from when I came to the laboratory (AP)	**0.507**	0.438				
Factor 2: Body ownership	Q2:It felt as if the virtual body I saw was someone else (-) (OW)		**0.813**				
	Q5:I felt as if the virtual body I saw when the game character looked in the mirror was another person (-) (OW)		**0.708**				-0.352
	Q4:I felt as if the virtual body I saw when looking in the mirror was my own body (OW)		**0.675**			0.382	
	Q1: I felt as if the virtual body I saw when I looked down was my body (OW)		**0.665**			0.497	
Factor 3: Tactile Sensations	Q11:It seemed as if I felt the trees when the game character touched them (TS)			**0.791**			
	Q10:It seemed as if I felt the touch of water when the game character was swimming. (TS)			**0.780**			
	Q13:It seemed as if I felt the virtual shot when I was felt the vibrations in the vest (TS)			**0.722**			
	Q18:At some point it felt as if my real body was starting to take on the posture or shape of the virtual body that I saw (AP)		**0.360**	**0.489**	0.384		0.325
	Q12:It seemed as if I felt the virtual shot when I was felt the vibrations in the vest (TS)	0.481		**0.483**	0.383		
Factor 4: Location of the body	Q8:I felt as if the movements of the virtual body were influencing my own movements (AG)			**0.371**	**0.675**		
	Q16:I felt as if my (real) body was drifting towards the virtual body or as if the virtual body was drifting towards my (real) body" (LO)		0.335		**0.655**		0.334
	Q15:I felt out of my body (LO)	0.413			**0.636**		
	Q14:I felt as if my body was located where I saw the virtual body (LO)	0.433			**0.530**		
	Q17:t felt as if my (real) body was turning into an 'avatar' body (AP)		0.447		**0.552**		
Factor 5: Agency and motor control	Q7:The movements of the virtual body were caused by my movements (AG)					**0.837**	
	Q6:It felt like I could control the virtual body as if it was my own body (AG)		0.302			**0.707**	
Factor 6: External appearance	Q9:I felt as if the virtual body was moving by itself (AG)		−0.303	0.323			**0.639**
	Q3:It seemed as if I might have more than one body (OW)						**0.835**
Note. Extraction	Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization. a. Rotation converged in 15 iterations. We suppressed a small coefficient (0.30).						

C

D Justin Beaver Stories: A conversational and empathic virtual animal in mixed reality technology

In this **Appendix D**, this work describes a framework for the creation of a conversational character in a mixed reality empathetic experience. The framework allows for the synchronization of emotional animations of the virtual character in line with the character's dialogue text, with the aim to improve the users' empathetic experience. The dialogue is driven by a Natural Language Processing (NLP) pipeline, including automatic speech recognition, chatbot, and text to speech generation micro-services. Within this framework, we present a holographic experience called "Justin Beaver Stories" using the Magic Leap one, Hololens or Nreal mixed reality goggles to project the virtual character into the user's field of vision. This can be used to evaluate the impact of bringing a beaver to the user's environment instead of bringing the user to the beaver's natural environment. Interaction occurs by humanizing the beaver through human communications abilities, resulting in a conversational virtual beaver. The storyline describes the beaver's lifestyle and problems, represented in a situation of distress. Positive experiences show the practical usability of the framework in the area of HCI.

D.1 Introduction

As animals do not communicate as humans do, people are incapable of speaking with animals, creating a language barrier between species. However, here we propose a framework that allows for the development of conversational virtual characters in a Mixed Reality (MR) environment, also allowing for animal shaped characters. The framework allows the generation and visualization of emotional responses in the virtual character to elicit empathetic reactions in humans towards the animal in an affective way (Zhao et al., 2019) using mixed reality devices. We describe the construction of a communication system for a virtual agent, which is capable of communicating, adapting, customizing, and offering a new "reality" to the users in their own environment. The effectiveness of the framework relies heavily on whether empathetic behavior can be realized or not. Empathy is defined here as the ability to share the feelings and thoughts of other people, including their psychological states of pain or distress (Batson et al., 1987b; Clark et al., 2019b). It facilitates the process of social interactions and can repress antisocial behavior

D

and aggression towards others (Stanger et al., 2012). In recent years, a new construct, "dispositional empathy with nature" (Tam, 2013), has been defined in terms of the dispositional tendency to understand and share the emotional experience of the natural world. The development of this tendency plays a crucial role in environmentalism and in the assessment of environmental educational programs.

D.2 Previous work

Recent evidence suggests that one may experience empathy not just towards humans or animals in real life, but also towards virtual agents or virtual characters (Paiva, 2011; Kano and Morita, 2019).On the one hand, studies investigate the use of virtual agents for emotional and social support for mental health of users. For instance, de Gennaro et al. (2020) investigated effects of an empathetic chatbot, which had more sensitive responses, such as "I'm sorry that this happened to you" compared to a control condition, and using more neutral responses as "Thank you for your feedback". The results show that the empathetic chatbot, compared to the control condition, elicited a more positive mood in the users. On the other hand, studies investigated the use of virtual agents in education. In a study with a virtual tutor called "Alice" (Oker et al., 2020), which uses empathetic feedback to stimulate motivation to learn, the conversational agent was manipulated on facial expressions and the textual feedback to the user on their learning performance. It showed that verbal feedback that was coherent with the agent's face resulted in higher levels of empathy towards the agent. The effect of expression of emotions in conversational character behavior with human forms has received much interest (Egges et al., 2004). It has been demonstrated that people respond more positively to agents that express emotions compared to those that do not (De Melo et al., 2014; de Gennaro et al., 2020). To improve the emotional appearance of the conversational agents in mixed and virtual reality setups, one should consider three key aspects, (Mensio et al., 2018): textual interaction, vocal interaction, and embodied agents. For textual interaction, the agent should simulate emotions related to the text in a coordinated style, allowing the interlocutor to generate meaningful responses, for instance, by extracting emotions from text. Regarding vocal interaction, the emotion should be recognized (speech-to-text) and manipulated (text-to-speech) while emphasizing human imitation of the tone modulation, and aiming to recreate human voice expressions. Embodied agents control the movements of the body to mimic emotions that are expressed in the agent's conversation. It is relevant to consider that, although these aspects have shown effective empathetic interactions in virtual human agents, it is unclear whether this has the same effect on virtual animal agents.

D.3 General Framework

We propose a framework to enable investigation in the effectiveness of emotional appearance in interactive narratives with empathetic virtual conversational characters. This allows for a wide range of interaction channels, such as voice or text to interact with the conversational characters, but also other (sensory) inputs such as video, temperature measurements, and eye tracking (Haag et al., 2004; Mirsamadi et al., 2017; Marinoiu et al., 2018). Using multiple input sources might help detect user's emotions during the experience, which in turn may improve the character's emotional response and delivery of the narrative.

The framework is designed to support interactive narratives. Interactive narratives should provide textual interaction, expressive speech synthesis, and animated embodied agents. With respect to the textual interaction, emotions can be delivered together with the text of the narrative (as illustrated in Table D.3.1). Furthermore, vocal interaction is ensured through expressive speech synthesis, which modulates the voice of the character by adjusting pitch, tone, cadence, or accent (Valle et al., 2020), adding a more natural expression. The virtual embodied character is also capable of expressing emotions using animations of the body synchronized with the narrative text.

Table D.3.1: Relation of emotions and text in the narrative.

Emotion	Sentence
Surprise	Ohhhh a beginner in this a job
Angry	The river is drying because of the humans
Neutral	Have you noticed the temperature change?
Fear	It is impossible! The forest is burning!
Sad	I want to cry! The trees are burning!
Happy	I can eat a willow tree and you?
Disgust	You are a very strange beaver

D

The framework consists of a pipeline of interchangeable conceptual modules that detect and deliver emotions as represented in Figure D.3.1. The pipeline starts with the user input acquisition, followed by the parallel execution of emotion extraction and Automatic Speech Recognition (ASR). At this point, the pipeline obtains a text tagged with emotions that serves as input to the Natural Language Processing (NLP) module. (Note that the framework supports multiple emotions per text fragment as illustrated in Figure D.3.2.) The NLP module produces a synchronized sequence of texts and corresponding emotions to send to the user. This is converted using a Text-to-Speech (TTS) module maintaining the synchronization of emotions. Finally, it delivers the voice to the user, synchronized with emotion animations with the aim to reinforce the users' experience and provide natural human-to-machine interaction. This pipeline can be understood as an extension to conversational AI pipelines solutions like Nemo (Kuchaiev et al., 2019) from NVidia or Par-LAI (Miller et al., 2017). The modular setup of the framework allows for the continuous improvement by incorporating new software pieces once they are available.

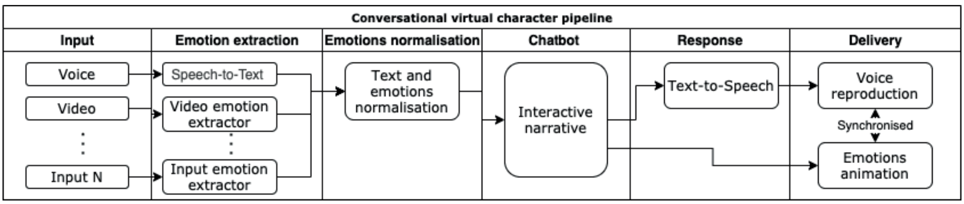


Figure D.3.1: Conversational virtual character pipeline, which is divided into five stages: source inputs, emotions extraction, emotions normalization, chatbot, response, and synchronized output. The first three handle input emotions and the last two handle output emotions.

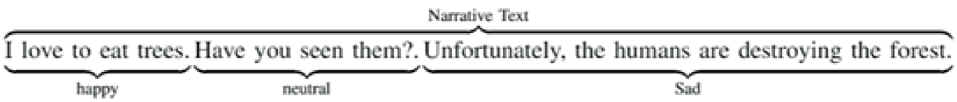


Figure D.3.2: Tagging of multiple emotions per text fragment.

D4 Case Study

While the overall framework is important for the overall emotional experience, the delivery system plays a crucial role in its effectiveness. Here, we present a case study that implements the framework which includes the delivery of emotions using an augmented reality device.

The case study fits in the larger Justin Beaver project. The main objective of this project is to educate people in the beaver's environmental problems by developing an empathetic relationship with the users. Justin is a virtual beaver that inhabits several virtual universes with different capabilities. Two versions already existed: a "regular" computer version gamifying the experiences and a virtual reality version.

In the "regular" computer version, the appearance of the beaver is manipulated in terms of emotional facial expressions and body appearance. The results show that people reacted empathetically when the animal had a biological body and displayed emotional facial expressions, and with a robot representation without facial expressions. This indicates that body shape, and facial expressions have an effect on empathy towards virtual animal characters.

Experiments in the "Justin Beaver VR" virtual reality environment bring users to the beaver's natural environment. In first person view, users perform beavers' daily activities embodied as Justin. The experience then highlights two problems: deforestation and hunting (in the form of a virtual shooting, emphasized through haptic feedback). This study shows that the appearance of the animal is important when it comes to immersion and the perception of pain (during the virtual shooting). This effect was stronger when users embodied a beaver with natural appearance, compared to conditions of more artificial appearances (e.g., robot beavers, or amorphous figures like a marshmallow). The results, however, show that in this version no empathetic response is illicited compared to the "regular" computer version.

Finally, in the "Justin Beaver MR" mixed reality environment, proposed here, we allow users to interact with Justin Beaver in its natural environment as a conversational virtual character (Smid and Pandzic, 2002). Through interactive story-telling, Justin provides information on the nutritional behavior of beavers and also talks about wildfires due to global warming. This allows, for instance, experimentation on the effects of two aspects: emotional expression (of the character), and language interaction. Emotional expressions are implemented as animations with both facial and bodily movements. These animations are synchronized with the narrative of Justin. The language interaction relates to the communication pipeline that allows users to communicate with Justin via speech, using Natural Language Processing (NLP) micro-services. As a whole, these capabilities are used to induce empathetic reactions in the users, with the aim to expand the results of the Justin "regular" computer and VR experiences.

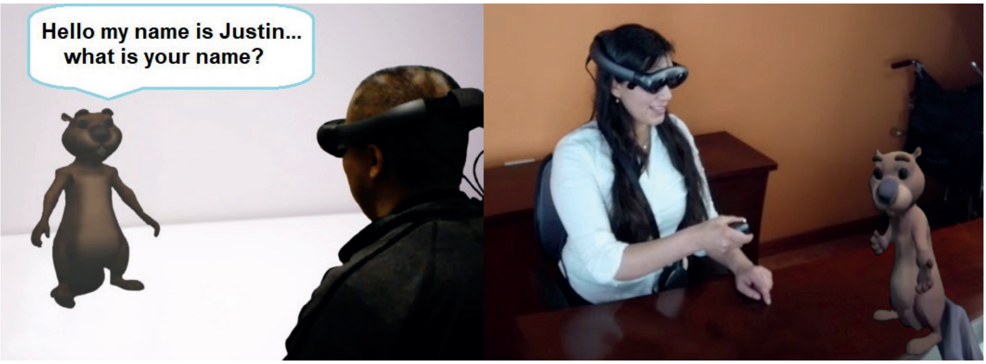


Figure D.4.1: Virtual animal character simulation with Magic Leap.

D4.1 Implementation

"Justin Beaver Stories" follows the pipeline using only voice as conversational input. Functionality is implemented using IBM Watson's micro-services, while including automatic speech recognition (Anusuya and Katti, 2010), and generating input to its chatbot (Setiaji and Wibowo, 2016), which simultaneously generates text for its text-to-speech system (Allen, 1976). The output is represented as a three dimensional hologram, placing the character in the user's real world. The hologram is projected using the Magic Leap mixed reality device (Bradski et al., 2019) utilizing its development framework. This device is capable of combining a virtual layer over the user's reality and providing real-time interaction, taking into account the environment (registered in three dimensions through motion detection and 3D reconstruction) (Molyneaux et al., 2019). The character's dialog delivery is enriched with the synchronization between emotional expressions and the generated speech (see Figure D.4.1).

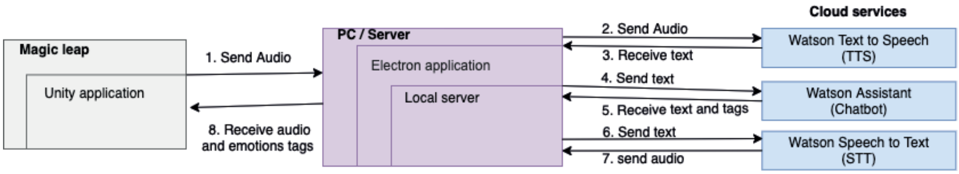


Figure D.4.2: Interaction handling process.

Practically, the experience is implemented using three tiers (see Figure D.4.2). This first tier is implemented using a unity application that runs on the Magic Leap one device, which presents a virtual layer and captures user's voice interactions. The second tier is an Electron web server, which receives the interactions and passes them through the third tier that relies on IBM Watson's natural language micro-services: Natural Language Understanding, Dialog Management, and Natural Language Generation modules (Park and Jeong, 2019). The results are sent back to the virtual agent. Finally, the agent outputs the audio combined with emotion animations.

Note that the animations of the face and the body of the animal (projecting the emotional content of the message) are synchronized in time with the dialogue. Table D.3.1 illustrates how the emotions are annotated in the dialog text. Animations corresponding to the emotions synchronize the movement of the virtual character's eyebrows and different body parts. The transition between emotions is performed using an automatically interpolated transition between the different animations. A natural transition between emotions is pivotal for a more empathetic experience.

D4.2 Visualization

In order to place the character in a virtual scene within the user's environment, we use the native 3D data framework from Magic Leap. The device recognizes the environment and builds a forest taking the physical constrains into account. The virtual world scales with respect to two constraints: the minimum size possible to facilitate control during the experience, and the maximum size of the hologram to fairly visualize the emotional body and facial expression animations.

The experience has been tested on five different scales (where scale indicates height, width, and depth in cm of the virtual world): 30, 50, 70, 100, and 150. After informal testing using laboratory members, an arbitrary selection of 70cm for the scale was made (as illustrated in Figure D.4.1).

D

D4.3 Communication capabilities

The virtual character is capable of communicating, by using the multitier architecture described earlier, and relying on several NLP microservices combined with an experience narrative that guides the user through an educational story on the beaver's life. Due to the interchangeability of the NLP modules, Justin can speak English and Spanish and can be easily ported to other languages by adding the relevant narrative.

As the aim of the narrative is to try to build a relationship between the beaver and the user to enable sharing information on the beaver's environment and dietary description. Each step in the narrative is composed of text (in the form of voice) and corresponding emotiondriven animations. Internally, the text is tagged with emotions (happy, sad, angry, disgusted, fear, surprised, and neutral) which are converted into corresponding animations.

D5 Award Recognition as the Best Research in VR/MR

As is illustrated in Figure D.5.1, the demonstration of this project had the award recognition of the best revolution research in immersive technology (2021) in Laval Virtual, France.

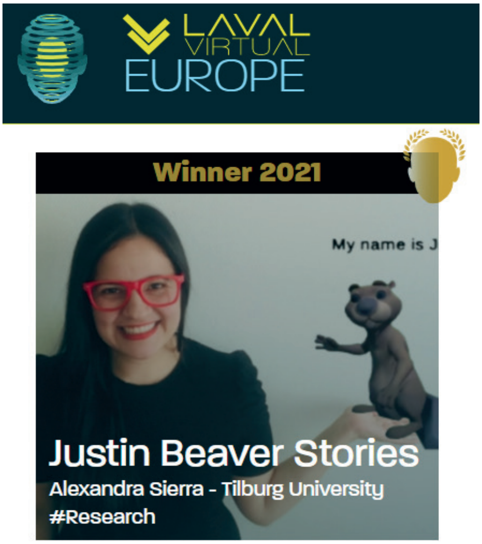


Figure D.5.1: Award recognition Laval Virtual.

D6 Conclusions and Future Work

We introduced a flexible framework that allows the development of interactive storytelling applications with a conversational character (in animal form) which incorporates emotional synchronization of visual and language channels. We have shown the effectiveness of the framework through a case study, "Justin Beaver Stories", which extends earlier computer and virtual reality versions. This allows for research on the impact of both visual and emotional choices on the immersion and emotional impact. For future work, we will focus on the effects of manipulations of different communication channels, (e.g., to investigate the effect of aligned versus misaligned, or presence versus absence of emotional animation related to the text). As the framework allows experiments in a mixed reality, we can also compare these results to the computer and virtual reality experiments.

We can improve virtual animals' performance using artificial intelligence. Artificial Intelligence could help this virtual animal be seen as a smart virtual agent capable of vocal interaction and creating interactive stories. Further research could be beneficial in exploring how such a virtual animal might help foster 21st-century learning skills in education through its active interactions with learners and empathetic reactions toward nature.

E Additional Statistics Chapter 3

Table E.0.1: Median, minimum, and maximum values for conditions of Participant-based ranking.

	Participant-based ranking																	
	Robot			Mechanical			Zombie			Stuffed			Photorealistic			Real		
	Median	Minimum	Maximum	Median	Minimum	Maximum	Median	Minimum	Maximum	Median	Minimum	Maximum	Median	Minimum	Maximum	Median	Minimum	Maximum
Familiarity still	3	1	9	3	1	9	2	1	9	6	1	9	8	1	9	9	1	9
Commonality still	3	1	8	3	1	9	2	1	8	5	1	9	7	1	9	8	1	9
Attractiveness still	4	1	9	5	1	9	3	1	9	5	1	9	6	1	9	8	1	9
Interestingness still	6	1	9	6	1	9	6	1	9	5,5	1	9	6	1	9	7	1	9
Naturalness still	2	1	9	2	1	9	2	1	9	5	1	9	7	1	9	9	1	9
Animateness still	5	1	9	5,5	1	9	5	1	9	5	1	9	5	1	9	6	1	9
Familiarity moving	3	1	9	3	1	9	2	1	9	5	1	9	6	1	9	7	1	9
Commonality moving	3	1	9	3	1	9	2	1	9	4	1	9	6	1	9	7	1	9
Attractiveness moving	5	1	9	5	1	9	3	1	9	6	1	9	6	1	9	7	1	9
Interestingness moving	5	1	9	6	1	9	5	1	9	6	1	9	6	1	9	7	1	9
Naturalness moving	2	1	9	2	1	9	2	1	9	4	1	9	6,5	1	9	7	1	9
Animateness moving	6	1	9	6	1	9	6	1	9	6	1	9	6	1	9	6	1	9

Table E.0.2: Median, minimum, and maximum values for conditions of Expert-based ranking.

	Expert-based ranking																	
	Mechanical			Stuffed			Robot			Robot			Photorealistic			Real		
	Median	Minimum	Maximum	Median	Minimum	Maximum	Median	Minimum	Maximum	Median	Minimum	Maximum	Median	Minimum	Maximum	Median	Minimum	Maximum
Familiarity still	3	1	9	6	1	9	3	1	9	2	1	9	8	1	9	9	1	9
Commonality still	3	1	9	5	1	9	3	1	8	2	1	8	7	1	9	8	1	9
Attractiveness still	5	1	9	5	1	9	4	1	9	3	1	9	6	1	9	8	1	9
Interestingness still	6	1	9	5,5	1	9	6	1	9	6	1	9	6	1	9	7	1	9
Naturalness still	2	1	9	5	1	9	2	1	9	2	1	9	7	1	9	9	1	9
Animateness still	5,5	1	9	5	1	9	5	1	9	5	1	9	5	1	9	6	1	9
Familiarity moving	3	1	9	5	1	9	3	1	9	2	1	9	6	1	9	7	1	9
Commonality moving	3	1	9	4	1	9	3	1	9	2	1	9	6	1	9	7	1	9
Attractiveness moving	5	1	9	6	1	9	5	1	9	3	1	9	6	1	9	7	1	9
Interestingness moving	6	1	9	6	1	9	5	1	9	5	1	9	6	1	9	7	1	9
Naturalness moving	2	1	9	4	1	9	2	1	9	2	1	9	6,5	1	9	7	1	9
Animateness moving	6	1	9	6	1	9	6	1	9	6	1	9	6	1	9	6	1	9

Table E.0.3: Kruskal-Wallis H for Participants-based ranking (Virtual Character appearance).

Variable	Kruskal-Wallis H	df	Asymp. Sig.
Familiarity still	412.310	5	<,001
Commonality still	434.438	5	<,001
Attractiveness still	210.907	5	<,001
Interestingness still	30.473	5	<,001
Naturalness still	419.731	5	<,001
Animateness still	3.362	5	.644
Familiarity moving	227.446	5	<,001
Commonality moving	222.322	5	<,001
Attractiveness moving	163.988	5	<,001
Interestingness moving	21.448	5	<,001
Naturalness moving	275.973	5	<,001
Animateness moving	2.316	5	.804

Table E.0.4: Test of Normality.

Tests of Normality						
	Kolmogorov-Smirnovn[a]			Shapiro-Wilk Statistic		
	Statistic	df	Sig	Statistic	df	Sig
Familiarity still	0,139	984	0,000	0,889	984	0,000
Commonality still	0,129	984	0,000	0,903	984	0,000
Attractiveness still	0,103	984	0,000	0,940	984	0,000
Interestingness still	0,129	984	0,000	0,937	984	0,000
Naturalness still	0,158	984	0,000	0,869	984	0,000
Animateness still	0,122	984	0,000	0,907	984	0,000
Familiarity moving	0,128	984	0,000	0,919	984	0,000
Commonality moving	0,149	984	0,000	0,928	984	0,000
Attractiveness moving	0,112	984	0,000	0,947	984	0,000
Interestingness moving	0,135	984	0,000	0,947	984	0,000
Naturalness moving	0,143	984	0,000	0,909	984	0,000
Animateness moving	0,120	984	0,000	0,926	984	0,000 n

a.Lilliefors Significance Correction.

E

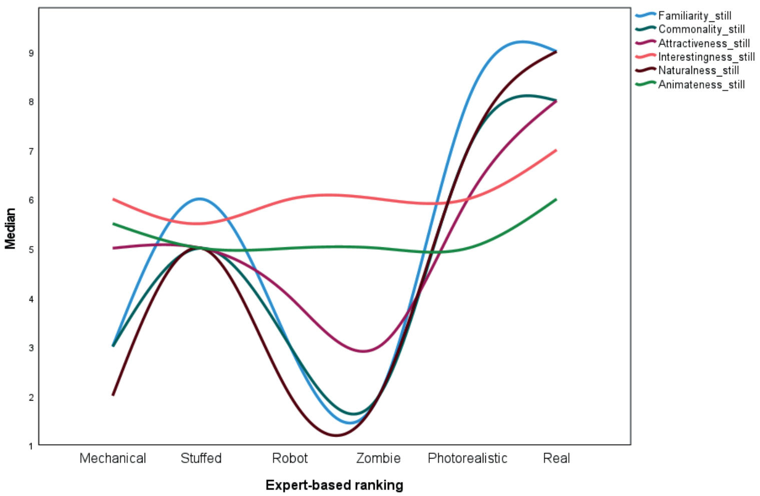


Figure E.0.1: Graphs of the perceived familiarity, commonality, naturalness, attractiveness, interestingness, and animateness for the six still stimuli organized on the x-axis according to expert-based ranking.

