BRINGING A SOCIOTECHNICAL LENS TO USER EXPERIENCE (UX) RESEARCH: ANALYZING AN AI-BASED SIMULATION

A Capstone Defense

Presented to

The Faculty of the School of Education and Human Development

University of Virginia

In Partial Fulfillment

Of the Requirements for the Degree

Doctor of Education

---

by

Michele L. Spotts, M.S.

March 29, 2024

Michele L. Spotts
Curriculum and Instruction
UVA School of Education and Human Development
University of Virginia
Charlottesville, VA

APPROVAL OF THE CAPSTONE PROJECT

This capstone, *Bringing A Sociotechnical Lens to User Experience (UX) Research: Analyzing an Ai-Based Simulation*, has been approved by the Graduate Faculty of the School of Education and Human Development in partial fulfillment of the requirements for the degree of Doctor of Education.

____
Jennie Chiu, Ph.D., Chair

____
Jim Bywater, Ph.D., Committee Member

____
Sarah McCorkle, Ph.D., Committee Member

____
Date

ABSTRACT

Recent growth in artificial intelligence (AI) based technologies adds complexity to designing and evaluating human-computer interaction (HCI) and learning design. Many research studies on the usability and user experience (UX) of learning technologies follow evaluation approaches originally designed for HCI. However, due to numerous differences between the fields, the usability tools often do not properly align. This study outlines the need for expert and user evaluation tools following the sociotechnical theory to guide research on the usability evaluation of an AI-based Conversational Agent (CA) simulated experience for preservice teachers to practice teaching mathematics. This study combines the TPACK framework (technology, pedagogical, and content knowledge) and the Community of Inquiry framework to evaluate the social, pedagogical, technical, and content aspects of a simulated learning experience, creating the Social-TPAC heuristic evaluation tool. The study uncovered how the technical and content aspects of a user experience impacted the pedagogical and social aspects of learning, while also uncovering the need for human presence to facilitate AI-based simulation education.

Keywords: Usability, Artificial Intelligence, Simulation-based Education, Conversational Agents, User Experience Design, Sociotechnical Theory, TPACK, Community of Inquiry

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

DEDICATION

In memory of my grandfather,

Norman R. Spotts Sr.

April 11, 1925 – March 28, 2024

I love and miss you, Pap. I hope you are relaxing and watching the Phillies.

ACKNOWLEDGMENTS

Frank Lloyd Wright said, “I know the price of success: dedication, hard work, and an unremitting devotion to the things you want to see happen.” While I agree with Wright and acknowledge the tremendous dedication, devotion, and hard work I put into completing this doctorate, I truly could not have done it without the *support of so many people*.

To *Jennie Chiu, Ph.D.*, I am beyond thankful for how you stepped into the role of being an advisor for so many of us during a difficult transition and supported me on this academic journey. I am grateful that you guided me in exploring non-traditional topics, such as learning analytics, simulation, and user experience research, which provided the groundwork for this research.

Additionally, thank you to *Jim Bywater, Ph.D.*, for serving on my committee and allowing me to explore the ACTS system from two different perspectives of UX research; your support and encouragement was so helpful and appreciated.

To *Sarah McCorkle, Ph.D.,* thank you for serving on my committee and providing insightful and thorough feedback that helped to shape and form this paper into what it is.

To my peers in this program, especially *Deb Jewell* and *Allyson Wharam*, I feel so lucky to have had the opportunity to learn and grow with both of you; I wish you all the success in your academics and careers.

To *Lauren DeChant, Ed.D.,* and *Cara DeLura,* thank you both for your leadership, support, and feedback on my papers and presentations, and for allowing me the flexibility to work and complete this degree simultaneously.

To *Sharon Griswold, MD*, thank you for encouraging me to enroll in a doctorate program and providing guidance and support on my academic journey; I appreciate you taking the time to read my papers and provide honest feedback.

Lastly, to my *family and friends*, I started this degree a month after moving to a new state for a new job at the beginning of a global pandemic. I could not have survived without your support from a distance; your phone calls, Zoom gatherings, and text messages were so helpful. I am also grateful to the friends I made while working on this degree; thank you for your understanding and support! Finally, to my dog *Gracie*, you were truly by my side for every word I typed, presentations I gave, and the general ups and downs of learning; thank you—I promise we will go on more hikes now!

CHAPTER I: INTRODUCTION

Background of the Problem: Usability of Learning Technologies

Usability refers to the ease of use and intuitiveness of an interactive system to perform tasks and achieve certain goals (Norman, 1986). An essential concept of *Human-Computer Interaction* (HCI), usability focuses on designing, implementing, and evaluating interactive systems to ensure they are usable and practical (Issa & Isaias, 2022). Within an educational setting, useable and accessible designs are often a focus of learning technologies. Multiple studies have shown that the usability of learning technologies can affect the pedagogical value (Squires & Preece, 1996; Zaharias & Poylymenakou, 2009), demonstrating a need to design and test for usability.

When designing effective *user interfaces* (UI) for HCI, evaluation methods are conducted at various stages of design and development and often include heuristic evaluations, think-aloud walk-throughs, questionnaires, and eye-tracking to create an intuitive and pleasing *user experience* (UX) (Issa & Isaias, 2022; Zardari et al., 2020). UX research examines human behaviors, preferences, and emotions before, during, and after using a product to understand the desire to adopt a technology (Law et al., 2014; Travis & Hodgson, 2019). Some researchers suggest that UX has replaced usability, while others view usability as one aspect of the UX (Rose & Turner, 2023). For instance, UX testing methods focus more on *hedonic qualities*, defined as a person's pleasure or discomfort using a product, with technical or utilitarian aspects of usability identified as the product being usable (Hornbæk & Hertzum, 2017).

Stepping away from usability towards UX could be due to the growing expertise and knowledge of professional designers in the commercial space who inherently incorporate usability principles into their designs (Alharoon et al., 2021). In parallel, most technological innovations are an evolving process within the HCI field to update or substitute an existing technology (Ayoko & Ashkanasy, 2019; Geels, 2004; Jones et al., 2013). *Sociotechnical Systems Theory* suggests that technology has *social* and *technical* aspects, viewed as independent parts that work together (Cherns, 1976). Geels (2002) considers innovations to be *Technological Transitions*—the co-evolution of technology and society substituting one technology with another (Geels, 2004). As younger generations grow up interacting with computers and smart devices, they gain intuitive knowledge of how to use technology and seek familiar yet enhanced environments (Ayoko & Ashkanasy, 2019; Geels, 2004; Jones et al., 2013). However, UX research examines the hedonic qualities of HCI, and learning technologies necessitate additional evaluation methods to understand the pedagogical value.

A large body of instructional design (ID) models provides educators, designers, and developers with frameworks and tools to guide the design and evaluation processes for creating learning technologies. Tessmer (1993) highlighted conducting expert and learner *formative evaluations* during the design, development, and implementation stages, but like other ID models there are no guidelines for usability testing (Branch, 2009; Dick et al., 2008; Morrison et al., 2019). Conversely, while ID and UX have a core framework around user-centered design, the UX testing tools do not automatically align with learning evaluation needs (Norman, 1986; Soloway et al., 1994). A recent systematic literature review on the usability and UX of learning technologies found that many studies are flawed by misinterpretations of evaluation tools developed for HCI, leading to inappropriate applications and unreliable results (Lu et al., 2022). Additionally, many learning design studies lack actionable insights into improving usability based on the results of the UX evaluations.

One possible reason is the commercial industry's emphasis on the user's first impressions and desire to use a technology to predict its adoption rate in a competitive market (Nielsen & Molich, 1990; Rogers, 2003). E.M. Rogers's Diffusion of Innovation (DOI) Theory, developed in 1962, provides a framework for understanding how a social group adopts new technology as a downstream direct-to-consumer trajectory based on influence and desire (Vargo et al., 2020) and considers non-adoption a failure (Rogers, 2003; Straub, 2009; Vargo et al., 2020). However, technology diffusion in a learning system is different, as learners are often excluded from the choice to adopt a technology and "do what they need or want to do" to meet course requirements (Issa & Isaias, 2022; Preece et al., 2002).

Another key difference is that UX evaluation focuses heavily on the perception of *aesthetic*s in creating a positive experience (Alharoon et al., 2021; Yablonski, 2020), while learning design often points to aesthetics as distractions that can increase a learner's *cognitive load* or working memory to complete a task (Mayer, 2020). *Learning Experience* (LX) design aimed to close the gap between UX and learning design (Banes & Behnke, 2019). However, researchers have noted that the LX model overlooks the UX knowledge required for effective technology-based learning design and evaluation (Abramenka-Lachheb, 2023; Punchoojit & Hongwarittorrn, 2017; Reigeluth & An, 2023).

Recently, *Simulation-based education* (SBE) for teacher education has moved to a computer-based learning setting, done at either a distance or through enhanced technologies that foster social learning (Kaufman & Ireland, 2016; Levin & Flavian, 2020). SBE allows learners to practice their skills and knowledge in a controlled and safe environment, often followed by instructor-led debriefing or feedback (Bauer et al., 2022; Cook et al., 2018). SBE allows preservice teachers to use their pedagogical and content knowledge to practice instruction and classroom management (Kaufman & Ireland, 2016). One way to conduct SBE for teacher training is by using *Conversational Agents* (CA).

CAs allow users to communicate through text-based or voice chat systems that respond based on user inputs (Mafra et al., 2022; McTear et al., 2016; Sansonnet et al., 2006). CAs are becoming increasingly popular in education, and research shows positive results utilizing such technology as a tutoring method (Datta et al., 2022; Kumar et al., 2023). Previously, CAs were developed using pre-determined responses to typical questions or statements, often creating a poor user experience when the system did not recognize the user's input (McTear et al., 2016). However, recent growth in *artificial intelligence* (AI) and Large Language Models, such as ChatGPT or Bard, allow for a high-fidelity *social experience* that mimics chatting with a human (Ashfaq et al., 2020; Sansonnet et al., 2006). Yet, the social human-like interaction of AI-based technologies complicates UX evaluation methods, especially for learning experiences (Celik, 2023). Researchers are calling for AI-based UX to be evaluated with a sociotechnical perspective to grasp the interaction and evolution between technology and society (Herrmann & Pfeiffer, 2022; Sartori & Theodorou, 2022; Sony & Naik, 2020). Yet, few tools or methods with a sociotechnical lens exist to evaluate the UX, not to mention how to effectively evaluate the learning experience of AI-based educational technologies or CAs (Farrow, 2023; Lu et al., 2022).

One such AI-based educational technology is the AI-based Classroom Teaching Simulation (ACTS), a CA that can be considered a *technological transition*, as it took a validated method of practice and evaluation and incorporated it into a technology-based learning experience. Traditionally, pre-service teachers would learn about a pedagogical strategy but have little opportunity to practice that strategy. Some pre-service teachers might practice tutoring a student in mathematics with an instructor present to observe and provide feedback. The instructor could use the *Instructional Quality Assessment* (IQA) Mathematics Toolkit to evaluate the pre-service teacher and provide feedback following the set of standards outlined in the tool (Boston, 2012). However, in this scenario the pre-service teacher may get very few practice opportunities as it is resource-intensive for a single methods instructor. Utilizing an AI-based CA provides many affordances, such as the ability for pre-service to practice their skills with automated feedback and a virtual student to practice with. The use of AI in ACTS creates a high-fidelity simulation for pre-service teachers to practice their mathematic questioning skills in their own space in their own time, allowing them to build self-efficacy prior to being evaluated in-person (Datta et al., 2022). The learning technology simulates a conversation with a student using AI and then provides the pre-service teacher with automated feedback on their tutoring quality based on the ACTS tool. In addition, ACTS' AI-based CA application also offers the ability for human-human interaction and a human-in-the-loop approach, where a human can step in if the CA fails to provide an effective learning experience, presenting a need to examine both human-nonhuman and human-human interactions (Datta et al., 2022). Further, ACTS utilizes an interactive *knowledge visualization* tool, which enables users to communicate, interpret, problem-solve, and reflect through an interactive diagram to help advance critical thinking and understanding of the mathematical problem (Arcavi, 2003; Datta et al., 2022)—adding a need to focus on the graphical and aesthetic design elements of ACTS. Although ACTS has been piloted with in-service and pre-service teachers with initial evidence of realism and utility (Phillips, 2022), it has not been subject to any formal usability or user experience testing. Thus, the problem of practice for this capstone study is to analyze the user experience of a text-based conversation agent with dynamic visualizations developed as a learning tool for pre-service teachers to practice teaching mathematics. How a learner engages with students and receives feedback has been changed due to the integration of an AI-based Conversational Agent. Borsci et al. (2021) acknowledged that more research is needed to understand user perception of interacting with CAs. While some tools have recently been established to evaluate human-nonhuman AI-based CAs, few align with the unique knowledge visualization and pedological aspects of ACTS (Borsci et al., 2021; Mafra et al., 2022). Further, few usability methods look at evaluation through a sociotechnical lens to understand the social and pedagogical aspects of learning with technology. In turn, there is a need for additional research to understand what evaluation methods are needed to analyze ACTS' usability effectively. Therefore, this study looks to uncover what are the essential aspects of a learning experience and how do they interact to influence the learning experience of ACTS. The study involves developing a heuristic evaluation protocol for analyzing a technology-based learning activity's user experience (UX) with human and nonhuman social interactions. The protocol outlines evaluation criteria of the learning experience (LX) to align with current UX and learning evaluation methods aligned to key performance indicators to provide feedback to the project team to create a more effective learning experience. The study answers the following research questions:

RQ1: What are the social, pedagogical, content, and technical indicators of user experience that emerge from expert review of ACTS?

RQ2: What are the social, pedagogical, content, and technical indicators of user experience that emerge from participants using ACTS?

RQ3: What are participants' perceptions of the social, pedagogical, content, and technical indicators of user experience after using ACTS?

Through this capstone study, I provide a framework to evaluate AI-based technologies from a sociotechnical perspective. I also provide user experience information to the ACTS research team that will inform subsequent revisions to the tool. In chapter 2, I outline the literature that has guided this study. In chapter 3, I discuss methods, data sources, and analysis techniques that I will use to answer my research questions.

CHAPTER II: LITERATURE REVIEW

As an essential concept of *Human-Computer Interaction* (HCI), *usability* refers to the ease of use of *User Interfaces* (UI) for interactive systems (Issa & Isaias, 2022; Norman, 1986). Usability focuses on designing intuitive experiences that consider a user's needs to perform tasks and achieve certain goals (Apple, 2024; Issa & Isaias, 2022; Law et al., 2014; Punchoojit & Hongwarittorrn, 2017; Travis & Hodgson, 2019). An intuitive experience occurs when a user's prior knowledge of using technology allows their subconscious to operate the system while they focus on the content and their goals (Hurtienne & Blessing, 2007).

Lu et al. (2022) found that many usability and UX studies of learning technologies were misaligned with the evaluation tools, pushing for further research on the best methods and protocols to evaluate learning technologies effectively. Designing a usable technology is an iterative user-centered design process that combines design and user evaluations, generally aligned to the performance goals, called *key performance indicators* (KPIs) (Olsen et al., 2009; Keates et al., 2000; Lim et al., 2013; Straub, 2009). KPIs are based on user needs or industry standards and should align to the tool and the stage of development, and help to understand if users will adopt a new technology and continue to use it; from a learning perspective, they help to determine if the tool meets the academic requirements and is effective in developing skills and transferring knowledge (Olsen et al., 2009; Lim et al., 2013; Straub, 2009). To better inform the evaluation process, this literature review will start with understanding *simulation-based education*, followed by an outline of the four essential components of ACTS used to create the simulated learning experience.

Simulation-Based Education

Simulation-based education (SBE) is used in both higher education and professional training and allows learners to practice their skills and knowledge in a controlled and safe environment, often aligned with professional skills and generally followed by instructor-led debriefing or feedback (Bauer et al., 2022; Cook et al., 2018). Within the field of teacher education, SBE is a relatively new practice that has recently grown as a method out of the need for more accountability in teaching quality (Kaufman & Ireland, 2016; Levin & Flavian, 2020). SBE allows preservice teachers to use their pedagogical and content knowledge to practice instruction and classroom management (Kaufman & Ireland, 2016). The amount of *deliberate practice* a preservice teacher can participate in can greatly improve their performance as an inservice and practicing teacher (Ericsson, 2012). Further, SBE allows preservice teachers to practice with challenging experiences and diverse students they may not experience during their inservice hours (Kaufman & Ireland, 2016). Exposure to more social interactions allows learners to expand their understanding, better preparing them for more real-world environments (Fosnot, 2005; Kaufman & Ireland, 2016; Mezirow, 1994).

In addition to practicing, the feedback provided by instructors is an essential aspect of SBE, and many educational researchers acknowledge that most of the learning takes place through good feedback (Boston, 2012; Cheng et al., 2020; Dreifuerst, 2012). Further, feedback allows preservice teachers to adjust their current instructional methods and experiment with suggested changes to receive additional feedback, working towards mastering their skills (Grossman et al., 2009). Within the field of healthcare, SBE is considered vital in preventing patient harm; in turn, they have conducted a large amount of research on how to properly run a simulation, with heightened attention on how to give feedback to learners through postsimulation debriefing sessions (Cheng et al., 2020; Dreifuerst, 2012). Further, they also place an emphasis on pre-briefing and properly facilitating an SBE session; pre-briefing prepares the learners for what they can expect, establishes psychological safety, and outlines how they will be evaluated (El Hussein et al., 2021; Levin & Flavian, 2020; Mezirow, 1994). Facilitating a simulation occurs throughout the process, including during the simulation, where they may need to cue a learner who might be unaware of an environmental difference between a simulated setting and real life (Hellaby, 2013). The pre-brief, simulation, and debrief or feedback occur in a social setting, generally in person. However, more recently, simulation has moved to a computer-based learning setting at a distance or through enhanced technologies that foster social learning (Cheng et al., 2020). Bringing social learning into an online setting can best be viewed through Garrison et al. (1999) *Community of Inquiry* (CoI) framework.

Community of Inquiry

The CoI framework was developed from Moore's (1989) *Interaction framework*, which looked at content, instructor, and learner interaction in distance education. With learning moving online, a need emerged to also understand the social influences of learners and instructors within an online learning environment (Cheng et al., 2020; Garrison et al., 1999; Swan et al., 2012). Originally, CoI was developed for asynchronous learning, yet more recently, the framework has also been applied to synchronous learning sessions, including SBE (Cheng et al., 2020; Fatani, 2020; Seckman, 2018; Wanstreet & Stein, 2011). CoI has three fundamental indicators of an engaging online learning experience: *cognitive presence,* which supports reflective dialogue, exchanging ideas and connecting themes; *social presence,* which fosters user engagement, interactivity, and emotions; and *teacher presence* for feedback, instruction, and creating a supportive environment; all coming together to create a wholistic learning experience (Garrison et al., 1999). High levels of social presence is an essential element in computer-based simulation and can increase satisfaction and perceived learning, demonstrated by participants feeling able to present their whole personality through emotional, risk-free expression, acknowledging others, and being encouraging and supportive (Cheng et al., 2020; Garrison et al., 1999; Richardson & Swan, 2003).

Learners must have a sustained level of cognitive presence to achieve a higher level of learning, such as critical thinking in an online setting (Kanuka & Garrison, 2004). To increase cognitive presence in the learning environment, learners must feel comfortable and free of distractions, such as those caused by poor usability, so their engagement in conversation can extend their knowledge and teaching skills (Cheng et al., 2020; Shea & Bidjerano, 2009). Further, a complex UI can interfere with a learner's ability to be cognitively present (Hollender et al., 2010). To better understand how social, cognitive, and instructor presence are used for SBE within the ACTS system, it is essential to understand the various components of the technologies used to create the learning experience.

AI-Based Classroom Teaching Simulation System

Due to the recent affordances of using technology for *computer-based simulation*, there has been a growth of SBE in the digital space using human or virtual actors (Kaufman & Ireland, 2016; Levin and Flavian, 2020). The AI-based classroom teaching simulation (ACTS) system provides a space where students can participate in SBE to practice essential teaching skills in a platform powered by AI to create an interactive experience with four main components, three of which are AI-based: an *AI-based Conversational Agent* (CA), an *AI-based Knowledge Visualization*, and *AI-generated feedback* for the learner. Below is an overview of the four components to provide a basic understanding of the ACTS system.

Setup Process

Currently, the ACTS system is being used with instructor oversight, and preservice teachers are assigned different problems to work through. A faculty member *familiar* with the system can select a problem, the type of feedback provided, and if the simulation will be human-human, human-AI, or human-AI with a human-in-the-loop, where a human can jump into the simulation if needed. The setup process will be evaluated for future use where a faculty member *unfamiliar* with the system can create a simulation and possibly where a learner can set up their own practice session without a faculty member.

Conversational Agents & Conversational AI

Conversational Agents (CAs) allow humans to engage with technology through text-based chatbots, text messages, voice chatbots, and voice assistants that portray a human (Mafra et al., 2022; McTear et al., 2016; Sansonnet et al., 2006). There has been recent growth in using CAs as a learning technology, but most tools available for teacher education follow a *rule-based platform* (Johannsen et al., 2023; Labadze et al., 2023). Rule-based CAs use pre-determined responses to typical questions or statements, which is sometimes frustrating to the end user if they ask a question that is not typical for the interaction and receive messages such as "I don't understand" in response (McTear et al., 2016). However, AI-based CAs, also referred to as *Conversational AI*, utilizes a Large Language Model (LLM), such as ChatGPT or Bard, to process text entered by a user to then logically respond, providing a high-fidelity interaction (Ashfaq et al., 2020; Kulkarni et al., 2019; Sansonnet et al., 2006). With more open access to LLMs, research on Conversational AI is expanding to various fields, including education, with most research focusing on the mechanics and flow of the conversation (Kulkarni et al., 2019). Unlike rule-based CAs, Conversational AI uses *Dialogue Management* to acknowledge human inputs, steer the conversation, and request additional information (Kulkarni et al., 2019). Hill et al. (2015) noted that the human language, even in a text-based format, is socially complex and emotional, yet the higher levels of fidelity with Conversational AI create a human-like presence in computer-based simulation that mimics human-to-human conversations (Ashfaq et al., 2020; Kim et al., 2023; Tun et al., 2015). For many, human-to-human text-based conversations have become a daily practice due to technologies such as text messaging or chat applications within social media. In turn, conversing with a Conversational AI can feel like a familiar and realistic experience for the user.

Knowledge Visualization

The system also uses *knowledge visualization*, which enables users to communicate, interpret, problem-solve, and reflect through an interactive diagram to help advance critical thinking and understanding of the mathematical problem (Arcavi, 2003; Datta et al., 2022). Knowledge visualizations are graphics that communicate information to enhance a learner's understanding of the content (Arcavi, 2003). The ACTS system has a knowledge visualization tool built into the user interface alongside the CA that allows learners to dynamically interact with a graphic to communicate with another human or AI (Arcavi, 2003; Datta et al., 2022). Fleming (1979) and Levie & Lentz (1982) emphasized that when illustrations and images are aligned with text in instructional materials, knowledge transfer increases compared to text without images or images without text. However, the research on using an AI-based knowledge visualization with a Conversational AI is limited, and few tools exist for evaluating its usability within a learning setting (Borsci et al., 2021; Mafra et al., 2022).

Feedback

As previously outlined, most of the learning in SBE occurs through feedback or debriefing (Boston, 2012; Cheng et al., 2020; Dreifuerst, 2012). CoI emphasizes the need for reflective learning during the feedback and debriefing portion of a simulation (Cheng et al., 2020). The ACTS system utilizes machine learning to incorporate the *Instructional Quality Assessment* (IQA) tool, which is a set of standards used to evaluate a teacher's instructional quality (Boston, 2012). ACTS utilizes the IQA tool to analyze the learner's conversation with a simulated student and provides text-based and graphical feedback at the end of the session. The ability to automatically generate feedback, creates a full simulation-based learning experience that can be conducted online without an instructor present. While some simulation research shows success in self-debriefing (Lapum et al., 2018), the CoI model emphasizes having instructors and students project themselves as real-life people and feel connected to one another, which is unclear if users will feel that way using an AI-based system (Salas et al., 2009).

Therefore, while engaging in a learning experience without a human present opens doors for learners to practice on their own time, the feedback can make or break a simulated experience and is vital to the overall user experience (Cheng, 2020), leading to a need to understand the overall experience from a usability and social lens, to understand affordances and potential use. Geels (2004) stated that adopting new technologies into practice is a prolonged process, often done in an incremental progression, and sometimes with setbacks. The ACTS system has gone through initial testing, but it is unclear if it is ready for user *adoption* due to the need for additional user experience and usability testing.

User Adoption of New Technologies

Geels (2004) suggested that technological innovations differ from the ordinary production of goods and services and, therefore, diffuse differently. E.M. Rogers's Diffusion of Innovation (DOI) Theory is widely credited for understanding the diffusion or adoption of products and goods, including technologies (Rogers, 2003). The theory, developed in 1962, looks at the stages in which a social group adopts a new or different product (or behavior) than what they had previously, comparing the advantages over existing technologies, compatibility, knowledge, and complexity to create a pattern of (1) early adopters; (2) the majority of adopters; and (3) late adopters (Rogers, 2003). The theory views technology adoption in a downstream direct-to-consumer view (Vargo et al., 2020) and considers non-adoption a failure (Rogers, 2003; Straub, 2009; Vargo et al., 2020). Geels (2004) believes the diffusion of innovative technologies focuses heavily on the developer's knowledge and less on the user's influence on the need and use of technology and its impact on society. AI, in particular, is heavily influenced and created by the end user due to machine learning, despite how many users are unaware that they contribute to AI through data collection (Kenny et al., 2022; Sartori & Theodorou, 2022).

Additionally, Geels (2004) recognized that different social groups or institutions have their own preferences, strategies, and goals for technology, impacting how technology diffuses. Within a learning setting, the instructor generally coordinates activities, and institutions may require the use of a system to meet a requirement, excluding the end users from the choice to adopt a product, as they must "do what they need or want to do" to fulfill a learning requirement (Geels, 2004; Issa & Isaias, 2022; Preece et al., 2002).

Geels (2004) stated that sometimes, as users integrate new technologies into their practices, the technologies must be 'tamed' to fit the organization's needs. This is often true within the field of education, as there are additional specific social needs per organization, course, and learner, which are called sub-functions within a sociotechnical system (Geels, 2004). Koehler and Mishra (2009, p. 62) recognized that technology “integration efforts should be creatively designed or structured for particular subject matter ideas in specific classroom contexts,” acknowledging that many popular software programs are designed with a business approach, and web-based technologies have an entertainment focus, leading them to create the *TPACK* framework for integrating technology into a learning setting.