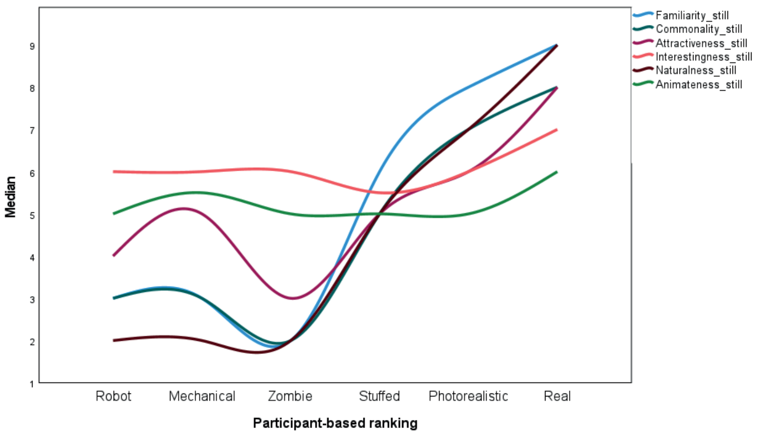


Figure E.0.2: Graphs of the perceived familiarity, commonality, naturalness, attractiveness, interestingness, and animateness for the six still stimuli organized on the x-axis according to participant-based ranking.

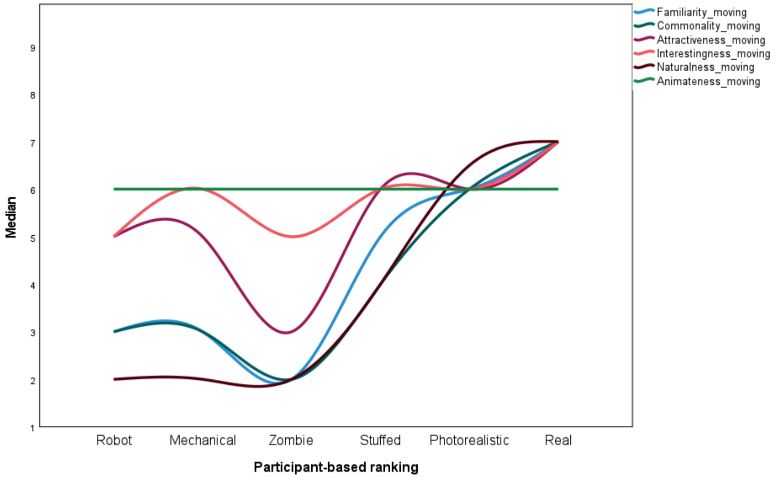


Figure E.0.3: Graphs of the perceived familiarity, commonality, naturalness, attractiveness, interestingness, and animateness for the six moving stimuli organized on the x-axis according to participant-based ranking.

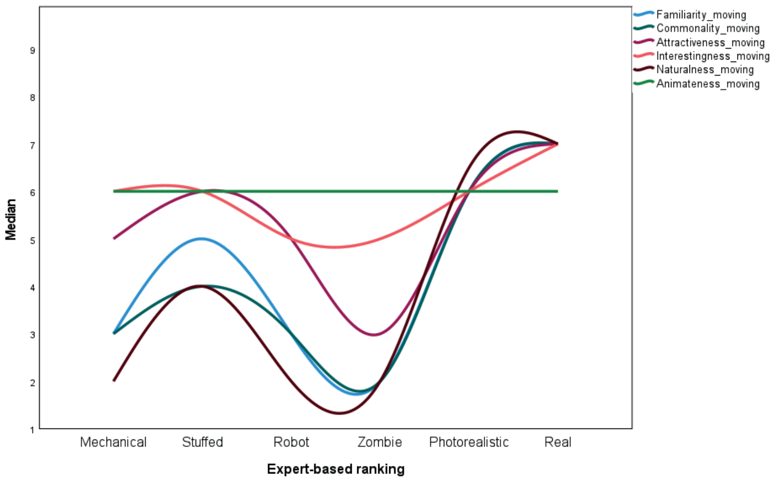


Figure E.0.4: Graphs of the perceived familiarity, commonality, naturalness, attractiveness, interestingness, and animateness for the six moving stimuli organized on the x-axis according to expert-based ranking.

F Additional Statistics Chapter 6

Table F.0.1: Median, minimum, and maximum values for conditions of the Appearance of the Instructor.

	Appearance of the Instructor								
	Virtual robot panda			Virtual Pand			Virtual Teacher		
	Median	Minimum	Maximum	Median	Minimum	Maximum	Median	Minimum	Maximum
Domain knowledge (pre-test)	2,0	0,0	4,0	1,5	0,0	3,0	1,0	0,0	4,0
Knowledge recall (post-test)	4,0	2,0	5,0	4,0	1,0	5,0	4,0	3,0	5,0
Concentration	4,0	2,0	5,0	4,0	2,0	5,0	4,0	2,0	5,0
Social presence	3,5	1,8	5,0	3,6	1,6	5,0	3,8	2,0	5,0
Interpersonal attraction	3,8	1,3	5,0	3,9	2,0	5,0	4,0	2,0	5,0
Presentation skills	4,3	2,7	5,0	4,3	2,0	5,0	4,7	3,0	5,0
Familiarity	3,0	1,0	5,0	4,0	1,0	5,0	4,0	1,0	5,0
Commonality	3,0	1,0	5,0	3,0	1,0	5,0	3,0	1,0	5,0
Attractiveness	3,0	1,0	5,0	3,0	1,0	5,0	4,0	2,0	5,0
Interestigness	4,0	2,0	5,0	4,0	1,0	5,0	5,0	2,0	5,0
Naturaless	3,0	1,0	5,0	3,5	1,0	5,0	4,0	2,0	5,0
Animateness	4,0	1,0	5,0	4,0	1,0	5,0	5,0	1,0	5,0

Table F.0.2: Median, minimum, and maximum values for conditions of the Topic.

	Topic					
	Netherlands			Robotic		
	Median	Minimum	Maximum	Median	Minimum	Maximum
Domain knowledge (pre-test)	1,0	0,0	4,0	2,0	0,0	4,0
Knowledge recall (post-test)	4,0	2,0	5,0	4,0	1,0	5,0
Concentration	4,0	2,0	5,0	4,0	2,0	5,0
Social presence	3,7	1,6	5,0	3,6	2,0	5,0
Interpersonal attraction	4,0	1,3	5,0	3,8	1,5	5,0
Presentation skills	4,7	2,7	5,0	4,3	2,0	5,0
Familiarity	3,0	1,0	5,0	4,0	1,0	5,0
Commonality	3,0	1,0	5,0	3,0	1,0	5,0
Attractiveness	4,0	1,0	5,0	4,0	1,0	5,0
Interestigness	4,0	1,0	5,0	4,0	2,0	5,0
Naturaless	4,0	1,0	5,0	4,0	1,0	5,0
Animateness	4,0	1,0	5,0	4,0	2,0	5,0

Table F.0.3: Test of Normality.

Tests of Normality						
	Kolmogorov-Smirnovn[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Domain knowledge (pre-test)	0,258	130	0,000	0,857	130	0,000
Knowledge recall (post-test)	0,273	130	0,000	0,834	130	0,000
Concentration	0,311	130	0,000	0,787	130	0,000
Social presence	0,111	130	0,000	0,970	130	0,006
Interpersonal attraction	0,129	130	0,000	0,951	130	0,000
Presentation skills	0,199	130	0,000	0,855	130	0,000
Familiarity	0,207	130	0,000	0,884	130	0,000
Commonality	0,210	130	0,000	0,902	130	0,000
Attractiveness	0,205	130	0,000	0,886	130	0,000
Interestigness	0,242	130	0,000	0,804	130	0,000
Naturaless	0,198	130	0,000	0,871	130	0,000
Animateness	0,259	130	0,000	0,820	130	0,000

a. Lilliefors Significance Correction

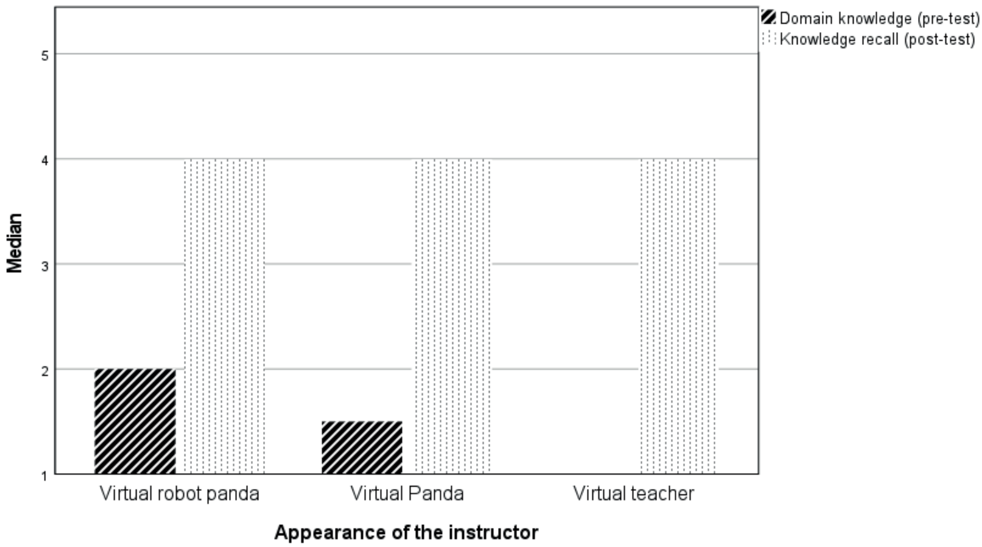


Figure F.0.1: Median Participants' domain knowledge (pre-test) and knowledge recall (post-test) in relation to the appearance of the instructors.

Table F.0.4: Kruskal-Wallis H for Appearance of the Instructor.

Test Statisticsn[a],n[b]			
	Kruskal-Wallis H	**df**	**Asymp. Sig.**
Domain knowledge (pre-test)	0,006	2	0,997
Knowledge recall (post-test)	6,533	2	0,038
Concentration	0,382	2	0,826
Social presence	3,830	2	0,147
Interpersonal attraction	5,506	2	0,064
Presentation skills	2,011	2	0,366
Familiarity	9,109	2	0,011
Commonality	2,190	2	0,335
Attractiveness	8,576	2	0,014
Interestigness	6,322	2	0,042
Naturaless	20,478	2	0,000
Animateness	6,942	2	0,031

a. Kruskal Wallis Test

b. Grouping Variable: Appearance of the instructor

F

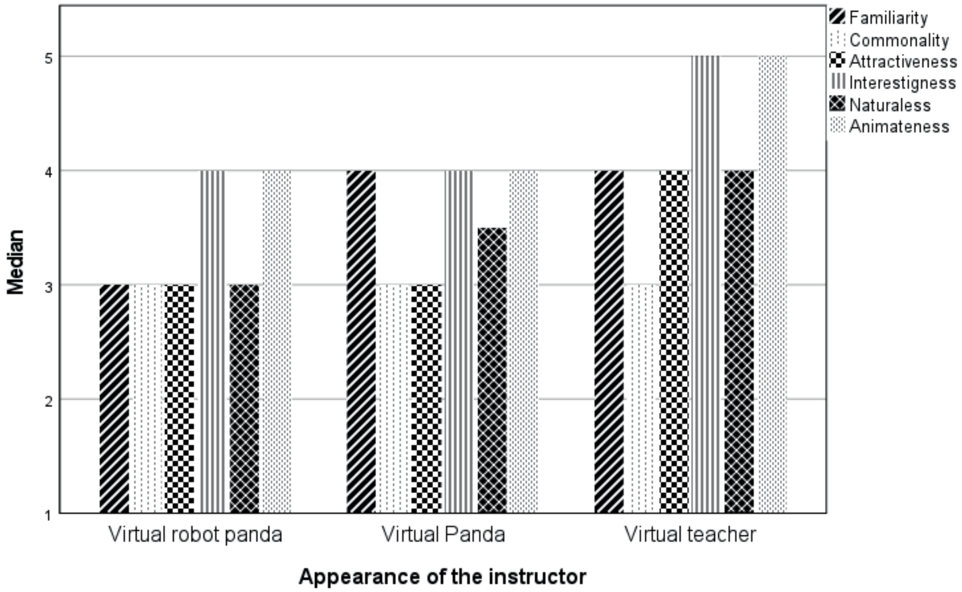


Figure F.0.2: Median scores of participants' perception of familiarity, commonality, attractiveness, interestingness, naturalness, and animateness in relation to the appearance of the instructor.

Table F.0.5: Kruskal-Wallis H for Topic

Test Statisticsn[a],n[b]	Kruskal-Wallis H	df	Asymp. Sig.
Domain knowledge (pre-test)	0,002	1	0,964
Knowledge recall (post-test)	7,243	1	0,007
Concentration	0,079	1	0,779
Social presence	0,060	1	0,807
Interpersonal attraction	1,088	1	0,297
Presentation skills	0,729	1	0,393
Familiarity	2,469	1	0,116
Commonality	0,324	1	0,569
Attractiveness	0,183	1	0,669
Interestigness	0,018	1	0,893
Naturaless	0,413	1	0,521
Animateness	1,574	1	0,210

a. Kruskal Wallis Test

b. Grouping Variable: Topic

G SIKS Dissertatiereeks

==== 2016 ====

2016-01 Syed Saiden Abbas (RUN) Recognition of Shapes by Humans and Machines

2016-02 Michiel Christiaan Meulendijk (UU) Optimizing medication reviews through decision support: prescribing a better pill to swallow

2016-03 Maya Sappelli (RUN) Knowledge Work in Context: User Centered Knowledge Worker Support

2016-04 Laurens Rietveld (VU) Publishing and Consuming Linked Data

2016-05 Evgeny Sherkhonov (UVA) Expanded Acyclic Queries: Containment and an Application in Explaining Missing Answers

2016-06 Michel Wilson (TUD) Robust scheduling in an uncertain environment

2016-07 Jeroen de Man (VU) Measuring and modeling negative emotions for virtual training

2016-08 Matje van de Camp (TiU) A Link to the Past: Constructing Historical Social Networks from Unstructured Data

2016-09 Archana Nottamkandath (VU) Trusting Crowdsourced Information on Cultural Artefacts

2016-10 George Karafotias (VUA) Parameter Control for Evolutionary Algorithms

2016-11 Anne Schuth (UVA) Search Engines that Learn from Their Users

2016-12 Max Knobbout (UU) Logics for Modelling and Verifying Normative Multi-Agent Systems

2016-13 Nana Baah Gyan (VU) The Web, Speech Technologies and Rural Development in West Africa An ICT4D Approach

2016-14 Ravi Khadka (UU) Revisiting Legacy Software System Modernization

2016-15 Steffen Michels (RUN) Hybrid Probabilistic Logics Theoretical Aspects, Algorithms and Experiments

2016-16 Guangliang Li (UVA) Socially Intelligent Autonomous Agents that Learn from Human Reward

2016-17 Berend Weel (VU) Towards Embodied Evolution of Robot Organisms

2016-18 Albert Meroo Peuela (VU) Refining Statistical Data on the Web

2016-19 Julia Efremova (Tu/e) Mining Social Structures from Genealogical Data

2016-20 Daan Odijk (UVA) Context Semantics in News Web Search 2016-21 Alejandro Moreno Clleri (UT) From Traditional to Interactive Playspaces: Automatic Analysis of Player Behavior in the Interactive Tag Playground

2016-22 Grace Lewis (VU) Software Architecture Strategies for Cyber-Foraging Systems
2016-23 Fei Cai (UVA) Query Auto Completion in Information Retrieval
2016-24 Brend Wanders (UT) Repurposing and Probabilistic Integration of Data; An Iterative and data model independent approach
2016-25 Julia Kiseleva (TU/e) Using Contextual Information to Understand Searching and Browsing Behavior
2016-26 Dilhan Thilakarathne (VU) In or Out of Control: Exploring Computational Models to Study the Role of Human Awareness and Control in Behavioural Choices, with Applications in Aviation and Energy Management Domains
2016-27 Wen Li (TUD) Understanding Geo-spatial Information on Social Media
2016-28 Mingxin Zhang (TUD) Large-scale Agent-based Social Simulation A study on epidemic prediction and control
2016-29 Nicolas Hning (TUD) Peak reduction in decentralised electricity systems -Markets and prices for flexible planning
2016-30 Ruud Mattheij (UvT) The Eyes Have It
2016-31 Mohammad Khelghati (UT) Deep web content monitoring 2016-32 Eelco Vriezekolk (UT) Assessing Telecommunication Service Availability Risks for Crisis Organisations
2016-33 Peter Bloem (UVA) Single Sample Statistics, exercises in learning from just one example
2016-34 Dennis Schunselaar (TUE) Configurable Process Trees: Elicitation, Analysis, and Enactment
2016-35 Zhaochun Ren (UVA) Monitoring Social Media: Summarization, Classification and Recommendation
2016-36 Daphne Karreman (UT) Beyond R2D2: The design of nonverbal interaction behavior optimized for robot-specific morphologies
2016-37 Giovanni Sileno (UvA) Aligning Law and Action a concep- tual and computational inquiry
2016-38 Andrea Minuto (UT) MATERIALS THAT MATTER - Smart Materials meet Art Interaction Design
2016-39 Merijn Bruijnes (UT) Believable Suspect Agents; Response and Interpersonal Style Selection for an Artificial Suspect
2016-40 Christian Detweiler (TUD) Accounting for Values in Design
2016-41 Thomas King (TUD) Governing Governance: A Formal Framework for Analysing Institutional Design and Enactment Governance

2016-42 Spyros Martzoukos (UVA) Combinatorial and Compositional Aspects of Bilingual Aligned Corpora

2016-43 Saskia Koldijk (RUN) Context-Aware Support for Stress Self-Management: From Theory to Practice

2016-44 Thibault Sellam (UVA) Automatic Assistants for Database Exploration

2016-45 Bram van de Laar (UT) Experiencing Brain-Computer In- terface Control

2016-46 Jorge Gallego Perez (UT) Robots to Make you Happy

2016-47 Christina Weber (UL) Real-time foresight - Preparedness for dynamic innovation networks

2016-48 Tanja Buttler (TUD) Collecting Lessons Learned

2016-49 Gleb Polevoy (TUD) Participation and Interaction in Projects. A Game-Theoretic Analysis

2016-50 Yan Wang (UVT) The Bridge of Dreams: Towards a Method for Operational Performance Alignment in IT-enabled Service Supply Chains

==== 2017 ====

2017-01 Jan-Jaap Oerlemans (UL) Investigating Cybercrime

2017-02 Sjoerd Timmer (UU) Designing and Understanding Forensic Bayesian Networks using Argumentation

2017-03 Danil Harold Telgen (UU) Grid Manufacturing; A Cyber- Physical Approach with Autonomous Products and Reconfigurable Manufacturing Machines

2017-04 Mrunal Gawade (CWI) MULTI-CORE PARALLELISM IN A COLUMN-STORE

2017-05 Mahdieh Shadi (UVA) Collaboration Behavior

2017-06 Damir Vandic (EUR) Intelligent Information Systems for Web Product Search

2017-07 Roel Bertens (UU) Insight in Information: from Abstract to Anomaly

2017-08 Rob Konijn (VU) Detecting Interesting Differences:Data Mining in Health Insurance Data using Outlier Detection and Subgroup Discovery

2017-09 Dong Nguyen (UT) Text as Social and Cultural Data: A Computational Perspective on Variation in Text

2017-10 Robby van Delden (UT) (Steering) Interactive Play Behavior 2017-11 Florian Kunneman (RUN) Modelling patterns of time and emotion in Twitter anticipointment

2017-12 Sander Leemans (TUE) Robust Process Mining with Guarantees

G

2017-13 Gijs Huisman (UT) Social Touch Technology - Extending the reach of social touch through haptic technology

2017-14 Shoshannah Tekofsky (UvT) You Are Who You Play You Are: Modelling Player Traits from Video Game Behavior

2017-15 Peter Berck, Radboud University (RUN) Memory-Based Text Correction

2017-16 Aleksandr Chuklin (UVA) Understanding and Modeling Users of Modern Search Engines

2017-17 Daniel Dimov (UL) Crowdsourced Online Dispute Resolution

2017-18 Ridho Reinanda (UVA) Entity Associations for Search 2017-19 Jeroen Vuurens (TUD) Proximity of Terms, Texts and Semantic Vectors in Information Retrieval

2017-20 Mohammadbashir Sedighi (TUD) Fostering Engagement in Knowledge Sharing: The Role of Perceived Benefits, Costs and Visibil- ity

2017-21 Jeroen Linssen (UT) Meta Matters in Interactive Story- telling and Serious Gaming (A Play on Worlds)

2017-22 Sara Magliacane (VU) Logics for causal inference under uncertainty

2017-23 David Graus (UVA) Entities of Interest— Discovery in Digital Traces

2017-24 Chang Wang (TUD) Use of Affordances for Efficient Robot Learning

2017-25 Veruska Zamborlini (VU) Knowledge Representation for Clinical Guidelines, with applications to Multimorbidity Analysis and Literature Search

2017-26 Merel Jung (UT) Socially intelligent robots that understand and respond to human touch

2017-27 Michiel Joosse (UT) Investigating Positioning and Gaze Behaviors of Social Robots: People's Preferences, Perceptions and Behav- iors

2017-28 John Klein (VU) Architecture Practices for Complex Contexts

2017-29 Adel Alhuraibi (UVT) From IT-BusinessStrategic Alignment to Performance: A Moderated Mediation Model of Social Innovation, and Enterprise Governance of IT

2017-30 Wilma Latuny (UVT) The Power of Facial Expressions 2017-31 Ben Ruijl (UL) Advances in computational methods for QFT calculations

2017-32 Thaer Samar (RUN) Access to and Retrievability of Content in Web Archives

2017-33 Brigit van Loggem (OU) Towards a Design Rationale for Software Documentation: A Model of Computer-Mediated Activity

2017-34 Maren Scheffel (OUN) The Evaluation Framework for Learning Analytics

2017-35 Martine de Vos (VU) Interpreting natural science spreadsheets

2017-36 Yuanhao Guo (UL) Shape Analysis for Phenotype Characterisation from High-throughput Imaging
2017-37 Alejandro Montes Garca (TUE) WiBAF: A Within Browser Adaptation Framework that Enables Control over Privacy
2017-38 Alex Kayal (TUD) Normative Social Applications
2017-39 Sara Ahmadi (RUN) Exploiting properties of the human auditory system and compressive sensing methods to increase noise robustness in ASR
2017-40 Altaf Hussain Abro (VUA) Steer your Mind: Computational Exploration of Human Control in Relation to Emotions, Desires and Social Support For applications in human-aware support systems"
2017-41 Adnan Manzoor (VUA) Minding a Healthy Lifestyle: An Exploration of Mental Processes and a Smart Environment to Provide Support for a Healthy Lifestyle
2017-42 Elena Sokolova (RUN) Causal discovery from mixed and missing data with applications on ADHD datasets
2017-43 Maaike de Boer (RUN) Semantic Mapping in Video Retrieval 2017-44 Garm Lucassen (UU) Understanding User Stories - Computational Linguistics in Agile Requirements Engineering
2017-45 Bas Testerink (UU) Decentralized Runtime Norm Enforcement
2017-46 Jan Schneider (OU) Sensor-based Learning Support
2017-47 Yie Yang (TUD) Crowd Knowledge Creation Acceleration 2017-48 Angel Suarez (OU) Collaborative inquiry-based learning

==== 2018 ====

2018-01 Han van der Aa (VUA) Comparing and Aligning Process Representations
2018-02 Felix Mannhardt (TUE) Multiperspective Process Mining 2018-03 Steven Bosems (UT) Causal Models For Well-Being: Knowledge Modeling, Model-Driven Development of Context-Aware Applications, and Behavior Prediction
2018-04 Jordan Janeiro (TUD) Flexible Coordination Support for Diagnosis Teams in Data-Centric Engineering Tasks
2018-05 Hugo Huurdeman (UVA) Supporting the Complex Dynamics of the Information Seeking Process
2018-06 Dan Ionita (UT) Model-Driven Information Security Risk Assessment of Socio-Technical Systems
2018-07 Jieting Luo (UU) A formal account of opportunism in multiagent systems

G

2018-08 Rick Smetsers (RUN) Advances in Model Learning for Software Systems

2018-09 Xu Xie (TUD) Data Assimilation in Discrete Event Simulations

2018-10 Julienka Mollee (VUA) Moving forward: supporting physical activity behavior change through intelligent technology

2018-11 Mahdi Sargolzaei (UVA) Enabling Framework for Service- oriented Collaborative Networks

2018-12 Xixi Lu (TUE) Using behavioral context in process mining 2018-13 Seyed Amin Tabatabaei (VUA) Using behavioral context in process mining: Exploring the added value of computational models for increasing the use of renewable energy in the residential sector 2018-14 Bart Joosten (UVT) Detecting Social Signals with Spatiotemporal Gabor Filters

2018-15 Naser Davarzani (UM) Biomarker discovery in heart failure 2018-16 Jaebok Kim (UT) Automatic recognition of engagement and emotion in a group of children

2018-17 Jianpeng Zhang (TUE) On Graph Sample Clustering 2018-18 Henriette Nakad (UL) De Notaris en Private Rechtspraak

2018-19 Minh Duc Pham (VUA) Emergent relational schemas for RDF

2018-20 Manxia Liu (RUN) Time and Bayesian Networks

2018-21 Aad Slootmaker (OUN) EMERGO: a generic platform for authoring and playing scenario-based serious games

2018-22 Eric Fernandes de Mello Arajo (VUA) Contagious: Modeling the Spread of Behaviours, Perceptions and Emotions in Social Networks 2018-23 Kim Schouten (EUR) Semantics-driven Aspect-Based Sentiment Analysis

2018-24 Jered Vroon (UT) Responsive Social Positioning Behaviour for Semi-Autonomous Telepresence Robots