TPACK

Koehler and Mishra (2009) developed the *Technology, Pedagogy, and Content Knowledge* (TPACK) framework after Shulman’s (1986) *Pedagogical Content Knowledge* (PCK), which emphasized the need to know the subject matter content and pedagogical methods to appropriately teach learners. TPACK emerged from the need for instructors to understand the nuances of selecting a technology that best integrates with the content and pedagogy (Koehler & Mishra, 2009). TPACK emphasizes the need for teachers to have equal knowledge of technology, pedagogy, and content to understand how they all function together (Koehler & Mishra, 2009). Due to technology constantly changing, it often causes a misalignment between technology and learning needs (Geels, 2002, 2004; Koehler & Mishra, 2009). AI-based learning tools, such as ACTS, are still being ethically examined, and few policies or regulations exist for using them in education systems, pushing researchers to also call for data collection through a *sociotechnical* lens to understand the implications of integrating the functional needs of society (e.g., education) to innovative technologies (e.g., AI, conversational agents) (Farrow, 2023; Geels, 2004; Sartori & Theodorou, 2022).

Sociotechnical Theory

Sociotechnical Theory looks at the design and improvement of technology through a lens where the *social* and *technical* aspects are viewed as independent parts of a larger system (Cherns, 1976). While it is important to view them separately, it is also vital to understand how they function together. Humans create technical aspects as tools with functional goals in mind, while social aspects focus on how humans use those tools and influence societal and technical changes (Akbarighatar et al., 2023; Kenny et al., 2022; Sarker et al., 2019). The same is true for AI, which relies heavily on social aspects, such as human interaction and data collection, to build large language models (Kenny et al., 2022). In turn, AI can be considered a sociotechnical system alone (Akbarighatar et al., 2023; Kenny et al., 2022; Sarker et al., 2019).

Geels (2004) explains that innovation is the co-evolution of *technology* and *society*—bringing together form and function, stating that most innovations are *Technological Transitions* that change technology by updating it or substituting one technology with another (Geels, 2002). As society interacts with technologies, they develop new behaviors and needs for technology; consequently, as society uses one technology, it influences the growth and evolution of another technology (Ayoko & Ashkanasy, 2019; Freeman & Perez, 1988; Geels, 2004; Jones et al., 2013; Koehler & Mishra, 2009). Therefore, society rarely adopts new or previously unknown technologies (Freeman & Perez, 1988). The same also applies to theories and methods of learning; for example, learning theories like CoI, once applied only to asynchronous learning, but are now being applied to synchronous and AI-based technologies (Cheng et al.; Qin et al., 2020).

With the recent hype around AI, it may appear to be a new technology. However, AI was first coined in 1956 at the *Dartmouth Workshop* and is currently going through a technological transition, referred to as a *spring*, as it is increasing in popularity (Sartori & Theodorou, 2022).

Further, some predict it may enter a *winter* as the media has created fear in AI, causing governments and institutions to push for regulations, slowing its growth (Geels, 2004; Sartori & Theodorou, 2022). Geels (2002) stated that technologies rarely follow a linear adoption pattern and don't necessarily *close* as they can later regain adopters due to societal changes and user preferences that co-evolve with technology as it continually transitions (Celik, 2023; Freeman & Perez, 1988; Geels, 2002). For example, as rule-based CAs caused frustration among users, the technology did not diffuse well within educational systems. Yet, as AI has become more widely available to developers and learning designs, it has opened the door for greater use of CAs for educational needs such as SBE, leading to a longer adoption period for CAs (Geels, 2004). Furthermore, when educators use AI-based applications for personal use, they become familiar with the technology and will comprehend the ethical implications and pedagogical affordances before considering adding it to their curriculum (Celik, 2023; Freeman & Perez, 1988; Geels, 2002).

Of note, while end users may reject adopting AI, it already exists on the back end of many technologies, often hidden from the end users, such as how AI-based bots can comment on social media posts, yet the majority of users have a limited ability to detect them (Akbarighatar et al., 2023; Kenny et al., 2022; Sarker et al., 2019; Sartori & Theodorou, 2022). Allowing for the diffusion of the technology in a non-linear format while allowing society to influence its evolution without knowingly adopting it demonstrates the need for understanding the sociotechnical theory and technical transitions when evaluating the learning experience.

User Experience (UX) Research

While not directly connected to the research, sociotechnical theory and technological transitions explain the need and growth of *User Experience* (UX) research. Geels (2004) recognized that while technologies are a co-evolution with society, he also recognized how mass production and global distribution caused developers to grow apart from the local user. In turn, there was a need for “inter-group coordination” to bring together developers and users while also crossing between different domains, cultures, policies, and organizations to create accessible and inclusive technologies (Geels, 2004). Norman (2013, p. 39) saw this role as bridging the gap between how the technology informs the usability and how a user interprets it, noting that users will blame themselves for “being stupid” instead of blaming the technology when it doesn't operate as expected. In turn, UX research facilitates the co-evolution of society and technology to create more user-centered and accessible technologies (Sony & Naik, 2020).

Frank Lloyd Wright was a renowned 20th-century architect known for designing user-centered and aesthetically pleasing architecture, bridging the gap between architecture design and societal needs (Steinfeld & Maisel, 2012, p. xi). In 1952, Wright designed an innovative and accessible home—forty years before the Americans with Disabilities Act was signed into law (Billock, 2020). The house not only accommodated the homeowner's disability, but the design gave the homeowner “emotional and spiritual fulfillment” (Billock, 2020, para 7). Wright believed that all aspects of a project should merge to appear and interact together as one (Betsky & Shapiro, 2021, p. 162; Cleary, 2009, p. 10), stating “form and function should be one, joined in a spiritual union” (Silver, 2007, “Pondering Form” section). Trist (1981) called the union of form and function *joint optimization* of equal consideration of technical and social human elements in a sociotechnical system; further stating that the technical and social aspects of a system must complement each other and attempting to optimize one without the other will lead suboptimization.

Wright's design and development process of understanding the homeowner's needs aligns with UX research as it aims to establish joint optimization of human behaviors, preferences, and emotions in alignment with technology. Morville (2014) created the user experience honeycomb (Figure 2.1) that emphasizes the optimization of seven elements of UX research: useful, usable, valuable, desirable, findable, credible, and accessible. It is important to note that Morville's honeycomb framework includes usability as a sub-function of the UX; while some researchers view usability and UX separately, other researchers see it as a paradigm shift in the field of HCI and call for UX to replace usability (Rose & Turner, 2023). Following a sociotechnical lens, this

Figure 2.1

User experience honeycomb

Note. A graphical recreation by the author of the UX Honeycomb from *Intertwingled: information changes everything,* by Morville, P. (2014).

paper views usability as a sub-function of the overall social experience with technology and views UX as the research required to bring technology and society together (Geels, 2004; Travis & Hodgson, 2019).

Viewing the UX Honeycomb from a research perspective helps to explain the differences between usability testing and UX research. As UX research goes beyond usability to focus on *hedonic qualities*, defined as a person's pleasure or discomfort using a technology, examining human behaviors, preferences, and emotions (Law et al., 2014; Travis & Hodgson, 2019). Rose and Turner's (2023) idea of stepping away from usability towards UX could be due to the growing expertise and knowledge of professional designers who inherently incorporate usability principles into their designs (Alharoon et al., 2021). As noted by Trist (1981), as sociotechnical systems transition and grow, so must the design, usability, and accessibility standards that users expect on their devices. Wright acknowledged the need to understand how things previously worked and looked, even if the appearance changes, they needed to remain *familiar* to the user (Betsky & Shapiro, 2021).

Familiarity

Familiarity, also known as Jakob's Law, is a UX theory that aligns with technological transitions and is an essential part of UX design (Nielsen & Molich, 1990; Yablonski, 2020). Familiarity emphasizes the need for users to immediately know how to interact with technology by using similar design features in the UI to make it easier for people to achieve their goals (Nielsen & Molich, 1990; Yablonski, 2020). Hurtienne and Blessing (2007) saw this as the user's subconscious knowledge of prior interaction with a technology that creates an intuitive experience (Lawry et al., 2019). For example, Apple's (2024) *Human-Interface Guidelines* state that when people are familiar with a system's colors, text sizes, and common layouts, a user “feels at home on [the device].” Norman (2013) used the term *skeuomorphic* to explain how when designs slowly transform from old to new, it allows for a familiar and comfortable UI, which makes it easier to learn. Therefore, when system settings drastically change on established technologies, it can cause resistance among users (Geels, 2004). Familiarity has also been tied to *Cognitive Load Theory* (CLT), which has shown that an unfamiliar UI has more effect on intuitive use than age-related cognitive decline (Lawry et al., 2019).

Cognitive Load Theory

CLT examines how humans process new information and construct knowledge into long-term memory, focusing on the limited memory available to process that information within a time frame (Sweller et al., 2018). Much of the literature on CLT has identified three types of cognitive load: *intrinsic*, *extraneous*, and *germane*. *Intrinsic load* is defined as the complexity of the working memory and how it aligns with the task and cannot be manipulated by the design (Sewell et al., 2018; Zagermann et al., 2016). *Germane load* is the cognitive process of constructing patterns to understand and process information for long-lasting knowledge (Mayer, 2020; Norman, 2013; Sewell et al., 2018; Zagermann et al., 2016). Lastly, *extraneous load* is defined as how information is presented, such as the user interface, environmental distractions, or emotional aspects, like time pressure to complete a task (Kosch et al., 2023; Sewell et al., 2018; Zagermann et al., 2016). The field of HCI refers to extraneous load as the relationship between the UI and the level of cognitive resources needed to interpret it, which they call *mental workload* (Kosch et al., 2023). Often, HCI usability measurements look at the efficiency or effectiveness of a UI, focusing on time on task and errors in relation to mental workload or extraneous load (Kosch et al., 2023).

Hollender et al. (2010) noted the extraneous load from a UI can come from the complexity of the technology, the design not following usability standards, or the level of expertise using the technology. In turn, designers or developers creating a technology will likely find even a complex UI they created intuitive; however, when a design does not follow design standards, it will be less intuitive to users. Norman (2013) used an analogy of a master chef cooking in an unfamiliar kitchen; the kitchen may be organized by the owner, but the chef will be confused and slowed down in the unfamiliar space. Yet, the chef will eventually find the tools to create an excellent meal. Of note, Frank Lloyd Wright's Taliesin Fellowship program included cooking to provide an understanding of the human experience in a kitchen to inform their architectural designs to create familiar and useful spaces (Friedland & Zellman, 2009, p.178).

Technologies like Apple Computers and Microsoft grew from independent systems into larger complex infrastructures as they began connecting through the Internet, in turn, they became *Baseline Technologies* within *Large Technical System*s (Demchak, 2012; Geels, 2004; Olsen et al., 2009). Once society began to create innovations utilizing the baseline systems, a need emerged to create *familiar* yet *enhanced* environments within the baseline technologies (Ayoko & Ashkanasy, 2019; Freeman & Perez, 1988; Geels, 2004; Jones et al., 2013; Koehler & Mishra, 2009).

Universal and Inclusive Design

Therefore, the suggestion to use device-equivalent design is not only for a familiar experience to decrease a user's extraneous load but also follows *universal usability/design* or *inclusive design* (Geels, 2004; Grelle & Gutierrez, 2019; Nielsen & Molich, 1990; Steinfeld & Maisel, 2012; Yablonski, 2020). *Universal Design* is a term primarily used in the United States and provides guidelines for designing spaces accessible to as many users as possible (Clarkson et al., 2003; Steinfeld & Maisel, 2012). Operating systems such as Windows, Chrome, Mac, and mobile devices with iOS, Android, and Windows all have accessibility settings built into the device (Edyburn, 2020). When a UI is designed to meet the device's system settings, it provides an *adaptable* experience and aligns with an individual's preferences and accessibility needs (Oppermann, 1994). *Universal Design for Learning* (UDL) was framed after Universal Design and "aims to change the design of the environment rather than to change the learner" (CAST, 2018, "What's the goal of UDL?" section). Yet, the guidelines put the onus on the instructor to provide accessible options, while this may work for a single classroom, it steps away from the core concept of universal design when designing technology (CAST, 2018). The term *Inclusive Design* emerged from Europe around the same time as Universal Design, primarily focusing on accessibility with a similar meaning of designing for all without compromising on the key goals of the technology (Clarkson et al., 2003; Steinfeld & Maisel, 2012). Yet, inclusive design in the U.S. goes beyond the utility of universal design, aiming to create user experiences that respect human characteristics such as gender, age, culture, and race through respectful communication and graphics to create a welcoming environment (Apple, 2024; Joyce, 2022). Podmajersky (2019) focuses on how voice and tone are used in UI for titles, buttons, labels, warning messages, and controls that guide a user through the system in an inclusive, friendly manner. Further, Apple (2024) acknowledges that designs should not be exclusive of cultures but rather represent them within designs by depicting human diversity through images that include different ages, racial or ethnic identities, body types, and physical abilities. Norman (2013) highlights that emotions and cognition often work together in tandem, demonstrating that design choices provide a feeling of belonging and assign value to a user experience.

Yet, recent research on the value of emotion within learning design (Mayer & Estrella, 2014) does not acknowledge basic *color theory* that has already been tied to emotion with a deeper context than using bright or vibrant colors (Fleming, 1967). The sociotechnical theory explains that the interplay between industries is a slow process, with industries making moves at each round of the process, sometimes with setbacks (Geels, 2004). This interplay can be seen between graphic design and learning design, as graphics, colors, or aesthetics within learning design seem to fall into two categories: distracting or enhancing learning (Mayer, 2020), leaving little room for adding value to the emotional aspects of design.

Principles of Design

Color theory examines the hue, saturation, or value of colors and recognizes how colors complement and contrast each other, commonly referred to as complementary colors, tetradic harmonies, analogous colors, and monochromatic colors (O'Connor, 2013; White, 2011, p. 66). Fleming (1967) looked at colors to find that illustrations with color engaged learners and increased knowledge transfer; more so, a minimal number of colors performed better. Using two to three colors in a design, follows color theories such as complementary colors or triadic harmonies (O'Connor, 2013; White, 2011), rather than Mayer's (2014) suggestion of selecting many bright colors to enhance emotion. Fleming's work aligns with design and color theory principles that go back to the 18th century and are part of many UX design guidelines (Lavie & Tractinsky, 2004; Yablonski, 2020). Other design theories, such as the *Gestalt theory*, are often mentioned in UX design and emphasize organized and clear designs by following the principles of continuation**,** closure**,** proximity**,** figure/ground**,** similarity**,** and symmetrical and asymmetrical balance (O'Connor, 2013; Yablonski, 2020). Fleming (1979) and Levie and Lentz (1982) highlighted the need for proximity and emphasized that when illustrations and images are aligned with text in instructional materials, knowledge transfer increases compared to text without images or images without text. Frank Lloyd Wright used continuation through flowing lines and surfaces to bring rooms together in an approachable manner to guide humans through a building (Betsky & Shapiro, 2021), similar to how small aesthetic lines can connect steps of a process together in a UI.

Ngo (2001) recognized that designers who followed design guidelines were more successful, highlighting the need for an evaluation framework to measure aesthetic elements of design based on design principles such as balance, density, rhythm, symmetry, and simplicity. However, these principles seem lost in recent research, perhaps due to decreasing respect for the field of graphic design and professional designers, a shift that came from computer design programs being readily available to the general public (Kaiser, 2019). Therefore, many usability and UX studies of learning technologies evaluate design and color preferences without referencing color theory or design principles, providing misleading findings and outcomes (Lazard & King, 2019; Mayer & Estrella, 2014; Reyna, 2013; Tomita, 2022). Bias (2011) stated that such poorly designed studies and misleading findings further skew amateur design practices, highlighting a need for more rigorous usability studies.

Kruse et al. (2022) refer to the friction and misalignment between fields such as graphic design, UX research, and learning, as the *not-invented-here* (NIH) syndrome, which causes researchers to adapt one model from another industry to their industry rather than finding joint optimization. Geels (2004) explained that different industries and social groups often have their own business models, cultures, and policies, leading to friction and misalignment when they attempt to integrate their frameworks; and can take years to fully integrate. So, while *Learning Experience* (LX) design aims to close the gap between UX and learning design (Banes &

Behnke, 2019), researchers have noted that the LX model overlooks the knowledge of UX required for effective technology-based learning design (Abramenka-Lachheb, 2023; Punchoojit & Hongwarittorrn, 2017; Reigeluth & An, 2023). Punchoojit and Hongwarittorrn (2017) acknowledged that designing an effective user interface requires understanding various disciplines, such as computer science and graphic design, skills a learning designer may not have expert-level knowledge needed for evaluation.

Additionally, it has been noted that cognitive biases often influence users' perceptions of design (Ciriello & Loss, 2023; Lavie & Tractinsky, 2004; Lim et al., 2013; Lu et al., 2022). To put this in perspective, one study explained that a can of beans on display in an art gallery may arouse feelings that would not exist when viewing the can of beans at a grocery store (Juslin et al., 2021). In turn, learners may perceive a design to be of higher value if they respect the instructor and institution. Paul Rand (2014), a pioneer in graphic design, stated:

> Even if it is true that the average man seems most comfortable with the commonplace and familiar, it is equally true that catering to bad taste…denies the reader one of the most easily accessible means for esthetic development and eventual enjoyment. (p. 95)

That is not to say that learners will not recognize a usable design when they see it. Lavie and Tractinsky (2004) found in their usability testing that users often describe layouts as *clean* and *usable* when the designs followed principles such as Gestalt and color theory. Further, following these and other design principles has demonstrated a decrease in cognitive load while using technology (Norman, 2013; Yablonski, 2020).

In contrast, the *aesthetic-usability effect* states that a good aesthetic design can mask usability problems, influencing users to believe a product works well (Yablonski, 2020). While many UX evaluation tools focus heavily on the perception of *aesthetic*s, stating that pleasing designs play a large role in creating a positive user experience, it is vital to look beyond user perception to remove bias and understand the depts of usability and design (Alharoon et al., 2021; Lim et al., 2013; Yablonski, 2020). Avoiding bias in design evaluation can sometimes be achieved by using an iterative design and evaluation process that aims at moving away from the design and developer's views to a more inclusive design aimed at larger populations of learners with each iteration (Clarkson et al., 1999; Tessmer, 1993).

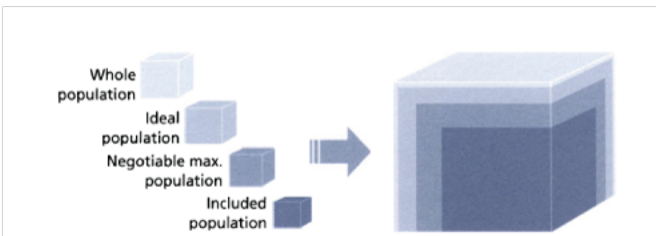
Designing and Evaluating for Usability

Clarkson et al. (2003) outlined a design and evaluation process for developing new and innovative software that follows a technical transitions approach towards an inclusive design. After identifying and understanding the need, a minimal viable product is developed for testing to see if the technology will meet the overall objectives, followed by initial user testing to understand user perception (Clarkson et al., 1999). During the initial testing stages of an educational technology, such as ACTS, it is common to have the designers or developers of the technology set up the technology for use and provide verbal directions for testing. Yet, for more users outside of the design and development to start using the technology, the UI must be updated to meet the usability needs of additional users (Clarkson et al., 1999). Tessmer (1993) outlined the formative evaluation process to include an expert review, followed by an evaluation conducted with learners in a one-to-one setting, a small group, and a field test or summative evaluation in a real-life setting, highlighting a focus on learner-centered design. Similarly, Barnum (2020) suggests evaluating usability through a Big U and little u approach; the Big U refers to formative usability testing that can be done by a designer or developer during the design stage, and the little u refers to testing done with users.

Each design evaluation informs the next, creating or modifying the system for the next iteration of design; with each design, the system opens to more users, leading to an eventual inclusive design (Clarkson et al., 2003). Clarkson et al. (2003) created the *Inclusive Design Cube*, to show how each iteration of the design opens it up for a greater target user group, (see Figure 2.2). Further, formative evaluations for HCI prioritize preventing user errors and are an

Figure 2.2

Inclusive Design Cube



Note. From *Inclusive design: Design for the whole population* by Clarkson, P., Coleman, R., Keates, S., & Lebbon, C., 2003, p. 92. Springer.

integral part of the design process and consider users' needs and accessibility standards, ensuring a safe and friendly system (Issa & Isaias, 2022; Law et al., 2014; Punchoojit & Hongwarittorrn, 2017; Travis & Hodgson, 2019).

The ACTS system has been tested and verified with inservice and preservice teachers for user perception, understanding, and user control (Phillips, 2022). However, the system currently requires a faculty member familiar with the technology to setup a simulation and give the learner instructions on how the technology works; but there is a need to expand ACTS to additional faculty and learners. Thus, the current target users will be for more faculty to create simulations for learners and use the simulation without verbal directions.

Designing and evaluating a product for usability is a process that is commonly aligned with *Key Performance Indicators* (KPIs), which are equivalent to objectives in learning design

(Olsen et al., 2009; Keates et al., 2000; Straub, 2009). From a UX research perspective, KPIs help to understand if users will adopt a new technology and continue to use it; from a learning perspective, they help to determine if the tool is effective in developing skills and transferring knowledge (Olsen et al., 2009; Lim et al., 2013; Straub, 2009). UX research and learning design acknowledge the need to identify KPIs and objectives that align with the learner or user. Geels (2004) acknowledged that it is often unclear what aspects of different social groups should be included in a technical system and how they should be arranged according to hierarchy in design plans for expanding to more users and immediate needs. Many researchers acknowledge that designing and evaluating usability for learning is still uniquely challenging (Fleury & Chaniaud, 2023; Jahnke et al., 2020; Lim et al., 2013). Straub (2009) highlights the complexity of technology adoption and evaluation in education and the friction between usability testing tools and educational technology, and many studies do not align usability tests with traditional methods of success in learning, such as learning outcomes through summative exams. While many principles of the UX Honeycomb model are valuable to technology-based learning research, the exclusion of learning creates a gap between KPIs for learning and HCI (Morville, 2014). Further, many usability and UX evaluation methods follow Roger's DOI theory and aim to understand adoption in a competitive market (Nielsen & Molich, 1990; Rogers, 2003; Straub, 2009). Further, Kosch et al. (2023) found that HCI has adapted tools for measuring cognitive load from other fields, including education, that do not properly align to HCI usability measurements. In turn, consideration must be made per-project on the specific need for KPIs and the indicators needed to evaluate them.

Identifying KPIs

Reviewing the literature, two educational frameworks emerged with a sociotechnical perspective for designing learning. Garrison's (1999) *Community of Inquiry* (CoI) model looks at social integration into technology, while Koehler and Mishra's (2009) TPACK framework looks at integrating technology into a social setting. Further, both models acknowledge a foundational concept from HCI and learning design of understanding the cognitive impact of bringing society and technology together.

Many researchers have adapted the TPACK framework, some of which relate to the ACTS system. The *Intelligent-TPACK* framework, is a scale to measure teachers' knowledge of AI and expanded an element for ethical considerations (Celik, 2023). Hadjerrouit (2017) used the TPACK model to understand a knowledge visualization tool's technical, pedagogical, and content affordance, aligning the content in the technology to support pedagogical aspects such as motivation, feedback, and critical thinking. Similarly, the Technological Pedagogical Content Design (TPCD) framework moves away from the knowledge aspect and views content as the learning materials, placing a heavy emphasis on how content and aesthetics relate to usability (Hosseini et al., 2022). Qin et al. (2020) applied the CoI framework to design a chatbot learning tool, aligning each section of the tool to the content needed for the curriculum. They identified social aspects of the chat through emojis; demonstrating an alignment of CoI to text-based chat to analyze the social aspects of learning using a CA. Yet, few studies had a full sociotechnical perspective of society's influence on technology and technology's influence on society. Jahnke et al. (2020) developed the sociotechnical-pedagogy framework that takes usability and adds social and pedagogical aspects. However, the framework does not follow a sociotechnical perspective outlined by Trist (1981) of acknowledging that society and technology are jointly connected, creating a need designing intuitive technologies that are familiar to users, instead Jahnke et al. (2020) proposes the need for instructions and guidelines for multimedia technologies and navigation. Additionally, the framework was designed for full courses as compared to a learning technology, in turn the framework does not align to a SBE system such as ACTS (Leacock & Nesbit, 2007). Yet, the article also provides an example of how learning design has a different perspective of the end user than UX design, where learning design emphasizes the need to instruct and guide a user through a system, and UX looks to create a jointly optimized experience between the user and the technology.

Social-TPAC

Therefore, this study creates a new framework to understand the KPIs needed for evaluating the ACTS systems. The TPACK and COI frameworks have many crossovers, especially in terms of pedagogical knowledge and teacher presence. While TPACK was previously adapted for an AI-based scale and as a UX-based scale; the models did not include cognitive load and cognitive presence, which are essential elements of learning and usability (Celik, 2023). Learning aims to process information, construct knowledge, and develop skills without increasing cognitive load during a learning activity (Sweller et al., 2018). In turn, four KPIs were identified from the TPACK and CoI frameworks: Technical Indicators (TI), Pedagogical Indicators (PI), Content Indicators (CI), and Social Indicators. As the sociotechnical theory demonstrates, these four KPIs are the co-evolution of technology, society, content, and pedagogy—bringing together form and function in one design to create the optimal learning experience (Garrison et al., 1999; Geels, 2002, 2004; Koehler & Mishra, 2009). While these four KPIs function separately, they are also brought together under one sociotechnical experience. Therefore, they must work together at an ideal level of cognitive presence to create an effective user experience (Geels, 2004). Yet, when poor usability or social interaction impacts a learner's extraneous cognitive load, it impacts their ability to be cognitively present for the simulation. Thus, the model also represents Cognitive Presence and Cognitive Load to address any areas that may increase or decrease the optimal level of cognitive thinking. The cognitive load will primarily look at extraneous load from technical usability errors, distracting aesthetic design elements, and content produced by AI that may not align with the learning goals. At the same time, the Cognitive Presence will look at the level of social pedagogical or teaching presence within ACTS. Figure 2.3 demonstrates how the KPIs must align and work together to create a learning experience that allows for cognitive presence. The cognitive load will primarily look at extraneous load from technical usability errors, distracting aesthetic design elements, and content produced by AI that may not align with the learning goals. At the same time, the Cognitive Presence will look at the level of social pedagogical or teaching presence within ACTS.