2018-25 Riste Gligorov (VUA) Serious Games in Audio-Visual Collections

2018-26 Roelof de Vries (UT) Theory-Based And Tailor-Made: Motivational Messages for Behavior Change Technology

2018-27 Maikel Leemans (TUE) Hierarchical Process Mining for Scalable Software Analysis

2018-28 Christian Willemse (UT) Social Touch Technologies: How they feel and how they make you feel

2018-29 Yu Gu (UVT) Emotion Recognition from Mandarin Speech 2018-30 Wouter Beek (VU) The ”K” in ”semantic web” stands for ”knowledge”: scaling semantics to the web

==== 2019 ====

2019-01 Rob van Eijk (UL) Web privacy measurement in real-time bidding systems. A graph-based approach to RTB system classification 2019-02 Emmanuelle Beauxis- Aussalet (CWI, UU) Statistics and Visualizations for Assessing Class Size Uncertainty

2019-03 Eduardo Gonzalez Lopez de Murillas (TUE) Process Mining on Databases: Extracting Event Data from Real Life Data Sources

2019-04 Ridho Rahmadi (RUN) Finding stable causal structures from clinical data

2019-05 Sebastiaan van Zelst (TUE) Process Mining with Streaming Data

2019-06 Chris Dijkshoorn (VU) Nichesourcing for Improving Access to Linked Cultural Heritage Datasets

2019-07 Soude Fazeli (TUD) Recommender Systems in Social Learning Platforms

2019-08 Frits de Nijs (TUD) Resource-constrained Multiagent Markov Decision Processes

2019-09 Fahimeh Alizadeh Moghaddam (UVA) Self-adaptation for energy efficiency in software systems

2019-10 Qing Chuan Ye (EUR) Multi-objective Optimization Methods for Allocation and Prediction

2019-11 Yue Zhao (TUD) Learning Analytics Technology to Understand Learner Behavioral Engagement in MOOCs

2019-12 Jacqueline Heinerman (VU) Better Together

2019-13 Guanliang Chen (TUD) MOOC Analytics: Learner Modeling and Content Generation

2019-14 Daniel Davis (TUD) Large-Scale Learning Analytics: Modeling Learner Behavior Improving Learning Outcomes in Massive Open Online Courses

2019-15 Erwin Walraven (TUD) Planning under Uncertainty in Constrained and Partially Observable Environments

2019-16 Guangming Li (TUE) Process Mining based on Object- Centric Behavioral Constraint (OCBC) Models

2019-17 Ali Hurriyetoglu (RUN) Extracting actionable information from microtexts

2019-18 Gerard Wagenaar (UU) Artefacts in Agile Team Communication

2019-19 Vincent Koeman (TUD) Tools for Developing Cognitive Agents 2019-20 Chide Groenouwe (UU) Fostering technically augmented human collective intelligence

2019-21 Cong Liu (TUE) Software Data Analytics: Architectural Model Discovery and Design Pattern Detection

2019-22 Martin van den Berg (VU) Improving IT Decisions with Enterprise Architecture

2019-23 Qin Liu (TUD) Intelligent Control Systems: Learning, Interpreting, Verification

2019-24 Anca Dumitrache (VU) Truth in Disagreement- Crowdsourcing Labeled Data for Natural Language Processing

2019-25 Emiel van Miltenburg (UVT) Pragmatic factors in (automatic) image description

2019-26 Prince Singh (UT) An Integration Platform for Synchro- modal Transport

2019-27 Alessandra Antonaci (OUN) The Gamification Design Process applied to (Massive) Open Online Courses

2019-28 Esther Kuindersma (UL) Cleared for take-off: Game-based learning to prepare airline pilots for critical situations

2019-29 Daniel Formolo (VU) Using virtual agents for simulation and training of social skills in safety-critical circumstances

2019-30 Vahid Yazdanpanah (UT) Multiagent Industrial Symbiosis Systems

2019-31 Milan Jelisavcic (VUA) Alive and Kicking: Baby Steps in Robotics

2019-32 Chiara Sironi (UM) Monte-Carlo Tree Search for Artificial General Intelligence in Games

2019-33 Anil Yaman (TUE) Evolution of Biologically Inspired Learning in Artificial Neural Networks

2019-34 Negar Ahmadi (TUE) EEG Microstate and Functional Brain Network Features for Classification of Epilepsy and PNES

2019-35 Lisa Facey-Shaw (OUN) Gamification with digital badges in learning programming

2019-36 Kevin Ackermans (OUN) Designing Video-Enhanced Rubrics to Master Complex Skills

2019-37 Jian Fang (TUD) Database Acceleration on FPGAs

2019-38 Akos Kadar (OUN) Learning visually grounded and multilingual representations

==== 2020 ====

2020-01 Armon Toubman (UL) Calculated Moves: Generating Air Combat Behaviour

2020-02 Marcos de Paula Bueno (UL) Unraveling Temporal Processes using Probabilistic Graphical Models

2020-03 Mostafa Deghani (UvA) Learning with Imperfect Supervision for Language Understanding

2020-04 Maarten van Gompel (RUN) Context as Linguistic Bridges 2020-05 Yulong Pei (TUE) On local and global structure mining 2020-06 Preethu Rose Anish (UT) Stimulation Architectural Thinking during Requirements Elicitation - An Approach and Tool Support 2020-07 Wim van der Vegt (OUN) Towards a software architecture for reusable game components

2020-08 Ali Mirsoleimani (UL) Structured Parallel Programming for Monte Carlo Tree Search

2020-09 Myriam Traub (UU) Measuring Tool Bias Improving Data Quality for Digital Humanities Research

2020-10 Alifah Syamsiyah (TUE) In-database Preprocessing for Pro- cess Mining

2020-11 Sepideh Mesbah (TUD) Semantic-Enhanced Training Data AugmentationMethods for Long-Tail Entity Recognition Models

2020-12 Ward van Breda (VU) Predictive Modeling in E-Mental Health: Exploring Applicability in Personalised Depression Treatment 2020-13 Marco Virgolin (CWI) Design and Application of Gene-pool Optimal Mixing Evolutionary Algorithms for Genetic Programming 2020-14 Mark Raasveldt (CWI/UL) Integrating Analytics with Relational Databases

2020-15 Konstantinos Georgiadis (OU) Smart CAT: Machine Learning for Configurable Assessments in Serious Games

2020-16 Ilona Wilmont (RUN) Cognitive Aspects of Conceptual Modelling

2020-17 Daniele Di Mitri (OU) The Multimodal Tutor: Adaptive Feedback from Multimodal Experiences

2020-18 Georgios Methenitis (TUD) Agent Interactions Mechanisms in Markets with Uncertainties: Electricity Markets in Renewable Energy Systems

2020-19 Guido van Capelleveen (UT) Industrial Symbiosis Recommender Systems

2020-20 Albert Hankel (VU) Embedding Green ICT Maturity in Organisations

2020-21 Karine da Silva Miras de Araujo (VU) Where is the robot?: Life as it could be

2020-22 Maryam Masoud Khamis (RUN) Understanding complex systems implementation through a modeling approach: the case of e- government in Zanzibar

2020-23 Rianne Conijn (UT) The Keys to Writing: A writing analytics approach to studying writing processes using keystroke logging

2020-24 Lenin da Nobrega Medeiros (VUA/RUN) How are you feeling, human? Towards emotionally supportive chatbots

2020-25 Xin Du (TUE) The Uncertainty in Exceptional Model Mining

2020-26 Krzysztof Leszek Sadowski (UU) GAMBIT: Genetic Algorithm for Model-Based mixed-Integer opTimization

2020-27 Ekaterina Muravyeva (TUD) Personal data and informed consent in an educational context

2020-28 Bibeg Limbu (TUD) Multimodal interaction for deliberate practice: Training complex skills with augmented reality

2020-29 Ioan Gabriel Bucur (RUN) Being Bayesian about Causal Inference

2020-30 Bob Zadok Blok (UL) Creatief, Creatieve, Creatiefst

2020-31 Gongjin Lan (VU) Learning better – From Baby to Better 2020-32 Jason Rhuggenaath (TUE) Revenue management in online markets: pricing and online advertising

2020-33 Rick Gilsing (TUE) Supporting service-dominant business model evaluation in the context of business model innovation

2020-34 Anna Bon (MU) Intervention or Collaboration? Redesigning Information and Communication Technologies for Development

2020-35 Siamak Farshidi (UU) Multi-Criteria Decision-Making in Software Production

==== 2021 ====

2021-01 Francisco Xavier Dos Santos Fonseca (TUD) Location-based Games for Social Interaction in Public Space

2021-02 Rijk Mercuur (TUD) Simulating Human Routines: Integrating Social Practice Theory in Agent-Based Models

2021-03 Seyyed Hadi Hashemi (UVA) Modeling Users Interacting with Smart Devices

2021-04 Ioana Jivet (OU) The Dashboard That Loved Me: Designing adaptive learning analytics for self-regulated learning

2021-05 Davide Dell'Anna (UU) Data-Driven Supervision of Autonomous Systems

2021-06 Daniel Davison (UT) "Hey robot, what do you think?" How children learn with a social robot

2021-07 Armel Lefebvre (UU) Research data management for open science

2021-08 Nardie Fanchamps (OU) The Influence of Sense-Reason-Act Programming on Computational Thinking

2021-09 Cristina Zaga (UT) The Design of Robothings. Non-Anthropomorphic and Non-Verbal Robots to Promote Childrens Collaboration Through Play

2021-10 Quinten Meertens (UvA) Misclassification Bias in Statistical Learning

2021-11 Anne van Rossum (UL) Nonparametric Bayesian Methods in Robotic Vision

2021-12 Lei Pi (UL) External Knowledge Absorption in Chinese SMEs

2021-13 Bob R. Schadenberg (UT) Robots for Autistic Children: Understanding and Facilitating Predictability for Engagement in Learning 2021-14 Negin Samaeemofrad (UL) Business Incubators: The Impact of Their Support

2021-15 Onat Ege Adali (TU/e) Transformation of Value Propositions into Resource Re-Configurations through the Business Services Paradigm

2021-16 Esam A. H. Ghaleb (MU) BIMODAL EMOTION RECOGNITION FROM AUDIO-VISUAL CUES

2021-17 Dario Dotti (UM) Human Behavior Understanding from motion and bodily cues using deep neural networks

2021-18 Remi Wieten (UU) Bridging the Gap Between Informal Sense-Making Tools and Formal Systems - Facilitating the Construction of Bayesian Networks and Argumentation Frameworks

2021-19 Roberto Verdecchia (VU) Architectural Technical Debt: Identification and Management

2021-20 Masoud Mansoury (TU/e) Understanding and Mitigating Multi-Sided Exposure Bias in Recommender Systems

2021-21 Pedro Thiago Timb Holanda (CWI) Progressive Indexes 2021-22 Sihang Qiu (TUD) Conversational Crowdsourcing

2021-23 Hugo Manuel Proena (LIACS) Robust rules for prediction and description

2021-24 Kaijie Zhu (TUE) On Efficient Temporal Subgraph Query Processing

2021-25 Eoin Martino Grua (VUA) The Future of E-Health is Mobile: Combining AI and Self-Adaptation to Create Adaptive E-Health Mobile Applications

2021-26 Benno Kruit (CWI VU) Reading the Grid: Extending Knowledge Bases from Human-readable Tables

G

2021-27 Jelte van Waterschoot (UT) Personalized and Personal Conversations: Designing Agents Who Want to Connect With You

2021-28 Christoph Selig (UL) Understanding the Heterogeneity of Corporate Entrepreneurship Programs

==== 2022 ====

2022-1 Judith van Stegeren (UT) Flavor text generation for role- playing video games

2022-2 Paulo da Costa (TU/e) Data-driven Prognostics and Logistics Optimisation: A Deep Learning Journey

2022-3 Ali el Hassouni (VUA) A Model A Day Keeps The Doctor Away: Reinforcement Learning For Personalized Healthcare

2022-4 Anal Aksu (UU) A Cross-Organizational Process Mining Framework

2022-5 Shiwei Liu (TU/e) Sparse Neural Network Training with In- Time Over-Parameterization

2022-6 Reza Refaei Afshar (TU/e) Machine Learning for Ad Publishers in Real Time Bidding

2022-7 Sambit Praharaj (OU) Measuring the Unmeasurable? Towards Automatic Co-located Collaboration Analytics

2022-8 Maikel L. van Eck (TU/e) Process Mining for Smart Product Design

2022-9 Oana Andreea Inel (VUA) Understanding Events: A Diversitydriven Human-Machine Approach

2022-10 Felipe Moraes Gomes (TUD) Examining the Effectiveness of Collaborative Search Engines

2022-11 Mirjam de Haas (UT) Staying engaged in child-robot interaction, a quantitative approach to studying preschoolers engagement with robots and tasks during second-language tutoring

2022-12 Guanyi Chen (UU) Computational Generation of Chinese Noun Phrases

2022-13 Xander Wilcke (VUA) Machine Learning on Multimodal Knowledge Graphs: Opportunities, Challenges, and Methods for Learning on Real-World Heterogeneous and Spatially-Oriented

2022-14 Michiel Overeem (UU) Evolution of Low-Code Platforms 2022-15 Jelmer Jan Koorn (UU) Work in Process: Unearthing Meaning using Process Mining

2022-16 Pieter Gijsbers (TU/e) Systems for AutoML Research

2022-17 Laura van der Lubbe (VUA) Empowering vulnerable people with serious games and gamification

2022-22 Alexandra Sierra Rativa (UT) Virtual Character Design and its potential to foster Empathy, Immersion, and 21st Century Learning Skills in Video Games and Virtual Reality Simulations

H Professional Biography

Alexandra Sierra Rativa was born in Bogota, Colombia. She obtained a bachelor's in Electronics at Universidad Pedagogica Nacional, Colombia and a master's in education at Universidad de Los Andes, Colombia.

H.1 Master Professional Achievements

In 2012, she was recognized by the Colombian Ministry of Education as one of the best Technology and Informatics teachers in High School Education. For this reason, she was selected to participate in an ICT training course in South Korea. In 2015, she was nominated among the best ten proposals for the Teacher Award (Premio Compartir al Maestro) in Colombia with the project Robotics, Art, and Technology.

H.2 PhD Professional Achievements

In 2014, she was selected among the first places for a doctoral scholar- ship abroad in Colciencias. From 2015 to 2022, she was a PhD student in Cognitive Sciences and Artificial Intelligence at Tilburg University, the Netherlands.

In 2016, she and her team won a Dutch Hackathon with an interactive boxing machine (i.e., car tires and Arduino) for children.

In 2019, her research "Justin Beaver VR" was presented at a conference on innovation in education called LINC 2019 conducted by the University of Massachusetts-MIT (Boston). One of the best universities in the world.

In 2021, she and her team received an award for revolutionary research in immersive technology at Laval Virtual, France, for her research project called "Justin Beaver Stories". For Laval, she showed a new version of this immersive technology, a hologram in mixed reality. In this demonstration, users can interact with Justin Beaver. Justin was capable of voice interaction, making interactive stories and smart conversations about science topics, nature, or describing objects/animals of a place as a guide tour animal, virtual instructor, or virtual animal. In 2022, she was selected to present her virtual reality movie "Justin Beaver Survivor" at Tokyo Sandbox in Japan.

In 2022, her PhD project was awarded 15.000 euros to transform a research idea into a video game. In the Crea Digital call, Justin Beaver Survivor's video game was selected by the Ministry of Culture in Colombia. The following year, Justin Beaver will be in Spanish and exhibited in Bogota, Colombia.

H.3 Collaboration between Colombia and The Netherlands

During her PhD Studies, she co-founded the Colombian Association of Immersive, Interactive, and Emerging Realities (XRCOL).

From 2020 to 2022, she was also responsible for organizing the Inter- national Conference on Immersive Technology (Virtual Reality, Augmented Reality, Mixed Reality, and AI) focus on education, research, and innovation companies. This conference is focus on teachers and researchers from Colombia and the Netherlands. These conferences were organized in an immersive virtual environment, where speakers had an avatar body. She found the possibility to work collaboratively with a Dutch University (Tilburg University), an Applied Sciences University (Breda University of Applied Sciences) and three Colombian Univerities (Pontificia Universidad Javeriana, Universidad El Bosque and Universidad EAFIT).

Finally, she is the Editor-in-Chief of the open-access publication "XR ACademia: Advances on Research in Virtual Reality, Augmented Reality, Mixed Reality, and Artificial Intelligence in Latin America and Europe".

15. Andersen, J. F., Andersen, P. A., & Jensen, A. D. (1979). The measurement of nonverbal immediacy. Journal of Applied Communication Research, 7(2), 153-180. https://doi.org/10.1080/00909887909365204
16. Andrews, C. (2013). Zoos of the future. *Engineering & Technology*, 8(7), 68-71. https://doi.org/10.1049/et.2013.0726
17. Angantyr, M., Hansen, E. M., Eklund, J. H., & Malm, K. (2016). Reducing sex differences in children's empathy for animals through a training intervention. *Journal of Research in Childhood Education*, 30(3), 273–281. https://doi.org/10.1080/02568543.2016.1178198
18. Anusuya, M. A. & Katti, S. K. (2010). Speech Recognition by Machine, a review. Proceedings ArXiv, International Journal of Computer Science and Information Security, 6(3), 181-205. https://doi.org/10.48550/arxiv.1001.2267
19. Arden, R., Chavez, R. S., Grazioplene, R., & Jung, R. E. (2010). Neuroimaging creativity: a psychometric view. *Behavioural Brain Research*, 214(2), 143–156. https://doi.org/10.1016/j.bbr.2010.05.015
20. Argyle, M., Salter, V., Nicholson, H., Williams, M., & Burgess, P. (1970). The communication of inferior and superior attitudes by verbal and non-verbal signals. *British Journal of Social and Clinical Psychology*, 9(3), 222–231. https://doi.org/10.1111/j.2044-8260.1970.tb00668.x
21. Asplund, J., Lopez, S. J., Hodges, T., & Harter, J. (2007). The Clifton StrengthsFinder 2.0 technical report: Development and validation (Publication No. 2). The Gallup Organization, Princeton, *NJ*. http://jobtransition.net/wp-content/uploads/2012/11/All-About-Strengthsfinders.pdf
22. Astutik, S., & Prahani, B. K. (2018). The Practicality and Effectiveness of Collaborative Creativity Learning (CCL) Model by Using PhET Simulation to Increase Students' Scientific Creativity. *International Journal of Instructio*n, 11(4), 409-424.
23. Atabek, O. (2019). Personal Differences and Social Networking: A Comparison of Two Countries. *Educational Policy Analysis and Strategic Research*, 14(4), 167-189. https://doi.org/10.29329/epasr.2019.220.10
24. Ayas, M. B., & Sak, U. (2014). Objective measure of scientific creativity: Psychometric validity of the Creative Scientific Ability Test. *Thinking Skills and Creativity*, 13, 195-205. https://doi.org/10.1016/j.tsc.2014.06.001
25. Azevedo, R. T., Macaluso, E., Avenanti, A., Santangelo, V., Cazzato, V., & Aglioti, S. M. (2013). Their pain is not our pain: brain and autonomic correlates of empathic resonance with the pain of same and different race individuals. *Human Brain Mapping*, 34(12), 3168– 3181. https://doi.org/10.1002/hbm.22133
26. Bachen, C. M., Hernández-Ramos, P. F., & Raphael, C. (2012). Simulating REAL LIVES: Promoting global empathy and interest in learning through simulation games. *Simulation & Gaming*, 43(4), 437-460. https://doi.org/10.1177/1046878111432108
27. Baer, J. & Kaufman, J. C. (2008). Gender differences in creativity. *The Journal of Creative Behavior*, 42(2), 75–105. https://doi.org/10.1002/j.2162-6057.2008.tb01289.x
28. Bailey, J. T., McDonald, F. J., & Claus, K. E. (1970). Evaluation of the development of creative behavior in an experimental nursing program. *Nursing Research*. https://doi.org/10.1097/00006199-197003000-00002

29. Baki, N. U., Rafik-Galea, S., & Nimehchisalem, V. (2016). Malaysian rural ESL students' critical thinking literacy level: A case study. *International Journal of Education and Literacy Studies*, 4(4), 71–80. http://dx.doi.org/10.7575/aiac.ijels.v.4n.4p.71
30. Balamuralithara, B. & Woods, P. C. (2009). Virtual laboratories in engineering education: The simulation lab and remote lab. *Computer Applications in Engineering Education*, 17(1), 108–118. https://doi.org/10.1002/cae.20186
31. Banakou, D., Groten, R., & Slater, M. (2013). Illusory ownership of a virtual child body causes overestimation of object sizes and implicit attitude changes. P*roceedings of the National Academy of Sciences*, 110 (31), 12846–12851. https://doi.org/10.1073/pnas.1306779110
32. Banks, M. R., Willoughby, L. M., & Banks, W. A. (2008). Animal assisted therapy and loneliness in nursing homes: use of robotic versus living dogs. *Journal of the American Medical Directors Association*, 9(3), 173–177. https://doi.org/10.1016/j.jamda.2007.11.007
33. Bar, M. A., Leurer, M. K., Warshawski, S., & Itzhaki, M. (2018). The role of personal resilience and personality traits of healthcare students on their attitudes towards interprofessional collaboration. *Nurse Education Today*, 61, 36–42. https://doi.org/10.1016/j.nedt.2017.11.005
34. Barin, M. (2019). Pre-Service Teachers' Attitudes towards English Courses and Their Critical Thinking Skills. *World Journal of Education*, 9(2), 46-55. https://doi.org/10.5430/wje.v9n2p46
35. Barrick, M. R., & Mount, M. K. (1991). The big five personality dimensions and job performance: a meta-analysis. *Personnel Psychology,* 44(1), 1-26. https://doi.org/10.1111/j.1744-6570.1991.tb00688.x
36. Bartneck, C., Kanda, T., Ishiguro, H., & Hagita, N. (2009, September). My robotic doppelgänger-A critical look at the uncanny valley. In RO-MAN 2009-*The 18th IEEE International Symposium on Robot and Human Interactive Communication*, 269-276, https://doi.org/10.1109/ROMAN.2009.5326351
37. Bataineh, R. F., & Zghoul, L. H. (2006). Jordanian TEFL graduate students' use of critical thinking skills (as measured by the Cornell Critical Thinking Test, Level Z). *International Journal of Bilingual Education and Bilingualism*, 9(1), 33-50. https://doi.org/10.1080/13670050608668629
38. Batey, M. & Furnham, A. (2008). The relationship between measures of creativity and schizotypy. *Personality and Individual Differences*, 45(8), 816–821. https://doi.org/10.1016/j.paid.2008.08.014
39. Batson, C. D., Fultz, J., & Schoenrade, P. A. (1987). Distress and empathy: Two qualitatively distinct vicarious emotions with different motivational consequences. *Journal of Personality*, 55(1), 19-39. https://doi.org/10.1111/j.1467-6494.1987.tb00426.x
40. Bautista, C., Revilla, E., Naves, J., Albrecht, J., Fernández, N., Olszańska, A., Adamec, M., Berezowska-Cnota, T., Ciucci, P., Groff, C. & Härkönen, S. (2019). Large carnivore damage in Europe: Analysis of compensation and prevention programs. *Biological Conservatio*n, 235, 308-316. https://doi.org/10.1016/j.biocon.2019.04.019
41. Bawtree, S., Hawkey, S., Holdsworth, R., Nyfield, G., & Saville, P. (1991). Critical Reasoning Test Battery. *Saville and Holdsworth Ltd*.

42. Bayram, D., Gökçe, K. U. R. T., & Derin, A. T. A. Y. (2019). The Implementation of webquest-supported critical thinking instruction in pre-service English teacher education: The Turkish context. *Participatory Educational Research*, 6(2), 144-157. https://doi.org/10.17275/per.19.18.6.2
43. Beck, D., Morgado, L., & O'Shea, P. (2020). Finding the gaps about uses of immersive learning environments: a survey of surveys. *Journal of Universal Computer Science*, 26, 1043-1073. http://hdl.handle.net/10400.2/10070
44. Becker-Asano, C., Prendinger, H., Ishizuka, M., & Wachsmuth, I. (2005). Empathy for max (preliminary project report). *Proceedings of the International Conference on Active Media Technology* (AMT 2005), 541–545. https://pub.uni-bielefeld.de/record/2610101
45. Beer, J. M., Smarr, C. A., Fisk, A. D., & Rogers, W. A. (2015). Younger and older users' recognition of virtual agent facial expressions. *International Journal of Human-Computer Studies*, 75, 1-20. https://doi.org/10.1016/j.ijhcs.2014.11.005
46. Behar-Horenstein, L. S., & Niu, L. (2011). Teaching critical thinking skills in higher education: A review of the literature. *Journal of College Teaching & Learnin*g (TLC), 8(2). https://doi.org/10.19030/tlc.v8i2.3554
47. Bellanca, J. A. (2010). 21st century skills: Rethinking how students learn. *Solution Tree Press.*
48. Belman, J. (2016). A study of empathetic play in serious games [Doctoral dissertation, New York University]. *ProQuest*. https://www.proquest.com/docview/1813802971?pq-origsite=gscholar&fromopenview=true
49. Belman, J., & Flanagan, M. (2010). Designing games to foster empathy. *International Journal of Cognitive Technology*, 15(1), 11.
50. Belpaeme, T., Kennedy, J., Ramachandran, A., Scassellati, B., & Tanaka, F. (2018). Social robots for education: A review. *Science robotics*, 3(21). https://www.science.org/doi/10.1126/scirobotics.aat5954
51. Bensley, D. A., Rainey, C., Murtagh, M. P., Flinn, J. A., Maschiocchi, C., Bernhardt, P. C., & Kuehne, S. (2016). Closing the assessment loop on critical thinking: The challenges of multidimensional testing and low test-taking motivation. *Thinking Skills and Creativity*, 21, 158-168. https://doi.org/10.1016/j.tsc.2016.06.006
52. Besemer, S. P., & O'Quin, K. (1999). Confirming the three-factor creative product analysis matrix model in an American sample. *Creativity Research Journal*, 12(4), 287-296. https://doi.org/10.1207/s15326934crj1204_6
53. Binkley, M., Erstad, O., Herman, J., Raizen, S., Ripley, M., Miller-Ricci, M., & Rumble, M. (2012). Defining twenty-first century skills. I*n Assessment and teaching of 21st century skills*, 17-66. Springer, Dordrecht.
54. Blake, J. B. (1978). The reliability and validity of the TEMPLIN- DARLEY Test of articulation with spastic cerebral-palsied children. New York University.
55. Blumenfeld, P. C., Marx, R. W., Soloway, E., & Krajcik, J. (1996). Learning with peers: From small group cooperation to collaborative communities. *Educational Researcher,* 25(8), 37-39. https://doi.org/10.3102/0013189X025008037
56. Bonk, C. J. (1988). The Effects of Convergent and Divergent Computer Software on Children's Critical and Creative Thinking. *In the Annual Meeting of the American Educational Research Association*, 1–29. ERIC.