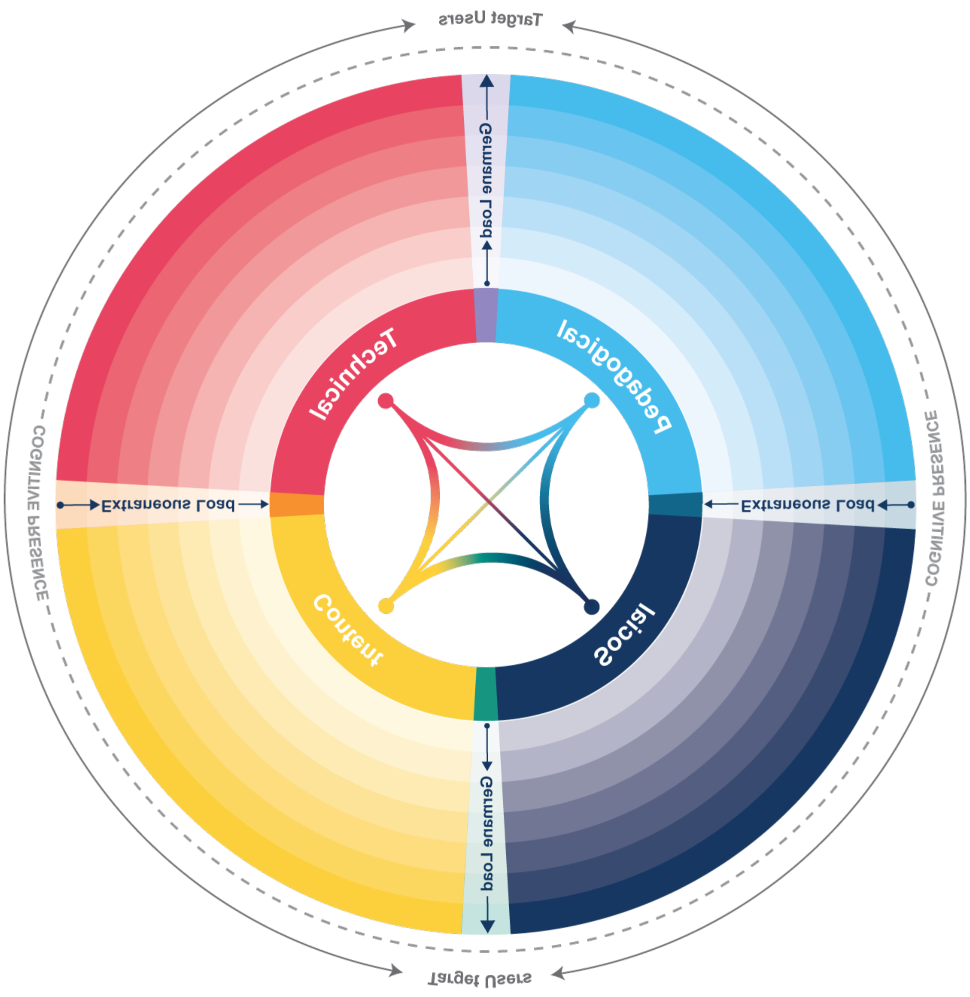
Creating Heuristics

Sociotechnical systems are often changing and fluctuating (Geels, 2004; Koehler & Mishra, 2009). Geels (2004) stated that there is often "substantial uncertainty in optimizing the optimal design heuristics, user preferences, behavioral patterns, and public policies," all of which align with using a new tool such as Conversational AI for learning. Keates et al. (2000) suggest creating criteria for each KPI that can be used during the evaluation process through *heuristic evaluations*. Heuristic evaluation is a usability testing method where a user is presented with a technology and asked to use it to identify problems. Generally, an evaluator observes and collects feedback from the user after the test (Nielsen & Molich, 1990). Multiple authors have proposed specific heuristics for specific domains, such as learning design, game design, website design, or mobile app, but few heuristics protocols align with more than one type of technology

(Borsci et al., 2021; Langevin et al., 2021; Quiñones & Rusu, 2017); Lu et al. (2022) found that many usability and UX studies of learning design had unreliable results, likely due to a misalignment of KPIs and heuristics.

Figure 2.3

Social-TPAC



Note. Social, pedagogical, content, and technical KPI are shown as intertwined elements in the center, with extraneous and germane load represented on the outside with a gradient circle indicating that when extraneous load is decreased, germane load increases, allowing for cognitive presence for the target users.

Quiñones and Rusu (2017) state that there is no reliable heuristic evaluation protocol for validating the usability of a product, but instead suggest a methodology for creating heuristics by extracting information from principles of design, existing heuristic guidelines, theories, or UX research interviews; then transforming that information into a set of heuristics that align to the evaluation methods. Thus, the literature review will assess existing design guidelines, heuristic evaluation tools, and learning criteria to create a set of heuristics, and indicators aligned to the KPIs identified in the Social-TPAC framework.

Nielsen's Heuristic Evaluation of User Interfaces

Nielsen and Molich's (1990) heuristics are the most well-known and useful criteria for evaluating a UI. Despite the sociotechnical theory outlining that technology is an ever-changing field, the heuristics offer the basics of HCI and the baseline requirements for evaluating a technology for usability. Concepts such as supporting undo and redo, the system reacting and providing instant feedback, and a user interface that is *adaptable*, *predictable*, *dependable*, and *intuitive* are all needed to create a familiar experience. Nielsen and Molich's (1990) also outline accessible heuristics, such as a high-level of contrast for text, buttons that are large enough to select, and clear labels for navigation. Further, the heuristics follow concepts of Gestalt such as continuation, proximity, and similarity, which are also mentioned in Yablonski's (2020) book *Laws of UX*.

Laws of UX

Yablonski's *Laws of UX* (2020) leans heavily on many of the same concepts from Nielsen and Molich's (1990) heuristics while also providing other concepts from HCI and psychology principles to outline the basics of creating a successful UX. Psychological theories from Gestalt's grouping principles, such as proximity, similarity, and continuance, are essential elements of aesthetic design that guide the user through a UI. The book also acknowledges theories related to biases that can affect user feedback, such as the aesthetic-usability effect, which is when an aesthetically pleasing design hides poor usability, and the peak-end rule, which states that users often base their opinion on a user experience based on their last known interaction (Yablonski, 2020). While theories on bias should not be included in heuristics, they outline the very need for heuristics to evaluate an experience and the need for an expert review to identify them in a design. Many of the principles outlined by Yablonski (2020) are often mentioned in design guidelines as well.

Apple's (2024) Human-Interface Guidelines

As noted by the sociotechnical theory, there is a need to consider the baseline technologies like Apple Computers and Microsoft to achieve familiar designs (Apple, 2024; Demchak, 2012; Geels, 2004; Olsen et al., 2009; Microsoft, 2016; Steinfeld & Maisel, 2012). While Microsoft (2016) offers an *inclusive* design guidebook, the design of the guide itself follows an exclusive pattern of design by using images void of skin color and diversity. Apple (2024) suggests depicting diversity by using diverse names and images representing a variety of races, ages, and genders. Therefore, it was decided to include Apple's (2024) *Human-Interface Guidelines* to consider a familiar and inclusive environment. However, it is important to consider the audience and baseline devices that learners will use when developing heuristics, should a classroom only have Microsoft-based computers, it would be wise to consider their guidelines to create heuristics.

Apple (2024) provides guidelines that align with color theory and design principles for typefaces, fonts, buttons, navigation, and layout (Fleming, 1967; O'Connor, 2013; White, 2011). Furthermore, they highlight the need for a *consistent* design across the technology, including the aforementioned design elements and using a neutral, approachable *voice* and tone (Apple, 2004). Another valuable concept they include is to use a grid layout and hierarchy to create an organized, *aesthetically* pleasing design. A grid-based design and hierarchy help to define *clarity* among the sections, subsections, and navigation on UI screens, often done through headings or Gestalt theories (Apple, 2024; Bringhurst, 2004; Kimball & Hawkins, 2007). Often, it is suggested to decide on hierarchy for text size by using the Golden Ratio, which is approximately 0.618 and often found throughout nature; when applied to design, it helps to create hierarchy in an eye-pleasing manner (Bringhurst, 2004; Lidwell et al., 2010). Lastly, they pointed out the need for *clarity* and *predictability* when interacting with the UI by explaining the action or task the button or link will do by first using a verb, such as "add a new simulation" or "remove a user" (Apple, 2024). Further adding brief explanations or help in context, allowing a user to use new or unfamiliar elements without having to read directions.

WCAG 2.1 Guidelines

Concerning accessibility, the *Web Content Accessibility Guidelines* (WCAG) version 2.1— released in 2018 by the World Wide Web Consortium (W3C), provides a full range of parameters that web developers and designers can reference to create websites that are compliant with baseline web browsers and HTML accessibility standards (Paul, 2022). The guidelines consider items such as color, contrast, readability, keyboard accessibility, and page layout, focusing on creating an accessible website that aligns with tools and systems that assist people with accessibility needs to ensure the greatest number of people can access information (Paul, 2022). For example, WCAG level AA requires a contrast ratio of at least 4.5:1 for normal-sized text and 3:1 for graphics, UI elements, and large text to assure readability. Further, WCAG provides resources that detail how to test ACTS for accessibility, such as tools for testing contrast levels of text for readability. Additionally, many browsers, such as Mozilla Firefox or Google Chrome, offer developer tools that follow the WCAG standards to test a website for accessibility issues. ACTS operates within a web browser and should align with many of the guidelines.

Strategic Writing for UX

Podmajersky's (2019) book on strategic writing for UX provides many tips on writing content that enhances the usability of technology, many that cross over with Apple's (2024) guidelines, such as the need for an *aesthetically* pleasing grid-based layout, hierarchy, and using an active *voice* for instructions and button labels to guide users through the UI (Podmajersky, 2019). Further explaining the active voice of creating directions that follow a "to do x, do y" action statement. The book also highlights an often overlooked UX design element known as *empty states* (Podmajersky, 2019). Empty states are when a screen is completely blank because the user has yet to add content to the system. For example, if a user recently created a faculty account in ACTS, they likely will not have created any simulations or added any learners to the system yet. In turn, the screen will be empty, which can cause confusion for a new user. Podmajersky (2019) suggests adding a sample of what will go there, guiding the user toward adding an item that would display in its place after creation, making the UI more *consistent* and usable.

Quality Matters and OSCQR

Within a learning setting, heuristics can often be seen in rubrics created to evaluate learning design. Two popular online learning design rubrics were reviewed, primarily for the pedagogical KPI of the social-TPAC model—the *Quality Matters Higher Education Rubric, Seventh Edition* (QM) rubric, and the *SUNY Online Course Quality Review Rubric* (OSCQR), (Online Learning Consortium, 2024), both which have an expert-level peer review process known for their validity and reliability in assessing online course design. Of note, OSCQR and QM are designed for assessing online courses and not SBE or the usability of technology (Gregory et al., 2020; Leacock & Nesbit, 2007; Legon, 2015), yet provide insights for evaluating ACTS. OSCQR follows some of Apple's (2024) and Podmajersky's (2019) guidelines, such as *consistency* across the technology and the need for hierarchy in design and using an icon set commonly used in technologies to create an *aesthetically* familiar UI. Further, they emphasize providing learners with multiple opportunities to track their learning progress and aligned feedback. QM and OSCQR also highlight the need for *high-quality instructions,* including a welcome message and clear directions on where to start. Providing instructions follows a learning design framework, yet the concept is echoed from a UX perspective in Podmajersky's (2019) avoidance of empty states and Apple's (2024) using an active voice to guide the user through navigating the UI without written instructions. Lastly, QM and OSCQR point to the need for specific and descriptive criteria on how they will be evaluated is made available to all learners. While ACTS does not grade learners on their performance, it does use the IQA tool to evaluate the learners, which should be evident to users to *support* their learning (Boston, 2012).

Simulation Based Education Literature

As noted, SBE is generally more successful when there is a pre-brief and a debrief to discuss feedback with learners; therefore, a variety of SBE literature provides a basis for creating heuristics related to the pedagogical KPI that outline a *quality of instruction* that is *supportive* and *encouraging* (Camarata & Slieman, 2020; Cheng et al., 2021; Cheng et al., 2020; Dreifuerst, 2012; Lapum et al., 2018; Runnacles et al., 2014). While debriefing in simulation is often a conversation and not direct feedback, the healthcare field has developed many tools to guide and evaluate effective debriefing that applies to providing feedback in a computer-based simulation (Cheng et al., 2020). Further, the literature ties the CoI framework to providing a high level of instructor and social presence when giving feedback online (Camarata & Slieman, 2020; Cheng et al., 2020; Lowenthal & Dunlap, 2020). Similar to Apple (2024), the debriefing literature mirrors the need for inclusivity, such as addressing learners independently of their gender, well-being, or age (Borsci et al., 2021; Cheng et al., 2020; Lowenthal & Dunlap, 2020). In a similar vein, they stress providing positive and negative feedback, and closing performance gaps by suggesting specific ways to improve (Cheng et al., 2021; Cheng et al., 2020). Cheng et al. (2020) further outlined that negative feedback should be delivered in the first person, such as "I noticed you," to frame the problem objectively. In line with the ACTS system that utilizes the IQA to provide feedback, Cheng et al. (2021) highlighted the need to use data to inform feedback. Similar to themes outlined by OSCQR and QM, SBE also focuses on welcoming the learners and orienting them to the environment while also communicating the rules and expectations and emphasizing the importance of confidentiality. Yet, it also highlights that students should be informed to refrain from using their phones or checking their emails (Cheng et al., 2020), something that does not need to be included in OSCQR or QM due to their alignment with semester-long courses.

BOT-Check

BOT-Check is a diagnostic checklist that analyzes conversations with AI-based CAs (Borsci et al., 2021). The BOT-Check is a formative tool for developers or designers to examine a CA without focusing on consumer marketing, which many previous chatbot usability evaluations concentrate on (Borsci et al., 2021; Cheng, 2018). Further, the BOT-Check steps away from shorter chat systems and reviews longer form conversations for credibility, cohesiveness, interaction, and accuracy of the AI responses, which aligns well with ACTS' needs. Studies focusing on social presence aligned with the CoI framework also provided insight into creating heuristics for the ACTS CA that aligned to *affective*, *interactive, cohesive*, and *credible* social presence (Liebrecht et al., 2021; Rourke et al., 1999).

Evaluation Methods

According to Parcell's (2013) article on digital.gov, usability tests the "ease of use and intuitiveness of a product" by evaluating users as they perform tasks to identify problems that prevent or hinder them from completing a goal. Travis and Hodgson (2019) stated that good design is a symptom of effective UX research and valuing a human-centered design. Further, stating that at its core UX research answers "can people use the thing we've designed to solve their problem" (Travis & Hodgson, 2019, p. 18). Travis and Hodgson (2019) acknowledged that as few as five participants in a usability study can uncover 85% of the usability problems, and the key to uncovering problems is through effective evaluation methods. One way to create an effective evaluation is to create a set of *heuristics* for usability testing (Nielsen & Molich, 1990).

Heuristic Evaluations

Heuristic evaluation is a usability testing method where evaluators are presented with an interface design and asked to use the technology to identify problems (Nielsen & Molich, 1990). According to Nielsen and Molich (1990), an evaluator must sit alone and use the product without interruption. They can then write down anything they see as an error or thought, or they can use the *Think Aloud* method and verbally speak about their thoughts as they progress through the system (Barnum, 2020). While observing a user, it is recommended to observe without any interruptions, including when a user makes an error, as this allows the researcher to understand how a user will recover from the mistake (Travis & Hodgson, 2019). Nielsen has stated the think aloud method is the number one way to test usability (Nielsen & Molich, 1990). Travis and Hodgson (2019) stated that summative and formative heuristics evaluations of task analysis provide the strongest evidence for UX research. Additionally, Norman (2005) has pointed to the need for experts to perform heuristic evaluations, and other studies have demonstrated success in utilizing expert analysis to improve the UI of learning technologies (Zardari et al., 2020).

Heuristic evaluations can also be accompanied by tools such as eye tracking to record a user's eye movements and gaze across a screen or device to measure visual attention (Carter & Luke, 2020). The more attention a user places on a certain element signifies cognitive thinking. Eye tracking has historically been expensive, and even with technical advances that are more affordable, few studies have utilized the technology for text-based chat. Yet, as noted by Barnum (2020), screen captures or capturing video of a learner using technology also provides usable data similar to eye-tracking that can be done with technology commonly found on most modern-day devices. Barnum (2020) pointed to recognizing facial movements as a way to identify cognitive load; for example, if a user pauses and tilts their head, it may indicate confusion with the technology. With the unique nature of ACTS's knowledge visualization, analyzing eye tracking data or video data would help to understand a learner's cognitive thinking to align with other testing methods.

Heuristic evaluations provide a detailed list of usability problems, which can sometimes be too difficult to update in a system to correct all problems. Therefore, rating each heuristic's severity helps identify the significance and impact on the user experience while creating a plan for recommended updates (Barnum, 2020). Hassenzahl (2000) acknowledged that severity levels can be data-driven or judgment-driven, with measures such as time to fix the problem (data) and user interests (judgment) used in determining the order to heuristics. Nielsen (1994) stated that the market impact should be considered with severity scales for the growing popularity of a technology. While this aligns with Roger's (2003) DoI theory and less on the sociotechnical approach, it also highlights the need to align usability to the concerns of the target user group and the Inclusive Design Cube (Clarkson et al., 2003; Lim et al., 2013).

Questionnaires

Questionnaires are a popular evaluation method due to their ease of use (Lu et al., 2022; Norman, 2013; Clarkson et al., 2003). Many usability questionnaires were developed by usability experts and rely on end-users to assess the usability of a technology (Marangunić & Granić, 2014; Schrepp et al., 2017). Four questionnaires are often mentioned in usability research; however, each model has deficits that do not properly align to the ACTS system. The *Technology Acceptance Model* (TAM) primarily focuses on a technology's perceived usefulness and ease-of-use (Davis, 1989; Marangunić & Granić, 2014); but is more suited for individual adoption and less for educational purposes (Cheng, 2018). The *Unified Theory of Acceptance and Use of Technology* (UTAUT) focuses on predicting if a user will adopt and accept a technology, yet it considers controversial elements of social influence, including gender and age (Venkatesh et al., 2003), which does not align with research on familiarity. The *System Usability Scale* (SUS) focuses on a technology's usability and can be easily adopted to many settings but does not include questions about the user's experience with the technology (Bangor et al., 2008). The *User Experience Questionnaire* (UEQ) is an adaptable questionnaire that provides a UX perspective (Schrepp et al., 2017); however, the tool does not align to learning technologies due to its lack of a pedagogical perspective. Yet, questionnaires allow for users to provide feedback outside of observations and rate their experience, demonstrating a need to create a questionnaire that aligns to the evaluation criteria identified for the UX research.

Social-TPAC

Lu et al. (2022) called for further research to understand the best methods and protocols to evaluate learning technologies for usability and UX effectively, noting that while many usability tools exist, they often aim to understand if a technology is marketable and do not consider the unique nature of learning tools. Yet, the sociotechnical theory explains that strategies, behaviors, policies, and heuristics are not static but ever-changing due to society and technology constantly reshaping their very structure (Geels, 2002). These changes in structure for innovations happen in a non-linear format, making it difficult to align UX and usability evaluation tools from one product to the next. Further, Geels (2004) underlines how the differences between social groups, such as HCI and learning design, require longer periods of time to integrate, a theme echoed by Koehler and Mishra's (2009) TPACK framework. While many researchers emphasize the need for utilizing validated tools, applying them to different fields of research or platforms often causes a misalignment that leads to misleading results. Therefore, there is a need to align evaluation not from the point of using a validated tool but to the objectives, needs, and KPIs using a sociotechnical lens that considers the technical (e.g., devices, systems) and social (e.g., instructors, learners) aspects of a learning experience using technology. Pushing the need to go beyond user-centered design to consider device-centered design, acknowledging that society and technology influence one another, forcing a need for *familiar* designs.

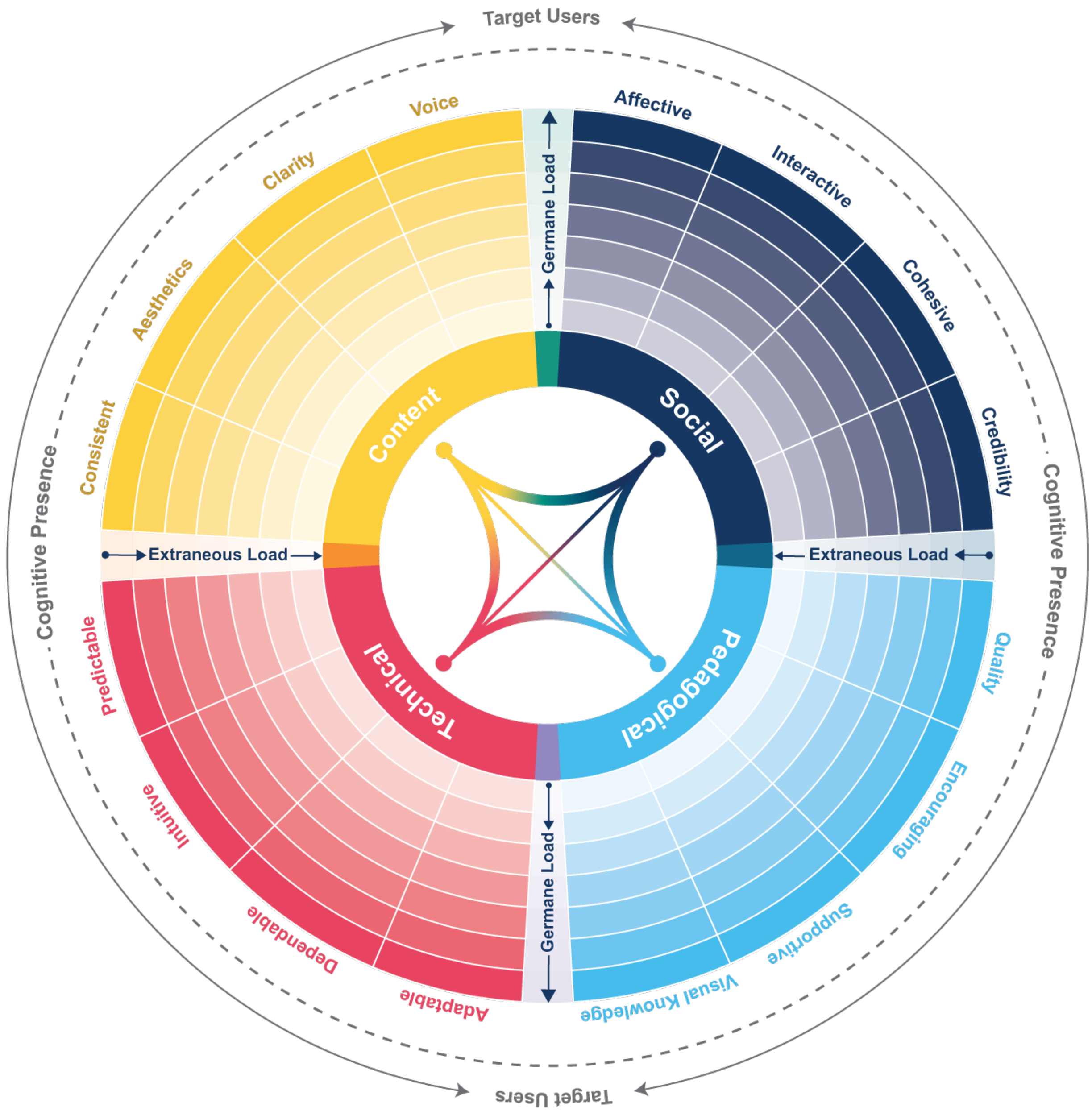
Sociotechnical theory acknowledges that friction between social groups to create their own structures and policies can sometimes cause setbacks in the field. While the field of learning has looked to understand how design or colors influence the cognitive experience of learning, they have backed away or attempted to recreate design principles that have existed for hundreds of years (Bias, 2011; Kaiser, 2019; Lazard & King, 2019; Mayer & Estrella, 2014; Ngo, 2001; Reyna, 2013; Tomita, 2022). Further, fields such as HCI and learning design cannot be easily combined by creating learning experience (LX) design, as the knowledge from multiple fields generally takes years and significant structuring to integrate fully (Abramenka-Lachheb, 2023; Geels, 2004; Punchoojit & Hongwarittorrn, 2017; Reigeluth & An, 2023). While some researchers look to replace usability with UX, it is important to understand that the role of a UX researcher is to understand the form and function of society and technology together, and usability is *one of many sub-functions* that create a user experience (Geels, 2004; Morville, 2014).

Therefore, to understand the ACTS system, it is necessary to create a set of heuristics taken from different fields without attempting to restructure them. Instead, they should be applied holistically to evaluate the learning experience of an innovative technology. Likewise, rather than attempting to reach validity, acknowledge that some tools like QMs and OSCQR were created for online semester-long curriculums and will not apply to a single exercise such as a simulation-based learning experience yet provide insight into the pedagogical lens of computer-based learning that is lacking from the field of HCI. Garrison's (1999) CoI framework provides a vital understanding of how teaching presence and social presence are needed to process a cognitive learning activity in computer-based education. The sociotechnical theory emulates the need to understand best practices from HCI, SBE, and learning design while acknowledging best practices of the baseline technologies to create a familiar and inclusive experience.

The Social-TPAC framework provides an outline for evaluating a simulation-based learning experience using conversational AI with knowledge visualization. Currently, few tools exist that align specifically with the evaluation needs of the ACTS system; therefore, while the Social-TPAC framework may apply to other systems, the specific heuristics must be aligned to the exact tool. In turn, this study will use the Social-TPAC framework to create the following groupings of heuristics: (1) Social: *affective, interactive, cohesive, credibility*; (2) Technical: *adaptable, dependable, intuitive, predictable*; (3) Pedagogical: *quality of instruction, visual knowledge, encouragement, supportive*; and (4) Content: *consistent, aesthetics, voice, clarity* (see Figure 2.4).

Figure 2.4

KPIs in the Social-TPAC Framework



CHAPTER III: METHODS

This study aims to understand the usability of a text-based Conversation Agent (CA) that is used as a learning tool for preservice teachers to practice teaching mathematics. While many usability tools exist, they do not fully align with the unique nature of learning tools (Lu et al., 2022). Further, few usability tools focus on the sociotechnical aspects of using AI within SBE to practice teaching math with a knowledge visualization tool, followed by AI-based feedback. This study looks to align usability research to a unifying framework that will allow for a more holistic approach to usability testing using heuristics pulled from different areas of research, including HCI, SBE, AI, online learning design, and from baseline technologies design guides to understand the impact of usability issues at the cognitive level. Then, the heuristics are aligned with *key performance indicators* (KPI) determined from theories in technology, pedagogy, social presence, and content design for a sociotechnical perspective.

This study will use expert analysis of design and UX, an observation analysis of users working with both human and non-human actors, and surveys of user experience. Specifically, the research questions are:

RQ1: What are the social, pedagogical, content, and technical indicators of user experience that emerge from expert review of ACTS?

RQ2: What are the social, pedagogical, content, and technical indicators of user experience that emerge from participants using ACTS?

RQ3: What are participants' perceptions of the social, pedagogical, content, and technical indicators of user experience after using ACTS?