57. Booth, A. J., Rodgers, J. D., Volker, M. A., Lopata, C., & Thomeer, M. L. (2019). Psychometric characteristics of the DANVA-2 in high-functioning children with ASD. *Journal of Autism and Developmental Disorder*s, 49(10), 4147-4158. https://doi.org/10.1007/s10803-019-04130-w

58. Bostrom, R. N., & Waldhart, E. S. (1983). The Kentucky comprehensive listening test. Kentucky Listening Research Center.

59. Botvinick, M., & Cohen, J. (1998). Rubber hands 'feel' touch that eyes see. *Nature,* 391(6669), 756-756. https://doi.org/10.1038/35784

60. Bouaziz, S., Wang, Y., & Pauly, M. (2013). Online modeling for realtime facial animation. *ACM Transactions on Graphics* (ToG), 32(4), 1-10. https://doi.org/10.1145/2461912.2461976

61. Boud, D. (2018). Assessment could demonstrate learning gains, but what is required for it to do so?. *Higher Education Pedagogies*, 3(1), 54-56. https://doi.org/10.1080/23752696.2017.1413671

62. Bowman, D. A., & McMahan, R. P. (2007). Virtual reality: how much immersion is enough?. *Computer*, 40(7), 36-43. https://doi.org/10.1109/MC.2007.257

63. Bozgeyikli, H. (2017). Big five personality traits as the predictor of teachers' organizational psychological capital. *Online Submission,* 8(18), 125–135.

64. Bradski, G. R., Miller, S. A., & Abovitz, R. (2019). U.S. Patent No. 10,203,762. *Washington, DC: U.S. Patent and Trademark Offic*e. https://patents.google.com/patent/US10203762B2/en

65. Brenton, H., Gillies, M., Ballin, D., & Chatting, D. (2005). The uncanny valley: does it exist. In Proceedings of conference of human computer interaction, workshop on human animated character interaction. *Citeseer*.

66. Broadleaf, C. (2020). Tools for measuring collaboration. *Technical report, Broadleaf Consulting.* https://broadleafconsulting.ca/uploads/3/4/0/8/3408103/tools_for_measuring_collaboration.pdf

67. Bronstein, L. R. (2002). Index of interdisciplinary collaboration. (Instrument Development). *Social Work Research*, 26(2), 113–127.

68. Brown, A. L. and Campione, J. C. (2002). In Woods P. (Ed.). Communities of learning and thinking, or a context by any other name. *Contemporary issues in teaching and learning,* 120–126. The Open University.

69. Brown, E., & Cairns, P. (2004). A grounded investigation of game immersion. *In CHI'04 Extended Abstracts on Human Factors in Computing Systems,* 1297-1300. https://doi.org/10.1145/985921.986048

70. Brown, T., Yu, M. l., & Etherington, J. (2020). Listening and inter- personal communication skills as predictors of resilience in occupational therapy students: A cross-sectional study. *British Journal of Occupational Therapy*, 6(2), 42–53. https://doi.org/10.1016/j.hpe.2020.01.001

71. Bruffee, K. A. (1987). The art of collaborative learning. Change. *The Magazine of Higher Learning*, 19(2), 42–47. https://doi.org/10.1080/00091383.1987.9939136

72. Bujic, M., Salminen, M., Macey, J., and Hamari, J. (2020). "Empathy machine": how virtual reality affects human rights attitudes. *Internet Research*, 30(5), 1407–1425. https://doi.org/10.1108/INTR-07-2019-0306

73. Bulgurcuoglu, A. N. (2016). Relationship between critical thinking lev- els and attitudes towards reading habits among pre-service physical education teachers. *Educational Research and Reviews*, 11(8), 708– 712. https://doi.org/10.5897/ERR2016.2713

74. Burleigh, T. J., & Schoenherr, J. R. (2015). A reappraisal of the uncanny valley: categorical perception or frequency-based sensitization?. *Frontiers in Psychology,* 5, 1488. https://doi.org/10.3389/fpsyg.2014.01488

75. Burleigh, T. J., Schoenherr, J. R., & Lacroix, G. L. (2013). Does the uncanny valley exist? An empirical test of the relationship between eeriness and the human likeness of digitally created faces. *Computers in Human Behavior,* 29(3), 759-771. https://doi.org/10.1016/j.chb.2012.11.021

76. Burns, R. B. (1991). Essential Psychology: For Students and Professionals in the Health and Social Services, chapter 15. *Verbal and non-verbal communication*, 223–235. Springer, Dordrecht.

77. Busch, P. S., & Davis, S. W. (2018). Inside out personal branding (IOPB): Using Gallup Clifton Strengthsfinder 2.0 and 360Reach. *Marketing Education Review*, 28(3), 187-202. https://doi.org/10.1080/10528008.2017.1367930

78. Butler, H. A. (2012). Halpern Critical Thinking Assessment predicts real-world outcomes of critical thinking. *Applied Cognitive Psychology*, 26(5), 721-729. https://doi.org/10.1002/acp.2851

79. Carpin, S., Lewis, M., Wang, J., Balakirsky, S., & Scrapper, C. (2007). USARSim: a robot simulator for research and education. In Proceedings 2007 IEEE International Conference on Robotics and Automation, *IEEE*, 1400-1405. https://doi.org/10.1109/ROBOT.2007.363180

80. Carroll, T. L. (1999). Multidisciplinary collaboration: A method for measurement. *Nursing Administration Quarterly,* 23(4), 86-90.

81. Carson, S. H., Peterson, J. B., & Higgins, D. M. (2003). Decreased latent inhibition is associated with increased creative achievement in high-functioning individuals. *Journal of Personality and Social Psychology*, 85(3), 499-506. https://doi.org/10.1037/0022-3514.85.3.499

82. Carter, C., Hass, R. W., Charfadi, M., & Dinzeo, T. J. (2019). Probing linear and nonlinear relations among schizotypy, hypomania, cognitive inhibition, and creativity. *Creativity Research Journal,* 31(1), 83–92. https://doi.org/10.1037/0022-3514.85.3.499

83. Castleberry, S. B. & Shepherd, C. D. (1993). Effective interpersonal listening and personal selling. *Journal of Personal Selling & Sales Management*, 13(1), 35–49. https://doi.org/10.1080/08853134.1993.10753935

84. Causo, A., Vo, G. T., Chen, I.-M., & Yeo, S. H. (2016). Design of robots used as education companion and tutor. Robotics and Mechatronics, *Springer,* 75–84. https://doi.org/10.1007/978-3-319-22368-1_8

85. Cenydd, L. A., & Teahan, B. (2013). An embodied approach to arthropod animation. *Computer Animation and Virtual Worlds*, 24(1), 65-83. https://doi.org/10.1002/cav.1436

86. Chalkiadaki, A. (2018). A systematic literature review of 21st century skills and competencies in primary education. *International Journal of Instruction*, 11(3), 1-16. https://doi.org/10.12973/iji.2018.1131a

87. Charmello, C. (1993). Self-questioning prediction strategy's effect on comprehension. [Master's thesis, Kean College of New Jersey]. ERIC.

88. Charyton, C., Basham, K. M., & Elliott, J. O. (2008). Examining gender with general creativity and preferences for creative persons in college students within the sciences and the arts. *The Journal of Creative Behavior*, 42(3), 216-222. https://doi.org/10.1002/j.2162-6057.2008.tb01296.x

89. Cheetham, M., Pedroni, A., Antley, A., Slater, M., & Jäncke, L. (2009). Virtual milgram: empathic concern or personal distress? Evidence from functional MRI and dispositional measures. *Frontiers in Human Neuroscience*, 3, 29. https://doi.org/10.3389/neuro.09.029.2009

90. Chen, C., Himsel, A., Kasof, J., Greenberger, E., & Dmitrieva, J. (2006). Boundless creativity: Evidence for the domain generality of individual differences in creativity. *The Journal of Creative Behavior*, 40(3), 179-199. https://doi.org/10.1002/j.2162-6057.2006.tb01272.x

91. Chen, Z. H., Liao, C. C., & Chan, T. W. (2007, March). My-pet-and-her-friends: Identifying educational roles of animal companions in game-based learning environment. In 2007 First IEEE International Workshop on Digital Game and Intelligent Toy Enhanced Learning (DIGITEL'07), IEEE, 104-108. https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=4148838

92. Cheung, C.-K., Rudowicz, E., Yue, X., & Kwan, A. S. (2003). Creativity of university students: What is the impact of field and year of study?. *The Journal of Creative Behavior*, 37(1), 42-63. https://doi.org/10.1002/j.2162-6057.2003.tb00825.x

93. Chi, S. A., Kim, S. H., & Kim, H. (2016). Problem behaviours of kindergartners: The affects of children's cognitive ability, creativity, and self-esteem. *South African Journal of Education*, 36(1). https://doi.org/10.15700/saje.v36n1a1177

94. Choi, H. G., Kim, H. J., & Lee, S. (2021). Mediating Effects of Empathy on Competence of Korean Nursing Students. *Iranian Journal of Public Health,* 50(4), 737. https://doi.org/10.18502/ijph.v50i4.5998

95. Choi, Y. (2017). Effects of dimensions of word knowledge and their accessibility on different levels of reading comprehension in adolescent EFL learners. [Doctoral dissertation, Michigan State University]. ProQuest LLC.

96. Ciechanowski, L., Przegalinska, A., Magnuski, M., & Gloor, P. (2019). In the shades of the uncanny valley: An experimental study of human–chatbot interaction. *Future Generation Computer Systems*, 92, 539-548. https://doi.org/10.1016/j.future.2018.01.055

97. Clapham, M. M. (2004). The convergent validity of the Torrance Tests of Creative Thinking and creativity interest inventories. *Educational and Psychological Measurement,* 64(5), 828-841. https://doi.org/10.1177/0013164404263883

98. Clark, M. A., Robertson, M. M., & Young, S. (2019). "I feel your pain": A critical review of organizational research on empathy. *Journal of Organizational Behavior*, 40(2), 166-192. https://doi.org/10.1002/job.2348

99. Coeckelbergh, M. (2011). Humans, animals, and robots. *International Journal of Social Robotics*, 3(2). https://doi.org/10.1007/s12369-010-0075-6

100. Coetzee, D., Lim, S., Fox, A., Hartmann, B., & Hearst, M. A. (2015). Structuring interactions for large-scale synchronous peer learning. *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, 1139-1152. https://doi.org/10.1145/2675133.2675251

101. Coffelt, T. A., Grauman, D., & Smith, F. L. (2019). Employers' perspectives on workplace communication skills: The meaning of communication skills. *Business and Professional Communication Quarterly*, 82(4), 418-439. https://doi.org/10.1177/2329490619851119

102. Coggins, J. V., Kim, J. K., & Briggs, L. C. (2017). Comparison of IRT and CTT Using Secondary School Reading Comprehension Assessments. *Research in the Schools,* 24(1), 80-93.

103. Coghlan, S., Waycott, J., Neves, B. B., & Vetere, F. (2018). Using robot pets instead of companion animals for older people: a case of reinventing the wheel'?. *In Proceedings of the 30th Australian Conference on Computer-Human Interaction*, 172-183. https://doi.org/10.1145/3292147.3292176

104. Criscuolo, N. M. (2019). Meatspace. [Master's thesis, University of Massachusetts]. Amherst. https://doi.org/10.7275/14427281

105. Cropley, A. J. (2000). Defining and measuring creativity: Are creativity tests worth using?. *Roeper Review*, 23(2), 72-79. https://doi.org/10.1080/02783190009554069

106. Cupach, W. R., & Spitzberg, B. H. (1983). Trait versus state: A comparison of dispositional and situational measures of interpersonal communication competence. *Western Journal of Speech Communication,* 47(4), 364-379. https://doi.org/10.1080/10570318309374131

107. Dal, N. (2018). The Big Five Personality Traits and Narcissism as the Predictors of Anxiety and Confidence before Archery Class Final Exam. *Universal Journal of Educational Research*, 6(12), 2875-2879. https://doi.org/10.13189/ujer.2018.061222

108. Das, D. A., Grimmer, K. A., Sparnon, A. L., McRae, S. E., & Thomas, B. H. (2005). The efficacy of playing a virtual reality game in modulating pain for children with acute burn injuries: a randomized controlled trial [ISRCTN87413556]. *BMC Pediatrics,* 5(1), 1-10. https://doi.org/10.1186/1471-2431-5-1

109. Davenport Sypher, B., Bostrom, R. N., & Hart Seibert, J. (1989). Listening, communication abilities, and success at work. *The Journal of Business Communicatio*n (1973), 26(4), 293-303. https://doi.org/10.1177/002194368902600401

110. Davis, G. A. (1989). Testing for creative potential. *Contemporary Educational Psychology,* 14(3), 257-274. https://doi.org/10.1016/0361-476X(89)90014-3

111. Davis, M. H. (1983). Measuring individual differences in empathy: evidence for a multidimensional approach. *Journal of personality and social psychology*, 44(1), 113–126. https://doi.org/10.1037/0022-3514.44.1.113

112. Davis, M. H. (2018). Empathy: A social psychological approach. *Routledge*. https://doi.org/10.4324/9780429493898

113. Davis, M. H. (1980). A multidimensional approach to individual differences in empathy. Technical report. *The University of Texas at Austin*. https://www.uv.es/friasnav/Davis_1980.pdf

114. Davis, M. H., Luce, C., & Kraus, S. J. (1994). The heritability of characteristics associated with dispositional empathy. *Journal of Personality*, 62(3), 369-391. https://doi.org/10.1111/j.1467-6494.1994.tb00302.x

115. De Corte, K., Buysse, A., Verhofstadt, L. L., Roeyers, H., Ponnet, K., & Davis, M. H. (2007). Measuring empathic tendencies: Reliability and validity of the Dutch version of the Interpersonal Reactivity Index. *Psychologica Belgica,* 47(4), 235-260. https://doi.org/10.5334/pb-47-4-235

116. De Droog, S. M., Buijzen, M., & Valkenburg, P. M. (2014). Enhancing children's vegetable consumption using vegetable-promoting picture books. The impact of interactive shared reading and character–product congruence. *Appetite,* 73, 73-80. https://doi.org/10.1016/j.appet.2013.10.018

117. De Droog, S. M., van Nee, R., Govers, M., & Buijzen, M. (2017). Promoting toddlers' vegetable consumption through interactive reading and puppetry. *Appetite*, 116, 75-81. https://doi.org/10.1016/j.appet.2017.04.022

118. De Gennaro, M., Krumhuber, E. G., & Lucas, G. (2020). Effectiveness of an empathic chatbot in combating adverse effects of social exclusion on mood. *Frontiers in Psychology,* 10, 3061. https://doi.org/10.3389/fpsyg.2019.03061

119. De Melo, C. M., Gratch, J., & Carnevale, P. J. (2014). Humans versus computers: Impact of emotion expressions on people's decision making. *IEEE Transactions on Affective Computing,* 6(2), 127-136. https://doi.org/10.1109/TAFFC.2014.2332471

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

121. Dede, C. (2010). Comparing frameworks for 21st century skills. In Bellanca, J. & Brandt R. (Eds.). *21st century skills: Rethinking how students learn*, 20(2010), 51-76. Solution Tree Press.

122. Delgado-Mata, C., Martinez, J. I., Bee, S., Ruiz-Rodarte, R., & Aylett, R. (2007). On the use of virtual animals with artificial fear in virtual environments. *New Generation Computing*, 25(2), 145-169.

123. Demiral, S. (2019). Critical Thinking and Problem-Solving Skills of Visually Impaired Female National Judo Team Athletes. *Journal of Education and Training Studies*, 7, 8-16. https://doi.org/10.11114/jets.v7i3s.4004

124. Demirbag, B., Unisen, A., & Yesilyurt, A. (2016). Training of critical thinking skills in teacher candidates and placebo effect: A quasi-experimental study. *Eurasian Journal of Educational Research*, 16(63), 375-392. https://files.eric.ed.gov/fulltext/EJ1138040.pdf

125. Demirci, F., & Özyürek, C. (2017). The effects of using concept cartoons in astronomy subjects on critical thinking skills among seventh grade student. *International Electronic Journal of Elementary Education,* 10(2), 243-254. https://www.iejee.com/index.php/IEJEE/article/view/369

126. Derman, B., & Yavuz, S. (2022). The effect of educational game design process on students' creativity. *Smart Learning Environments*, 9(1). http://dx.doi.org/10.1186/s40561-022-00188-9

127. Diel, A., & MacDorman, K. F. (2021). Creepy cats and strange high houses: Support for configural processing in testing predictions of nine uncanny valley theories. *Journal of Vision*, 21(4), 1-1. https://doi.org/10.1167/jov.21.4.1

128. Dill, V., Flach, L. M., Hocevar, R., Lykawka, C., Musse, S. R., & Pinho, M. S. (2012). Evaluation of the uncanny valley in CG characters. *International Conference on Intelligent Virtual Agents Springer, Berlin, Heidelberg,* 511-513. https://doi.org/10.1007/978-3-642-33197-8_62

129. Dodd, N., & Bayne, R. (2007). Internal reliability and item analysis of the Keirsey Temperament Sorter II. *Journal of Psychological Type*, 67(8), 71–83.

130. Dollinger, S. J., Urban, K. K., & James, T. A. (2004). Creativity and openness: Further validation of two creative product measures. *Creativity Research Journal,* 16(1), 35-47. https://doi.org/10.1207/s15326934crj1601_4

131. Domsch, S. (2017). Dialogue in video games. Jarmila M., & Browen T., (Eds.), Dialogue across media, 28, 251-271. John Benjamins Publishing Company.

132. Dostál, D., Plháková, A., & Záškodná, T. (2017). Domain-specific creativity in relation to the level of empathy and systemizing. *The Journal of Creative Behavior,* 51(3), 225-239. https://doi.org/10.1002/jocb.103

133. Dougherty, M. B., & Larson, E. (2005). A review of instruments measuring nurse-physician collaboration. JONA: *The Journal of Nursing Administration*, 35(5), 244-253.

134. Ducarme, F., Luque, G. M., & Courchamp, F. (2013). What are "charismatic species" for conservation biologists. BioSciences Master Reviews, 10 (2013), 1-8.

135. Dumitru, D. (2019). Creating meaning. The importance of Arts, Humanities and Culture for critical thinking development. *Studies in Higher Education*, 44(5), 870-879. https://doi.org/10.1080/03075079.2019.1586345

136. Dyck, M., Winbeck, M., Leiberg, S., Chen, Y., Gur, R. C., & Mathiak, K. (2008). Recognition profile of emotions in natural and virtual faces. *PloS One*, 3(11). https://doi.org/10.1371/journal.pone.0003628

137. Egges, A., Kshirsagar, S., & Magnenat-Thalmann, N. (2004). Generic personality and emotion simulation for conversational agents. *Computer Animation and Virtual Worlds*, 15(1), 1-13. https://doi.org/10.1002/cav.3

138. Eguchi, A. (2014). Educational robotics for promoting 21st century skills. *Journal of Automation, Mobile Robotics and Intelligent Systems*, 5-11.

139. Eguchi, A. (2016). RoboCupJunior for promoting STEM education, 21st century skills, and technological advancement through robotics competition. *Robotics and Autonomous Systems,* 75, 692-699. https://doi.org/10.1016/j.robot.2015.05.013

140. Eisenberg, N., Fabes, R.A., Murphy, B., Karbon, M., Maszk, P., Smith, M., O'Boyle, C. and Suh, K., (1994). The relations of emotionality and regulation to dispositional and situational empathy-related responding. *Journal of Personality and Social Psychology*, 66(4), 776.

141. Eisenberg, N., Miller, P. A., Schaller, M., Fabes, R. A., Fultz, J., Shell, R., & Shea, C. L. (1989). The role of sympathy and altruistic personality traits in helping: A reexamination. *Journal of Personality*, 57(1), 41-67. https://doi.org/10.1111/j.1467-6494.1989.tb00760.x

142. Eisenberg, N. and Strayer, J. (1990). Empathy and its development. (1st ed.). Cambridge Studies in Social & Emotional Development.

143. Ekman, P. (1992). Facial expressions of emotion: New findings, new questions. *Psychological Science*. 3(4), 34-38. https://doi.org/10.1111/j.1467-9280.1992.tb00253.x

144. Ellis, R. J. B., Carmon, A. F., & Pike, C. (2016). A review of immediacy and implications for provider–patient relationships to support medication management. *Patient Preference and Adherence*, 10, 9. https://doi.org/10.2147/PPA.S95163

145. Ennis, R. H. (1993). Critical thinking assessment. *Theory into practice*, 32(3), 179–186. https://doi.org/10.1080/00405849309543594

146. Ennis, R.H. (2015). Critical Thinking: A Streamlined Conception. In: Davies, M., Barnett, R. (eds). The Palgrave Handbook of Critical Thinking in Higher Education. *Palgrave Macmillan, New York*. https://doi.org/10.1057/9781137378057_2

147. Ercan, H. (2017). The relationship between resilience and the big five personality traits in emerging adulthood. *Eurasian Journal of Educational Research*, 17(70), 83-103. http://dx.doi.org/10.14689/ejer.2017.70.5

148. Erdogan, F. (2019). Effect of cooperative learning supported by reflective thinking activities on students' critical thinking skills. *Eurasian Journal of Educational Research,* 19 (80), 89-112. http://dx.doi.org/10.14689/ejer.2019.80.5

149. Eroglu, S., & Bektas, O. (2022). The Effect of STEM Applications on the Scientific Creativity of 9th-Grade Students. Journal of Education in Science, *Environment and Health*, 8(1), 17-36. https://doi.org/10.21891/jeseh.1059124

150. Facione, P. A. (1990). The California Critical Thinking Skills Test--College Level. Technical Report# 1. *Experimental Validation and Content Validity*, 3-21. ERIC.

151. Facione, P. A. (2011). Critical thinking: What it is and why it counts. *Insight Assessment,* 2007(1), 1-23.

152. Facione, P. A., Sanchez, C. A., Facione, N. C., & Gainen, J. (1995). The disposition toward critical thinking. *The Journal of General Education*, 44(1), 1-25.

153. Fadel, C. and Trilling, B. (2010). 21st century skills: Learning for life in our times. Education Review. Jossey-Bass.

154. Fadhlullah, A., & Ahmad, N. (2017). Thinking outside of the box: Determining students' level of critical thinking skills in teaching and learning. *Asian Journal of University Education (AJUE)*, 13(2), 51-70.