Plan of Inquiry

Clarkson et al. (2003) outlined an evaluation process for innovative software that follows an approach towards an inclusive design. The process allows for a gradual approach to enhancing a technology to include more users with each iteration. The ACTS system has been tested and verified for user perception but needs additional usability testing (Phillips, 2022). Currently, ACTS requires developers and researchers familiar with the system to set up simulations but hopes to open it up to more faculty to increase learner participation. The literature review revealed that many usability questionnaires do not align with learning design, plus they were created to understand the market value of technology, which is currently not the aim of the ACTS system. Therefore, a heuristic evaluation process was chosen to set design goals to include more learner and faculty users. The heuristic evaluation follows the methods set by Nielsen and Molich (1990) and outlined by Barnum (2020), following a Big U and little u approach by first conducting a heuristic evaluation led by a UX researcher, followed by user testing. Norman (2005) pointed to the need for a Big U approach by having experts perform heuristic evaluations, and other studies have demonstrated success in utilizing expert analysis to improve the UI of learning technologies (Zardari et al., 2020). While the little u approach can be conducted by having an evaluator observe and collects feedback from the user (Nielsen & Molich, 1990). Thus, participants were asked to interact with the ACTS system to complete a series of tasks while on a Zoom recording to capture their experience. Users were then asked to provide any verbal feedback, followed by completing a questionnaire that aligns directly to the KPIs to compare to the heuristic evaluation.

The literature review identified that sociotechnical systems are often in flux and, therefore, difficult to optimize heuristics that perfectly align with each technology and its goals for the users who interact with the technology (Geels, 2004; Koehler & Mishra, 2009). Further, multiple studies pointed to creating heuristics and indicators that align to KPIs of the technology to use for heuristic evaluations (Keates et al., 2000; Quiñones et al., 2018). Multiple heuristics were identified that met the needs of specific domains, such as learning design, application design, conversational agents, or web design, but few heuristics protocols aligned with ACTS (Borsci et al., 2021; Quiñones & Rusu, 2017). Quiñones and Rusu (2017) state that there is no reliable heuristic evaluation protocol for validating the usability of a product, but instead suggest a methodology for creating heuristics by extracting information from principles of design, existing heuristic guidelines, theories, or UX research interviews; then transforming that information into a set of heuristics that align to the evaluation methods. Therefore, this study follows the methodology framework outlined by Quiñones et al. (2018) for developing heuristics and criteria for evaluations. Jahnke et al. (2021) followed this methodology to create sociotechnical-pedagogical heuristics to evaluate learning management systems (LMS). While Jahnke et al.'s (2021) heuristics for an LMS greatly vary from ACTS, the study provided insight into creating heuristics with a sociotechnical lens. Quiñones et al. (2018) created an eight-step iterative process, yet acknowledged that since many steps overlap, they may be performed together, and following a sociotechnical perspective, also acknowledged that some steps are optional depending on the specific technology and social use. The literature review for this study largely covered the first three steps of understanding the problem, identifying resources and heuristics, and therefore the next section will outline how evaluation criteria were selected and aligned to the heuristics and KPIs.

Identifying Heuristics and Evaluation Criteria

The literature review highlighted several studies with heuristic evaluations directly related to the social, technical, pedagogical, and content aspects of the Social-TPAC framework. However, some concepts outlined in the literature review, such as Gestalt principles and UX design, were not represented in heuristics or evaluation rubrics; therefore, based on the literature review, two user experience books and a set of design guidelines were reviewed to create additional criteria (Podmajersky, 2019; Yablonski, 2020). Further, additional research studies and protocols were included to provide guidance on groupings (Lowenthal & Dunlap, 2020; Schrepp et al., 2017). Lastly, following a sociotechnical approach, the study also references the design guidelines from Apple Computers to provide an understanding of baseline technologies (Apple, 2024).

In total, twelve studies, protocols, and guidelines were chosen to develop heuristics and criteria for evaluating the ACTS system (Apple, 2023; Borsci et al., 2021; Camarata & Slieman, 2020; Cheng et al., 2021; Cheng et al., 2020; Liebrecht et al., 2021; Nielsen & Molich, 1990; Online Learning Consortium, 2024; Podmajersky, 2019; *Quality Matters Higher Education Rubric, Seventh Edition*, 2023; Runnacles et al., 2014; Yablonski, 2020). The criteria were then analyzed by developing a codebook in *Microsoft Excel* (see Appendix A) with lists for priori codes for the social, technical, pedagogical, and content aspects of the ACTS system. Some protocols, such as the *Chatbot Usability Scale*, applied only to the content generated in the Conversational agent, while heuristics, such as *Nielsen's heuristics,* applied to all aspects of ACTS (Borsci et al., 2021; Nielsen & Molich, 1990). Therefore, the first step was to identify the functional aspects of ACTS in alignment with the KPIs (see Table 3.1). While the main aspect of ACTS is the conversational agent, knowledge visualization, and feedback, opening and setting up a simulation in the system requires various steps previously identified as complex, and therefore included in the study.

Table 3.1

KPIs – ACTS System

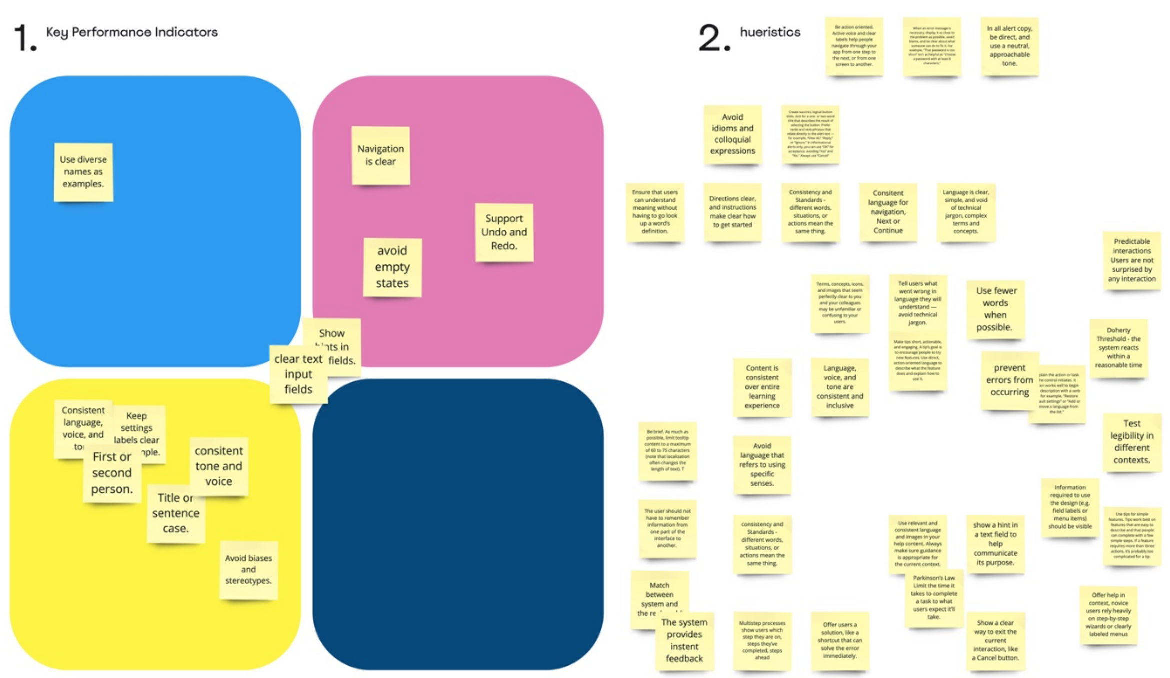
KPI	ACTS System
Social	Conversational Agent, Feedback from AI
Pedagogical	Setup Process, Knowledge Visualization, Feedback from AI
Technical	Setup Process, Conversational Agent, Knowledge Visualization, Feedback from AI
Content	Setup Process, Conversational Agent, Knowledge Visualization, Feedback from AI

After identifying the elements of the system, criteria that only pertained to one aspect of the system were added to a list for the appropriate KPI. Criteria that met various applications were correlated into one list for additional analysis. While some criteria were already grouped accordingly in step two, others applied to the entire experience with ACTS and, therefore, needed additional analysis. The list of criteria was analyzed and coded using priori codes for the KPIs and system; additionally, they were coded with priori codes identified from the literature review or themes identified during the analysis. Criteria that did not apply to any aspects of ACTS or KPI were removed from the list. After the first pass of the analysis, the criteria were sorted to identify themes and initial groupings that would become the heuristics.

The groupings were then moved into a *Miro* board to sort and identify additional overlap using a concept mapping technique. Each indicator statement was added to a Miro board as a sticky note and then was sorted into boxes that represented the KPI (see Figure 3.1). Sticky notes with cross-over were initially grouped in the middle to be analyzed after identifying themes.

Figure 3.1

Miro Board



.

After the criteria were aligned to their respective KPI, they were analyzed to identify heuristic groupings. The groupings were formed to clearly define the objective and definition of sub-functions that comprise the Social-TPAC framework in alignment with the sociotechnical theory (Cherns, 1976).

Content Heuristics

There were a lot of crossovers in the criteria related to the technical and content KPIs, and therefore sorting criteria to identify indicators for each heuristic was particularly thought-provoking. Yet, it is important to separate content from technical aspects to emphasize the tone and clarity of words and how they appear (aesthetics) on a page can make or break an effective user experience. Of note, while voice and tone emerged across various tools to create an approachable experience, Apple's (2024) *Human-Interface Guidelines* focused on a need for inclusive language and designs, a topic often mentioned in education but generally absent from many UX studies. Yet, since an approachable experience is also an inclusive experience, many aspects of voice and tone fell under *Inclusivity, including* aspects of aesthetics*,* further demonstrating the sociotechnical aspects of technology. Therefore, the following definitions were assigned to Consistent, Aesthetics, Inclusivity, and Clarity (see Table 3.2).

Table 3.2

Content Heuristics and Definitions

KPI	Heuristic	Definition
Content	Consistent	Content is consistent over the entire learning experience.
	Aesthetics	Overall appearance of the technology represents a well-designed platform.
	Inclusivity	Language, voice, and tone are consistent and inclusive.
	Clarity	Content is clear, simple, and void of technical jargon, complex terms and concepts.

Note. Heuristics and definitions based on literature review.

Technical Heuristics

Technical heuristics focus on creating a a familiar experience. As noted by the sociotechnical theory and technological changes, users seek out a familiar user experience that is intuitive. Therefore, when users interact with a technology, they expect it to act and perform according to how they would have previously interacted with a technology. Thus, many of the criteria selected for the heuristics related to these concepts. Additionally, the idea of developing technology with a device-centered approach is also emphasized in the heuristics, many usability and learning design studies focus on user-centered design with no mention of designing for baseline technologies, allowing design to step away from a familiar user experience. See Table 3.3 for a list of technical heuristics and definitions.

Table 3.3

Technical Heuristics and Definitions

KPI	Heuristic	Definition
Technical	Adaptable	The technology follows platform and industry conventions to create a familiar user experience.
	Dependable	Users are in control of the interaction and can avoid and recover from errors.
	Intuitive	The UI can be used at a subconscious level without instructions, allowing users to focus on the task.
	Predictable	Users know what will happen before performing or interacting with any element.

Note. Heuristics and definitions based on literature review.

Social Heuristics

Garrison's (1999) *Community of Inquiry* informs the *social presence* aspect in the Social-TPAC conceptual framework. Under CoI, social presence was defined as emotional expression, open communication, and group cohesion and was later redefined as affective responses, interactive responses, and cohesive responses, as shown in Table 3.4 (Rourke et al., 1999).

The categories identified by Rourke et al. (1999) directly informed the heuristics for the social heuristics, while the criteria were included alongside attributes from the BOT-Check, a diagnostic checklist that concentrates on analyzing a conversation with AI (Borsci et al., 2021). The BOT-Check was developed as a formative tool for developers to examine a CA without focusing on consumer marketing, which many previous chatbot usability tools focused on (Borsci et al., 2021). Unlike the Rourke et al. (1999) study, the BOT-Check reviews the credibility, appropriateness, and accuracy of the AI responses to the human, which were categorized under *credibility* in the heuristics of the Social-TPAC (see Table 3.5).

Table 3.4

Social Presence (Rourke et al., 1999).

Heuristic	Indicators
Affective responses	Expression of emotions Use of humor Self-disclosure
Interactive responses	Continuing a thread Quoting from other messages Referring explicitly to other messages Asking questions Complimenting, Expressing agreement
Cohesive responses	Vocatives Addresses or refers to the group using inclusive pronouns Phatics or salutations

Table 3.5

Social Heuristics and Definitions

KPI	**Heuristic**	**Definition**
Social	Affective Responses	The AI or human generated text to demonstrates emotional expressions.
	Interactive Responses	There was a high level of interactivity between the messages exchanged by participants, AI, or human.
	Cohesive Responses	The AI or human demonstrates a genuine interest in the other participant, demonstrates acceptance and approval.
	Credibility	The AI or human demonstrates clear, accurate, and contextually appropriate communication.

Note. Heuristics and definitions based on literature review.

Pedagogical Heuristics

Cheng et al. (2020) utilized the CoI to outline a framework for debriefing a healthcare simulation online. While debriefing in simulation is often a conversation and not direct feedback, the healthcare field has developed many tools to guide and evaluate effective debriefing (Cheng et al., 2020). ACTS is a simulation tool, and therefore, debriefing guidelines and evaluation tools were extracted and evaluated for criteria aligned to the pedagogical heuristics. Further, the *Quality Matters Higher Education Rubric, Seventh Edition* (QM) and *the SUNY Online Course Quality Review Rubric* (OSCQR) rubrics were analyzed to provide criteria on an appropriate online learning experience. Additionally, as identified in the literature review, graphics under a UX model are viewed differently from a learning perspective; therefore, criteria related to *visual knowledge* were aligned under pedagogical rather than under content, yet again demonstrating a need for viewing content from a usability and pedagogical approach (Alharoon et al., 2021; Fleming, 1979; see Table 3.6).

Table 3.6

Pedagogical Heuristics and Definitions

KPI	Heuristic	Definition
Pedagogical	Quality of Instruction	The technology demonstrates clear directions, rules, and evaluation criteria.
	Visual Knowledge	The application demonstrates instructional images that are clear, cohesive, and intuitive.
	Encouraging	The learner is encouraged through relevant, appreciation, validation, and non-judgmental feedback.
	Supportive	The learner is supported in their learning through appropriate feedback.

Note. Heuristics and definitions based on literature review.

After all criteria were aligned to the heuristics they were edited and narrowed down to no more than seven performance indicators per heuristic. Each indicator was then numbered for clarity. Appendix A shows the final list of indicators and how the KPIs of the Social-TPAC model form together to provide an understanding of how each element can impact the overall learning experience with ACTS. If all the heuristics rate high on a 7-point severity scale, it will demonstrate a high level of usability for the ACTS system, allowing for the social and pedagogical experiences to flourish, creating the joint optimization that Trist (1981) identified as a need for a successful sociotechnical experience. In turn, this gives learners the cognitive presence needed to process information and build their skills as teachers. However, suppose the evaluation demonstrates one of the KPIs is performing lower than others; in that case, it will increase the learner's extraneous cognitive load, decreasing their germane load and the overall learning experience. For example, if content scores lower, it demonstrates that a learner needs to concentrate on clarifying and understanding an element of the system rather than on building their teaching skills. Yet, when a system provides usability content that is short, familiar, and easy to understand, a user can operate the technology at a subconscious level and focus on their learning.

Participants

This study looked to understand the usability and overall user experience of a system from a student and faculty perspective. Therefore, former and current teachers were selected to test the usability of an improper fractions scenario. The aim was to obtain at least three participants to interact with ACTS and the setup process. Three participants were identified for participation in the study. Due to technical issues with the connection to the ChatGPT system, additional participants were recruited, leading the study to look at human-human and human-AI interaction. Therefore, three participants examined the human-human interaction and the setup process and two additional participants were recruited to participate in the human-AI portion of the study and were not asked to go through the setup process due to sufficient findings from the first three participants.

All five participants identified as female and described themselves as White. The participants ranged in ages from 22-55+ at the time of the study; two participants identified as 22-34; one identified as 35-55; and two identified as 55+. While all participants had previous teaching experience, only one participant identified as a current teacher. Two participants currently work in the field of education, in outreach, and as a STEM researcher, while two identified that they no longer work in the field of education. Two participants had 10+ years of teaching experience, and two had 3-10 years of experience. Thus, participants had more experience than the target users of preservice teachers but met the target group of faculty usability. Further, one participant identified as having 1-3 years of teaching experience, which more closely aligns with novice teachers who may use the ACTS system. Having five participants that range in age, experience, and digital skills, aligns with Travis and Hodgson (2019) recommendation of increasing the probability of uncovering usability problems with as few as five participants.

Of note, two participants expressed knowledge of older models of the ACTS system, and one used the older system. However, due to significant changes made between versions, their prior knowledge did not appear to impact their understanding surrounding the usability of the system.

Data Sources

Four methods of data collection were used for this study. The first part was done through an expert heuristic evaluation of the ACTS system using the Social-TPAC heuristics and indicators. Norman (2005) pointed to the need for experts to perform heuristic evaluations to identify details an end user may not identify during use. Barnum (2020) also highlighted the need to conduct heuristic evaluation through observation, and therefore, all five participants were observed using the ACTS system. To understand their perspective and to add trustworthiness to the study, users were also asked to provide feedback through a semi-structured interview and by filling out a brief questionnaire that directly related to the heuristics and definitions of the Social-TPAC heuristics (see Appendix C).

Heuristic Evaluation

Each screen was looked at holistically using the list of heuristics and indicators, identifying any problems or elements needing improvement. Any findings were marked with an x next to the indicators, and notes were added explaining the problem. Additionally, the section of the system where the finding occurred was noted, and findings were rated for their level of ease in fixing the problem, such as a simple color change versus a more complex issue with the conversational AI. However, the developers will have the final insight into the complexity of updating the system to correct any usability issues. To understand accessibility issues noted by the WCAG, such as color contrast, the *Mozilla Firefox Developers Tools Accessibility Inspector* was used to inspect the screen (Mozilla, 2024). In addition to Firefox, the system was tested in Chrome and Safari browsers.

User Testing

Each participant was emailed to arrange a scheduled time for user experience testing. Two participants were not provided with any information about the math problem, but three participants were provided with the math problem ahead of time to prepare for the session after the initial participants expressed anxiety about not having the problem ahead of time. All participants completed a brief survey detailing their demographic information and consented to the study before participating.

Pre-brief

When participants signed onto the Zoom session for the study, they were given additional details, including information about ACTS and an outline of the steps that would happen during the study. Participants were also provided with the option of turning the camera off if they wished; two out of five participants chose to turn their cameras off during the study.

Task Analysis

The participants were provided a link to the system and asked to share their screens before recording. After the recording started, they were given a username and password to login to the system. They were then asked to navigate to the simulation and begin the simulation. Users were provided with as few details as possible to simulate an experience with a new user. All users were notified that they could ask for help at any point and were notified that they could not do anything wrong, and if they ran into any difficulties, it was a problem with the system and had no reflection on their capabilities. As suggested by Travis and Hodgson (2019), if users were observed making mistakes they were not interrupted or corrected and instead observed to understand how they would recover from the error. Further, they were informed that they could end the simulation at any point and review the provided feedback. Users were also allowed to try the simulation for a second time, which no participant chose to do. Three participants used the human-human simulation, and two used the human-AI simulation. All participants received AI-generated feedback. Additionally, three participants were asked to set up a simulation based on the settings provided.

Users were observed going through the system, and any usability findings were marked according to Barnum's (2020) metrics: (1) help assists; (2) confirmation of click or data entry; (3) mouse-overs or extra clicks; (4) expressing frustration, and (5) selecting the wrong item to understand the user's mental model while using the system.

User Feedback

Participants were asked to share their thoughts after completing the simulation and reviewing the feedback. If any parts of the system were not utilized by the participants, questions were asked to determine the reason. Further, any metrics noticed during the observation were questioned for further understanding. Participants had the opportunity to review functions that were not used after completing the task.

Questionnaire

After a brief conversation, the participants were sent a link to fill out a questionnaire aligned with the heuristics. All participants were notified that they could ask for clarification on any of the questions if needed; additionally, they were again asked if they had any additional thoughts after completing the questionnaire.

Data Analysis

Expert Review

Findings identified from the heuristic walkthrough were compiled in Microsoft Excel. Each finding was labeled with the location of the finding, the ease of fixing the heuristic, a suggested update, if participants were affected, and lastly, the target audience for which the finding would need to be fixed to meet the needs of that audience.

Observations

Notes were taken during the observations and while reviewing the recordings of each participant. Notes were made when participants asked for assistance, confirmed any clicks or data entries, or expressed frustrations. Notes were then aligned to the heuristics and added as a new finding or marked as affecting participants. The notes taken from the observations further informed the semi-structured interviews to confirm any struggles or pauses a participant experienced while completing the tasks.

Semi-Structured Interviews

The data analysis was completed by first developing a codebook (Appendix D) and coding the interview data using priori codes found during the observations and any codes that emerged from the notes taken. A different code was assigned if the data did not align with the codes from the initial review. The first pass of codes was used to create the main themes from the interviews and uncovered three main concepts —a need for a pre-brief to explain the learning goals, mentioning past experiences with other software, and problems relating to the knowledge visualization. Coded data was then aligned to the heuristics to uncover additional findings and note any previous findings that affected participants. A second review of the interview data and heuristics was conducted to understand the impact on the participants further and was noted in the final heuristic findings (see Appendix B). Results were than rated on their severity level.

Researchers have noted that heuristics can be rated based on severity levels, which are either data-driven or judgment-driven (Hassenzahl, 2000). In turn, findings from the heuristics, observations, and interviews were compiled for each heuristic and were rated based on (1) frequency, how often they occurred within the ACTS system (2) impact, how many participants were impacted; and (3) target audience (Hassenzahl, 2000; Nielsen, 1994). Most severity scales call for the persistence of the problem; however, to meet the goals of this study, the target audience was chosen instead. Nielsen (1994) acknowledges that the market impact should be considered with severity scales; while this concerns the growing popularity of a technology, it also aligns with the Inclusive Design Cube, matching user needs to an iterative design process (Clarkson et al., 2003). Thus, the target audience was categorized according to the Inclusive Design Cube as having no user impact, no facilitator present, faculty only, and all users (see Table 3.7).

Table 3.7

Severity Rating Scale

Severity Scale Items	1	0.66	0.33	0
Frequency	Does not occur	1 component	2-3 components	All components
Impact	No users	Only researcher noticed	1-2 users	3-5 users
Target audience	No users	No facilitator	Faculty	All users

Note. Items were rated on a 1-point scale, with 0 having the highest severity.

Of note, while most impact severity scales recognize cosmetic problems as a lower severity level, this study considers aesthetics an essential element of usability (Barnum, 2020; Hassenzahl, 2000; Nielsen, 1994). Thus, it was removed from the rating scale, and cosmetics elements were rated at the same level as all other findings. A mean of all three severity ratings was found and multiplied by 7 to compare to the 7-point Likert scale of the questionnaire.

Questionnaire Results

All participants were asked to fill out the survey after providing verbal feedback; results from the questionnaires were analyzed to find the mean for each of the 16 heuristics, and the results were then applied to the Social-TPAC model to visualize the sociotechnical impact on the overall learner experience. Due to the low number of questionnaires results further data analysis was not conducted.

CHAPTER IV: FINDINGS

This capstone project uncovered usability problems that may affect the learning experience with ACTS utilizing user experience (UX) research methods. Specifically, the research looked to understand (1) what social, pedagogical, content, and technical key performance indicators (KPI) emerge from an expert review of ACTS; (2) what social, pedagogical, content, and technical KPI of user experience emerge from participants using ACTS; and (3) what are participants' perceptions of the user experience for the social, pedagogical, content, and technical KPIs after using ACTS.

After analysis of the expert heuristic evaluation, observation, and data collected, the results uncovered several usability problems that aligned to the social, pedagogical, content, and technical KPIs. Additionally, results aligned with the literature review, which outlined the need for systems to consider how technology evolves through a sociotechnical lens, pushing for familiar and inclusive designs (Geels, 2002, 2004). Lastly, as found in the literature review, a number of usability problems were found in relation to Gestalt design principles, emphasizing that aesthetics plays a larger role in learning and usability than acknowledged by current research (Mayer, 2020).

Summary of Heuristic Findings

Overall, 209 heuristics were noted for the entire system in relation to the following sections of ACTS: home, admin settings, edit user roles, your simulations, welcome to the simulation page, simulation with improper fractions page, and feedback page. Of these, 15 were social, 66 were technical, 34 were pedagogical, and 94 were content. 63 of the findings directly related to the setup area only faculty would view, not the average learner going through a simulation. Additionally, it should be noted that some findings, such as participants not immediately finding the home button, were marked under multiple indicators such as link color, lack of icon, and having a similar text as the name of the system, which all are possible reasons why the participant would not have recognized where to click.

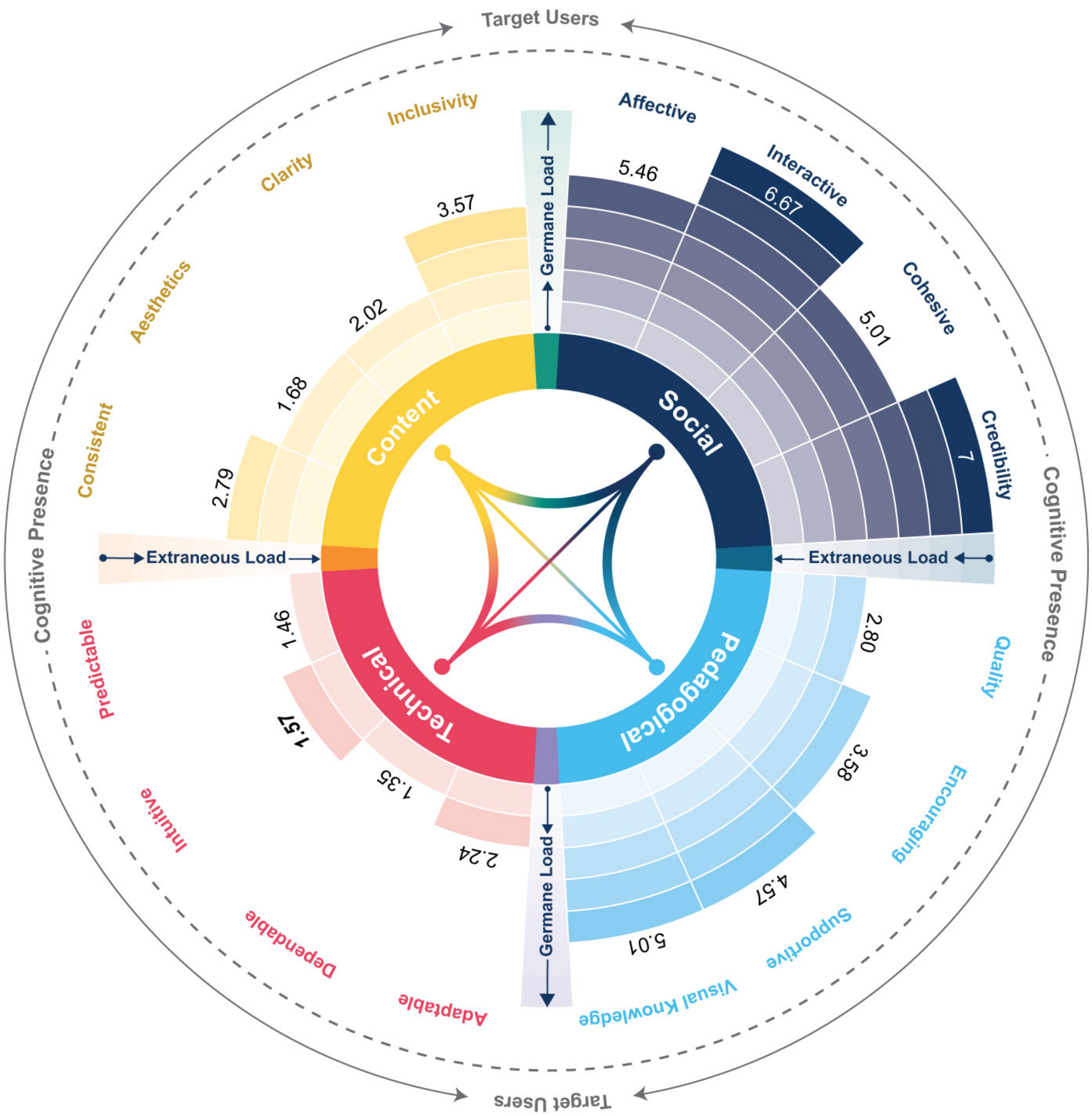
Overall, the results demonstrated that most of the findings aligned with the *Content* KPI, with 94 findings. However, when analyzed for their level of severity on a scale of 0-1, the data point out that the *Technical* KPI findings having the greatest impact on the overall system, with a final rating of 1.655 on a seven-point scale with *dependable* (1.35), *predictable* (1.46), and *intuitive* (1.57) performing the lowest. The *Content* KPI performed at a rate of (2.52), with *aesthetics* (1.68) and *clarity* (2.02) performing the lowest, (see Figure 4.1). In contrast, the highest rated indicators were found under the Social KPI, with no findings under *credibility* (7) and one finding under *interactive* with a 6.67 severity scale.

Technical and Content

Under the Technical and Content KPIs, there was a significant cross-over of findings related to the navigation. While Nielsen & Molich's (1990) heuristics have *clear navigation*, which is *indicator* T3.2 in the Social-TPAC heuristics, indicators such as T3.1: items with similar functions are the same color, shape, size, and orientation; and C4.7: buttons are short one- or two-word labels that describe the result of selecting the button (see Appendix B) helped identify the problem more deeply. Further, indicator T1.2 indicates the need for a high level of contrast, which was also uncovered in the ACTS system. WCAG 2.0 level AA calls for a ratio of at least 4.5:1 for normal text sizes, yet the links in ACTS against white are at 3.98:1 contrast, while links on a green background are at 3.20:1. While this finding is vital to fix, it is also an easier fix that can be applied across the system. Additionally, the C2.3 indicator states that layouts follow a visual hierarchy and reading order of left to right and top to bottom, with

Figure 4.1

Heuristic Findings

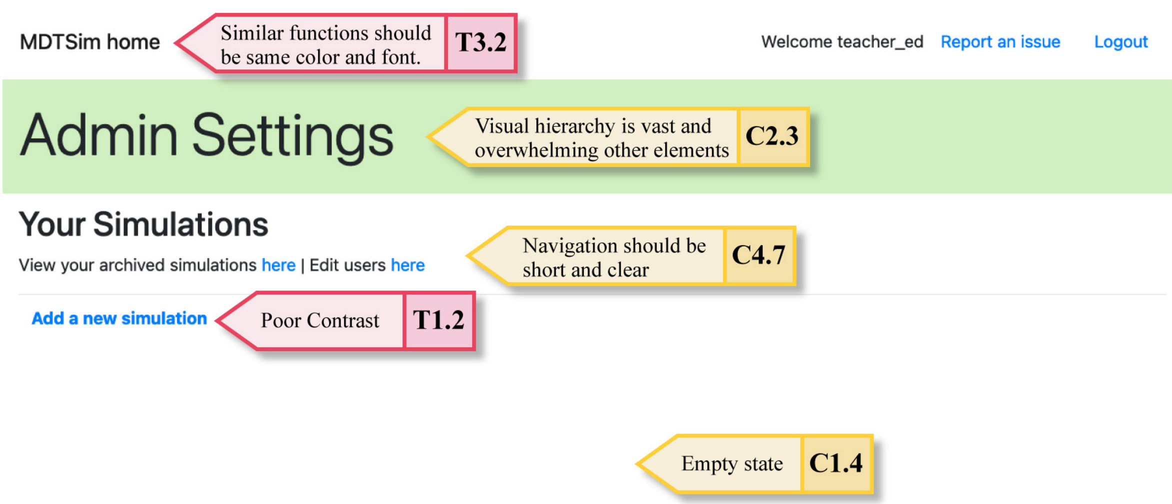


important information at the top. Moreover, it acknowledges using the Golden Ratio to help determine text sizes. Figure 4.2 shows that the Admin Settings header overpowers the other elements on the page and could cause the navigation to become hidden during a quick glance.

Lastly, C1.4 states that screens should be void of empty states. Podmajersky (2019) explains that empty states are blank screens, generally apparent in newly created accounts, and can be avoided by adding a blank item that will appear in that space, triggering a person to add content. This is similar to creating a new Facebook account and having Mark Zuckerberg as a friend. While some users found that aspect annoying, it provided guidance to the user on where their friends would appear.

Figure 4.2

Admin Page Findings



Note. The Admin Settings page in ACTS prior to adding any simulations.

Not shown in Figure 4.2 is any way to get help from the system. Under the predictable heuristic, many indicators point to having a context to help guide users through a design. *Predictable* indicators T4.2 – T4.6 all relate to in-context help; however, the ACTS system does not provide any in-context help to users, primarily due to the nature of having faculty familiar with the system setting up simulations and researchers and faculty facilitating the simulations, and therefore serving as in context assistance. While this works now, this update will need to happen to open the system up to more users.

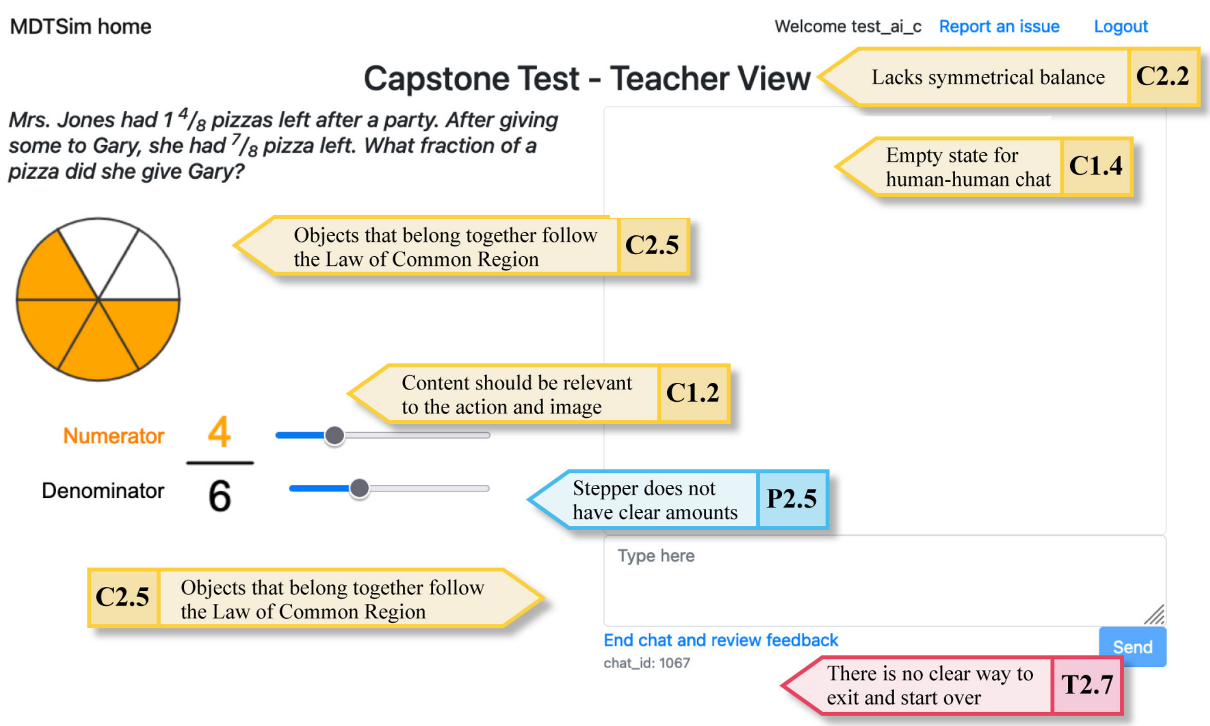
Pedagogical

Under the Pedagogical KPI, *visual knowledge* had a severity scale of 5.01, performing well. However, during the observations, all participants struggled to use the knowledge visualization in the simulation. All five participants completed the simulation without fully using the interaction, two never interacting with the element, and three only partially utilizing the tool. While it is suggested that the design of the knowledge visualization be improved, many of the indicators pointing to why the users did not use the tool fall under different aspects of the Social-TPAC KPIs; thus, from a pedagogical viewpoint, the knowledge visualization works well. However, as demonstrated in Figure 4.3, various heuristics were identified, bringing insight into why many participants overlooked the tool. One participant familiar with the ACTS system understood that a visual component was being added but still overlooked how to interact with it until prompted by the human student. Yet, having the system function without a facilitator is a goal for a later iteration; in turn, it lowers the severity level.

A lack of Gestalt principles is a potential cause of users not recognizing the tool as interactive. Indicator C2.5 points to the need for objects that belong together to follow the *Law of Common Region* (Yablonski, 2020). The Law of Common Region demonstrates the need for a bounding box that brings elements of one interaction together by using a similar background or border. Likewise, adding small labels to the stepper can help guide a user to using the knowledge visualization. Of note, Firefox's Developer Tools also identified having no labels on the stepper as an accessibility problem, and the stepper does not allow for full keyboard-based interaction. Lastly, some participants started teaching to the 4/6 because they assumed the learner

Figure 4.3

Simulation Page Findings



Note. The simulation screen and problem used for testing the ACTS system, with notes on identified problems.

was somehow looking at that equation or entered that information. To a user who is fully aware of how the knowledge visualization works, this may not matter, but to others it does not appear to be relevant to the problem as stated by indicator C1.2.

Another issue that surfaced was the confusion caused by the empty state; indicator C1.4 was identified by participants who were unsure how to interact with the chat element at the start of the simulation. However, participants who used the AI-based chat did not experience this because the Conversational AI immediately generated a message. Furthermore, the inability to see if the other participant was typing caused an additional empty state for users in the human-human interaction.

Lastly, as previously highlighted in the findings, the navigation changes from page to page; further, there is no element of continuance on the admin pages (Yablonski, 2020; see Figure 4.3). Continuance is Gestalt theory that connects steps using small graphical elements, which can help improve navigation by allowing users to understand where they are in the system and how to go back (Yablonski, 2020).

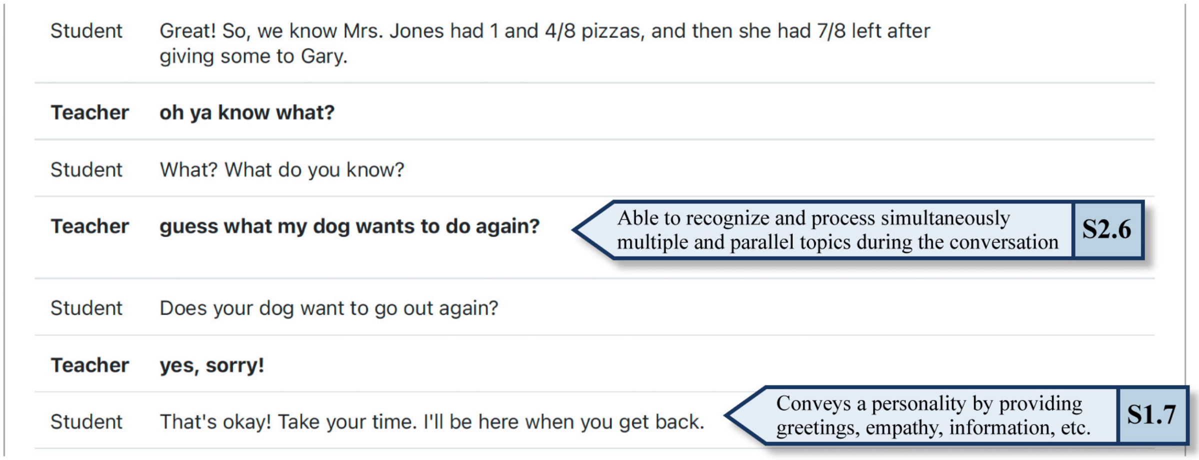
Social

Testing done for the social heuristics aimed to understand the Conversational AI ‘s ability to perform the identified criteria. In turn, testing was more direct than a typical encounter, thus, the Conversational AI could respond directly. For example, indicators S2.6: the ability to recognize and process multiple and parallel topics simultaneously during the conversation, and S3.7: which maintains a conversational theme once introduced and keeps track of the context to understand the user’s utterances, were tested through a series of messages sent to the Conversational AI (Borsci et al., 2021). Early into the conversation, a message was sent: “Can you wait while I take my dog out?” The Conversational AI responded, “Sure, I can wait.” Later in the conversation, the following exchange happened (see Figure 4.4).

The interaction demonstrated that the Conversational AI could keep track of a parallel conversation while demonstrating a small level of empathy. However, after further prompting, the Conversational AI kept pushing to work on the math problem, such as “Let's focus on how to find out how much pizza Gary got.” In turn, the Conversational AI began to teach the problem and solve it without assistance from the teacher, and not fulfilling indicator S3.5: responding in an informative way without adding too much information or explaining (Borsci et al., 2021).

Figure 4.4

Test Chat Transcript



Note. Type your note content here.

While the Conversational AI demonstrated a higher level of social presence when tested, one participant expressed, "I didn't feel…any emotional expression" while completing the questionnaire. Therefore, providing a need for user feedback through observation and feedback.

Observation Findings

Overall, four main themes emerged from the observations: (1) a need for a familiar user experience, (2) self-blame when errors occur, (3) clarity on using Knowledge Visualization, (4) the need for pedagogical presence to assist with pre-briefing and feedback. While most findings pertained to the pedagogical KPI, they also aligned with content and technology, providing information on improving interactions. For example, one participant ended the chat early with the Conversational AI out of frustration; yet the reasoning had less to do with the CA but more to do with the system's functionality and lack of facilitation of the simulation.

Familiar User Experience

Two learners who used human-human simulation expressed feeling unclear if the human student was typing, and one participant referenced Apple's "iMessaging where you can, like, see if the other person's typing," to alert the other user that they are responding. Similarly, another participant mentioned that a lack of knowing if the student was responding made her nervous, "If you could see that I'm typing, you would wait to respond." Further, the participants pointed out that it left them feeling unclear if the student was struggling to think through the problem on their own before responding. When participants are nervous or feeling confused, it increases their cognitive load and thus decreases their ability to be cognitively present (Sweller et al., 2019).

Another participant who entered the simulation first attempted to teach by verbally speaking instead of typing, after realizing she later stated, "I've done mixed reality simulations in the past, so I was expecting that sort of interface; so it was way different than what I was anticipating." In contrast, two participants who previously interacted with the ACTS system acknowledged that the system has greatly improved. Yet despite their prior knowledge, both still noted usability problems with the system, demonstrating the sociotechnical theory and the subconscious skills humans develop by using technology (Norman, 2013).

One participant added a space to the username and was presented with an error that was not immediately recognized; the participant later stated that "sometimes [websites] will actually scroll back to the place where you need to make the change," reinforcing the need for a familiar user experience even regarding errors.

One user acknowledged teaching over Zoom during the COVID-19 Pandemic and was able to interact with the ACTS system at a similar level to other participants but said her age likely made it more difficult to use; however, Lawry et al. (2019) noted that an unfamiliar UI has more effect on intuitive use than age-related cognitive decline. This also aligned with research showing that users are more likely to blame themselves for not being able to use a system than the UI (Norman, 2103).

Self-Blame

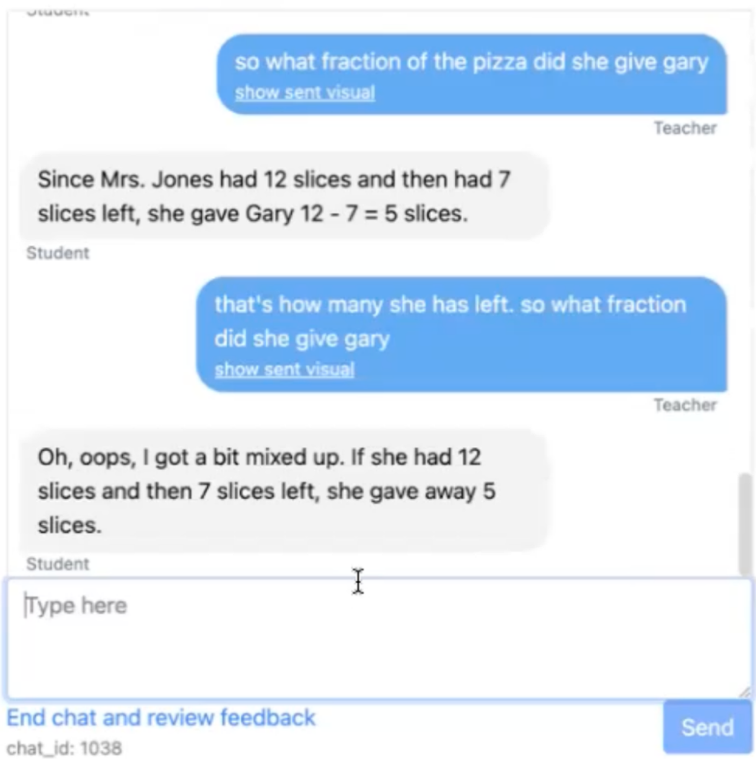
One participant commented after the simulation, “It'd be nice to have a, like, a let me try this again (feature); I have a new idea,” also stating, “like how do I get back because we were at ‘number of slices’ and not at ‘fraction of the pizzas’.” However, after two prompts asking about fractions*,* the student Conversational AI continued to talk about slices, (see Figure 4.5). This interaction in the simulation caused the participant to say, “Alright, I might have to give up here,” and ended the simulation. While the word *fraction* was not mentioned earlier in the chat, the Conversational AI may have been responding as trained to focus on slices. Yet, the participant felt she “got backed into a corner at some point…I can't get out of this because I was giving [the Conversational AI student] the wrong advice up to now” (see Figure 4.5).

Seeing that the problem given to the participant to work on with the student was part of their learning on fractions and the question asked, “what fraction of pizza was left,” the learner may have assumed that the student understood the word ‘fraction’ and felt less empathy for understanding if the student was lost considering it was a Conversational AI and not a human. Another participant stated, “I forgot to treat the chatbot like it was a person, like, I would never just walk up to a student and be like, what answer did you get?”

As a learning app, the ability to start over and avoid getting stuck is valuable to the experience, yet it is unclear if the Conversational AI was not responding clearly and the participant was blaming herself when she was not at fault. In contrast, both participants who worked with the Conversational AI acknowledge that, at times, “it just dropped off;” therefore, when the error was a problem with the Conversational AI, they did not experience self-blame.

Figure 4.5

Chat with AI-Based Student



Note. The student is an Conversational AI chatting with a study participant.

The ACTS system can have a human-in-the-loop to assist in facilitating a simulation and, depending on the goal of the exercise, could save the simulation from ending in a similar situation through cueing (Hellaby, 2013). Cueing by a facilitator in a simulated experience prompts a user who may be unaware of an environmental difference that creates an unfamiliar situation and, thus, a lapse of action (Hellaby, 2013; Norman, 2013). While acting as a student in the human-human interactions, attempts were made to cue the participants to use knowledge visualization, such as saying, "I updated the image," which did not always work. As outlined in the heuristic findings, the knowledge visualization was often overlooked; however, the testing was done without giving any guidance on how to interact with the system to understand user interaction.

Pedagogical Presence

Currently, a faculty member or facilitator is present to prepare a learner to use the ACTS system. The role of a facilitator in simulation not only provides guidance on how to interact with the system, and cue them towards an interaction, but they also guides a learner to *suspend their disbelief* so they can interact with the system, knowing it may have errors but it is a learning experience and thus may not match a real-world setting (Wittmann-Price & Wilson, 2014). In line with the need to suspend disbelief, one participant stated:

> You're imagining a conversation with somebody that you're with, as opposed to like imagining like your instant messaging a student, so I wonder if that would help just to be like, imagine that you're standing with a student solving a problem or something like that, because it felt remote to me because we were like chatting.

Additionally, participants experienced a great deal of confusion on the feedback screen, with many unaware of the simulation's goals and how they aligned with the feedback.

> I mean, I just don't necessarily know what the goal is. Like, I don't know if it's trying to tell me that I should have asked more of a certain type of question. Or, I mean, I could see that being the case, but I just don't know that I know how to look at this and improve my questioning skills. So I assume that probing and expository are better than procedural is that what the goal is, or is there not a goal?

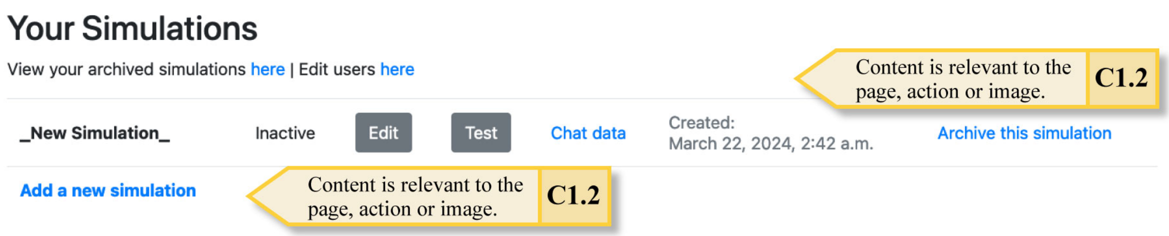
Had a faculty member set up the simulation for the learner, they likely would have guided them on the meaning behind the feedback. Likewise, they would have explained how the learner could improve their questioning skills and allowed them to attempt the simulation again. Thus, many of the heuristics under pedagogy closely relate to how to provide feedback to the learner properly; if done verbally with an instructor present, the ACTS system has more of a sociotechnical approach of bringing together society and technology. However, as a stand-alone system, there is a need for additional pedagogical presence within ACTS.

Admin Settings

Three of the five participants were asked to test the admin section of the ACTS system during the observations, which helped to identify additional heuristic findings. To test users on the admin screen, they were provided criteria to set up a simulation; by doing so, in some ways made it easier for users to navigate the system. However, all the participants asked for confirmation on what they were about to click on; further, despite having the exact criteria sent to them in a Zoom chat message, many still asked what they should click on. Additionally, participants struggled with the navigation, and all three jumped around between screens trying to find where to add learners to a simulation. Many heuristics were identified for the admin section of the ACTS system, as demonstrated in Figure 4.6. The add a simulation page was particularly challenging, starting with when you click “Add a simulation,” it does not immediately take you to a new screen; instead, it adds a row in the admin screen that a user must select to edit to create a new simulation (see Figure 4.6). Yet, items are added in alphabetical order and the default naming convention does not place the item at the top of the list. On top of that, the date and time created does not align with the current time zone, making it even more difficult to see which simulation should be edited.

Figure 4.6

Add a Simulation

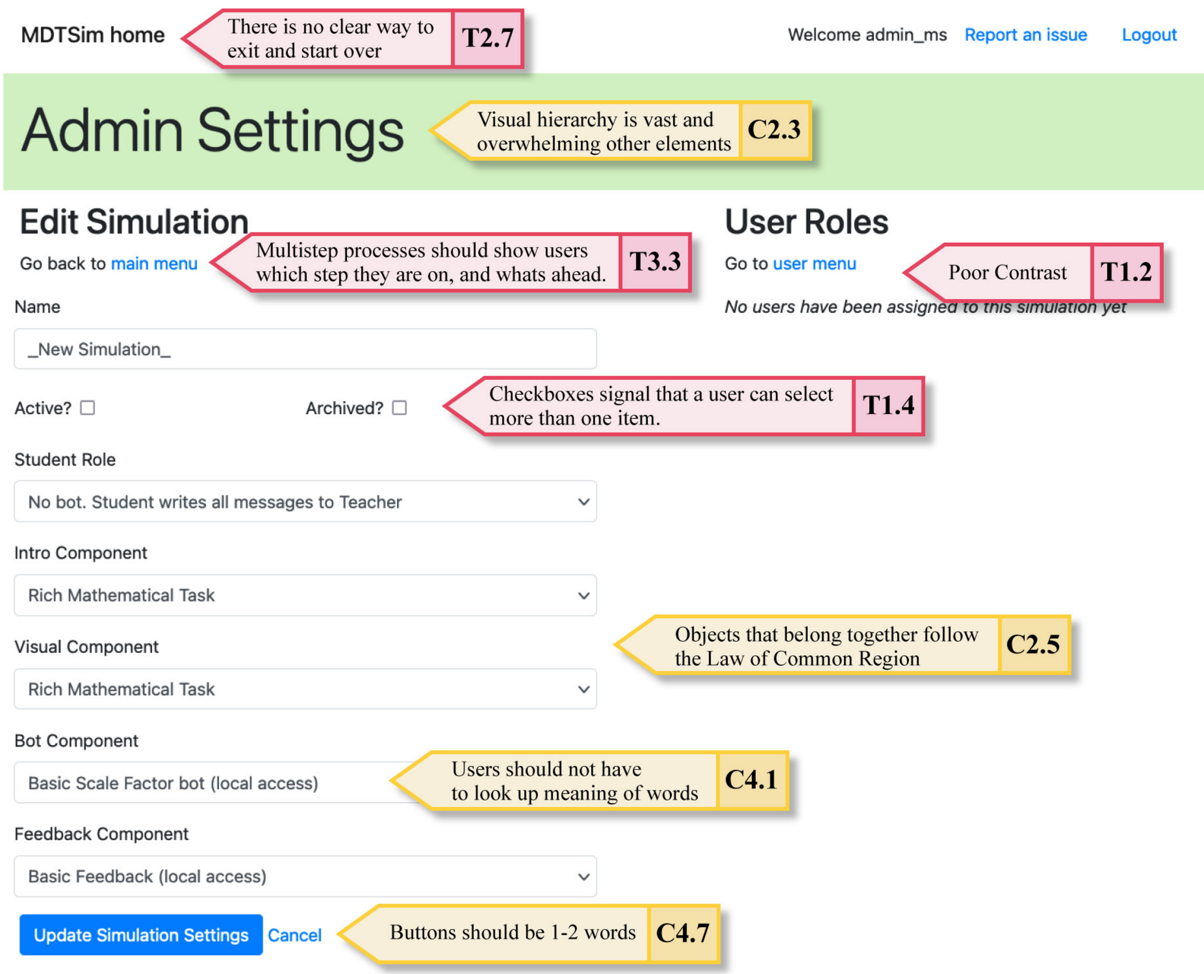


Note. The add a simulation text adds a row to the screen that the user must then select edit to update.