155. Fairclough, S. H., Stamp, K., Dobbins, C., & Poole, H. M. (2020). Computer games as distraction from PAIN: Effects of hardware and difficulty on pain tolerance and subjective Immersion. *International Journal of Human-Computer Studies,* 139 (102427), 1-12. https://doi.org/10.1016/j.ijhcs.2020.102427

156. Faulkner, N. (2018). "Put yourself in their shoes": Testing empathy's ability to motivate cosmopolitan behavior. *Political Psychology*, 39(1), 217-228. https://doi.org/10.1111/pops.12411

157. Fee, A., & Gray, S. J. (2012). The expatriate-creativity hypothesis: A longitudinal field test. *Human Relations,* 65(12), 1515-1538. https://doi.org/10.1177/0018726712454900

158. Fernaeus, Y., Håkansson, M., Jacobsson, M., & Ljungblad, S. (2010). How do you play with a robotic toy animal? A long-term study of Pleo. *Proceedings of the 9th international Conference on interaction Design and Children,* 39-48. https://doi.org/10.1145/1810543.1810549

159. Fernández, A. M., Dufey, M., & Kramp, U. (2011). Testing the psychometric properties of the Interpersonal Reactivity Index (IRI) in Chile. *European Journal of Psychological Assessment*, 179–185. https://doi.org/10.1027/1015-5759/a000065

160. Fettahlioglu, P., & Kaleci, D. (2018). Online Argumentation Implementation in the Development of Critical Thinking Disposition. *Journal of Education and Training Studies*, 6(3), 127-136. https://doi.org/10.11114/jets.v6i3.2904

161. Field, A. (2013). Discovering statistics using IBM SPSS statistics. (3rd ed). London Sage.

162. Fisher, A. (2011). Critical thinking: An introduction. *In Carnichael M.* (Eds). Cambridge. University Press.

163. Fisher, J.A. (2017). Empathic Actualities: Toward a Taxonomy of Empathy in Virtual Reality. In: Nunes, N., Oakley, I., Nisi, V. (eds). Interactive Storytelling. ICIDS 2017. Lecture Notes in Computer Science, 10690. *Springer,* Cham. https://doi.org/10.1007/978-3-319-71027-3_19

164. Fitch-Hauser, M., & Hughes, M. A. (1992). The conceptualization and measurement of listening. International Listening Association. *Journal,* 6(1), 6-22. https://doi.org/10.1080/10904018.1992.10499105

165. Fortis, E. (2019). Manager's personality and their professional career. *Acta Educ. Gen*, 9, 20-38.

166. Fribourg, R., Argelaguet, F., Lécuyer, A., & Hoyet, L. (2020). Avatar and sense of embodiment: Studying the relative preference between appearance, control and point of view. *IEEE Transactions on Visualization and Computer Graphics*, 26(5), 2062-2072. https://doi.org/10.1109/TVCG.2020.2973077

167. Friedman, H. S., Prince, L. M., Riggio, R. E., & DiMatteo, M. R. (1980). Understanding and assessing nonverbal expressiveness: the affective communication test. *Journal of Personality and Social Psychology*, 39(2), 333–351. https://doi.org/10.1037/0022-3514.39.2.333

168. Furnham, A., & Bachtiar, V. (2008). Personality and intelligence as predictors of creativity. *Personality and Individual Differences,* 45(7), 613-617. https://doi.org/10.1016/j.paid.2008.06.023

169. Furnham, A., Batey, M., Booth, T. W., Patel, V., & Lozinskaya, D. (2011). Individual difference predictors of creativity in Art and Science students. *Thinking Skills and Creativity,* 6(2), 114-121. https://doi.org/10.1016/j.tsc.2011.01.006

170. Fusaro, M., Tieri, G., & Aglioti, S. M. (2016). Seeing pain and pleasure on self and others: behavioral and psychophysiological reactivity in immersive virtual reality. *Journal of Neurophysiology,* 116(6), 2656-2662. https://doi.org/10.1152/jn.00489.2016

171. Gadzella, B. M., & Baloğlu, M. (2003). Psychometric properties of Watson-Glaser Critical Thinking Appraisal for a sample of education majors. *Psychological Reports,* 92(3), 1249-1254. https://doi.org/10.2466/pr0.2003.92.3c.1249

172. Gadzella, B. M., Hogan, L., Masten, W., Stacks, J., Stephens, R., & Zascavage, V. (2006). Reliability and Validity of the Watson-Glasere Critical Thinking Appraisal-Forms for Different Academic Groups. *Journal of Instructional Psychology,* 33(2).

173. Garandeau, C. F., Laninga-Wijnen, L., & Salmivalli, C. (2021). Effects of the Kiva anti-bullying program on affective and cognitive empathy in children and adolescents. *Journal of Clinical Child & Adolescent Psychology*, 1-15. https://doi.org/10.1080/15374416.2020.1846541

174. Garcia-Barrera, M. A., Karr, J. E., Trujillo-Orrego, N., Trujillo-Orrego, S., & Pineda, D. A. (2017). Evaluating empathy in Colombian ex-combatants: Examination of the internal structure of the Interpersonal Reactivity Index (IRI) in Spanish. *Psychological assessment,* 29(1), 116–122. https://doi.org/10.1037/pas0000331

175. García-Carbajal, S., Pipa-Muniz, M., & Múgica, J. L. (2020). Using String Metrics to Improve the Design of Virtual Conversational Characters: Behavior Simulator Development Study. *JMIR Serious Games,* 8(1). https://doi.org/10.2196/15349

176. Gedney, J. (1994). Development of an instrument to measure collaboration and satisfaction about care decisions. *Journal of Advanced Nursing*, 20(1), 176-182. https://doi.org/10.1046/j.1365-2648.1994.20010176.x

177. Geller, T. (2008). Overcoming the uncanny valley. *IEEE Computer Graphics and Applications,* 28(4), 11-17. https://doi.org/10.1109/MCG.2008.79

178. Geographic, N. (2019). *Biodiversity*. https://www.nationalgeographic.org/encyclopedia/biodiversity

179. Germaine, R., Richards, J., Koeller, M., & Schubert-Irastorza, C. (2016). Purposeful use of 21st century skills in higher education. *Journal of Research in Innovative Teaching,* 9(1).

180. Gerschütz, B., Fechter, M., Schleich, B., & Wartzack, S. (2019). A review of requirements and approaches for realistic visual perception in virtual reality. In Proceedings of the Design Society: International Conference on Engineering Design, 1(1), 1893-1902, *Cambridge University Press*. https://doi.org/10.1017/dsi.2019.195

181. Gibbs, J. K., Gillies, M., & Pan, X. (2022). A comparison of the effects of haptic and visual feedback on presence in virtual reality. *International Journal of Human-Computer Studies,* 157, 102717. https://doi.org/10.1016/j.ijhcs.2021.102717

182. Gilet, A. L., Mella, N., Studer, J., Grühn, D., & Labouvie-Vief, G. (2013). Assessing dispositional empathy in adults: A French validation of the Interpersonal Reactivity Index (IRI). Canadian Journal of Behavioural Science/ *Revue Canadienne Des Sciences Du Comportemen*t, 45(1), 42–48. https://doi.org/10.1037/a0030425

183. Giusto, M., & Ehri, L. C. (2019). Effectiveness of a partial read-aloud test accommodation to assess reading comprehension in students with a reading disability. *Journal of Learning Disabilities*, 52(3), 259-270. https://doi.org/10.1177/0022219418789377

184. Glover, J. A., Ronning, R. R., and Reynolds, C. R. (2013). Glover, J., Ronning, R., & Reynolds, C., (Eds). Handbook of creativity. *Springer Science & Business Media.*

185. Glăveanu, V. P. (2011). Creativity as cultural participation. *Journal for the Theory of Social Behaviour,* 41(1), 48-67, https://doi.org/10.1111/j.1468-5914.2010.00445.x

186. Goldstein, G., Allen, D. N., & DeLuca, J. (2019). Historical perspectives. Handbook of Psychological Assessment (Fourth Ed), 3-27. Academic Press, *Elsevier*. https://doi.org/10.1016/B978-0-12-802203-0.00001-8

187. Gomez, J. G. (2007). What Do We Know about Creativity?. *Journal of Effective Teaching*, 7(1), 31-43. https://eric.ed.gov/?id=EJ1055657

188. Gonzalez-Franco, M. and Peck, T. C. (2018). Avatar embodiment. towards a standardized questionnaire. *Frontiers in Robotics and AI*, 5, 1–9. https://doi.org/10.3389/frobt.2018.00074

189. Goodwin, K. A., & Goodwin, C. J. (2016). *Research in Psychology: Methods and Design.* John Wiley & Sons.

190. Goris, K., Saldien, J., & Lefeber, D. (2009). Probo, a testbed for human robot interaction. In 2009 4th ACM/ *IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 253-254). IEEE. https://ieeexplore.ieee.org/abstract/document/6256050

191. Gosling, S. D., Rentfrow, P. J., & Swann Jr, W. B. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality,* 37(6), 504-528. https://doi.org/10.1016/S0092-6566(03)00046-1

192. Götz, I. L. (1981). On defining creativity. *The Journal of Aesthetics and Art Criticism*, 39(3), 297-301. https://doi.org/10.2307/430164

193. Gough, H. G. (1979). A creative personality scale for the Adjective Check List. *Journal of Personality and Social Psychology,* 37(8), 1398. https://doi.org/10.1037/0022-3514.37.8.1398

194. Gough, H. G. (1984). A managerial potential scale for the California Psychological Inventory. *Journal of Applied Psychology,* 69(2), 233. https://doi.org/10.1037/0021-9010.69.2.233

195. Greitemeyer, T. (2018). The spreading impact of playing violent video games on aggression. *Computers in Human Behavior,* 80, 216-219. https://doi.org/10.1016/j.chb.2017.11.022

196. Greitemeyer, T. (2019). The contagious impact of playing violent video games on aggression: Longitudinal evidence. *Aggressive behavior,* 45(6), 635-642. https://doi.org/10.1002/ab.21857

197. Griffin, G. R., & Hopson, J. A. (1978). An Evaluation of the Omnibus Personality Inventory in the Prediction of Attrition in Naval Aviation Training. *Naval Aerospace Medical Research Lab Pensacola FL.* https://apps.dtic.mil/sti/citations/ADA065001

198. Griffin, P., Care, E., & McGaw, B. (2012). The changing role of education and schools. In Assessment and teaching of 21st century skills (pp. 1-15). *Springer,* Dordrecht. https://doi.org/10.1007/978-94-007-2324-5_1

199. Gül, M., & Akcay, H. (2020). Structuring a new socioscientific issues (SSI) based instruction model: Impacts on pre-service science teachers (PSTs) critical thinking skills and dispositions. *International Journal of Research in Education and Science,* 6(1), 141-159. https://www.ijres.net/index.php/ijres

200. Guo, S. (2020). Synchronous versus asynchronous online teaching of physics during the COVID-19 pandemic. *Physics Education*, 55(6), 065007. https://iopscience.iop.org/article/10.1088/1361-6552/aba1c5/meta

201. Haag, A., Goronzy, S., Schaich, P., & Williams, J. (2004). Emotion recognition using bio-sensors: First steps towards an automatic system. In Tutorial and research workshop on affective dialogue systems, 36-48. *Springer, Berlin, Heidelberg.* https://doi.org/10.1007/978-3-540-24842-2_4

202. Håkansson, J., & Montgomery, H. (2003). Empathy as an interpersonal phenomenon. *Journal of Social and Personal Relationships*, 20(3), 267-284. https://doi.org/10.1177/0265407503020003001

203. Häkkinen, P., Järvelä, S., Mäkitalo-Siegl, K., Ahonen, A., Näykki, P., & Valtonen, T. (2017). Preparing teacher-students for twenty-first-century learning practices (PREP 21): a framework for enhancing collaborative problem-solving and strategic learning skills. *Teachers and Teaching,* 23(1), 25-41. https://doi.org/10.1080/13540602.2016.1203772

204. Halpern, D. F. (2016). Manual Halpern Critical Thinking Assessment. *Modling: Schucfried. Austria, Vienna Test System Edition.*

205. Hamlen, K. R. (2013). Trends in children's video game play: practical but not creative thinking. *Journal of Educational Computing Research*, 49(3), 277-291. https://doi.org/10.2190/EC.49.3.a

206. Han, K. S. (2003). Domain-specificity of creativity in young children: How quantitative and qualitative data support it. *The Journal of Creative Behavior*, 37(2), 117-142. https://doi.org/10.1002/j.2162-6057.2003.tb00829.x

207. Happ, C., Melzer, A., & Steffgen, G. (2013). Superman vs. BAD man? The effects of empathy and game character in violent video games. Cyberpsychology, *Behavior, and Social Networking,* 16(10), 774-778. https://doi.org/10.1089/cyber.2012.0695

208. Hee Kim, K. (2006). Is creativity unidimensional or multidimensional? Analyses of the Torrance Tests of Creative Thinking. *Creativity Research Journal,* 18(3), 251-259. https://doi.org/10.1207/s15326934crj1803_2

209. Heerink, M., Kröse, B. J., Wielinga, B., & Evers, V. (2009). Measuring the influence of social abilities on acceptance of an interface robot and a screen agent by elderly users. *People and Computers XXIII Celebrating People and Technology,* 430–439.

210. Heider, F., & Simmel, M. (1944). An experimental study of apparent behavior. *The American Journal of Psychology,* 57(2), 243-259. https://doi.org/10.2307/1416950

211. Heist, P. A., Yonge, G. D., Webster, H., and McConnell, T. R. (1968). Omnibus Personality Inventory: Form F Manual [by] Paul Heist [and] George Yonge. *Psychological Corporation.*

212. Helson, R. (1965). Childhood interest clusters related to creativity in women. *Journal of Consulting Psychology,* 29(4), 352. https://doi.org/10.1037/h0022401

213. Henneman, E. A., Lee, J. L., & Cohen, J. I. (1995). Collaboration: a concept analysis. *Journal of Advanced Nursing,* 21(1), 103-109. https://doi.org/10.1046/j.1365-2648.1995.21010103.x

214. Hensley, W. E. (1986). A new look at nonverbal expressiveness: The Affective Communication Test (act-10, Report). https://eric.ed.gov/?id=ED274008

215. Higuchi, T., & Troutt, M. D. (2004). Dynamic simulation of the supply chain for a short life cycle product—Lessons from the Tamagotchi case. *Computers* & *Operations Research*, 31(7), 1097-1114. https://doi.org/10.1016/S0305-0548(03)00067-4

216. Ho, C. C., & MacDorman, K. F. (2010). Revisiting the uncanny valley theory: Developing and validating an alternative to the Godspeed indices. *Computers in Human Behavior,* 26(6), 1508-1518. https://doi.org/10.1016/j.chb.2010.05.015

217. Hoorn, J., Eliëns, A., Huang, Z., Van Vugt, H. C., Konijn, E. A., & Visser, C. T. (2004). Agents with character: Evaluation of empathic agents in digital dossiers. *Emphatic Agents, AAMAS.* https://www.cs.vu.nl/~eliens/rif/onderwijs/research/cv/research/multimedia/@archive/publications/FinalVersionHoornetal.pdf

218. Chin-Chang, H., & MacDorman, K. F. (2017). Measuring the Uncanny Valley Effect. *International Journal of Social Robotic*s, 9(1), 129-139. https://doi.org/10.1007/s12369-016-0380-9

219. Hocevar, D. (1981). Measurement of creativity: Review and critique. *Journal of Personality Assessment,* 45(5), 450-464. https://doi.org/10.1207/s15327752jpa4505_1

220. Hodgins, J., Jörg, S., O'Sullivan, C., Park, S. I., & Mahler, M. (2010). The saliency of anomalies in animated human characters. *ACM Transactions on Applied Perception* (TAP), 7(4), 1-14. https://doi.org/10.1145/1823738.1823740

221. Hoffmann, H., Kessler, H., Eppel, T., Rukavina, S., & Traue, H. C. (2010). Expression intensity, gender and facial emotion recognition: Women recognize only subtle facial emotions better than men. *Acta Psychologica*, 135(3), 278-283. https://doi.org/10.1016/j.actpsy.2010.07.012

222. Hofstede, G. J., De Caluwé, L., & Peters, V. (2010). Why simulation games work-in search of the active substance: A synthesis. *Simulation & Gaming,* 41(6), 824-843. https://doi.org/10.1177/1046878110375596

223. Hoge, R. D., & Butcher, R. (1984). Analysis of teacher judgments of pupil achievement levels. J*ournal of Educational Psychology,* 76(5), 777. https://doi.org/10.1037/0022-0663.76.5.777

224. Hojat, M. (2016). Empathy in health professions education and patient care, 57–68. Switzerland. *Springer Nature.* https://doi.org/10.1007/978-3-319-27625-0

225. Hollis, H., Rachitskiy, M., Van der Leer, L., & Elder, L. (2020). Validity and reliability testing of the International Critical Thinking Essay Test form A (ICTET-A), *Psychological Reports.*

226. Holmgren, R. A., Eisenberg, N., & Fabes, R. A. (1998). The relations of children's situational empathy-related emotions to dispositional prosocial behaviour. *International Journal of Behavioral Development,* 22(1), 169-193. https://doi.org/10.1080/016502598384568

227. Holton, J.A. (2001), "Building trust and collaboration in a virtual team". *Team Performance Management*, 7 (3), 36-47. https://doi.org/10.1108/13527590110395621

228. Hou, J., Nam, Y., Peng, W., & Lee, K. M. (2012). Effects of screen size, viewing angle, and players' immersion tendencies on game experience. *Computers in Human Behavior,* 28(2), 617-623. https://doi.org/10.1016/j.chb.2011.11.007

229. Howard, A., & Choi, M. (2000). How Do You Assess a Manager's Decision-Making Abilities? The Use of Situational Inventories. *International Journal of Selection and Assessment,* 8(2), 85-88. https://doi.org/10.1111/1468-2389.00136

230. Hu, W., & Adey, P. (2002). A scientific creativity test for secondary school students. *International Journal of Science Education*, 24(4), 389-403. https://doi.org/10.1080/09500690110098912

231. Hu, W., Wu, B., Jia, X., Yi, X., Duan, C., Meyer, W., & Kaufman, J. C. (2013). Increasing students' scientific creativity: The "learn to think" intervention program. *The Journal of Creative Behavior, 47*(1), 3-21. https://doi.org/10.1002/jocb.20

232. Huang, C. F., & Wang, K. C. (2019). Comparative analysis of different creativity tests for the prediction of students' scientific creativity. *Creativity Research Journal,* 31(4), 443-447. https://doi.org/10.1080/10400419.2019.1684116

233. Huang, P. S., Peng, S. L., Chen, H. C., Tseng, L. C., & Hsu, L. C. (2017). The relative influences of domain knowledge and domain-general divergent thinking on scientific creativity and mathematical creativity. *Thinking Skills and Creativity,* 25, 1-9. https://doi.org/10.1016/j.tsc.2017.06.001

234. Huhn, K., Black, L., Jensen, G. M., & Deutsch, J. E. (2011). Construct validity of the health science reasoning test. *Journal of allied health,* 40(4), 181-186.

235. Hullman, G. A., Planisek, A., McNally, J. S., & Rubin, R. B. (2010). Competence, personality, and self-efficacy: Relationships in an undergraduate interpersonal course. *Atlantic Journal of Communication,* 18(1), 36-49. https://doi.org/10.1080/15456870903340506

236. Hur, J. W., Shen, Y. W., & Cho, M. H. (2020). Impact of intercultural online collaboration project for pre-service teachers. *Technology, Pedagogy and Education*, 29(1), 1-17. https://doi.org/10.1080/1475939X.2020.1716841

237. Ickes, W., & Decety, J. (2009). Introduction: Seeking to understand the minds (and brains) of people who are seeking to understand other people's minds. *The Social Neuroscience of Empathy.*

238. Intons-Peterson, M. J. (1981). Constructing and using unusual and common images. *Journal of Experimental Psychology: Human Learning and Memory,* 7(2), 133. https://doi.org/10.1037/0278-7393.7.2.133

239. Ishiguro, H., Nishio, S. (2018). Building Artificial Humans to Understand Humans. In: Ishiguro, H., Dalla Libera, F. (eds) Geminoid Studies. *Springer, Singapore*. https://doi.org/10.1007/978-981-10-8702-8_2

240. Jackson, L. A., & Games, A. I. (2015). Video games and creativity. Video games and creativity, 3-38. Academic Press. *Elsevier.* https://doi.org/10.1016/B978-0-12-801462-2.00001-1

241. Jenkins, J. R., & Jewell, M. (1993). Examining the validity of two measures for formative teaching: Reading aloud and maze. *Exceptional Children*, 59(5), 421-432. https://doi.org/10.1177/001440299305900505

242. Jennett, C., Cox, A. L., Cairns, P., Dhoparee, S., Epps, A., Tijs, T., & Walton, A. (2008). Measuring and defining the experience of immersion in games. *International Journal of Human-Computer Studies,* 66(9), 641-661. https://doi.org/10.1016/j.ijhcs.2008.04.004

243. John, D. (1997). How We Think. *Dovey Publications,* Mineola, New York.

244. John-Steiner, V., Weber, R. J., & Minnis, M. (1998). The challenge of studying collaboration. *American Educational Research Journal*, 35(4), 773-783. https://doi.org/10.3102/00028312035004773

245. Kabeel, A. R., & Eisa, S. A. E. M. M. (2016). The Correlation of Critical Thinking Disposition and Approaches to Learning among Baccalaureate Nursing Students. *Journal of Education and Practice,* 7(32), 91-103.

246. Kalisch, B. J. (1973). What is empathy?. *The American Journal of Nursing,* 1548-1552. https://doi.org/10.2307/3422614

247. Kandel, K., Huettmann, F., Suwal, M. K., Regmi, G. R., Nijman, V., Nekaris, K. A. I., ... & Subedi, T. R. (2015). Rapid multi-nation distribution assessment of a charismatic conservation species using open access ensemble model GIS predictions: Red panda (Ailurus fulgens) in the Hindu-Kush Himalaya region. *Biological Conservation*, 181, 150-161. https://doi.org/10.1016/j.biocon.2014.10.007

248. Kano, Y., & Morita, J. (2019, September). Factors Influencing Empathic Behaviors for Virtual Agents: -Examining about the Effect of Embodiment. Proceedings of the 7th International Conference on Human-Agent Interaction, 236-238. https://doi.org/10.1145/3349537.3352777

249. Karduz, F. F. A., & Şar, A. H. (2019). The Effect of Psycho-Education Program On Increase The Tendency To Forgive and Five Factor Personality Properties Of Forgiveness Tendency. *International Journal of Psychology and Educational Studies,* 6(3), 89-105. https://doi.org/10.17220/ijpes.2019.03.010

250. Karpova, E., Marcketti, S. B., & Barker, J. (2011). The efficacy of teaching creativity: Assessment of student creative thinking before and after exercises. *Clothing and Textiles Research Journal,* 29(1), 52-66. https://doi.org/10.1177/0887302X11400065

251. Katrimpouza, A., Tselios, N., & Kasimati, M. C. (2019). Twitter adoption, students' perceptions, Big Five personality traits and learning outcome: Lessons learned from 3 case studies. *Innovations in Education and Teaching International,* 56(1), 25-35. https://doi.org/10.1080/14703297.2017.1392890

252. Kätsyri, J., Förger, K., Mäkäräinen, M., & Takala, T. (2015). A review of empirical evidence on different uncanny valley hypotheses: support for perceptual mismatch as one road to the valley of eeriness. *Frontiers in Psychology,* 390. https://doi.org/10.3389/fpsyg.2015.00390

253. Katz, J. S., & Martin, B. R. (1997). What is research collaboration?. *Research Policy,* 26(1), 1-18. https://doi.org/10.1016/S0048-7333(96)00917-1

254. Katz, L. J. and Slomka, G. T. (2000). Handbook of Psychological Assessment. Achievement testing, Oxford, 149–182. *Elsevier* Science Kidlington

255. Kaufman, S. B., Quilty, L. C., Grazioplene, R. G., Hirsh, J. B., Gray, J. R., Peterson, J. B., & DeYoung, C. G. (2016). Openness to experience and intellect differentially predict creative achievement in the arts and sciences. *Journal of Personality,* 84(2), 248-258. https://doi.org/10.1111/jopy.12156

256. Ke, F., & Moon, J. (2018). Virtual collaborative gaming as social skills training for high-functioning autistic children. *British Journal of Educational Technology,* 49(4), 728-741. https://doi.org/10.1111/bjet.12626

257. Kearney, P., & Beatty, M. J. (2020). Measures of instructional communication. (1st Ed). *Communication Research Measures,* 7-20. Routledge.

258. Keen, S. (2006). A theory of narrative empathy. *Narrative,* 14(3), 207-236.

259. Kelley, D. L. (1999). Measurement made accessible: A research approach using qualitative, quantitative and quality improvement methods, United Kingdom. *Sage Publications.*

260. Kelly, K. R., & Jugovic, H. (2001). Concurrent validity of the online version of the Keirsey Temperament Sorter II. *Journal of Career Assessment,* 9(1), 49-59. https://doi.org/10.1177/106907270100900104

261. Kelly, M., Svrcek, C., King, N., Scherpbier, A., & Dornan, T. (2020). Embodying empathy: A phenomenological study of physician touch. *Medical Education*, 54(5), 400-407. https://doi.org/10.1111/medu.14040

262. Kertész, C., & Turunen, M. (2019). *Exploratory analysis of Sony AIBO users. Ai & Society,* 34(3), 625-638. https://doi.org/10.1007/s00146-018-0818-8

263. Khalili, H., & HOSSEIN, Z. M. (2003). Investigation of reliability, validity and normality Persian version of the California Critical Thinking Skills Test; Form B (CCTST). *Journal of Medical Education*, 3, 1; 29 - 32.

264. Khalili, H., & Soleymani, M. (2003). Determination of reliability, validity and norm of California critical thinking skills test, form B. *Journal of Babol University of Medical Sciences,* 5(2), 84–90.

265. Khoda, M. K., Saeid, A. M., Alavi, M. H., Yaghmaei, F., & Shahabi, M. (2007). Translation and psychometric properties of California critical thinking skills test (form B), 16 (55), 12-19. https://www.sid.ir/en/Journal/ViewPaper.aspx?ID=109702

266. Keirsey, D., & Bates, M. M. (1984). Please understand me: Character & temperament types, 3-4. Del Mar. *Prometheus Nemesis Book Company*.