When users click on 'Edit,' they are brought to the page shown in Figure 4.7. Likely, to the research team who created this page, the settings are intuitive, but as mentioned, many users struggled to use this page. Several findings were identified such as indicator T1.4, which shows using a checkbox rather than a radio button. A radio button signals to the user that only one item can be selected; it is assumed that a simulation can be active or archived but *not* both, yet using a checkbox allows users to select both if they wish. In addition, the *intro component* and the *visual component* must be the same, yet there is no visual representation, in-context guidance, or error prevention to alert users to select the same item for both. Lastly, there is no guidance to add learners to the simulation. Instead, users must leave the screen and navigate to the user settings to add users to the simulation. While the system may need to have these steps on separate screens, a few updates to the UI may help make the process clearer to the user.

Figure 4.7

Add a Simulation Page



Biased AI

While not a theme among participants it is important to note that one participant stated that the Conversational AI was biased in its response during the simulation, “it seemed to call me by the name I gave it, but it originally assumed I was a Mrs. when I could have been a Mr.,…but yeah, it just assumed.” Further adding a possible solution that it “should start with a question, like when it says, ‘I need help’, maybe like, ‘who are you?’ or ‘how should I call you?’” If possible, having the Conversational AI ask about the participant's name, honorifics, or pronouns may create a more inclusive simulated experience.

Questionnaire Results

The questionnaire results identified participant frustration with knowledge visualization under pedagogical presence, as this was found to be the lowest-performing heuristic, with a mean of 3 from the overall results (see Table 4.1). Figure 4.8 shows the questionnaire results in Social-TPAC framework.

Table 4.1

Questionnaire Results

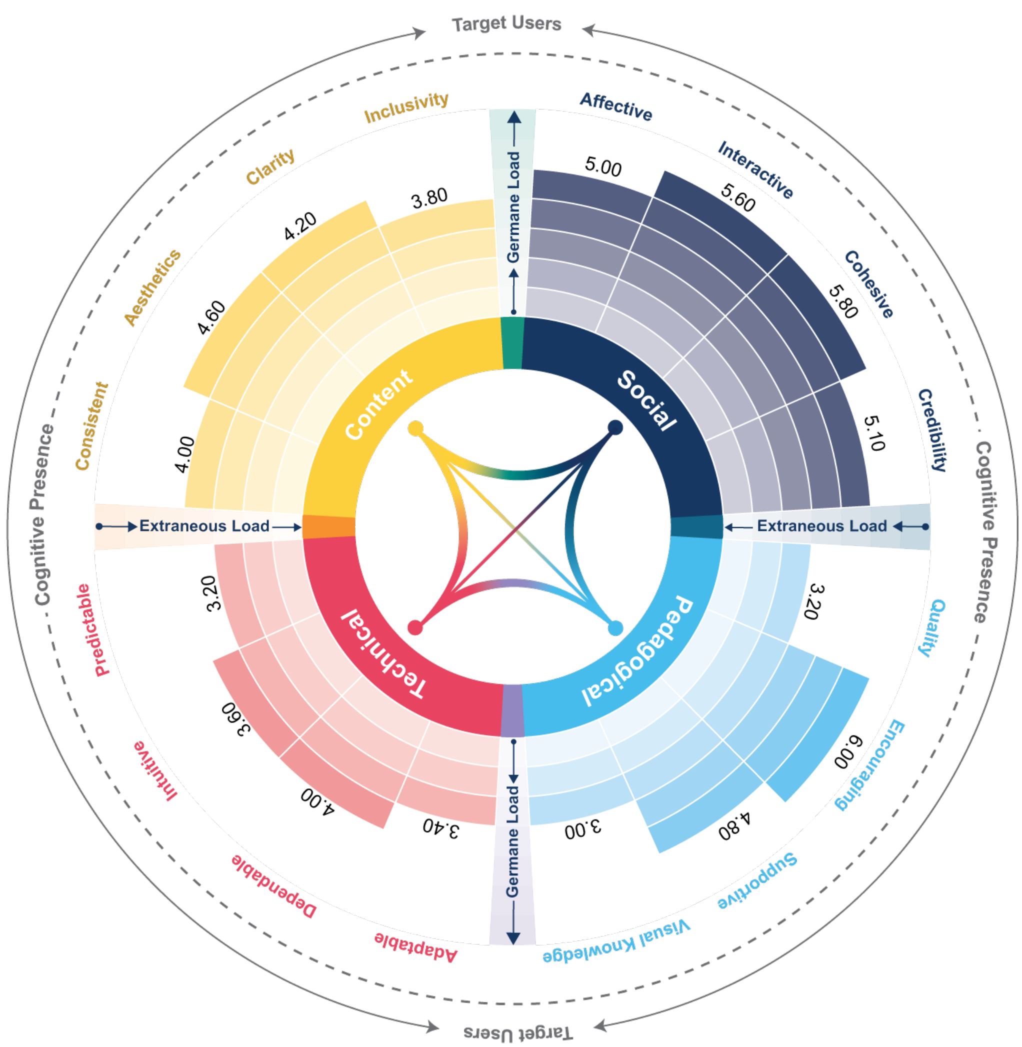
Social-TPAC	Mean	Human/Admin	AI-CA/No Admin
Affective	5	5	5
Interactive	5.6	6	5
Cohesive	5.8	5.66	6
Credibility	5.1	5.66	4.25
Adaptable	3.4	3.33	3.5
Dependable	4	4.33	3.5
Intuitive	3.6	3.66	3.5
Predictable	3.2	3.33	3
Quality	3.2	3	3.5
Visual Knowledge	3	3.66	2
Encouragement	6	6	6
Supportive	4.8	5.66	3.5
Consistent	4	4.66	3
Aesthetics	4.6	4.66	4.5
Inclusive	3.8	4.33	3
Clarity	4.2	4.33	4

However, the overall results reflect a slightly higher rating, except the *Social* KPI which performed more closely to the heuristics. Testing done for the social heuristics aimed to understand the Conversational AI's ability to perform the identified criteria. In turn, testing was more direct than a normal encounter; thus, the CA was able to respond directly to the prompts. Yet, the perceived social experience and knowledge of interacting with a Conversational AI can feel less familiar to end users. \

All participants were asked to fill out the survey after providing verbal feedback; they were also asked to fill out the survey while still in the Zoom meeting. This method highlighted user biases when answering questionnaires (Ciriello & Loss, 2023; Lavie & Tractinsky, 2004; Lu et al., 2022). Juslin et al. (2021) explained that participants who are familiar with the faculty or researcher may rate a learning experience more positively than if there was no relation. The

Figure 4.8

Overall Mean of Survey Responses



Note. Overall questionnaire results are graphically represented in the Social-TPAC framework.

participants in this study all had a relationship with the University or the researcher, creating a potential for bias. The bias was emphasized as participants spoke poorly of the system and then rated it positively just moments later. One participant expressed, “I didn't feel the emotional, any emotional expression,” while completing the questionnaire, yet gave it a 4 out of 7 rating. In addition, one participant provided feedback on her struggles using the admin screen yet rated the technical and content aspects of the Social-TPAC highly.

Discussion

Overall, the study aligned with the need to look at usability studies through a sociotechnical lens to understand how using computer-based technology has become a subconscious function of human behavior. Thus, when the technical or content aspects of a user experience are inadequate, it can greatly impact the pedagogical and social aspects of learning. Further, the need for teaching presence with learning technologies can occur through a mix of social and technical approaches. The Social-TPAC framework highlighted the role a facilitator plays in establishing goals, introducing how the technology functions, cueing participants, and debriefing them to understand their mistakes and how to improve their teaching methods. The ACTS system’s ability to have a human-in-the-loop approach not only assists with the social aspects of Conversational AI but also allows for a pedagogical approach to cue a learner to suspend their disbelief or assist with the technical and content aspects to create a more meaningful experience. In contrast, a Conversation AI may provide an experience that prevents a facilitator from over cueing a participant to meet a learning objective of the exercise, thus removing bias from the experience to provide an equal experience for all students. While this study unexpectedly had a human-AI and human-human approach, it was not the research goal to understand the differences. Yet, having adults act as students can bring varied learning experiences between actors. Meanwhile, conversational AI may allow learners to play on an even playing field. That said, as one participant acknowledged, "I forgot to treat the chatbot like it was a person," which also adds variability in how learners interact with the technology. That said, as technology evolves, how humans interact with AI will likely change, further highlighting the need to approach AI-based learning experience evaluations with a sociotechnical lens.

CHAPTER V: RECOMMENDATIONS

The heuristic evaluation, participant observations, and feedback provided numerous usability findings. The findings were analyzed, sorted, and rated based on severity, presumed ease of updating, and the target audience. Thus, the recommendations are presented in three phases: (Phase One) updates that are easily fixed by adjusting words, colors, and text sizes that will improve the UI for all users; (Phase Two) updates that require more robust design and development that will allow the system to become more user friendly for all learners and open the system up to more faculty to use to create their own simulations and add users; lastly (Phase Three) looks to further expand the audience to learners who wish to practice teaching on their own without a facilitator or faculty member present.

Phase One Suggested Updates

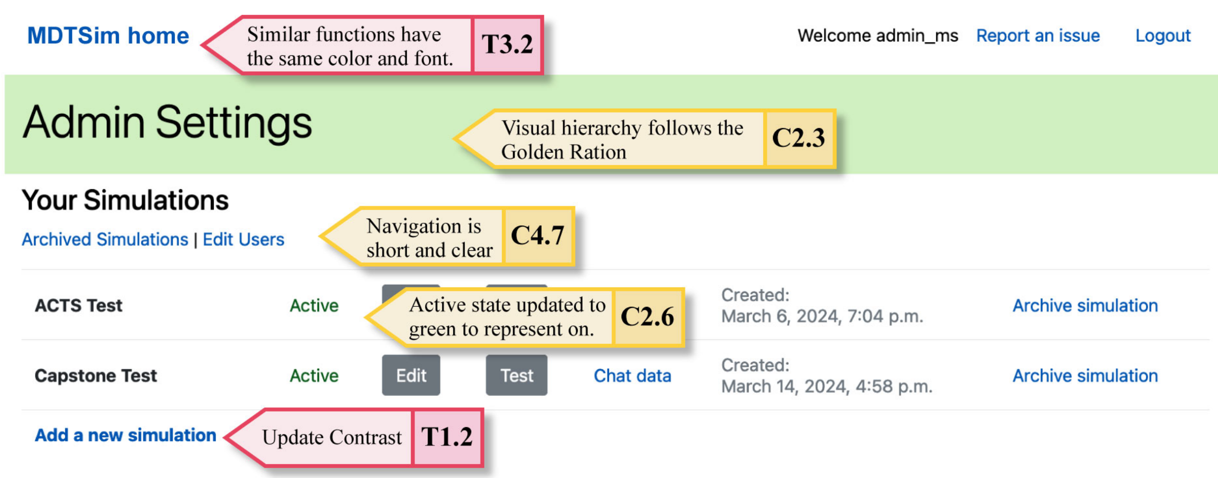
The findings uncovered several changes that affect not only usability but also accessibility. Many of the links within the ACTS system did not follow WCAG standards for color contrast. Colors are easily corrected by updating the Cascading Style Sheet (CSS) that the system uses to inform the stylistic parameters of ACTS. As shown in Figure 5.1, the colors are updated to have an increased contrast level. Therefore, the home function was updated to the same color as the links to cue users to click on home to go back to the main page, exit the simulation, and review the feedback.

On top of that, the navigation was simplified for the admin settings by removing extra words and providing a standard form of navigation that is familiar to users. The visual hierarchy across the screen was updated by utilizing the golden ratio, which is 1.618, and multiplying to the baseline text to find a more balanced increase in text sizes to demonstrate hierarchy without overwhelming other areas of the page and, in turn, decreasing the cognitive load of the user.

Lastly, the word *Active* was previously presented in red text. While indicator T1.7 states that interactions should not rely entirely on color to convey information, utilizing a color that aligns with the word's meaning will assist most users. Red is often associated with *stop* or *inactive*, while green often refers to *go* or *active*. Therefore, *Active* was updated to the color green to represent an active state. All updates were done using the developer tools in Mozilla Firefox, which allows for live adjustments to the CSS, without access to the actual code to assist in the development process of aligning code to baseline web browsers. A process that demonstrates the ease of updating the styles within the system without any extensive development work.

Figure 5.1

Phase One Changes



Note. Updates are noted according to the indicator that aligned to the usability finding.

Phase Two Suggested Updates

Updates for phase two focus on expanding the system to be usable by more faculty while also increasing usability for more users across the system. Thus, changes are concentrated on the admin page, and the knowledge visualization is improved to be more intuitive to users. Multiple study participants were noticed jumping around between pages in the admin section, attempting to add a new simulation and users. The heuristic evaluation identified that the navigation did not follow standard design patterns. Further, participants admitted needing to remember what page they were on as they went back and forth between the simulation and user admin pages. Testing was done with new accounts, meaning no users or simulations were in the system. In turn, it is unclear if more information would add confusion but it could increase a user's cognitive load even after they become familiar with the system. Indicator T3.6 states that users should not have to remember information from one part of the interface to another. Thus, there was a need to combine creating a simulation and adding users by leading a user to the next screen to add learners by adding a button to the next step. Further, the current system has the user add a simulation and then edit it. The suggested design immediately takes users to the edit a simulation page. Figure 5.2 shows a suggested design for adding a simulation according to the suggested design solution. However, some design suggestions can be utilized on the current page without taking users immediately to edit a simulation form or connecting the steps to add a simulation and user to the system.

The updated design also shows several design elements that would be applied to more of the system. Of importance, there was an attempt to meet the current structure of the ACTS system with the design, allowing for easier integration with the system but providing an updated design. The current system has a navigation area at the top of the page. However, not all sections of the system are visible. Hence indicator C4.7 demonstrates an updated navigation and includes a sub-navigation menu within the admin settings.

As noted by participants, they needed clarification on what page they were on last; indicator T3.3 states that a multistep process shows users which step they are on, the steps they have completed, and what steps are ahead (Nielsen & Molich, 1990). Yablonski (2020) suggests having continuance demonstrated through graphics to connect steps in a process. Indicator T1.4 shows an update to the form to alert users as to the type of action they can take by selecting either *Active* or *Archived*, removing the ability to select both.

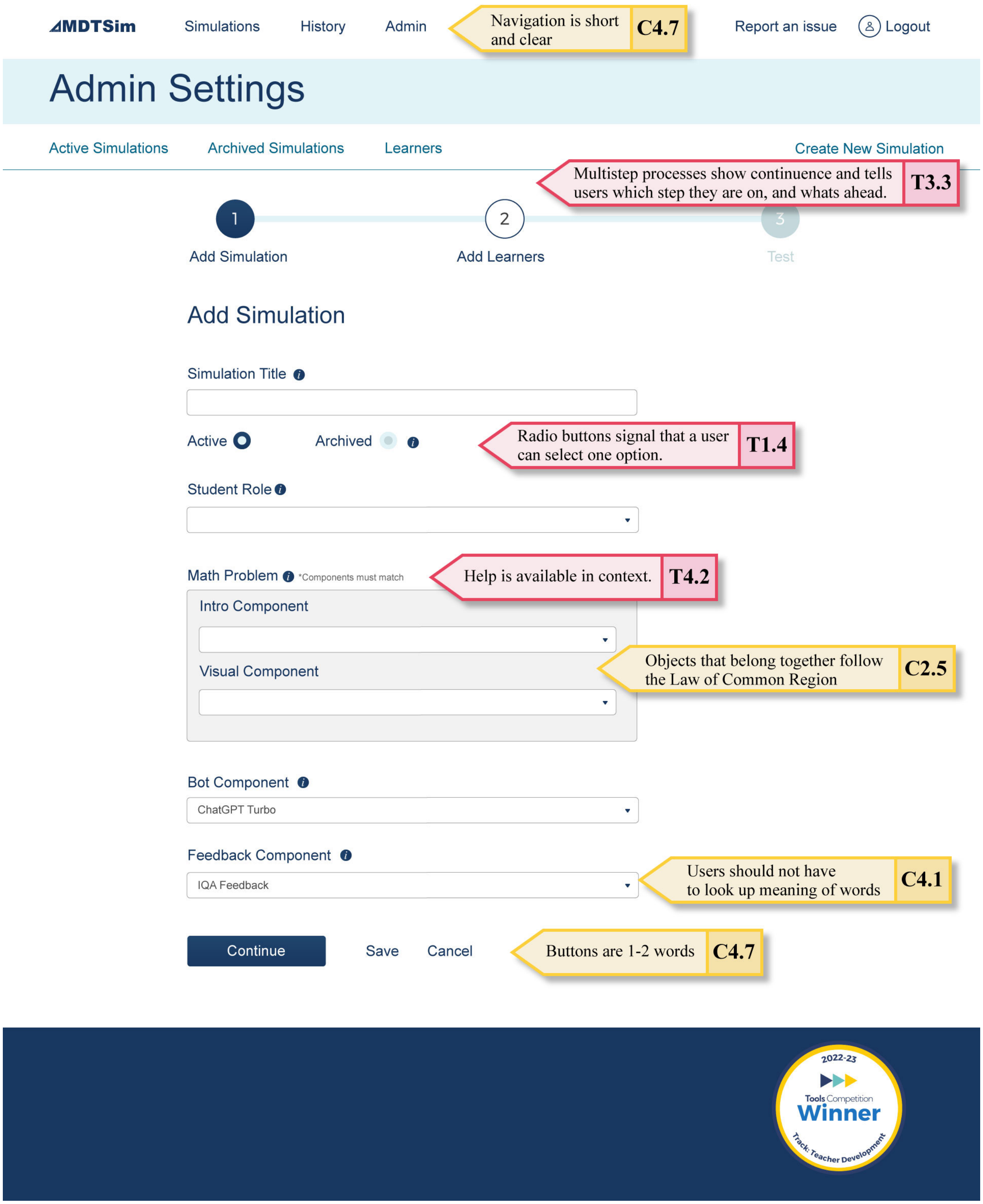
Indicators T4.2 and C2.5 focused on alerting users to align their choices for the intro and visual components. Two approaches were used to correct the finding: first, in-context help was added to the design, and second, following the law of common region, a light-shaded box was added to further bring the items together. It is also worth noting that this could be accomplished by a vertical line to bring the items together.

Findings from the observations and indicator C4.1 alerted the need to simplify the choices in the dropdown menus for clarity. However, as the ACTS tool is used primarily for research, menu items may need to be kept more complex to identify the complexity of the systems. Therefore, as with any other change, the suggested change should only be applied where applicable.

Lastly, Figure 5.2 utilizes shortened content for navigation buttons and connects adding a simulation to adding users (see Figure 5.3). The need for ease of navigation was identified because users had to click back and forth to add users and verify that they were added to the correct simulation. Notable, this design does not fully account for editing a simulation. Should the system not allow for a separate page to edit a simulation, then the design may need to account for that functionality. Figure 5.3 also demonstrates indicator C3.2, which uses an approachable tone and voice, and indicator T3.7 from the OSCQR rubric suggests using familiar icons to alert users of specific actions.

Figure 5.2

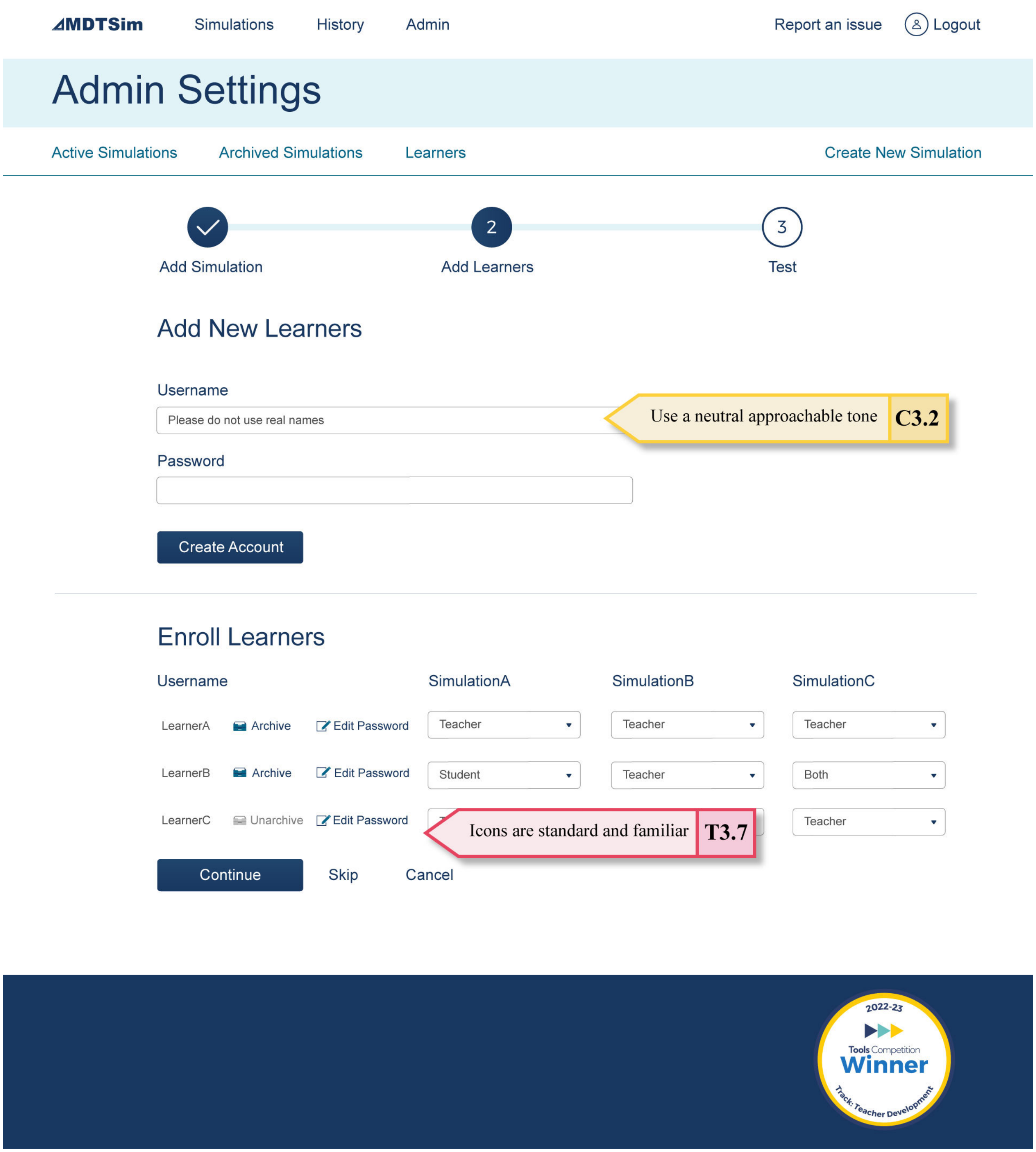
Admin Add Simulation Page



Note. Suggested design to improve the usability of the admin pages.

Figure 5.3

Simulation Add Users Page



Note. Updated add and enroll learners page.

The observations uncovered a previously unrealized usability problem with the knowledge visualization tool and text-based CA, with most users not utilizing the knowledge visualization tool. First, due to the question being already presented to the learners on the previous screen, it was separated from the simulation. The question now appears above the simulated experience for learners to reference. To connect the elements visually, a box was added around the chat box and knowledge visualization (see Figure 5.4). Next, the fraction was set to a number in the problem. As noted in the findings, some users tried to teach to the mismatched fraction presented by default, stating that they assumed the learner was attempting to align that fraction to the problem. Labels were added to the stepper to note the alignment to the knowledge visualization in hopes of further connecting the ability to interact with the graphic (see Figure 5.4).

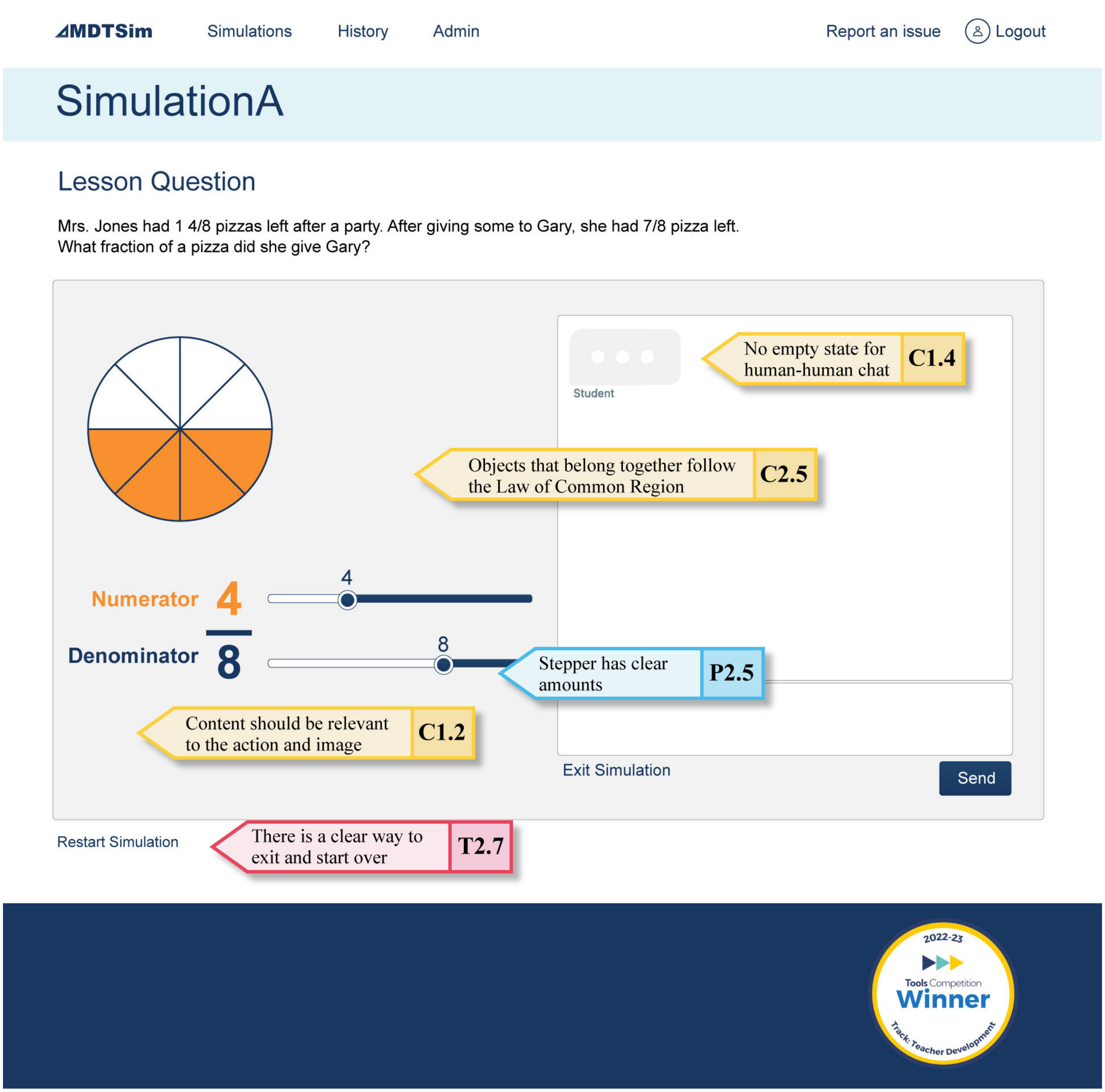
Lastly, the design shows a typing indicator to remove the problematic empty state for the human-human interaction (see Figure 5.4). Further, a link to restart the simulation was added under the simulation so that if a learner feels truly frustrated, they can start over. While exiting a simulation may not lead to improved learning, learners should never feel trapped in a simulation.

Phase Three Updates

The need to exit a simulation would likely be needed when no facilitator can manage the simulation through the human-in-the-loop feature. Many of the findings related to the pedagogical aspects of the Social-TPAC model related to the need for teacher presence online. However, following a sociotechnical approach, while the usability of the system can be updated to operate without the need for human presence, allowing users to practice their questioning skills with only the Conversational AI, it is unclear if they can suspend disbelief enough to interact with a CA without a human providing a thorough pre-brief and debrief of the learning experience. Thus,

Figure 5.4

Simulation Page Updates



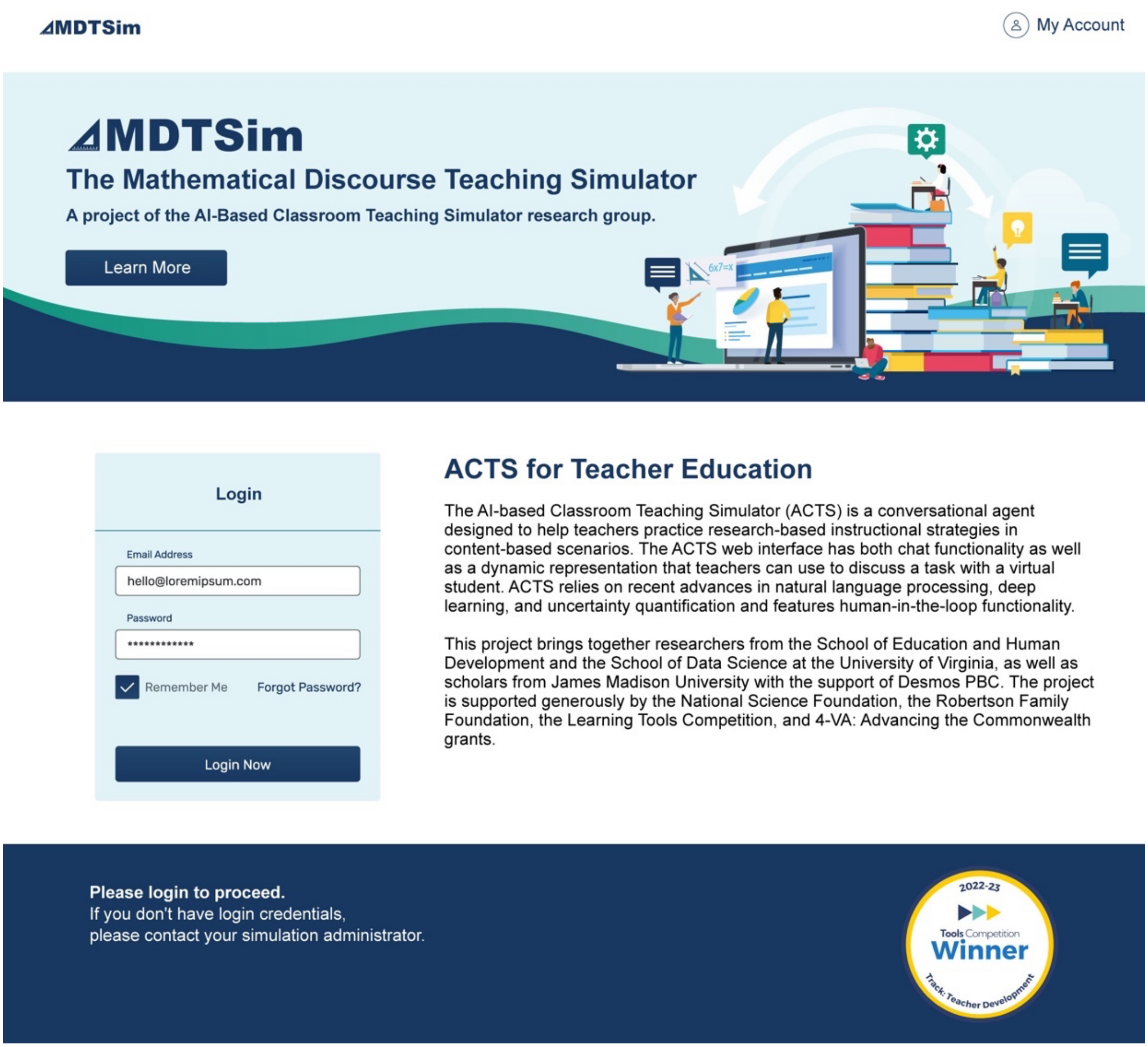
Note. Updates to the Simulation page.

additional research and technical advancements utilizing AI as a learning tool are still needed to increase the number of users. Yet, if the system were found to support these features safely, special attention would need to be given to how a simulation is presented digitally to a learner, alerting them to what they can expect, the goals of the simulation, and how they will be evaluated, which must be done using a supportive, inclusive, and encouraging voice. Further, the feedback must not only explain the types of questions a learner asked during the simulation but also guide them on how to improve their learning by validating their experience and highlighting both positive and negative feedback (Borsci et al., 2021; Cheng, 2020; Lowenthal & Dunlap, 2020).

Three heuristics from the Social-TPAC framework highlight the need for a quality, supportive, and encouraging environment. Since the ACTS system ultimately aims to create stronger teacher-student interactions, stepping fully away from an experience that is somewhat void of a sociotechnical interaction may not be optimal. Yet, once the interaction improves with the Conversational AI, the system can allow more learners to freely engage in a simulated learning experience independently. Thus, creating an inviting and inclusive system will be necessary; Figure 5.5 provides a design that gives insight into the system before logging in. Figures 5.2 - 5.4 demonstrate how aesthetic design elements can enhance the usability of a system. In contrast, Figure 5.5 illustrates how design and graphics can represent an inclusive environment and be more inviting to more users. As innovative sociotechnical systems evolve for learning, it is vital to consider all aspects of the user experience to understand how to improve the technology to be more inclusive. Creating a technology that is usable for the greatest number of people is easier than making a technology everyone can use. Thus, it is essential to align usability studies of learning technologies with the target users and understand the sociotechnical relationship of how humans and technologies evolve together to increase usability while creating familiar experiences.

Figure 5.5

ACTS Login Page Updates

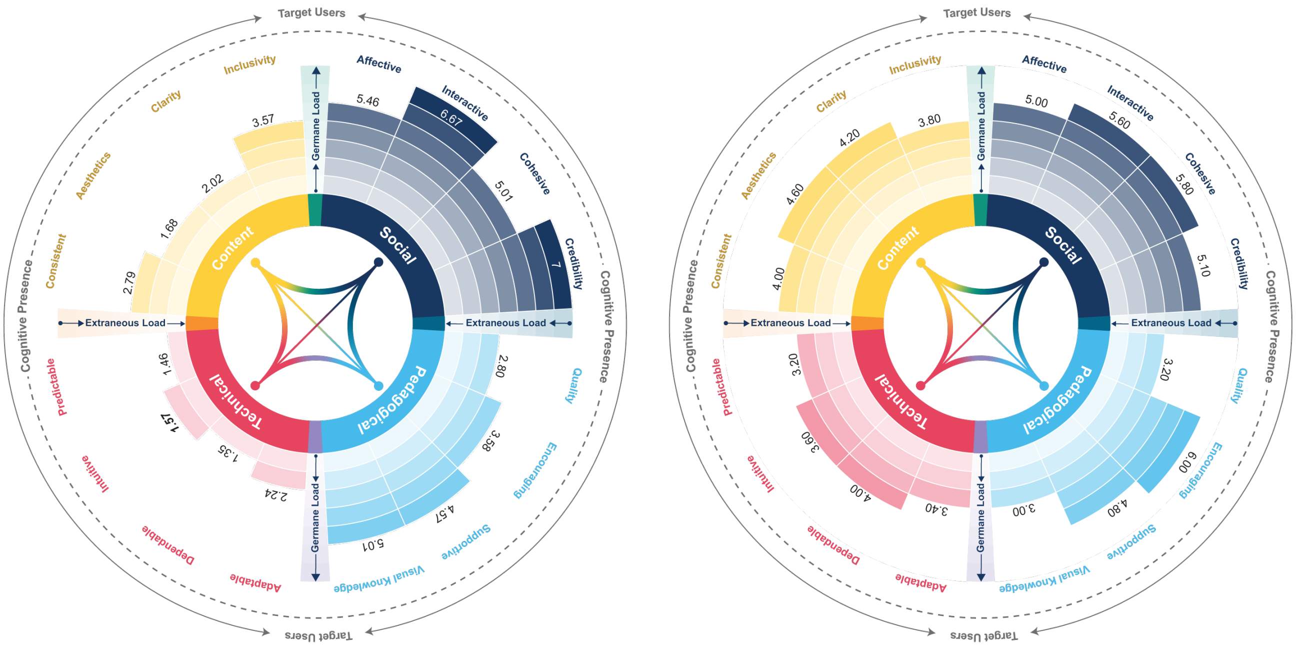


Note. Suggested login page for the ACTS system.

Lastly, Figure 5.6 demonstrates the differences between a heuristic evaluation done by an expert and a user's perception. As noted in the literature, users often have a positive bias when responding to a questionnaire about something they have a connection to, such as a project, researcher, or institution (Ciriello & Loss, 2023; Lavie & Tractinsky, 2004; Lu et al., 2022). Nevertheless, the results give a similar look into the user experience of ACTS, with the technical KPI performing the lowest and the social KPI performing the highest on both models. The Social-TPAC heuristics were not only detailed, but they also demonstrated a need to look at how pedagogical approaches can impact the user experience and how the user experience can impact the social aspects of a Conversational AI. Just as Frank Lloyd Wright believed that no home should ever be built *on* a hill, "it should be of the hill, belonging to it" (Betsky & Shapiro, 2021, p. 56), learning experiences should not simply be built in a system, but as a sociotechnical experience—where the social and technical aspects are of equal parts, belonging together. Thus, the Social-TPAC framework provides a sociotechnical base that looks at how all aspects of a learning experience interact and can be used as the base for creating an evaluation tool to understand the social and technical experiences of an AI-based simulation.

Figure 5.6

Survey vs Heuristic Evaluations



Hueristic Results

Questionnaire Results

Note. Heuristic results vs Questionnaire results.

Role of the Researcher

This capstone brings together my educational knowledge from my undergraduate studies in Art and Computer Science, my Master of Science in Communications, Culture, and Media, and primarily the research I completed as a Doctor of Education student, where I focused on understanding how to improve the quality of oral communication using technology-based education. As a learning experience designer for a tertiary care medical center, I primarily create education on patient safety and communication methods to decrease medical errors. I am interested in understanding educational techniques to improve communication among interprofessional healthcare employees, including practicing methods to speak up and address potential errors. I have twenty years of experience designing and developing digital technologies, such as websites, instructional videos, interactive simulations, and educational modules. Yet, I strongly believe that by following the design and color theories outlined in the Social-TPAC framework, novice learning experience designers can evaluate the nuances that affect usability and create updated designs.

Trustworthiness

In order to maximize the trustworthiness of this study, I employed triangulation through an iterative review process of the data collected through the expert heuristic review, user observations, and user feedback through verbal feedback and questionnaires. The iterative process provided an opportunity to compare the broad range of key indicators from the expert review to a narrow set of key indicators provided by the learner feedback and the observation logs. My knowledge and experience in graphic and instructional design brings credibility to the study by allowing for an objective review. Additionally, user perceptions allowed for contrasting points of view to confirm or refute the expert findings, such as found with the knowledge visualization tool. Lastly, combining a top-down and bottom-up approach to the heuristic evaluation and observations provides two perspectives, making the study more dependable (Barnum, 2020).

Ethical Considerations

All participants in this study were required to provide consent to participate in the research study. The research was conducted over a secure *Zoom* session, and the videos will be exported and stored to a secure Box account where only other members of ACTS research team will have access to the data. Any personal information that could identify a participant will be removed from the data and the findings.

Limitations

There were multiple limitations to this study. While the study aimed to provide an unbiased review of the system by using a detailed set of heuristics, without proper validation by other expert reviewers, it led to a single perception, decreasing the validity of the tool. Next, while this study evokes multiple user perspectives collected through observation and feedback, the relatively small sample size of users decreases the study's validity. Further, many UX researchers acknowledge the need for a diverse group of users to understand the cultural aspects of usability (Podmajersky, 2019). This paper also has delimitations; due to the nature of the Capstone project, only one researcher reviews the ACTS system and data gathered from the study. While I had no interaction in the design and development of ACTS, I offer an outside perspective. Yet, at the core of usability and UX research is stepping away from an internal opinion to understand other perspectives, which includes validating the expert-level opinions by having other expert reviewers participate in the evaluation process.

REFERENCES

Abramenka-Lachheb, V. (2023). Theory-driven and practice oriented perspectives on instructional design and learning experience design. *The Journal of Applied Instructional Design*. https://doi.org/10.59668/515.12998

Akbarighatar, P., Pappas, I., & Vassilakopoulou, P. (2023). A sociotechnical perspective for responsible ai maturity models: Findings from a mixed-method literature review. *International Journal of Information Management Data Insights*, *3*(2), 100193. https://doi.org/10.1016/j.jjimei.2023.100193

Alharoon, D., Gillan, D. J., & Lei, C. (2021). Scientific and design understandings of aesthetics. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *65*(1), 380–384. https://doi.org/10.1177/1071181321651242

Apple. (2023). *Human interface guidelines*. Apple Developer Documentation. Retrieved 2023, from https://developer.apple.com/design/human-interface-guidelines/

Arcavi, A. (2003). The role of visual representations in the learning of mathematics. *ducational studies in mathematics*, *52*(3), 215–241.

Ashfaq, M., Yun, J., Yu, S., & Loureiro, S. (2020). I, chatbot: Modeling the determinants of users' satisfaction and continuance intention of ai-powered service agents. *Telematics and Informatics*, *54*, 101473. https://doi.org/10.1016/j.tele.2020.101473

Ayoko, O. B., & Ashkanasy, N. M. (Eds.). (2019). *Organizational behaviour and the physical environment*. Routledge. https://doi.org/10.4324/9781315167237

Banes, D., & Behnke, K. (2019). The potential evolution of universal design for learning (udl) through the lens of technology innovation. In *Universal access through inclusive* *instructional design* (pp. 323–331). Routledge. https://doi.org/10.4324/9780429435515-43

Bangor, A., Kortum, P. T., & Miller, J. T. (2008). An empirical evaluation of the system usability scale. *International Journal of Human-Computer Interaction*, *24*(6), 574–594. https://doi.org/10.1080/10447310802205776

Barnum, C. M. (2020). *Usability testing essentials: Ready, set ...test!: Ready, set...test!* (2nd ed.). Morgan Kaufmann.

Bauer, E., Heitzmann, N., & Fischer, F. (2022). Simulation-based learning in higher education and professional training: Approximations of practice through representational scaffolding. *Studies in Educational Evaluation*, *75*, 101213. https://doi.org/10.1016/j.stueduc.2022.101213

Betsky, A., & Shapiro, G. F. (2021). *50 lessons to learn from frank lloyd wright*. Rizzoli.

Bias, R. G. (2011). The importance of rigor in usability studies. In *Lecture notes in computer science* (pp. 255–258). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-21708-1_29

Billock, J. (2020, July 24). *This frank lloyd wright home was a trailblazing example of accessible design*. Smithsonian Magazine. https://www.smithsonianmag.com/travel/this-frank-lloyd-wright-home-was-trailblazing-example-accessible-design-180975383/

Borsci, S., Malizia, A., Schmettow, M., van der Velde, F., Tariverdiyeva, G., Balaji, D., & Chamberlain, A. (2021). The chatbot usability scale: The design and pilot of a usability scale for interaction with ai-based conversational agents. *Personal and Ubiquitous Computing*, *26*(1), 95–119. https://doi.org/10.1007/s00779-021-01582-9

Boston, M. (2012). Assessing instructional quality in mathematics. *The Elementary School Journal*, *113*(1), 76–104. https://doi.org/10.1086/666387

Bringhurst, R. (2004). *The elements of typographic style* (3rd ed.). Hartley and Marks Publishers.

Camarata, T., & Slieman, T. A. (2020). Improving student feedback quality: A simple model using peer review and feedback rubrics. *Journal of Medical Education and Curricular Development*, *7*, 238212052093660. https://doi.org/10.1177/2382120520936604

Carter, B. T., & Luke, S. G. (2020). Best practices in eye tracking research. *International Journal of Psychophysiology*, *155*, 49–62. https://doi.org/10.1016/j.ijpsycho.2020.05.010

CAST. (2018). *Universal Design for Learning Guidelines version 2.2*. Retrieved 2023, from http://udlguidelines.cast.org

Celik, I. (2023). Towards intelligent-tpack: An empirical study on teachers' professional knowledge to ethically integrate artificial intelligence (ai)-based tools into education. *Computers in Human Behavior*, *138*, 107468. https://doi.org/10.1016/j.chb.2022.107468

Cheng, A., Eppich, W., Epps, C., Kolbe, M., Meguerdichian, M., & Grant, V. (2021). Embracing informed learner self-assessment during debriefing: The art of plus-delta. *Advances in Simulation*, *6*(1). https://doi.org/10.1186/s41077-021-00173-1

Cheng, A., Kolbe, M., Grant, V., Eller, S., Hales, R., Symon, B., Griswold, S., & Eppich, W. (2020). A practical guide to virtual debriefings: Communities of inquiry perspective. *Advances in Simulation*, *5*(1). https://doi.org/10.1186/s41077-020-00141-1

Cheng, E. L. (2018). Choosing between the theory of planned behavior (tpb) and the technology acceptance model (tam). *Educational Technology Research and Development*, *67*(1), 21–37. https://doi.org/10.1007/s11423-018-9598-6

Cherns, A. (1976). The principles of sociotechnical design. *Human Relations*, *29*(8), 783–792. https://doi.org/10.1177/001872677602900806

Ciriello, R. F., & Loss, S. (2023). The yea-paradox: Cognitive bias in technology acceptance surveys. *Information and Software Technology*, *161*, 107253. https://doi.org/10.1016/j.infsof.2023.107253

Clarkson, P., Coleman, R., Keates, S., & Lebbon, C. (2003). *Inclusive design: Design for the whole population* (2003rd ed.). Springer.

Clarkson, P., Keates, S., & Dowland, R. (1999). The design and control of assistive devices. *in Proceedings of the International Conference on Engineering Design (ICED)*, *1*, 425–428.

Cleary, R. (2009). *Frank lloyd wright: From within outward* (First ed., First Printing). Skira Rizzoli.

Cook, D. A., Andersen, D. K., Combes, J. R., Feldman, D. L., & Sachdeva, A. K. (2018). The value proposition of simulation-based education. *Surgery*, *163*(4), 944–949. https://doi.org/10.1016/j.surg.2017.11.008

Datta, D., Phillips, M., Bywater, J. P., Lilly, S., Chiu, J., Watson, G. S., & Brown, D. E. (2022). Human-in-the-loop data collection and evaluation for improving mathematical conversations. In *Artificial intelligence in education. posters and late breaking results, workshops and tutorials, industry and innovation tracks, practitioners' and doctoral consortium* (pp. 551–554). Springer International Publishing. https://doi.org/10.1007/978-3-031-11647-6_113

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, *13*(3), 319. https://doi.org/10.2307/249008

Demchak, C. C. (2012). Resilience and cyberspace: Recognizing the challenges of a global socio-cyber infrastructure (gsci). *Journal of Comparative Policy Analysis: Research and Practice*, *14*(3), 254–269. https://doi.org/10.1080/13876988.2012.687619

Dick, W., Carey, L., & Carey, J. O. (2008). *The systematic design of instruction* (7th ed.). Pearson.

Dreifuerst, K. (2012). Using debriefing for meaningful learning to foster development of clinical reasoning in simulation. *Journal of Nursing Education*, *51*(6), 326–333. https://doi.org/10.3928/01484834-20120409-02

Edmondson, A. C. (2003). Speaking up in the operating room: How team leaders promote learning in interdisciplinary action teams. *Journal of Management Studies*, *40*(6), 1419–1452. https://doi.org/10.1111/1467-6486.00386

Edyburn, D. L. (2020). Universal usability and universal design for learning. *Intervention in School and Clinic*, *56*(5), 310–315. https://doi.org/10.1177/1053451220963082

El Hussein, M., Harvey, G., & Kilfoil, L. (2021). Pre-brief in simulation-based experiences: A scoping review of the literature. *Clinical Simulation in Nursing*, *61*, 86–95. https://doi.org/10.1016/j.ecns.2021.08.003

Ericsson, K. (2012). The influence of experience and deliberate practice on the development of superior expert performance. In *The cambridge handbook of expertise and expert performance* (pp. 683–704). Cambridge University Press. https://doi.org/10.1017/cbo9780511816796.038

Farrow, R. (2023). The possibilities and limits of xai in education: A socio-technical perspective. *Learning, Media and Technology*, *48*(2), 266–279. https://doi.org/10.1080/17439884.2023.2185630

Fatani, T. H. (2020). Student satisfaction with videoconferencing teaching quality during the covid-19 pandemic. *BMC Medical Education*, *20*(1). https://doi.org/10.1186/s12909-020-02310-2

Fleming, M. (1967). Classification and analysis of instructional illustrations. *AV communication review*, *15*(3), 246–258. https://doi.org/10.1007/bf02768609

Fleming, M. E. (1979). *Instructional message design: Principles from the behavioral sciences*. Educational Technology Publications.

Fleury, S., & Chaniaud, N. (2023). Multi-user centered design: Acceptance, user experience, user research and user testing. *Theoretical Issues in Ergonomics Science*, 1–16. https://doi.org/10.1080/1463922x.2023.2166623

Fosnot, C. T. (2005). *Constructivism: Theory, perspectives, and practice* (2nd ed.). Teachers College Press.

Freeman, C., & Perez, C. (1988). Structural crises of adjustment, business cycles and investment behaviour. In *Technical change and economic theory* (pp. 38–73). Londres: Pinter.

Garrison, D., Anderson, T., & Archer, W. (1999). Critical inquiry in a text-based environment: Computer conferencing in higher education. *The Internet and Higher Education*, *2*(2-3), 87–105. https://doi.org/10.1016/s1096-7516(00)00016-6

Geels, F. W. (2002). Technological transitions as evolutionary reconfiguration processes: A multi-level perspective and a case-study. *Research Policy*, *31*(8-9), 1257–1274. https://doi.org/10.1016/s0048-7333(02)00062-8

Geels, F. W. (2004). From sectoral systems of innovation to socio-technical systems. *Research Policy*, *33*(6-7), 897–920. https://doi.org/10.1016/j.respol.2004.01.015

Gregory, R. L., Rockinson-Szapkiw, A. J., & Cook, V. S. (2020). Community college faculty perceptions of the quality matters™ rubric. *Online Learning*, *24*(2). https://doi.org/10.24059/olj.v24i2.2052

Grelle, D., & Gutierrez, S. (2019). Developing device-equivalent and effective measures of complex thinking with an information processing framework and mobile first design principles. *Personnel Assessment and Decisions*, *5*(3). https://doi.org/10.25035/pad.2019.03.004

Grossman, P., Hammerness, K., & McDonald, M. (2009). Redefining teaching, re-imagining teacher education. *Teachers and Teaching*, *15*(2), 273–289. https://doi.org/10.1080/13540600902875340

Hadjerrouit, S. (2017). Assessing the affordances of simreal+ and their applicability to support the learning of mathematics in teacher education. *Issues in Informing Science and Information Technology*, *14*, 121–138. https://doi.org/10.28945/3692

Hassenzahl, M. (2000). Prioritizing usability problems: Data-driven and judgement-driven severity estimates. *Behaviour & Information Technology*, *19*(1), 29–42. https://doi.org/10.1080/014492900118777

Hellaby, M. (2013). *Healthcare simulation in practice*. M&K Publishing.

Hendricks, V. F., Pedersen, S. A., & Olsen, J.-K. B. (2009). *A companion to the philosophy of technology*. John Wiley & Sons.

Herrmann, T., & Pfeiffer, S. (2022). Keeping the organization in the loop: A socio-technical extension of human-centered artificial intelligence. *AI & SOCIETY*, *38*(4), 1523–1542. https://doi.org/10.1007/s00146-022-01391-5

Hill, J., Randolph Ford, W., & Farreras, I. G. (2015). Real conversations with artificial intelligence: A comparison between human–human online conversations and human–chatbot conversations. *Computers in Human Behavior*, *49*, 245–250. https://doi.org/10.1016/j.chb.2015.02.026

Hollender, N., Hofmann, C., Deneke, M., & Schmitz, B. (2010). Integrating cognitive load theory and concepts of human–computer interaction. *Computers in Human Behavior*, *26*(6), 1278–1288. https://doi.org/10.1016/j.chb.2010.05.031

Hornbæk, K., & Hertzum, M. (2017). Technology acceptance and user experience. *ACM Transactions on Computer-Human Interaction*, *24*(5), 1–30. https://doi.org/10.1145/3127358

Hosseini, Z., Hytönen, K., & Kinnunen, J. (2022). Technological pedagogical content design (TPCD) for a user-centered website:a case study in finland. *Journal of e Learning and Knowledge Society*, *18*(1), 45–58. https://doi.org/10.20368/1971 8829/ 1135490

Hurtienne, J., & Blessing, L. (2007). Design for Intuitive Use-Testing image schema theory for user interface design. *DS 42: Proceedings of ICED 2007, the 16th International Conference on Engineering Design, Paris, France*, 829–830.