267. Kilic, I., Yazici, T., & Topalak, S. I. (2017). Critical thinking disposition of music teachers. *Eurasian Journal of Educational Research,* 17(72), 185-202. https://doi.org/10.14689/ejer.2017.72.10

268. Kilteni, K., Groten, R., & Slater, M. (2012). The sense of embodiment in virtual reality. *Presence: Teleoperators and Virtual Environments,* 21(4), 373-387. https://doi.org/10.1162/PRES_a_00124

269. Kim, B., Bruce, M., Brown, L., de Visser, E., & Phillips, E. (2020). A Comprehensive Approach to Validating the Uncanny Valley using the Anthropomorphic RoBOT (ABOT) *Database, 2020 Systems and Information Engineering Design Symposium* (SIEDS), 2020, 1-6. doi: https://doi.org/10.1109/SIEDS49339.2020.9106675

270. Kim, K. H. (2006). Can we trust creativity tests? A review of the Torrance Tests of Creative Thinking (TTCT). *Creativity Research Journal,* 18(1), 3-14. https://doi.org/10.1207/s15326934crj1801_2

271. Kim, K. H. (2019). Demystifying creativity: What creativity isn't and is?. *Roeper Review,* 41(2), 119-128. https://doi.org/10.1080/02783193.2019.1585397

272. Kim, S. J., Kim, K. H., & Mattila, P. (2012). The role of fashion in the characters of online games. *Journal of Global Fashion Marketing,* 3(2), 81-88. https://doi.org/10.1080/20932685.2012.10593110

273. Kim, Y., & Lee, H. (2022). Falling in love with virtual reality art: a new perspective on 3D immersive virtual reality for future sustaining art consumption. *International Journal of Human–Computer Interaction*, 38(4), 371-382. https://doi.org/10.1080/10447318.2021.1944534

274. Kirkagac, S., & Öz, H. (2017). The Role of Big Five Personality Traits in Predicting Prospective EFL Teachers' Academic Achievement. *Online Submission,* 4(4), 317-328.

275. Kivunja, C. (2014). Innovative pedagogies in higher education to become effective teachers of 21st century skills: Unpacking the learning and innovations skills domain of the new learning paradigm. *International Journal of Higher Education*, 3(4), 37-48. http://dx.doi.org/10.5430/ijhe.v3n4p37

276. Kizilet, A. (2017). Investigation of Critical Thinking Attitudes and Reading Habits of Teacher Candidates. *Educational Research and Reviews,* 12(6), 323-328. https://doi.org/10.5897/ERR2016.2939

277. Klebig, B., Goldonowicz, J., Mendes, E., Miller, A. N., & Katt, J. (2016). The combined effects of instructor communicative behaviors, instructor credibility, and student personality traits on incivility in the college classroom. *Communication Research Reports,* 33(2), 152-158. https://doi.org/10.1080/08824096.2016.1154837

278. Kline, P. (2000). The handbook of psychological testing. London. *Routledge.* https://doi.org/10.4324/9781315812274

279. Kober, S. E., & Neuper, C. (2013). Personality and presence in virtual reality: Does their relationship depend on the used presence measure?. *International Journal of Human-Computer Interaction,* 29(1), 13-25. https://doi.org/10.1080/10447318.2012.668131

280. Kocur, M., Graf, S., and Schwind, V. (2020). The impact of missing fingers in virtual reality. *26th ACM Symposium on Virtual Reality Software and Technology,* 1–5. https://doi.org/10.1145/3385956.3418973

281. Kokkinara, E., & Slater, M. (2014). Measuring the effects through time of the influence of visuomotor and visuotactile synchronous stimulation on a virtual body ownership illusion. *Perception,* 43(1), 43-58. https://doi.org/10.1068/p7545

282. Konrath, S. H., O'Brien, E. H., & Hsing, C. (2011). Changes in dispositional empathy in American college students over time: A meta-analysis. *Personality and Social Psychology Review,* 15(2), 180-198. https://doi.org/10.1177%2F1088868310377395

283. Kors, M. J., Van Der Spek, E. D., Bopp, J. A., Millenaar, K., Van Teutem, R. L., Ferri, G., & Schouten, B. A. (2020). The curious case of the transdiegetic cow, or a mission to foster other-oriented empathy through virtual reality. I*n Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems,* 1-13. https://doi.org/10.1145/3313831.3376748

284. Köseoglu, Y. (2016). To What Extent Can the Big Five and Learning Styles Predict Academic Achievement. J*ournal of Education and Practice,* 7(30), 43-51.

285. Kotzé, M., & Griessel, L. (2008). Predicting the Academic Performance of MBA Students: A South African Study. *International Journal of the Humanities,* 6(1).

286. Kralova, Z., Tirpakova, A., & Skorvagova, E. (2018). Personality Factors and Foreign Language Pronunciation Anxiety: The Effect of Psycho-Social Training. *European Journal of Contemporary Education,* 7(4), 728-740. https://doi.org/10.13187/ejced.2018.4.728

287. Krekhov, A., Cmentowski, S., Emmerich, K., & Krüger, J. (2019a). Beyond human: Animals as an escape from stereotype avatars in virtual reality games. *In Proceedings of the Annual Symposium on Computer-Human Interaction in Play,* 439-451. https://doi.org/10.1145/3311350.3347172

288. Krekhov, A., Cmentowski, S., & Krüger, J. (2019b). The illusion of animal body ownership and its potential for virtual reality games. In 2019 IEEE Conference on Games (CoG) (pp. 1-8). *IEEE*. https://doi.org/10.1109/CIG.2019.8848005

289. Ku, K. Y. (2009). Assessing students' critical thinking performance: Urging for measurements using multi-response format. *Thinking Skills and Creativity*, 4(1), 70-76. https://doi.org/10.1016/j.tsc.2009.02.001

290. Kuchaiev, O., Li, J., Nguyen, H., Hrinchuk, O., Leary, R., Ginsburg, B., ... & Cohen, J. M. (2019). Nemo: a toolkit for building ai applications using neural modules. *ArXiv.* https://doi.org/10.48550/arXiv.1909.09577

291. Kühn, S., Kugler, D. T., Schmalen, K., Weichenberger, M., Witt, C., & Gallinat, J. (2019). Does playing violent video games cause aggression? A longitudinal intervention study. *Molecular Psychiatry*, 24(8), 1220-1234. https://doi.org/10.1038/s41380-018-0031-7

292. Kulig, T. C., Cullen, F. T., Wilcox, P., & Chouhy, C. (2019). Personality and adolescent school-based victimization: do the big five matter?. *Journal of School Violence*, 18(2), 176-199. https://doi.org/10.1080/15388220.2018.1444495

293. Kuncel, N. R., Credé, M., & Thomas, L. L. (2007). A meta-analysis of the predictive validity of the Graduate Management Admission Test (GMAT) and undergraduate grade point average (UGPA) for graduate student academic performance. *Academy of Management Learning & Education,* 6(1), 51-68. https://doi.org/10.5465/amle.2007.24401702

294. Kuncel, N. R., Credé, M., Thomas, L. L., Klieger, D. M., Seiler, S. N., & Woo, S. E. (2005). A meta-analysis of the validity of the Pharmacy College Admission Test (PCAT) and grade predictors of pharmacy student performance. *American Journal of Pharmaceutical Education*, 69(3).

295. Kupsala, S., Jokinen, P., & Vinnari, M. (2013). Who cares about farmed fish? Citizen perceptions of the welfare and the mental abilities of fish. *Journal of Agricultural and Environmental Ethics*, 26(1), 119-135. https://doi.org/10.1007/s10806-011-9369-4

296. Kurbakova, S., Volkova, Z., & Kurbakov, A. (2020). Virtual learning and educational environment: New opportunities and challenges under the COVID-19 pandemic. 2020 *The 4th International Conference on Education and Multimedia Technology*, 167-171. https://doi.org/10.1145/3416797.3416838

297. Kyllonen, P. C. (2012). Measurement of 21st century skills within the common core state standards. *In Invitational Research Symposium on Technology Enhanced Assessment*s, 7-8.

298. Laakasuo, M., Palomäki, J., & Köbis, N. (2021). Moral Uncanny Valley: A robot's appearance moderates how its decisions are judged. *International Journal of Social Robotics*, 13(7), 1679-1688.

299. Lamer, S. A., Dvorak, P., Biddle, A. M., Pauker, K., & Weisbuch, M. (2022). The transmission of gender stereotypes through televised patterns of nonverbal bias. *Journal of Personality and Social Psychology.* https://doi.apa.org/doi/10.1037/pspi0000390

300. Lanier, J. (1992). Virtual reality: The promise of the future. *Interactive Learning International,* 8(4), 275-79.

301. Larson, L. C., & Miller, T. N. (2011). 21st century skills: Prepare students for the future. *Kappa Delta Pi Record,* 47(3), 121-123. https://doi.org/10.1080/00228958.2011.10516575

302. Leach, S. M., Immekus, J. C., French, B. F., & Hand, B. (2020). The factorial validity of the Cornell Critical Thinking Tests: A multi-analytic approach. *Thinking Skills and Creativity,* 37, 100676. https://doi.org/10.1016/j.tsc.2020.100676

303. Lee, H., & Kim, K. H. (2011). Can speaking more languages enhance your creativity? Relationship between bilingualism and creative potential among Korean American students with multicultural link. *Personality and Individual Differences,* 50(8), 1186-1190. https://doi.org/10.1016/j.paid.2011.01.039

304. Leite, I., Pereira, A., Martinho, C., & Paiva, A. (2008). Are emotional robots more fun to play with?. I*n RO-MAN 2008-The 17th IEEE International Symposium on Robot and Human Interactive Communication,* 77-82. IEEE. https://doi.org/10.1109/ROMAN.2008.4600646

305. Lennartsson, A. K., Horwitz, E. B., Theorell, T., & Ullén, F. (2017). Creative artistic achievement is related to lower levels of alexithymia. *Creativity Research Journal*, 29(1), 29-36. https://doi.org/10.1080/10400419.2017.1263507

306. Li, J. (2015). The benefit of being physically present: A survey of experimental works comparing copresent robots, telepresent robots and virtual agents. *International Journal of Human-Computer Studies,* 77, 23-37. https://doi.org/10.1016/j.ijhcs.2015.01.001

307. Li, J., & Chignell, M. (2011). Communication of emotion in social robots through simple head and arm movements. *International Journal of Social Robotic*s, 3(2), 125-142. https://doi.org/10.1007/s12369-010-0071-x

308. Li, J., Kizilcec, R., Bailenson, J., & Ju, W. (2016). Social robots and virtual agents as lecturers for video instruction. *Computers in Human Behavio*r, 55, 1222-1230. https://doi.org/10.1016/j.chb.2015.04.005

309. Liberman, Z., Woodward, A. L., Keysar, B., & Kinzler, K. D. (2017). Exposure to multiple languages enhances communication skills in infancy. *Developmental Scienc*e, 20(1). https://doi.org/10.1111/desc.12420

310. Linvill, D. L. (2019). Relationships between student traits and instructor feedback orientations. *College Teaching,* 67(4), 233-242. https://doi.org/10.1080/87567555.2019.1650706

311. Liu, L. C. (1998). An exploratory study of reading intervention in an inner-city third grade classroom. *San Jose State University.*

312. Liu, O. L., Mao, L., Frankel, L., & Xu, J. (2016). Assessing critical thinking in higher education: the HEIghten™ approach and preliminary validity evidence. *Assessment & Evaluation in Higher Education*, 41(5), 677-694. https://doi.org/10.1080/02602938.2016.1168358

313. Liu, O.L., Shaw, A., Gu, L., Li, G., Hu, S., Yu, N., Ma, L., Xu, C., Guo, F., Su, Q. & Kardanovaj, E. (2018). Assessing college critical thinking: preliminary results from the Chinese HEIghten® Critical Thinking assessment. *Higher Education Research & Development*, 37(5), 999-1014. https://doi.org/10.1080/07294360.2018.1467381

314. Liu, X., Qian, Y., & Evans, M. (2017). If you were a tree. Virtual reality opens up a novel perspective in storytelling and engagement. It's what seeded our VR film projects TreeSense and Tree. https://www.media.mit.edu/posts/tree-treesense

315. Löffler, D., Dörrenbächer, J., & Hassenzahl, M. (2020). The uncanny valley effect in zoomorphic robots: The U-shaped relation between animal likeness and likeability. *In Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction,* 261-270. https://doi.org/10.1145/3319502.3374788

316. Looije, R., Neerincx, M. A., & Cnossen, F. (2010). Persuasive robotic assistant for health self-management of older adults: Design and evaluation of social behaviors. *International Journal of Human-Computer Studies,* 68(6), 386-397. https://doi.org/10.1016/j.ijhcs.2009.08.007

317. Lu, D., & Xie, Y. (2019). The effects of a critical thinking oriented instructional pattern in a tertiary EFL argumentative writing course. *Higher Education Research & Development*, 38(5), 969-984. https://doi.org/10.1080/07294360.2019.1607830

318. Lubart, T. (1994). Thinking and problem solving. *Creativity,* 289–332. Academic Press.

319. Luecht, R. M., Madsen, M. K., Taugher, M. P., & Petterson, B. J. (1990). Assessing professional perceptions: Design and validation of an interdisciplinary education perception scale. *J Allied Health,* 19(2), 181-191.

320. Lugrin, J.L., Ertl, M., Krop, P., Klüpfel, R., Stierstorfer, S., Weisz, B., Rück, M., Schmitt, J., Schmidt, N. and Latoschik, M.E., (2018). Any body there? avatar visibility effects in a virtual reality game. *In 2018 IEEE conference on virtual reality and 3D user interfaces* (VR), 17-24. https://ieeexplore.ieee.org/abstract/document/8446229

321. Lugrin, J. L., Latt, J., & Latoschik, M. E. (2015). Anthropomorphism and Illusion of Virtual Body Ownership. *In ICAT-EGVE.* 1-8. https://diglib.eg.org/handle/10.2312/egve20151303

322. MacDorman, K. F. (2006). Subjective ratings of robot video clips for human likeness, familiarity, and eeriness: An exploration of the uncanny valley. *In ICCS/CogSci-2006 long symposium: Toward Social Mechanisms of Android Science,* 26-29.

323. MacGinitie, W. H., MacGinitie, R. K., Cooter, R. B., & Curry, S. (1989). Assessment: Gates-Macginitie Reading Tests. *The Reading Teacher,* 43(3), 256-258.

324. Mackay, S. (2004). The role perception questionnaire (RPQ): a tool for assessing undergraduate students' perceptions of the role of other professions. *Journal of Interprofessional Care,* 18(3), 289-302. https://doi.org/10.1080/13561820410001731331

325. Makransky, G., Terkildsen, T. S., & Mayer, R. E. (2019). Adding immersive virtual reality to a science lab simulation causes more presence but less learning. *Learning and Instruction*, 60, 225-236. https://doi.org/10.1016/j.learninstruc.2017.12.007

326. Maloa, F., & Bux, C. (2015). Psychometric Testing as a Predictor of Academic Performance: A South African Experience. *In 5th International Conference on Engaged Management Scholarship: Baltimore,* Maryland. https://dx.doi.org/10.2139/ssrn.2676348

327. Maltepe, S. (2016). Critical Thinking Dispositions of Pre-Service Turkish Language Teachers and Primary Teachers. *Journal of Education and Training Studies,* 4(6), 44-50. http://dx.doi.org/10.11114/jets.v4i6.1353

328. Marinoiu, E., Zanfir, M., Olaru, V., & Sminchisescu, C. (2018). 3d human sensing, action and emotion recognition in robot assisted therapy of children with autism. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,* 2158-2167.

329. Marsh, R. L., Landau, J. D., & Hicks, J. L. (1996). How examples may (and may not) constrain creativity. *Memory & Cognition*, 24(5), 669-680. https://doi.org/10.3758/BF03201091

330. Matamala-Gomez, M., Donegan, T., Bottiroli, S., Sandrini, G., Sanchez-Vives, M. V., & Tassorelli, C. (2019). Immersive virtual reality and virtual embodiment for pain relief. *Frontiers in Human Neuroscience*, 279. https://doi.org/10.3389/fnhum.2019.00279

331. Matravers, D. (2017). Empathy (Key Concepts in Philosophy). UK (1st Ed.). Malden. https://www.amazon.com/Empathy-Concepts-Philosophy-Derek-Matravers/dp/074567075X

332. Matsumoto, D., LeRoux, J., Wilson-Cohn, C., Raroque, J., Kooken, K., Ekman, P., Yrizarry, N., Loewinger, S., Uchida, H., Yee, A. and Amo, L. (2000). A new test to measure emotion recognition ability: Matsumoto and Ekman's Japanese and Caucasian Brief Affect Recognition Test (JACBART). *Journal of Nonverbal Behavior,* 24(3), 179-209. https://doi.org/10.1023/A:1006668120583

333. Mattessich, P. W., & Monsey, B. R. (1992). Collaboration: what makes it work. *A review of research literature on factors influencing successful collaboration.* Amherst.

334. Mattick, K., Bligh, J., Bluteau, P., & Jackson, A. (2009). Readiness for interprofessional learning scale. *Interprofessional Education Making It Happen,* 125, 142.

335. Mäyrä, F., & Ermi, L. (2011). Fundamental components of the gameplay experience. *Digarec Series,* (6), 88-115.

336. McClure, R. F., & Mears, F. G. (1984). Video game players: Personality characteristics and demographic variables. *Psychological Reports*, 55(1), 271-276. https://doi.org/10.2466%2Fpr0.1984.55.1.271

337. McCrae, R. R., Costa Jr, P. T., & Piedmont, R. L. (1993). Folk concepts, natural language, and psychological constructs: The California Psychological Inventory and the five-factor model. *Journal of Personality*, 61(1), 1-26. https://doi.org/10.1111/j.1467-6494.1993.tb00276.x

338. McCroskey, J. C. (2009). Communication apprehension: What have we learned in the last four decades. *Human Communication,* 12(2), 157-171.

339. McKibben, W. B., & Silvia, P. J. (2017). Evaluating the distorting effects of inattentive responding and social desirability on self-report scales in creativity and the arts. *The Journal of Creative Behavior,* 51(1), 57-69. https://doi.org/10.1002/jocb.86

340. McQuiggan, S. W., Rowe, J. P., & Lester, J. C. (2008). The effects of empathetic virtual characters on presence in narrative-centered learning environments. *In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1511-1520. https://doi.org/10.1145/1357054.1357291

341. Mensio, M., Rizzo, G., & Morisio, M. (2018). The rise of emotion-aware conversational agents: threats in digital emotions. *In Companion Proceedings of the Web Conference* 2018, 1541-1544. https://doi.org/10.1145/3184558.3191607

342. Milcent, A. S., Kadri, A., & Richir, S. (2022). Using facial expressiveness of a virtual agent to induce empathy in users. *International Journal of Human–Computer Interaction,* 38(3), 240-252. https://doi.org/10.1080/10447318.2021.1938387

343. Miller, A.H., Feng, W., Fisch, A., Lu, J., Batra, D., Bordes, A., Parikh, D. & Weston, J., (2017). *Parlai: A dialog research software platform*. ArXiv. https://doi.org/10.48550/arXiv.1705.06476

344. Miralles, A., Raymond, M., & Lecointre, G. (2019). Empathy and compassion toward other species decrease with evolutionary divergence time. *Scientific Reports,* 9(1), 1–8.

345. Mirsamadi, S., Barsoum, E., & Zhang, C. (2017). Automatic speech emotion recognition using recurrent neural networks with local attention. *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP),* 2227-2231. https://doi.org/10.1109/ICASSP.2017.7952552

346. Misselhorn, C. (2009). Empathy with inanimate objects and the uncanny valley. *Minds and Machines*, 19(3), 345-359. https://doi.org/10.1007/s11023-009-9158-2

347. Molyneaux, D. G., Steinbrucker, F. T., Wu, Z., Wei, X., Min, J., & Zhang, Y. (2019). Caching and updating of dense 3d reconstruction data. *Library Catalog: Google Patents.* https://patentscope.wipo.int/search/en/detail.jsf?docId=WO2020237089

348. Moore, C. (2016). The future of work: What Google shows us about the present and future of online collaboration. *TechTrends,* 60(3), 233-244. https://doi.org/10.1007/s11528-016-0044-5

349. Morgeson, F. P., Reider, M. H., & Campion, M. A. (2005). Selecting individuals in team settings: The importance of social skills, personality characteristics, and teamwork knowledge. *Personnel Psychology,* 58(3), 583-611. https://doi.org/10.1111/j.1744-6570.2005.655.x

350. Mori, M. (1970). Bukimi no tani [the uncanny valley]. Energy, 7:33–35.

351. Mori, M., MacDorman, K. F., & Kageki, N. (2012). The uncanny valley [from the field]. *IEEE Robotics & Automation Magazine,* 19(2), 98-100. https://doi.org/10.1109/MRA.2012.2192811

352. Moroz, M., & Krol, K. (2018). VR and empathy: the bad, the good, and the paradoxical. In 2018 IEEE *Workshop on Augmented and Virtual Realities for Good (VAR4Good),* 1-4. https://doi.org/10.1109/VAR4GOOD.2018.8576883

353. Morrison, J. A., & Shriberg, L. D. (1992). Articulation testing versus conversational speech sampling. *Journal of Speech, Language, and Hearing Research,* 35(2), 259-273. https://doi.org/10.1044/jshr.3502.259

354. MotionCow (2018). Photo realistic panda bear: rigged & animation ready. https://motioncow.com/index.php/product/panda-bear

355. Mount, M. K., Barrick, M. R., & Strauss, J. P. (1994). Validity of observer ratings of the big five personality factors. *Journal of Applied Psychology,* 79(2), 272. https://psycnet.apa.org/doi/10.1037/0021-9010.79.2.272

356. Mousas, C., Koilias, A., Rekabdar, B., Kao, D., & Anastaslou, D. (2021). Toward understanding the effects of virtual character appearance on avoidance movement behavior. *2021 IEEE Virtual Reality and 3D User Interfaces* (VR), 40-49. https://doi.org/10.1109/VR50410.2021.00024
357. Moutafi, J., Furnham, A., & Paltiel, L. (2005). Can personality factors predict intelligence?. *Personality and Individual Differences,* 38(5), 1021-1033. https://doi.org/10.1016/j.paid.2004.06.023
358. Mubin, O., Stevens, C. J., Shahid, S., Al Mahmud, A., & Dong, J. J. (2013). A review of the applicability of robots in education. *Journal of Technology in Education and Learning,* 1(209-0015), 13. https://doi.org/10.2316/Journal.209.2013.1.209-0015
359. Mundiri, A., Mahmud, M. E., Ubaidillah, M., Azizah, N., Zuhro, L., & Hasanah, F. (2021). Learning agility during pandemic; outstanding strategy in language learning by using zoom application. *In Journal of Physics: Conference Series,* 1779 (1), 012066. IOP Publishing. https://doi.org/10.1088/1742-6596/1779/1/012066
360. Mwalongo, A. I. (2014). Student teacher and lecturer perceptions of the use of asynchronous discussion forums, quizzes and uploaded resources for promoting critical thinking. [Doctoral dissertation, University of Waikato]. Hamilton, New Zealand. https://hdl.handle.net/10289/8848
361. Nacke, L., & Lindley, C. (2008). Boredom, immersion, flow: A pilot study investigating player experience. IADIS International Conference Gaming 2008: Design for engaging experience and social interaction. *IADIS Press.*
362. Nacke, L. E., Stellmach, S., & Lindley, C. A. (2011). Electroencephalographic assessment of player experience: A pilot study in affective ludology. *Simulation & Gaming,* 42(5), 632-655. https://doi.org/10.1177/1046878110378140
363. Neimark, E. D. (1984). A cognitive style change approach to the modification of thinking in college students. *Technical Report, The State University. Rutgers.*
364. Newgent, R. A., Parr, P. H., Newman, I., & Wiggins, K. K. (2004). The Riso-Hudson Enneagram type indicator: Estimates of reliability and validity. *Measurement and evaluation in Counseling and Development,* 36(4), 226-237. https://doi.org/10.1080/07481756.2004.11909744
365. Ng, S. Y., Cheung, K., & Cheng, H. L. (2022). Critical Thinking Cognitive Skills and Their Associated Factors in Chinese Community College Students in Hong Kong. *Sustainability,* 14(3), 1127. https://doi.org/10.3390/su14031127
366. Niay, B., Olivier, A. H., Zibrek, K., Pettré, J., & Hoyet, L. (2020). Walk ratio: Perception of an invariant parameter of human walk on virtual characters. *ACM Symposium on Applied Perception* 2020, 1-9. https://doi.org/10.1145/3385955.3407926
367. Nickerson, R. S., Butler, S. F., & Carlin, M. (2009). Empathy and knowledge projection. *The Social Neuroscience of Empathy*, 43-56.

368. Niemi, H., Harju, V., Vivitsou, M., Viitanen, K., Multisilta, J., & Kuokkanen, A. (2014). Digital storytelling for 21st-century skills in virtual learning environments. *Creative Education.* https://doi.org/10.4236/ce.2014.59078

369. Niewiadomski, R., Ochs, M., & Pelachaud, C. (2008). Expressions of empathy in ECAs. In International *Workshop on Intelligent Virtual Agents,* 37-44. Springer, Berlin, Heidelberg.

370. Niharika, P., Reddy, N. V., Srujana, P., Srikanth, K., Daneswari, V., & Geetha, K. S. (2018). Effects of distraction using virtual reality technology on pain perception and anxiety levels in children during pulp therapy of primary molars. *Journal of Indian Society of Pedodontics and Preventive Dentistry,* 36(4), 364-369.

371. Nowak, K. L., & Biocca, F. (2003). The effect of the agency and anthropomorphism on users' sense of telepresence, copresence, and social presence in virtual environments. *Presence: Teleoperators & Virtual Environments*, 12(5), 481-494. https://doi.org/10.1162/105474603322761289

372. Nowicki, S., & Duke, M. P. (1994). Individual differences in the nonverbal communication of affect: The Diagnostic Analysis of Nonverbal Accuracy Scale. *Journal of Nonverbal Behavior,* 18(1), 9-35. https://doi.org/10.1007/BF02169077

373. Ochs, M., Pelachaud, C., & Sadek, D. (2008). An empathic virtual dialog agent to improve human-machine interaction. *In Proceedings of the 7th International Joint Conference on Autonomous Agents and Multiagent Systems,* 1, 89-96.

374. Oker, A., Pecune, F., & Declercq, C. (2020). Virtual tutor and pupil interaction: A study of empathic feedback as extrinsic motivation for learning. *Education and Information Technologies,* 25(5), 3643-3658. https://doi.org/10.1007/s10639-020-10123-5

375. Oncu, E. C. (2016). Improved Creative Thinkers in a Class: A Model of Activity Based Tasks for Improving University Students' Creative Thinking Abilities. *Educational Research and Reviews,* 11(8), 517-522. https://doi.org/10.5897/ERR2015.2262

376. O'Quin, K., & Besemer, S. P. (1989). The development, reliability, and validity of the revised creative product semantic scale. *Creativity Research Journa*l, 2(4), 267-278. https://doi.org/10.1080/10400418909534323

377. O'Quin, K., & Besemer, S. P. (2006). Using the creative product semantic scale as a metric for results-oriented business. *Creativity and Innovation Management,* 15(1), 34-44. https://doi.org/10.1111/j.1467-8691.2006.00367.x

378. Ordem, E. (2017). Developing Critical-Thinking Dispositions in a Listening/Speaking Class. *English Language Teaching,* 10(1), 50-55. http://dx.doi.org/10.5539/elt.v10n1p50

379. Orhan, A. (2022). California Critical Thinking Disposition Inventory: Reliability Generalization Meta-Analysis. *Journal of Psychoeducational Assessment,* 40(2), 202-220. https://doi.org/10.1177/07342829211048962

380. Otterbacher, J., Ang, C. S., Litvak, M., & Atkins, D. (2017). Show me you care: Trait empathy, linguistic style, and mimicry on Facebook. *ACM Transactions on Internet Technology* (TOIT), 17(1), 1-22. https://doi.org/10.1145/2996188

381. Paiva, A. (2011). Empathy in social agents. *International Journal of Virtual Reality,* 10(1), 1-4. https://doi.org/10.20870/IJVR.2011.10.1.2794

382. Paiva, A., Dias, J., Sobral, D., Aylett, R., Woods, S., Hall, L., & Zoll, C. (2005). Learning by feeling: Evoking empathy with synthetic characters. *Applied Artificial Intelligence,* 19(3-4), 235-266. https://doi.org/10.1080/08839510590910165

383. Paiva, A., Leite, I., Boukricha, H., & Wachsmuth, I. (2017). Empathy in virtual agents and robots: A survey. *ACM Transactions on Interactive Intelligent Systems* (TiiS), 7(3), 1-40. https://doi.org/10.1145/2912150

384. Pallavicini, F., Pepe, A., Caragnano, C. C., & Mantovani, F. (2020). Video games to foster empathy: A critical analysis of the potential of Detroit: Become Human and The Walking Dead. International Conference on Human-Computer Interaction, 212-228. *Springer,* Cham. https://doi.org/10.1007/978-3-030-49108-6_16

385. Pan, X., & Hamilton, A. F. D. C. (2018). Why and how to use virtual reality to study human social interaction: The challenges of exploring a new research landscape. *British Journal of Psychology,* 109(3), 395-417. https://doi.org/10.1111/bjop.12290

386. Panda, A. (2017). Effect of virtual reality distraction on pain perception during dental treatment in children. *Int J Oral Res,* 5, 278-281. https://doi.org/10.5005/jp-journals-10051-0115

387. Papanastasiou, G., Drigas, A., Skianis, C., Lytras, M., & Papanastasiou, E. (2019). Virtual and augmented reality effects on K-12, higher and tertiary education students' twenty-first century skills. *Virtual Reality,* 23(4), 425-436. https://doi.org/10.1007/s10055-018-0363-2

388. Pardede, P. (2020). Integrating the 4Cs into EFL *Integrated Skills Learning. Journal of English Teaching,* 6(1), 71-85. https://doi.org/10.33541/jet.v6i1.190

389. Park, K., & Jeong, Y. S. (2019). Indoor Dialog Agent in Mixed Reality (video). Proceedings of the 17th Annual International Conference on Mobile Systems, Applications, and Services. Seoul, Republic of Korea. *Association for Computing Machinery,* 708-709. https://doi.org/10.1145/3307334.3328595

390. Parker-Oliver, D., Bronstein, L. R., & Kurzejeski, L. (2005). Examining variables related to successful collaboration on the hospice team. *Health & Social Work,* 30(4), 279-286. https://doi.org/10.1093/hsw/30.4.279

391. Parmar, D., Olafsson, S., Utami, D., Murali, P., & Bickmore, T. (2022). Designing empathic virtual agents: manipulating animation, voice, rendering, and empathy to create persuasive agents. *Autonomous Agents and Multi-Agent Systems,* 36(1), 1-24. https://doi.org/10.1007/s10458-021-09539-1

392. Parong, J., & Mayer, R. E. (2018). Learning science in immersive virtual reality. *Journal of Educational Psychology,* 110(6), 785–797. https://doi.org/10.1037/edu0000241

393. Pastor, B. L., & David, L. T. (2017). Relationship between creativity, intelligence and academic achievement amongst primary education students. Bulletin of the Transilvania University of Brasov. *Series VII, Social Sciences and Law.,* 10(2), 123-132.

394. Patel, P., Ivanov, D., Bhatt, S., Mastorakos, G., Birckhead, B., Khera, N., & Vittone, J. (2020). Low-cost virtual reality headsets reduce perceived pain in healthy adults: a multicenter randomized crossover trial. *Games for Health Journal*, 9(2), 129-136. https://doi.org/10.1089/g4h.2019.0052

395. Patient, D. L., & Skarlicki, D. P. (2010). Increasing interpersonal and informational justice when communicating negative news: The role of the manager's empathic concern and moral development. *Journal of Management,* 36(2), 555-578. https://doi.org/10.1177/0149206308328509

396. Paul, E. S. (2000). Empathy with animals and with humans: Are they linked?. *Anthrozoös,* 13(4), 194-202. https://doi.org/10.2752/089279300786999699

397. Paul, R., & Elder, L. (2006). Critical thinking reading & writing test. *Tomales, CA: Foundation for Critical Thinking.* https://www.criticalthinking.org/data/pages/70/e76672d48b191e2978c961904ec604a74e96157b5f0e6.pdf

398. Paulsen, V. H., & Kolstø, S. D. (2022). Students' reasoning when faced with test items of challenging aspects of critical thinking. *Thinking Skills and Creativity*, 43. https://doi.org/10.1016/j.tsc.2021.100969

399. Pease, A., & Colton, S. (2011). On impact and evaluation in computational creativity: A discussion of the Turing test and an alternative proposal. *In Proceedings of the AISB Symposium on AI and Philosophy,* 39, 1–7.

400. Pedwell, C. (2014). Affective relations: The transnational politics of empathy. *Springer.*

401. Pepe, O. (2018). The Impact of the Creativity Levels of the Students Who Study at the School of Physical Education and Sport on Their Critical Thinking Dispositions. *World Journal of Education,* 8(5), 185-191. https://doi.org/10.5430/wje.v8n5p185

402. Perach, R., & Wisman, A. (2019). Can creativity beat death? A review and evidence on the existential anxiety buffering functions of creative achievement. *The Journal of Creative Behavior,* 53(2), 193-210. https://doi.org/10.1002/jocb.171

403. Pereira, A., Martinho, C., Leite, I., & Paiva, A. (2008). iCat, the chess player: the influence of embodiment in the enjoyment of a game. Proceedings of the 7th International Joint *Conference on Autonomous Agents and Multiagent Systems*, 3, 1253-1256. https://www.ifaamas.org/Proceedings/aamas08/proceedings/pdf/paper/AAMAS08_0483.pdf

404. Perry, G. (2020). What Happened to Robin Williams: Creativity, Psychopathology, and Resilience. *Proceedings of the National Conference On Undergraduate Research (NCUR)* 2020, 1-9. http://libjournals.unca.edu/ncur/wp-content/uploads/2021/01/3351-Perry-Grace-FINAL.pdf

405. Pianzola, F., Bálint, K., & Weller, J. (2019). Virtual reality as a tool for promoting reading via enhanced narrative absorption and empathy. *Scientific Study of Literature,* 9(2), 163-194. https://doi.org/10.1075/ssol.19013.pia

406. Pimentel, D., & Kalyanaraman, S. (2021). How Multi-Sensory Animal Embodiment in Virtual Reality Influences Environmental Threat Processing and Conservation Behaviors. *Research Square,* 1–31. https://doi.org/10.21203/rs.3.rs-957911/v1

407. Pitterman, H., & Nowicki Jr, S. (2004). A test of the ability to identify emotion in human standing and sitting postures: The Diagnostic Analysis of Nonverbal Accuracy-2 Posture Test (DANVA2-POS). *Genetic, Social, and General Psychology Monographs*, 130(2), 146-162. https://doi.org/10.3200/MONO.130.2.146-162

408. Plax, T. G., Kearney, P., McCroskey, J. C., & Richmond, V. P. (1986). Power in the classroom VI: Verbal control strategies, nonverbal immediacy and affective learning. *Communication Education*, 35(1), 43-55. https://doi.org/10.1080/03634528609388318

409. Porssut, T., Blanke, O., Herbelin, B., & Boulic, R. (2022). Reaching articular limits can negatively impact embodiment in virtual reality. *Plos One,* 17(3). https://doi.org/10.1371/journal.pone.0255554

410. Prendinger, H., Mori, J., & Ishizuka, M. (2005). Using human physiology to evaluate subtle expressivity of a virtual quizmaster in a mathematical game. *International Journal of Human-Computer Studies,* 62(2), 231-245. https://doi.org/10.1016/j.ijhcs.2004.11.009

411. Ramirez, E. (2017). Empathy and the limits of thought experiments. *Metaphilosophy,* 48(4), 504-526. https://doi.org/10.1111/meta.12249

412. Rapuano, M., Sbordone, F. L., Borrelli, L. O., Ruggiero, G., & Iachini, T. (2021). The effect of facial expressions on interpersonal space: a gender study in immersive virtual reality. *In Progresses in Artificial Intelligence and Neural Systems,* 477-486. Springer, Singapore. https://doi.org/10.1007/978-981-15-5093-5_40

413. Rear, D. (2019). One size fits all? The limitations of standardised assessment in critical thinking. *Assessment & Evaluation in Higher Education*, 44(5), 664-675. https://doi.org/10.1080/02602938.2018.1526255

414. Rebol, M., Güti, C., & Pietroszek, K. (2021). Passing a non-verbal turing test: Evaluatina gesture animations generated from speech. *2021 IEEE Virtual Reality and 3D User Interfaces (VR*), 573-581. https://doi.org/10.1109/VR50410.2021.00082

415. Reid, A., & Petocz, P. (2004). Learning domains and the process of creativity. *The Australian Educational Researcher,* 31(2), 45-62.

416. Richmond, V. P., Gorham, J. S., & McCroskey, J. C. (1987). The relationship between selected immediacy behaviors and cognitive learning. *Annals of the International Communication Association*, 10(1), 574-590. https://doi.org/10.1080/23808985.1987.11678663

417. Rieger, S., Göllner, R., Spengler, M., Trautwein, U., Nagengast, B., & Roberts, B. W. (2017). Social cognitive constructs are just as stable as the Big Five between grades 5 and 8. *AERA Open,* 3(3). https://doi.org/10.1177/2332858417717691

418. Riek, L. D., Rabinowitch, T. C., Chakrabarti, B., & Robinson, P. (2009). How anthropomorphism affects empathy toward robots. *In Proceedings of the 4th ACM/IEEE International Conference on Human Robot Interaction*, 245-246. https://doi.org/10.1145/1514095.1514158

419. Rios, J. A., Ling, G., Pugh, R., Becker, D., & Bacall, A. (2020). Identifying critical 21st-century skills for workplace success: A content analysis of job advertisements. *Educational Researcher,* 49(2), 80-89. https://doi.org/10.3102/0013189X19890600

420. Riso, D. R. & Hudson, R. (1996). Personality types: Using the Enneagram for self-discovery. New York. *Houghton Mifflin Harcourt.*

421. Ritter, S. M., & Mostert, N. (2017). Enhancement of creative thinking skills using a cognitive-based creativity training. *Journal of Cognitive Enhancement*, 1(3), 243-253. https://doi.org/10.1007/s41465-016-0002-3

422. Rivers, A., Wickramasekera, I. E., Pekala, R. J., & Rivers, J. A. (2016). Empathic features and absorption in fantasy role-playing. *American Journal of Clinical Hypnosis*, 58(3), 286-294. https://doi.org/10.1080/00029157.2015.1103696

423. Roberts, C. V. (1986). A validation of the Watson-Barker Listening Test. *Communication Research Reports*, 3(1).

424. Robinson, H., MacDonald, B., Kerse, N., & Broadbent, E. (2013). The psychosocial effects of a companion robot: a randomized controlled trial. *Journal of the American Medical Directors Association*, 14(9), 661-667. https://doi.org/10.1016/j.jamda.2013.02.007

425. Rodriguez, C. M. (2013). Analog of parental empathy: Association with physical child abuse risk and punishment intentions. *Child Abuse & Neglect*, 37(8), 493-499. https://doi.org/10.1016/j.chiabu.2012.10.004

426. Rosen, A. S., & Schalling, D. (1974). On the validity of the California Psychological Inventory Socialization Scale: A multivariate approach. *Journal of Consulting and Clinical Psychology*, 42(6), 757–765. https://doi.org/10.1037/h0037566

427. Rosenthal-von der Pütten, A. M., Krämer, N. C., Hoffmann, L., Sobieraj, S., & Eimler, S. C. (2013). An experimental study on emotional reactions towards a robot. I*nternational Journal of Social Robotics,* 5(1), 17-34. https://doi.org/10.1007/s12369-012-0173-8

428. Rosip, J. C., & Hall, J. A. (2004). Knowledge of nonverbal cues, gender, and nonverbal decoding accuracy. J*ournal of Nonverbal Behavior,* 28(4), 267-286. https://doi.org/10.1007/s10919-004-4159-6

429. Rosson, H., & Weeks, P. P. (2018). "I am different, not less": Temple Grandin and strengths-based leadership education. *Journal of Leadership Education,* 17(3), 231-240. https://doi.org/10.12806/V17/I3/A4

430. Royse, P., Lee, J., Undrahbuyan, B., Hopson, M., & Consalvo, M. (2007). Women and games: Technologies of the gendered self. *New media & society,* 9(4), 555-576. https://doi.org/10.1177%2F1461444807080322

431. Rubin, R. B., & Martin, M. M. (1994). Development of a measure of interpersonal communication competence. *Communication Research Reports*, 11(1), 33-44. https://doi.org/10.1080/08824099409359938

432. Rubin, R. B., Martin, M. M., Bruning, S. S., & Powers, D. E. (1993). Test of a self-efficacy model of interpersonal communication competence. *Communication Quarterly,* 41(2), 210-220. https://doi.org/10.1080/01463379309369880

433. Runco, M. A., Noble, E. P., Reiter-Palmon, R., Acar, S., Ritchie, T., & Yurkovich, J. M. (2011). The genetic basis of creativity and ideational fluency. *Creativity Research Journal,* 23(4), 376-380. https://doi.org/10.1080/10400419.2011.621859

434. Ryokai, K., Lee, M. J., & Breitbart, J. M. (2009). Children's storytelling and programming with robotic characters. *In Proceedings of the seventh ACM conference on Creativity and Cognition*, 19-28. https://doi.org/10.1145/1640233.1640240

435. Saglam, A. Ç., Çankaya, İ., Üçer, H., & Çetin, M. (2017). The Effect of Information Literacy on Teachers' Critical Thinking Disposition. *Journal of Education and Learning*, 6(3), 31-40. http://doi.org/10.5539/jel.v6n3p31

436. Sak, U. and Ayas, M. B. (2013). Creative Scientific Ability Test (C- SAT): A new measure of scientific creativity. *Psychological Test and Assessment Modeling,* 55(3), 316-329. https://doi.org/10.1109/TAFFC.2019.2958657

437. Salminen, M., Járvelá, S., Ruonala, A., Harjunen, V., Jacucci, G., Hamari, J., & Ravaja, N. (2019). Evoking physiological synchrony and empathy using social VR with biofeedback. *IEEE Transactions on Affective Computing.* 13 (2), 746-755. https://doi.org/10.1109/TAFFC.2019.2958657

438. Sanabria, A. A., Restrepo, M. A., Walker, E., & Glenberg, A. (2022). A Reading Comprehension Intervention for Dual Language Learners With Weak Language and Reading Skills. *Journal of Speech, Language, and Hearing Research,* 1-22. https://doi.org/10.1044/2021_JSLHR-21-00266

439. Sánchez, J., & Olivares, R. (2011). Problem solving and collaboration using mobile serious games. *Computers & Education*, 57(3), 1943-1952. https://doi.org/10.1016/j.compedu.2011.04.012

440. Sanguna, J. A. F. (2021). Developing 21st Century Learning Skills in English Learners: A Unit of Study on Social Issues [Doctoral dissertation, Greensboro College]. *ProQuest.* https://www.proquest.com/openview/c1db74c0701f18050890f78daac73ad0/1?pq-origsite=gscholar&cbl=18750&diss=y

441. Saville and Ltd, H. (1991). Management and Graduate Item Bank Manual and User's Guide. *Saville and Holdsworth Ltd Thames Ditton, UK.* https://ptc.bps.org.uk/test-review/management-graduate-item-bank

442. Scassellati, B., Brawer, J., Tsui, K., Nasihati Gilani, S., Malzkuhn, M., Manini, B., Stone, A., Kartheiser, G., Merla, A., Shapiro, A. and Traum, D. (2018). Teaching language to deaf infants with a robot and a virtual human. *In Proceedings of the 2018 CHI Conference on Human Factors in Computing System*s (pp. 1-13). https://doi.org/10.1145/3173574.3174127

443. Schaefer, C. E. (1972). Predictive validity of the Biographical Inventory Creativity: Five-year follow-up study. *Psychological Reports,* 30(2), 471-476. https://doi.org/10.2466/pr0.1972.30.2.471

444. Schaefer, C. E., & Anastasi, A. (1968). A biographical inventory for identifying creativity in adolescent boys. *Journal of Applied Psychology,* 52(1p1), 42. https://doi.org/10.1037/h0025328

445. Schleicher, A. (2012). Preparing teachers and developing school leaders for the 21st century: Lessons from around the world. *OECD* Publishing. 2, rue Andre Pascal, F-75775 Paris Cedex 16, France. http://dx.doi.org/10.1787/9789264174559-en

446. Schloss, I., O. Bailey, J., & Tripathi, S. (2021 "I'm in his belly!": Children's Responses to Different Types of Characters in Virtual Reality. *In Interaction Design and Children*, 43-48. https://doi.org/10.1145/3459990.3460723

447. Schmidt, F. L., Hunter, J. E., & Pearlman, K. (1981). Task differences as moderators of aptitude test validity in selection: A red herring. *Journal of Applied Psychology,* 66(2), 166. https://doi.org/10.1037/0021-9010.66.2.166

448. Schneider, E., Wang, Y., & Yang, S. (2007). Exploring the Uncanny Valley with Japanese Video Game Characters. *In DiGRA Conference,* 546-549. http://www.digra.org/wp-content/uploads/digital-library/07312.11004.pdf

449. Schreiner, L. A. (2006). A technical report on the Clifton StrengthsFinder with college students*,* 1-27. Gallup®, Clifton StrengthsFinder®.

450. Schutte, N. S., Malouff, J. M., Hall, L. E., Haggerty, D. J., Cooper, J. T., Golden, C. J., & Dornheim, L. (1998). Development and validation of a measure of emotional intelligence. *Personality and Individual Differences,* 25(2), 167-177. https://doi.org/10.1016/S0191-8869(98)00001-4

451. Schutte, N. S., & Stilinović, E. J. (2017). Facilitating empathy through virtual reality. *Motivation and emotion*, 41(6), 708-712. https://doi.org/10.1007/s11031-017-9641-7

452. Schwind, V., Knierim, P., Chuang, L., & Henze, N. (2017). " Where's Pinky?" The Effects of a Reduced Number of Fingers in Virtual Reality. *In Proceedings of the Annual Symposium on Computer-Human Interaction in Play*, 507-515. https://doi.org/10.1145/3116595.3116596

453. Schwind, V. (2018). Implications of the uncanny valley of avatars and virtual characters for human-computer interaction. [Master's thesis, University of Stuttgart]. https://elib.uni-stuttgart.de/handle/11682/9953

454. Schwind, V., Leicht, K., Jäger, S., Wolf, K., & Henze, N. (2018a). Is there an uncanny valley of virtual animals? A quantitative and qualitative investigation. I*nternational Journal of Human-Computer Studies,* 111, 49-61. https://doi.org/10.1016/j.ijhcs.2017.11.003

455. Schwind, V., Wolf, K., & Henze, N. (2018b). Avoiding the uncanny valley in virtual character design. *Interactions,* 25(5), 45-49. https://doi.org/10.1145/3236673

456. Selvaratnam, R. (2021). The link between learning spaces and employability outcomes. *Journal of Learning Spaces,* 10(2). http://libjournal.uncg.edu/jls/article/view/1991

457. Serhan, D. (2020). Transitioning from face-to-face to remote learning: Students' attitudes and perceptions of using Zoom during COVID-19 pandemic. *International Journal of Technology in Education and Science,* 4(4), 335-342.

458. Setiaji, B., & Wibowo, F. W. (2016). Chatbot using a knowledge in database: human-to-machine conversation modeling. In 2016 7th International Conference on Intelligent Systems, *Modelling and Simulation (ISMS)*, 72-77. https://doi.org/10.1109/ISMS.2016.53

459. Sfetsos, P., Stamelos, I., Angelis, L., & Deligiannis, I. (2006). Investigating the impact of personality types on communication and collaboration-viability in pair programming–an empirical study. In International Conference on Extreme Programming and Agile Processes in Software Engineering, 43-52. *Springer,* Berlin, Heidelberg. https://doi.org/10.1007/11774129_5

460. Shabahang, M. J., & Amani, M. (2016). The relationship between personality factors and organizational commitment of Iranian primary school principals. *International Journal of Psychology and Educational Studies,* 3(3), 50-59. https://doi.org/10.17220/ijpes.2016.03.006

461. Shearer, P. J. (1976). Relationship of" r" articulatory proficiency to speech sound discrimination [Doctoral dissertation, Kansas State University]. https://krex.k-state.edu/dspace/bitstream/handle/2097/11445/LD2668T41976S52.pdf?sequence=1

462. Shelton, R. L., Johnson, A. F., & Arndt, W. B. (1977). Delayed judgment speech-sound discrimination and/r/or/s/articulation status and improvement. *Journal of Speech and Hearing Research,* 20(4), 704-717. https://doi.org/10.1044/jshr.2004.704

463. Shen, L. (2010). On a scale of state empathy during message processing. *Western Journal of Communication,* 74(5), 504-524. https://doi.org/10.1080/10570314.2010.512278

464. Sherman, D., & Geith, A. (1967). Speech sound discrimination and articulation skill. *Journal of Speech and Hearing Research,* 10(2), 277-280. https://doi.org/10.1044/jshr.1002.277

465. Sherry, J. L. (2001). The effects of violent video games on aggression: A meta-analysis. *Human Communication Research*, 27(3), 409-431. https://doi.org/10.1111/j.1468-2958.2001.tb00787.x

466. Shinozawa, K., Naya, F., Yamato, J., & Kogure, K. (2005). Differences in effect of robot and screen agent recommendations on human decision-making. *International Journal of Human-Computer Studies*, 62(2), 267-279. https://doi.org/10.1016/j.ijhcs.2004.11.003

467. Sierra Rativa, A. (2018). How can we Teach Educational Robotics to Foster 21st Learning Skills through PBL, Arduino and S4A?. In Lepuschitz W., Merdan M., Koppensteiner G., Balogh R., & Obdrzalek D. (Eds.), RiE 2018: Robotics in Education, vol 1316. Advances in Intelligent Systems and Computing, 149–161. *Springer.* https://doi.org/10.1007/978-3-319-97085-1_ 15

468. Sierra Rativa A., Postma M., & van Zaanen M. (2020). The Uncanny Valley of the Virtual (Animal) Robot. In Merdan M., Lepuschitz W., Koppensteiner G., Balogh R., & Obdrzalek D. (Eds.), RiE 2019: Robotics in Education, Vol 1023. Advances in Intelligent Systems and Computing, 419-4279. *Springer*. https://link.springer.com/chapter/10.1007/978-3-030-26945-6_38

469. Sierra Rativa A., Vasquez C.C., Martinez F., Orejuela Ramirez W., Postma M., & van Zaanen M. (2021). The Effectiveness of a Robot Animal as a Virtual Instructor. In Lepuschitz W., Merdan M., Koppensteiner G., Balogh R., & Obdrˇz´alek D. (Eds.), RiE 2020: Robotics in Education,Vol 1316. Advances in Intelligent Systems and Computing, 329–338. *Springer.* https://link.springer.com/chapter/10.1007/978-3-030-67411-3_30

470. Sierra Rativa, A. S., Postma, M., & Van Zaanen, M. (2020). Can virtual reality act as an affective machine? The wild animal embodiment experience and the importance of appearance. In Urrea C. (Eds.). Proceedings of the MIT LINC 2019 Conference, Vol 3. EPiC *Series in Education Scienc*e, 214–223. EasyChair. https://easychair.org/publications/paper/RFSQ

471. Sierra Rativa, A., Bakker, G. A., & Sierra Rativa, A. G. (2022). Animal Embodiment: Embodying a beaver in immersive virtual environments to create empathy and teach about the impact of global warming in a playful way. In Richir S.(Eds.). Proceedings of Virtual Reality International Conference (VRIC), 33–36. [Poster presentation], Laval, France. Laval Virtual.
472. Sierra Rativa A., Postma M., & van Zaanen M. (2022). The Uncanny valley of the virtual animals. *Computer Animation and Virtual Worlds Journal,* 33(2), 1–21. https://doi.org/10.1002/cav.2043
473. Sierra Rativa, A., Postma, M., & Van Zaanen, M. (2020). The influence of game character appearance on empathy and immersion: Virtual non-robotic versus robotic animals. *Simulation & Gaming,* 51(5), 685–711. https://doi.org/10.1177/1046878120926694
474. Silva, E. (2009). Measuring skills for 21st-century learning. *Phi Delta Kappan,* 90(9), 630-634.
475. Singer, T., Seymour, B., O'doherty, J., Kaube, H., Dolan, R. J., & Frith, C. D. (2004). Empathy for pain involves the affective but not sensory components of pain. *Science*, 303(5661), 1157-1162. https://doi.org/10.1126/science.1093535
476. Siu, A. M., & Shek, D. T. (2005). Validation of the interpersonal reactivity index in a Chinese context. *Research on Social Work Practice,* 15(2), 118-126. https://doi.org/10.1177/1049731504270384
477. Škola, F., & Liarokapis, F. (2021). Study of full-body virtual embodiment using noninvasive brain stimulation and imaging. *International Journal of Human–Computer Interaction,* 37(12), 1116-1129. https://doi.org/10.1080/10447318.2020.1870827
478. Slater, M., Pérez Marcos, D., Ehrsson, H., & Sanchez-Vives, M. V. (2008). Towards a digital body: the virtual arm illusion. *Frontiers in Human Neuroscience,* 2(6), 1-8. https://doi.org/10.3389/neuro.09.006.2008
479. Smid, K. and Pandzic, I. (2002). Conversational virtual character for the Web. *In Proceedings of Computer Animation 2002* (CA 2002), 240–247. https://doi.org/10.1109/CA.2002.1017543
480. Smith, A. R., & Gregg, T. R. (2020). Effects of Strength-Based Development on Resilience in Undergraduate Nursing Students, 1-68. Point Loma Nazarene University.
481. Soni, P., & Bakhru, K. M. (2019). Understanding triangulated collaboration of work-life balance, personality traits and eudaimonic well-being. *Problems and Perspectives in Management,* 17(2), 63. http://dx.doi.org/10.21511/ppm.17(2).2019.05
482. Soto, C. J., & John, O. P. (2009). Using the California Psychological Inventory to assess the Big Five personality domains: A hierarchical approach. *Journal of Research in Personality,* 43(1), 25-38. https://doi.org/10.1016/j.jrp.2008.10.005
483. Spell, L. A., & Frank, E. (2000). Recognition of nonverbal communication of affect following traumatic brain injury. *Journal of Nonverbal Behavior,* 24(4), 285-300. https://doi.org/10.1023/A:1006675230193
484. Spreng, R. N., McKinnon*, M. C., Mar, R. A., & Levine, B. (2009). The Toronto Empathy Questionnaire: Scale development and initial validation of a factor-analytic solution to multiple empathy measures. *Journal of Personality Assessment,* 91(1), 62-71. https://doi.org/10.1080/00223890802484381

485. Stanger, N., Kavussanu, M., & Ring, C. (2012). Put yourself in their boots: Effects of empathy on emotion and aggression. *Journal of Sport and Exercise Psychology,* 34(2), 208-222. https://journals.humankinetics.com/view/journals/jsep/34/2/article-p208.xml
486. Sternberg, R. J., & Lubart, T. I. (1999). The concept of creativity: Prospects and paradigms. Handbook of creativity, USA, 1, 3-15. *Cambridge University Press.*
487. Sternod, L., & French, B. (2016). Test Review: Watson, G., & Glaser, EM (2010). *Watson-Glaser™ II Critical Thinking Apprisa*l. https://doi.org/10.1177/0734282915622855
488. Sterzer, P., Stadler, C., Poustka, F., & Kleinschmidt, A. (2007). A structural neural deficit in adolescents with conduct disorder and its association with lack of empathy. *Neuroimage,* 37(1), 335-342. https://doi.org/10.1016/j.neuroimage.2007.04.043
489. Stevens, S. S. (1950). Introduction: A definition of communication. *The Journal of the Acoustical Society of America,* 22(6), 689-690. https://doi.org/10.1121/1.1906670
490. Stone, C. A., Davidson, L. J., Evans, J. L., & Hansen, M. A. (2001). Validity evidence for using a general critical thinking test to measure nursing students' critical thinking. *Holistic Nursing Practice*, 15(4), 65-74.
491. Stone, G. A., Duffy, L. N., Pinckney, H. P., & Templeton-Bradley, R. (2017). Teaching for critical thinking: preparing hospitality and tourism students for careers in the twenty-first century. *Journal of Teaching in Travel & Tourism,* 17(2), 67-84.
492. Sullivan, C., Schneider, P. E., Musselman, R. J., Dummett Jr, C. O., & Gardiner, D. (2000). The effect of virtual reality during dental treatment on child anxiety and behavior. *ASDC Journal of Dentistry for Children,* 67(3), 193-6.
493. Sustekova, E., Kubiatko, M., & Usak, M. (2019). Validation of critical thinking test on slovak conditions. EURASIA Journal of Mathematics, *Science and Technology Education,* 15(12). https://doi.org/10.29333/ejmste/112295
494. Syed, M., Eriksson, P. L., Frisén, A., Hwang, C. P., & Lamb, M. E. (2020). Personality development from age 2 to 33: Stability and change in ego resiliency and ego control and associations with adult adaptation. *Developmental Psychology,* 56(4), 815. https://doi.org/10.1037/dev0000895
495. Tabak, B.A., Gupta, D., Sunahara, C.S., Alvi, T., Wallmark, Z., Lee, J., Fulford, D., Hudson, N.W. and Chmielewski, M. (2022). Environmental sensitivity predicts interpersonal sensitivity above and beyond Big Five personality traits. *Journal of Research in Personality,* 98, 104210. https://doi.org/10.1016/j.jrp.2022.104210
496. Takayanagi, K., Kirita, T., & Shibata, T. (2014). Comparison of verbal and emotional responses of elderly people with mild/moderate dementia and those with severe dementia in responses to seal robot, PARO. *Frontiers in Aging Neuroscience,* 6 (257), 1-5. https://doi.org/10.3389/fnagi.2014.00257
497. Tam, K. P. (2013). Dispositional empathy with nature. *Journal of Environmental Psychology,* 35, 92-104. https://doi.org/10.1016/j.jenvp.2013.05.004
498. Taube, K. T. (1997). Critical thinking ability and disposition as factors of performance on a written critical thinking test. *The Journal of General Education*, 46(2), 129-164. https://www.jstor.org/stable/27797335?seq=1

499. Templin, M. (1943). A study of sound discrimination ability of elementary school pupils. J*ournal of Speech Disorders*, 8(2), 127-132. https://doi.org/10.1044/jshd.0802.127

500. Templin, M. C., & Darley, F. L. (1960). The Templin-Darley Tests of Articulation: A manual and discussion of the screening and diagnostic tests. Bureau of Educational Research and Service, *Extension Division, State University of Iowa.*

501. Olson, J. M., & Zanna, M. P. (1993). Attitudes and attitude change. *Annual Review of Psychology,* 44 (117), 479–523. https://www.proquest.com/openview/fe8ba6c7ac6d750ed7842a9dfb8eebe0/1?pq-origsite=gscholar&cbl=24892

502. Tesser, A. and Shaffer, D. R. (1990). Attitudes and attitude change.

503. Annual review of psychology, 41(1), 479–523.

504. Thaler, A., Bieg, A., Mahmood, N., Black, M. J., Mohler, B. J., & Troje, N. F. (2020). Attractiveness and confidence in walking style of male and female virtual characters. *In 2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW),* 678-679. https://doi.org/10.1109/VRW50115.2020.00190

505. Thannhauser, J., Russell-Mayhew, S., & Scott, C. (2010). Measures of interprofessional education and collaboration. *Journal of Interprofessional Care,* 24(4), 336-349. https://doi.org/10.3109/13561820903442903

506. Thepsoonthorn, C., Ogawa, K. I., & Miyake, Y. (2021). The exploration of the uncanny valley from the viewpoint of the robot's nonverbal behaviour. *International Journal of Social Robotics,* 13(6), 1443-1455. https://doi.org/10.1007/s12369-020-00726-w

507. Thézé, R., Gadiri, M. A., Albert, L., Provost, A., Giraud, A. L., & Mégevand, P. (2020). Animated virtual characters to explore audio-visual speech in controlled and naturalistic environments. *Scientific Reports*, 10(1), 1-12. https://doi.org/10.1038/s41598-020-72375-y

508. Thissen, D. E., & Wainer, H. E. (2001). Test scoring. *Lawrence Erlbaum Associates Publishers*. https://psycnet.apa.org/record/2001-01226-000

509. Thomas, F., Johnston, O., & Thomas, F. (1995). The illusion of life: Disney animation. *New York: Hyperion.* https://social.cs.uiuc.edu/class/cs498kgk/assignments/04.02.09/das10.pdf

510. Thonney, T., & Montgomery, J. C. (2019). Defining critical thinking across disciplines: An analysis of community college faculty perspectives. *College Teaching,* 67(3), 169-176. https://doi.org/10.1080/87567555.2019.1579700

511. Ting, H. N., Yunus, J., & Nordin, M. M. (2006). Assessing auditory discrimination skill of Malay children using computer-based method. *In 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference,* 2498-2501. https://doi.org/10.1109/IEMBS.2005.1616976

512. Tinwell, A., Grimshaw, M., Nabi, D. A., & Williams, A. (2011). Facial expression of emotion and perception of the Uncanny Valley in virtual characters. *Computers in Human Behavior,* 27(2), 741-749. https://doi.org/10.1016/j.chb.2010.10.018

513. Tiruneh, D. T., De Cock, M., Weldeslassie, A. G., Elen, J., & Janssen, R. (2017). Measuring critical thinking in physics: Development and validation of a critical thinking test in electricity and magnetism. *International Journal of Science and Mathematics Education,* 15(4), 663-682. https://doi.org/10.1007/s10763-016-9723-0

514. Tobe, F. and Greenberg, R. (2004). Podcast: our relationship with the uncanny valley. *Digital Publication.* https://www.therobotreport.com/our-relationship-with-the-uncanny-valley
515. Torrance, E. P. (1972). Predictive validity of the torrance tests of creative thinking. *The Journal of Creative* Behavior. https://doi.org/10.1002/j.2162-6057.1972.tb00936.x
516. Torres Martín, C., Acal, C., El Honrani, M., & Mingorance Estrada, Á. C. (2021). Impact on the virtual learning environment due to COVID-19. S*ustainability*, 13(2), 582. https://doi.org/10.3390/su13020582
517. Trilling, B., & Fadel, C. (2009). 21st century skills: Learning for life in our times. *John Wiley & Sons.*
518. Tsai, K. C. (2016). Quasi-experts' assessments of creative products: An evaluation using a sample of design students. *British Journal of Education, Society & Behavioral Science,* 12 (3), 1-7. https://doi.org/10.9734/BJESBS/2016/21057
519. Tsai, Y. F. (2008). The effects of interacting with a computer-simulated virtual pet dog on children's empathy and humane attitudes. [Doctoral dissertation, Faculty of Education-Simon Fraser University].
520. Tsai, Y. F., & Kaufman, D. (2014). Interacting with a computer-simulated pet: Factors influencing children's humane attitudes and empathy. *Journal of Educational Computing Research,* 51(2), 145-161. https://doi.org/10.2190/EC.51.2.a
521. University, N. T. (2021). PhD student Alexandra Sierra wins the Laval Virtual Europe 2021 award for educational empathic Justin Beaver applications. https://www.tilburguniversity.edu/current/news/more-news/alexandra-sierra-won-laval-virtual-europe-2021-award-justin-beaver
522. Unlu, Z. K., & Dokme, I. (2017). Science Teacher Candidates' Epistemological Beliefs and Critical Thinking Disposition. *Eurasian Journal of Educational Researc*h, 17(72), 203-220. https://dergipark.org.tr/en/pub/ejer/issue/42492/511913
523. Uslu, S. (2020). Critical Thinking Dispositions of Social Studies Teacher Candidates. *Asian Journal of Education and Training,* 6(1), 72-79. https://doi.org/10.20448/journal.522.2020.61.72.79
524. Vaida, M. F. (2019). Collaborative education teams development using alternative methodologies. *In Proceedings of the 2019 11th International Conference on Education Technology and Computers,* 223-227. https://doi.org/10.1145/3369255.3369280
525. Vaida, M. F., & Pop, P. G. (2014). Grouping strategy using Enneagram typologies. In 2014 IEEE International Conference on Automation, *Quality and Testing, Robotics,* 1-6. https://doi.org/10.1109/AQTR.2014.6857844
526. Valle, R., Shih, K., Prenger, R., & Catanzaro, B. (2020). Flowtron: an autoregressive flow-based generative network for text-to-speech synthesis. *ArXiv*. https://doi.org/10.48550/arXiv.2005.05957
527. Van der Merwe, R. P. (2002). Psychometric testing and human resource management. *SA Journal of Industrial Psychology,* 28(2), 77-86. https://hdl.handle.net/10520/EJC88900
528. Van Laar, E., van Deursen, A. J., van Dijk, J. A., & de Haan, J. (2020). Determinants of 21st-century skills and 21st-century digital skills for workers: A systematic literature review. *Sage Open,* 10(1). https://doi.org/10.1177/2158244019900176

529. Van Vugt, H. C., Hoorn, J. F., Konijn, E. A., & de Bie Dimitriadou, A. (2006). Affective affordances: Improving interface character engagement through Interaction. *International Journal of Human-Computer Studies,* 64(9), 874-888. https://doi.org/10.1016/j.ijhcs.2006.04.008

530. Van Vugt, H. C., Konijn, E. A., Hoorn, J. F., & Veldhuis, J. (2009a). When too heavy is just fine: Creating trustworthy e-health advisors. *International Journal of Human-Computer Studies*, 67(7), 571-583. https://doi.org/10.1016/j.ijhcs.2009.02.005

531. Van Vugt, H. C., Hoorn, J. F., & Konijn, E. A. (2009b). Interactive engagement with embodied agents: An empirically validated framework. *Computer Animation and Virtual Worlds,* 20(2-3), 195-204. https://doi.org/10.1002/cav.312

532. Venable, B. (1994). A Philosophical Analysis of Creativity Measurement. https://eric.ed.gov/?id=ED374140

533. Verburgh, A., François, S., Elen, J., & Janssen, R. (2013). The assessment of critical thinking critically assessed in higher education: A validation study of the CCTT and the HCTA. *Education Research International,* 2013. https://doi.org/10.1155/2013/198920

534. Vinayagamoorthy, V., Brogni, A., Gillies, M., Slater, M., & Steed, A. (2004). An investigation of presence response across variations in visual realism. Paper present *In The 7th Annual International Presence Workshop,* 148-155. https://www.researchgate.net/profile/Christoph-Guger/publication/220049115_Navigating_virtual_reality_by_thought_First_steps/links/0fcfd50c1a6ccc1e88000000/Navigating-virtual-reality-by-thought-First-steps.pdf#page=148

535.

536. Vinayagamoorthy, V., Gillies, M., Steed, A., Tanguy, E., Pan, X., Loscos, C., & Slater, M. (2006). Building expression into virtual characters. Proceedings of Eurographics Conference. *State of the Art Report*. https://research.gold.ac.uk/id/eprint/398/1/expressivevirtualcharacters.pdf

537. Voogt, J., & Roblin, N. P. (2010). 21st century skills. [Discussion paper]. University of Twente. http://opite.pbworks.com/w/file/fetch/61995295/White%20Paper%2021stCS_Final_ENG_def2.pdf

538. Wagner, D., Billinghurst, M., & Schmalstieg, D. (2006). How real should virtual characters be?. *Proceedings of the 2006 ACM SIGCHI International Conference on Advances in Computer Entertainment Technology*, 57. https://doi.org/10.1145/1178823.1178891

539. Wagner, T. A., & Harvey, R. J. (2006). Development of a new critical thinking test using item response theory. *Psychological Assessment,* 18(1), 100. https://doi.org/10.1037/1040-3590.18.1.100

540. Wagner-Döbler, R. (2001). Continuity and discontinuity of collaboration behaviour since 1800 from a bibliometric point of view. *Scientometrics*, 52(3), 503-517. https://doi.org/10.1023/A:1014208219788

541. Walton, A. P. (2003). The impact of interpersonal factors on creativity, *International Journal of Entrepreneurial Behavior & Research,* 9 (4), 146-162. https://doi.org/10.1108/13552550310485120

542. Wandari, G. A., Wijaya, A. F. C., & Agustin, R. R. (2018). The Effect of STEAM-Based Learning on Students' Concept Mastery and Creativity in Learning Light and Optics. *Journal of Science Learning,* 2(1), 26-32. https://doi.org/10.17509/jsl.v2i1.12878

543. Wang, C. C., Ho, H. C., Cheng, C. L., & Cheng, Y. Y. (2014). Application of the Rasch Model to the measurement of creativity: The Creative Achievement Questionnaire. *Creativity Research Journal,* 26(1), 62-71. https://doi.org/10.1080/10400419.2013.843347

544. Wang, I. and Ruiz, J. (2021). Examining the use of nonverbal communication in virtual agents. International *Journal of Human–Computer Interaction,* 37(17), 1648–1673. https://doi.org/10.1080/10447318.2021.1898851

545. Waters, K., & Levergood, T. (1995). DECface: A system for synthetic face applications. *Multimedia Tools and Applications,* 1(4), 349-366. https://doi.org/10.1007/BF01215883

546. Watson, G. (1980). Watson-Glaser critical thinking appraisal. *San Antonio, TX: Psychological Corporation.* https://www.assessmentday.co.uk/free/watson-glaser/freetest1/Inferences/index.php?extratime=99999999999999999999

547. Watson, K. W., & Barker, L. L. (1988). Listening assessment: The Watson-Barker listening test. *International Listening Association. Journa*l, 2(1), 20-32. https://doi.org/10.1080/10904018.1988.10499095

548. Wechsler, S. (2006). Validity of the Torrance Tests of Creative Thinking to the Brazilian culture. *Creativity Research Journal,* 18(1), 15-25. https://doi.org/10.1207/s15326934crj1801_3

549. Weisman, W. D., & Peña, J. F. (2021). Face the Uncanny: The Effects of Doppelganger Talking Head Avatars on Affect-Based Trust Toward Artificial Intelligence Technology are Mediated by Uncanny Valley Perceptions. *Cyberpsychology, Behavior, and Social Networking,* 24(3), 182-187. https://doi.org/10.1089/cyber.2020.0175

550. Weldon, A., Ma, W. W., Ho, I. M., & Li, E. (2021). Online learning during a global pandemic: Perceived benefits and issues in higher education. *Knowledge Management & E-Learning: An International Journal,* 13(2), 161-181. https://doi.org/10.34105/j.kmel.2021.13.009

551. Wesseldijk, L. W., Mosing, M. A., & Ullén, F. (2019). Gene–environment interaction in expertise: The importance of childhood environment for musical achievement. *Developmental psychology,* 55(7), 1473. https://doi.org/10.1037/dev0000726

552. West, D. R. (2016). School Counselors and Ethical Decision Making. [Doctoral dissertation, University of Wyoming]. https://eric.ed.gov/?id=ED577207

553. Wiederhold, M. D., Gao, K., & Wiederhold, B. K. (2014). Clinical use of virtual reality distraction system to reduce anxiety and pain in dental procedures. *Cyberpsychology, Behavior, and Social Networkin*g, 17(6), 359-365. https://doi.org/10.1089/cyber.2014.0203

554. Wild, S., & Alvarez, S. (2020). Cooperative Education in the Higher Education System and Big Five Personality Traits in Germany. *International Journal of Work-Integrated Learning,* 21(1), 37-49. https://eric.ed.gov/?id=EJ1241224

555. Winn, J., & Blanton, L. (2005). The call for collaboration in teacher education. *Focus on Exceptional Children*, 38(2), 1-10. https://www.academia.edu/50417659/The_Call_for_Collaboration_in_Teacher_Education?from=cover_page

556. Wismeijer, A. A., & Vingerhoets, A. J. (2005). The use of virtual reality and audiovisual eyeglass systems as adjunct analgesic techniques: a review of the literature. *Annals of Behavioral Medicine,* 30(3), 268-278. https://doi.org/10.1207/s15324796abm3003_11

557. Wittenberg-Lyles, E., Parker Oliver, D., Demiris, G., & Regehr, K. (2010). Interdisciplinary collaboration in hospice team meetings. *Journal of Interprofessional Care,* 24(3), 264-273. https://doi.org/10.3109/13561820903163421

558. Wolfradt, U., & Pretz, J. E. (2001). Individual differences in creativity: Personality, story writing, and hobbies. *European Journal of Personality*, 15(4), 297-310. https://doi.org/10.1002/per.409

559. Wood, D. J., & Gray, B. (1991). Toward a comprehensive theory of collaboration. *The Journal of Applied Behavioral Science,* 27(2), 139-162. https://doi.org/10.1177/0021886391272001

560. World, S. A. (2013). *Oso panda*. Blog. http://saveanimalsworld.blogspot.com/2013/09/oso-panda.html

561. Worthington, D. L., Keaton, S., Cook, J., Fitch-Hauser, M., & Powers, W. G. (2014). Assessing the Watson-Barker Listening Test (WBLT)-Form C in Measuring Listening Comprehension of Post-Secondary Hispanic-American Students. *Journal of Research in Education,* 24(2), 192-204. https://eric.ed.gov/?id=EJ1098281

562. Wu, J., Ghosh, S., Chollet, M., Ly, S., Mozgai, S., & Scherer, S. (2018). Nadia: Neural network driven virtual human conversation agents. *In Proceedings of the 18th International Conference on Intelligent Virtual Agent*s, 173-178. https://doi.org/10.1145/3267851.3267860

563. Yamane, K., Ariki, Y., & Hodgins, J. (2010, July). Animating non-humanoid characters with human motion data. *In Proceedings of the 2010 ACM SIGGRAPH/Eurographics Symposium on Computer Animation,*169-178. http://dx.doi.org/10.2312/SCA/SCA10/169-178

564. Yang, G.Z., Bellingham, J., Dupont, P.E., Fischer, P., Floridi, L., Full, R., Jacobstein, N., Kumar, V., McNutt, M., Merrifield, R. and Nelson, B.J. (2018). *The grand challenges of science robotics Sci. Robotics,* 3(14). https://doi.org/10.1126/scirobotics.aar7650

565. Yang, K. K., Lin, S. F., Hong, Z. R., & Lin, H. S. (2016). Exploring the assessment of and relationship between elementary students' scientific creativity and science inquiry. *Creativity Research Journal*, 28(1), 16-23. https://doi.org/10.1080/10400419.2016.1125270

566. Yeh, M. L. (2002). Assessing the reliability and validity of the Chinese version of the California Critical Thinking Disposition Inventory. *International Journal of Nursing Studies,* 39(2), 123-132. https://doi.org/10.1016/S0020-7489(01)00019-0

567. Yeh, Y. C. (2009). Integrating e-learning into the direct-instruction model to enhance the effectiveness of critical-thinking instruction. *Instructional Science*, 37(2), 185-203. https://doi.org/10.1007/s11251-007-9048-z

568. Young, C., Durham, P., Miller, M., Rasinski, T. V., & Lane, F. (2019). Improving reading comprehension with readers theater. *The Journal of Educational Research*, 112(5), 615-626. https://doi.org/10.1080/00220671.2019.1649240
569. Yükselir, C. (2020). A Quantitative Study of Turkish EFL Learners' Perceptions about Critical Thinking Dispositions. *International Online Journal of Education and Teaching*, 7(1), 126-139. https://eric.ed.gov/?id=EJ1244248
570. Zabelina, D. L., Zaonegina, E., Revelle, W., & Condon, D. M. (2021). Creative achievement and individual differences: Associations across and within the domains of creativity. P*sychology of Aesthetics, Creativity, and the Arts*. https://doi.org/10.1037/aca0000439
571. Zampetakis, L. A. (2010). Unfolding the measurement of the creative personality. *The Journal of Creative Behavior,* 44(2), 105-123. https://doi.org/10.1002/j.2162-6057.2010.tb01328.x
572. Zhao, Z., Han, F., & Ma, X. (2019). Live emoji: A live storytelling vr system with programmable cartoon-style emotion embodiment. *In 2019 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*, 251-2511. https://doi.org/10.1109/AIVR46125.2019.00057
573. Ziv, N., & Keydar, E. (2009). The relationship between creative potential, aesthetic response to music, and musical preferences. *Creativity Research Journa*l, 21(1), 125-133. https://doi.org/10.1080/10400410802633764
574. Zlatkin-Troitschanskaia, O., Pant, H. A., & Coates, H. (2016). Assessing student learning outcomes in higher education: Challenges and international perspectives. *Assessment & Evaluation in Higher Education,* 41(5), 655-661. https://doi.org/10.1080/02602938.2016.116950
575. Zlotowski, J. A., Sumioka, H., Nishio, S., Glas, D. F., Bartneck, C., & Ishiguro, H. (2018). Persistence of the uncanny valley. In Geminoid Studies, 163–187: *Springe*r. http://www.bartneck.de/publications/2018/persistence-uncanny-valley/zlotowskiBartneckGeminoid.pdf

R