Issa, T., & Isaias, P. (2022). *Sustainable design*. Springer London. https://doi.org/10.1007/978-1-4471-7513-1

Jahnke, I., Riedel, N., Singh, K., & Moore, J. (2021). Advancing sociotechnical-pedagogical heuristics for the usability evaluation of online courses for adult learners. *Online Learning*, *25*(4). https://doi.org/10.24059/olj.v25i4.2439

Jahnke, I., Schmidt, M., Pham, M., & Singh, K. (2020). Sociotechnical-pedagogical usability for designing and evaluating learner experience in technology-enhanced environments. In

Learner and user experience research: An introduction for the field of learning design & technology. EdTech Books. https://edtechbooks.org/ux/sociotechnical_pedagogical_usability

Johannsen, F., Knipp, M., Loy, T., Mirbabaie, M., Möllmann, N. J., Voshaar, J., & Zimmermann, J. (2023). What impacts learning effectiveness of a mobile learning app focused on first-year students? *Information Systems and e-Business Management*, *21*(3), 629–673. https://doi.org/10.1007/s10257-023-00644-0

Jones, A. I., Artikis, A., & Pitt, J. (2013). The design of intelligent socio-technical systems. *Artificial Intelligence Review*, *39*(1), 5–20. https://doi.org/10.1007/s10462-012-9387-2

Joyce, A. (2022). *Inclusive design*. Nielsen Norman Group. Retrieved March, 2024, from https://www.nngroup.com/articles/inclusive-design/

Juslin, P. N., Ingmar, E., & Danielsson, J. (2021). Supplemental material for aesthetic judgments of music: Reliability, consistency, criteria, self-insight, and expertise. *Psychology of Aesthetics, Creativity, and the Arts*. https://doi.org/10.1037/aca0000403.supp

Kaiser, Z. (2019). Creativity as computation: Teaching design in the age of automation. *Design and Culture*, *11*(2), 173–192. https://doi.org/10.1080/17547075.2019.1609279

Kanuka, H., & Garrison, D. (2004). Cognitive presence in online learning. *Journal of Computing in Higher Education*, *15*(2), 21–39. https://doi.org/10.1007/bf02940928

Kaufman, D., & Ireland, A. (2016). Enhancing teacher education with simulations. *TechTrends*, *60*(3), 260–267. https://doi.org/10.1007/s11528-016-0049-0

Keates, S., Clarkson, P., Harrison, L. A., & Robinson, P. (2000). Towards a practical inclusive design approach. *Proceedings on the 2000 conference on Universal Usability*, 45–52.

Kenny, R., Fischhoff, B., Davis, A., Carley, K. M., & Canfield, C. (2022). Duped by bots: Why some are better than others at detecting fake social media personas. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 001872082110726. https://doi.org/10.1177/00187208211072642

Kim, J., Li, Y., & Choi, J. (2023). Understanding the continuance intention to use chatbot services. *Asia Marketing Journal*, *25*(3), 99–110. https://doi.org/10.53728/2765-6500.1613

Kimball, M. A., & Hawkins, A. R. (2007). *Document design: A guide for technical communicators* (First ed.). Bedford/St. Martin's.

Koehler, M., & Mishra, P. (2009). What is technological pedagogical content knowledge (TPACK)?. *Contemporary issues in technology and teacher education,*, *9*(1), 60–70.

Kosch, T., Karolus, J., Zagermann, J., Reiterer, H., Schmidt, A., & Woźniak, P. W. (2023). A survey on measuring cognitive workload in human-computer interaction. *ACM Computing Surveys*, *55*(13s), 1–39. https://doi.org/10.1145/3582272

Kruse, L., Purao, S., & Seidel, S. (2022). How designers use design principles: Design behaviors and application modes. *Journal of the Association for Information Systems*, *23*(5), 1235–1270. https://doi.org/10.17705/1jais.00759

Kulkarni, P., Sirsikar, N., Mahabaleshwarkar, A., Gadgil, K., & Kulkarni, M. (2019). Conversational AI: An overview of methodologies, applications & future scope. *In 2019 5th International conference on computing, communication, control and automation (ICCUBEA)*, 1–7. https://doi.org/10.1109/ICCUBEA47591.2019.9129347

Kumar, H., Rothschild, D. M., Goldstein, D. G., & Hofman, J. (2023). Math education with large language models: Peril or promise? *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4641653

Labadze, L., Grigolia, M., & Machaidze, L. (2023). Role of ai chatbots in education: Systematic literature review. *International Journal of Educational Technology in Higher Education*, *20*(1). https://doi.org/10.1186/s41239-023-00426-1

Langevin, R., Lordon, R. J., Avrahami, T., Cowan, B. R., Hirsch, T., & Hsieh, G. (2021). Heuristic evaluation of conversational agents. *n Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–15. https://doi.org/10.1145/3411764.3445312

Lapum, J. L., Verkuyl, M., Hughes, M., Romaniuk, D., McCulloch, T., & Mastrilli, P. (2018). Self-debriefing in virtual simulation. *Nurse Educator*, 1. https://doi.org/10.1097/00006223-900000000-99544

Lavie, T., & Tractinsky, N. (2004). Assessing dimensions of perceived visual aesthetics of web sites. *International Journal of Human-Computer Studies*, *60*(3), 269–298. https://doi.org/10.1016/j.ijhcs.2003.09.002

Law, E.-C., van Schaik, P., & Roto, V. (2014). Attitudes towards user experience (ux) measurement. *International Journal of Human-Computer Studies*, *72*(6), 526–541. https://doi.org/10.1016/j.ijhcs.2013.09.006

Lawry, S., Popovic, V., Blackler, A., & Thompson, H. (2019). Age, familiarity, and intuitive use: An empirical investigation. *Applied Ergonomics*, *74*, 74–84. https://doi.org/10.1016/j.apergo.2018.08.016

Lazard, A. J., & King, A. J. (2019). Objective design to subjective evaluations: Connecting visual complexity to aesthetic and usability assessments of ehealth. *International Journal of Human–Computer Interaction*, *36*(1), 95–104. https://doi.org/10.1080/10447318.2019.1606976

Leacock, T. L., & Nesbit, J. C. (2007). A framework for evaluating the quality of multimedia learning resources. *Educational Technology & Society*, *10*(2), 44–59.

Legon, R. (2015). Measuring the impact of the quality matters rubric™: A discussion of possibilities. *American Journal of Distance Education*, *29*(3), 166–173. https://doi.org/10.1080/08923647.2015.1058114

Levie, W., & Lentz, R. (1982). Effects of text illustrations: A review of research. *ECTJ*, *30*(4), 195–232. https://doi.org/10.1007/bf02765184

Levin, O., & Flavian, H. (2020). Simulation-based learning in the context of peer learning from the perspective of preservice teachers: A case study. *European Journal of Teacher Education*, *45*(3), 373–394. https://doi.org/10.1080/02619768.2020.1827391

Lidwell, W., Holden, K., & Butler, J. (2010). *Universal prinicples of design* (2nd ed.). Rockport Publishers, Inc.

Liebrecht, C., Sander, L., & van Hooijdonk, C. (2021). Too informal? how a chatbot's communication style affects brand attitude and quality of interaction. In *Chatbot research and design* (pp. 16–31). Springer International Publishing. https://doi.org/10.1007/978-3-030-68288-0_2

Lim, Y., Ayesh, A., & Chee, K. (2013). Socio-demographic differences in the perceptions of learning management system (lms) design. *International Journal of Software Engineering & Applications*, *4*(5), 15–35. https://doi.org/10.5121/ijsea.2013.4502

Lowenthal, P. R., & Dunlap, J. C. (2020). Social presence and online discussions: A mixed method investigation. *Distance Education*, *41*(4), 490–514. https://doi.org/10.1080/01587919.2020.1821603

Lu, J., Schmidt, M., Lee, M., & Huang, R. (2022). Usability research in educational technology: A state-of-the-art systematic review. *Educational technology research and development*, *70*(6), 1951–1992. https://doi.org/10.1007/s11423-022-10152-6

Mafra, M., Nunes, K., Castro, A., Lopes, A., Oran, A., Braz Junior, G., Almeida, J., Paiva, A., Silva, A., Rocha, S., Viana, D., Melo, A., Barreto, R., & Rivero, L. (2022). Defining requirements for the development of useful and usable chatbots: An analysis of quality attributes from academy and industry. In *Human-computer interaction. user experience and behavior* (pp. 479–493). Springer International Publishing. https://doi.org/10.1007/978-3-031-05412-9_33

Marangunić, N., & Granić, A. (2014). Technology acceptance model: A literature review from 1986 to 2013. *Universal Access in the Information Society*, *14*(1), 81–95. https://doi.org/10.1007/s10209-014-0348-1

Mayer, R. E. (2020). *Multimedia learning* (3rd ed.). Cambridge University Press.

Mayer, R. E., & Estrella, G. (2014). Benefits of emotional design in multimedia instruction. *Learning and Instruction*, *33*, 12–18. https://doi.org/10.1016/j.learninstruc.2014.02.004

McTear, M., Callejas, Z., & Griol, D. (2016). *The conversational interface: Talking to smart devices* (1st ed., 2016). Springer.

Mezirow, J. (1994). Understanding transformation theory. *Adult Education Quarterly*, *44*(4), 222–232. https://doi.org/10.1177/074171369404400403

Microsoft. (2016). *Inclusive 101 guidebook* [PDF]. Microsoft Inclusive Design. Retrieved March, 2024, from https://inclusive.microsoft.design/tools-and-activities/Inclusive101Guidebook.pdf

Miro. (2024). Miro. https://miro.com/

Moore, M. G. (1989). Editorial: Three types of interaction. *American Journal of Distance Education*, *3*(2), 1–7. https://doi.org/10.1080/08923648909526659

Morrison, G. R., Ross, S. J., Morrison, J. R., & Kalman, H. K. (2019). *Designing effective instruction* (8th ed.). Wiley.

Morville, P. (2014). *Intertwingled: information changes everything.* Semantic Studios.

Mozilla. (2024). *Accessibility inspector — firefox source docs documentation*. Accessibility Inspector. Retrieved March, 2024, from https://firefox-source-docs.mozilla.org/devtools-user/accessibility_inspector/index.html

Ngo, D. (2001). Measuring the aesthetic elements of screen designs. *Displays*, *22*(3), 73–78. https://doi.org/10.1016/s0141-9382(01)00053-1

Nielsen, J. (1994). *Severity ratings for usability problems*. Nielsen Norman Group. Retrieved March, 2024, from https://www.nngroup.com/articles/how-to-rate-the-severity-of-usability-problems/

Nielsen, J., & Molich, R. (1990). Heuristic evaluation of user interfaces. *In Proceedings of the SIGCHI conference on Human factors in computing systems*, 249–256. https://doi.org/10.1145/97243.97281

Norman, D. (1986). *User centered system design: New perspectives on human-computer interaction* (1st ed.). CRC Press.

Norman, D. (2013). *The design of everyday things* (Revised and Expanded ed.). Basic Books.

Norman, D. A. (2005). Human-centered design considered harmful. *Interactions*, *12*(4), 14–19. https://doi.org/10.1145/1070960.1070976

O'Connor, Z. (2013). Colour, contrast and Gestalt theories of perception: The impact in contemporary visual communications design. *Color Research & Application*, *40*(1), 85–92. https://doi.org/10.1002/col.21858

Online Learning Consortium. (2024). New OLC OSCQR course design review. https://onlinelearningconsortium.org/consult/oscqr-course-design-review/

Oppermann, R. (1994). Adaptively supported adaptability. *International Journal of Human-Computer Studies*, *40*(3), 455–472. https://doi.org/10.1006/ijhc.1994.1021

Parcell, J. (2013, May 25). *Functionality and usability testing resources*. Digital.gov. https://digital.gov/2013/05/25/functionality-and-usability-testing-resources/

Paul, S. (2022). Accessibility analysis using wcag 2.1: Evidence from indian e-government websites. *Universal Access in the Information Society*, *22*(2), 663–669. https://doi.org/10.1007/s10209-021-00861-9

Phillips, M. (2022). Moving the dialogue forward: Virtual student conversational agent design in low data environments. Doctoral Dissertation, University of Virginia.

Podmajersky, T. (2019). *Strategic writing for UX: Drive engagement, conversion, and retention with every word*. O'Reilly Media.

Preece, J., Rogers, Y., & Sharp, H. (2002). *Interaction design: Beyond human-computer interaction* (1st ed.). Wiley.

Punchoojit, L., & Hongwarittorrn, N. (2017). Usability studies on mobile user interface design patterns: A systematic literature review. *Advances in Human-Computer Interaction*, *2017*, 1–22. https://doi.org/10.1155/2017/6787504

Qin, C., Huang, W., & Hew, K. (2020). Using the community of inquiry framework to develop an educational chatbot: lesson learned from a mobile instant messaging learning environment. *Proceedings of the 28th international conference on computers in education.*

Quality Matters Higher Education Rubric, Seventh Edition. (2023). Quality Matters. Used under license. All rights reserved. https://www.qualitymatters.org/sites/default/files/PDFs/StandardsfromtheQMHigherEducationRubric.pdf

Quiñones, D., & Rusu, C. (2017). How to develop usability heuristics: A systematic literature review. *Computer Standards & Interfaces*, *53*, 89–122. https://doi.org/10.1016/j.csi.2017.03.009

Quiñones, D., Rusu, C., & Rusu, V. (2018). A methodology to develop usability/user experience heuristics. *Computer Standards & Interfaces*, *59*, 109–129. https://doi.org/10.1016/j.csi.2018.03.002

Rand, P. (2014). *Thoughts on design* (Reissue ed.). Chronicle Books LLC.

Reigeluth, C., & An, Y. (2023). What's the difference between learning experience design and instructional design? *The Journal of Applied Instructional Design*. https://doi.org/10.59668/515.12897

Reyna, J. (2013). The importance of visual design and aesthetics in e-learning. *Training and Development*, *40*(5), 28–31.

Richardson, J., & Swan, K. (2003). Examining social presence in online courses in relation to students' perceived learning and satisfaction. *Journal of Asynchronous Learning Networks*, *7*, 68–88.

Rogers, E. M. (2003). *Diffusion of innovations, 5th edition* (5th ed.). Free Press.

Rose, E. J., & Turner, H. N. (2023). The paradigm shift to ux and the durability of usability in tpc. *Technical Communication Quarterly*, 1–12. https://doi.org/10.1080/10572252.2023.2274067

Rourke, L., Anderson, T., Garrison, R. D., & Archer, W. (1999). Assessing social presence in asynchronous text-based computer conferencing. *Journal of Distance Education.*, *14*(2), 50–71, 22p.

Runnacles, J., Thomas, L., Sevdalis, N., Kneebone, R., Arora, S., & Cooper, M. (2014). Development of a tool to improve performance debriefing and learning: The paediatric objective structured assessment of debriefing (osad) tool. *Postgraduate Medical Journal*, *90*(1069), 613–621. https://doi.org/10.1136/postgradmedj-2012-131676

Salas, E., Goodwin, G. F., & Burke, C. S. (2009). Overcoming Barriers to Collaboration: Psychological Safety and Learning in Diverse Teams. In *Team effectiveness in complex organizations: Cross-disciplinary perspectives and approaches (siop organizational frontiers series)* (1st ed.). Routledge.

Sansonnet, J.-P., Leray, D., & Martin, J.-C. (2006). Architecture of a framework for generic assisting conversational agents. In *Intelligent virtual agents* (pp. 145–156). Springer Berlin Heidelberg. https://doi.org/10.1007/11821830_12

Sarker, S., Chatterjee, S., Xiao, X., & Elbanna, A. (2019). The sociotechnical axis of cohesion for the is discipline: Its historical legacy and its continued relevance. *MIS Quarterly*, *43*(3), 695–719. https://doi.org/10.25300/misq/2019/13747

Sartori, L., & Theodorou, A. (2022). A sociotechnical perspective for the future of ai: Narratives, inequalities, and human control. *Ethics and Information Technology*, *24*(1). https://doi.org/10.1007/s10676-022-09624-3

Schrepp, M., Hinderks, A., & Thomaschewski, J. (2017). Design and evaluation of a short version of the user experience questionnaire (ueq-s). *International Journal of Interactive Multimedia and Artificial Intelligence*, *4*(6), 103. https://doi.org/10.9781/ijimai.2017.09.001

Seckman, C. (2018). Impact of interactive video communication versus text-based feedback on teaching, social, and cognitive presence in online learning communities. *Nurse Educator*, *43*(1), 18–22. https://doi.org/10.1097/nne.0000000000000448

Sewell, J. L., Maggio, L. A., ten Cate, O., van Gog, T., Young, J. Q., & O'Sullivan, P. S. (2018). Cognitive load theory for training health professionals in the workplace: A beme review of studies among diverse professions: Beme guide no. 53. *Medical Teacher*, *41*(3), 256–270. https://doi.org/10.1080/0142159x.2018.1505034

Shea, P., & Bidjerano, T. (2009). Community of inquiry as a theoretical framework to foster "epistemic engagement" and "cognitive presence" in online education. *Computers & Education*, *52*(3), 543–553. https://doi.org/10.1016/j.compedu.2008.10.007

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. Educational Researcher, 15(2), 4–14. https://doi.org/10.3102/0013189x015002004

Silver, K. (2007). *What puts the design in interaction design*. https://www.uxmatters.com/mt/archives/2007/07/what-puts-the-design-in-interaction-design.php

Soloway, E., Guzdial, M., & Hay, K. E. (1994). Learner-centered design. *Interactions*, *1*(2), 36–48. https://doi.org/10.1145/174809.174813

Sony, M., & Naik, S. (2020). Industry 4.0 integration with socio-technical systems theory: A systematic review and proposed theoretical model. *Technology in Society*, *61*, 101248. https://doi.org/10.1016/j.techsoc.2020.101248

Squires, D., & Preece, J. (1996). Usability and learning: Evaluating the potential of educational software. *Computers & Education*, *27*(1), 15–22. https://doi.org/10.1016/0360-1315(96)00010-3

Steinfeld, E., & Maisel, J. (2012). *Universal design: Creating inclusive environments* (1st ed.). Wiley.

Straub, E. T. (2009). Understanding technology adoption: Theory and future directions for informal learning. *Review of Educational Research*, *79*(2), 625–649. https://doi.org/10.3102/0034654308325896

Swan, K., Matthews, D., Bogle, L., Boles, E., & Day, S. (2012). Linking online course design and implementation to learning outcomes: A design experiment. *The Internet and Higher Education*, *15*(2), 81–88. https://doi.org/10.1016/j.iheduc.2011.07.002

Sweller, J., van Merriënboer, J. G., & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review*, *31*(2), 261–292. https://doi.org/10.1007/s10648-019-09465-5

Tessmer, M. (1993). *Planning and conducting formative evaluations*. Routledge. https://doi.org/10.4324/9780203061978

Tomita, K. (2022). Visual design as a holistic experience: How students' emotional responses to the visual design of instructional materials are formed. *Educational technology research and development*, *70*(2), 469–502. https://doi.org/10.1007/s11423-022-10088-x

Travis, D., & Hodgson, P. (2019). *Think like a ux researcher* (1st ed.). Taylor & Francis.

Trist, E. L. (1981). *The evolution of socio-technical systems* (Vol 2, p. 1981). Toronto: Ontario Quality of Working Life Centre.

Tun, J., Alinier, G., Tang, J., & Kneebone, R. L. (2015). Redefining simulation fidelity for healthcare education. *Simulation & Gaming*, *46*(2), 159–174. https://doi.org/10.1177/1046878115576103

Vargo, S. L., Akaka, M., & Wieland, H. (2020). Rethinking the process of diffusion in innovation: A service-ecosystems and institutional perspective. *Journal of Business Research*, *116*, 526–534. https://doi.org/10.1016/j.jbusres.2020.01.038

Venkatesh, Morris, Davis, & Davis. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, *27*(3), 425. https://doi.org/10.2307/30036540

Wanstreet, C. E., & Stein, D. S. (2011). Presence over time in synchronous communities of inquiry. *American Journal of Distance Education*, *25*(3), 162–177. https://doi.org/10.1080/08923647.2011.590062

White, A. W. (2011). *The elements of graphic design* (Second ed.). Allworth.

Wittmann-Price, R. A., & Wilson, L. (2014). *Review manual for the certified healthcare simulation educator exam* (1st ed.). Springer Publishing Company.

Yablonski, J. (2020). *Laws of ux* (1st ed.). O'Reilly Media, Inc.

Zagermann, J., Pfeil, U., & Reiterer, H. (2016). Measuring cognitive load using eye tracking technology in visual computing. *Proceedings of the Sixth Workshop on Beyond Time and*

Errors on Novel Evaluation Methods for Visualization, 78–85. https://doi.org/10.1145/2993901.2993908

Zaharias, P., & Poylymenakou, A. (2009). Developing a usability evaluation method for e-learning applications: Beyond functional usability. *International Journal of Human-Computer Interaction*, *25*(1), 75–98. https://doi.org/10.1080/10447310802546716

Zardari, B., Hussain, Z., Arain, A., Rizvi, W. H., & Vighio, M. (2020). Quest e-learning portal: Applying heuristic evaluation, usability testing and eye tracking. *Universal Access in the Information Society*, *20*(3), 531–543. https://doi.org/10.1007/s10209-020-00774-z

Zoom. (2021). (Version 5.5.2 (12513.0205)) [Computer software]. Zoom Video Communications. https://zoom.us/

APPENDIX A

Indicators of the Social-TPAC Model

Code	Indicator Criteria	Examples	Citation
S1.1	Expression of emotion - Use descriptive words that indicate feelings, conventional or unconventional expressions of emotions.	love, sad, hate, silly	(Rourke et al., 1999).
S1.2	Use of humor	joking, teasing, cajoling, irony, sarcasm, understatement	(Rourke et al., 1999).
S1.3	Self-disclosure - Sharing personal information and expressing vulnerability or feelings.	I made that mistake before	(Rourke et al., 1999).
S1.4	Emoticons	:(:-)	(Liebrecht et al., 2021)
S1.5	Sound mimicking	awww, ooh, oops	(Liebrecht et al., 2021)
S1.6	Contractions and Shortenings	LOL, got it	(Liebrecht et al., 2021)
S1.7	The AI conveys a personality by providing greetings, self-introductory, empathy, information, etc.	Hello, my name is, I understand	(Borsci et al., 2021)
S2.1	Referring to other messages Process tracking and follow up. Chatbot seems to be able to inform and update users about their status and progresses toward the achievement of the goal.		(Borsci et al., 2021; Rourke et al., 1999)
S2.2	Asking questions, gathering additional information		(Rourke et al., 1999).
S2.3	Complimenting, expressing appreciation or approval of messages or content.	thank you, thanks, that makes sense,	(Rourke et al., 1999).
S2.4	Expressing agreement - Expressing agreement or disagreement with other's messages		(Rourke et al., 1999).

S2.5	Response Time. The chatbot is perceived as able to respond in a timely manner to requests.		(Borsci et al., 2021)
S2.6	Able to recognize and process simultaneously multiple and parallel topics during the conversation.		(Borsci et al., 2021)
S2.7	Provides a relevant and appropriate contribution to people's needs at each stage		(Borsci et al., 2021)
S3.1	Addressing or referring to each other by name		(Rourke et al., 1999).
S3.2	Expresses inclusivity, able to meet needs of the users independently form their gender, well-being, age, etc.	she, he, them, Mr., Ms.	(Borsci et al., 2021; Rourke et al., 1999)
S3.3	Able to initiate conversation (or to offer cues) for further discussion by offering suggestions, etc.	hi, how are you?, how can I help you?	(Borsci et al., 2021; Rourke et al., 1999)
S3.4	Embracing opinions	I understand, can you clarify, I am not sure I know what you are saying	(Rourke et al., 1999).
S3.5	Responds in an informative way without adding too much information, over explaining.		(Borsci et al., 2021)

S3.6	Handles messages gracefully, including unexpected events such as communication mismatch, or a broken line of conversation, etc.	(Borsci et al., 2021)
S3.7	Maintains a conversational theme once introduced and keep track of the context to understand the user's utterances	(Borsci et al., 2021)
S4.1	Conveys correct statements and information.	(Borsci et al., 2021)
S4.2	Messages with purpose clear without ambiguity	(Borsci et al., 2021)
S4.3	Uses appropriate and accurate language style for the context.	(Borsci et al., 2021)
S4.4	Seems able to respond in different and appropriate ways to similar or repeated requests.	(Borsci et al., 2021)
S4.5	Able to exhibit knowledge that it is out of its immediate domain during a conversation.	(Borsci et al., 2021)
S4.6	Able to appropriately recognize the mood of the user from the conversation and to respond accordingly.	(Borsci et al., 2021)
S4.7	Sensitivity to safety, and able to recognize and respond to social concerns and to refer a user to helpline if needed	(Borsci et al., 2021)
P1.1	Directions are clear, and instructions make clear how to get started. Course includes Welcome and Getting Started content.	(Apple, 2024; QM; OSCQR)
P1.2	Specific and descriptive criteria are provided for the evaluation of learners' work, and their connection to the course grading policy is clearly explained.	QM
P1.3	Uses graceful language that skillfully communicates meaning to readers with clarity and fluency and is error-free.	(Cheng, 2020)

P1.4	Orient learners to the collaborative environment. Provide learners with an orientation to the features of the online environment.	Overlays, information	(Cheng, 2020)
P1.5	Communicate rules of engagement, emphasizing the importance of privacy and confidentiality, establishing an expectation of messaging quality, frequency		(Cheng, 2020)
P1.6	Encourage learners to refrain from activities that are not directly related to the learning activity, suggest muting phone, closing email or other messaging programs, turn off computer notifications.		(Cheng, 2020)
P1.7	Ensure that the help documentation is easy to search.		(Nielsen & Molich, 1990)
P2.1	Objects that are near or proximate to each other, are meant to be grouped together. (Law of Proximity)		(Yablonski, 2020)
P2.2	Illustrations and images are aligned with text in the instructional materials.		(Fleming, 1979)
P2.3	A minimal number of colors are used.		(Fleming, 1979)
P2.4	Labels are readable.		(Apple, 2024; Podmajersky, 2019)
P2.5	Any stepper used to change the values, should be clear in and obvious what values will change.		(Borsci et al., 2021)
P2.6	Numeric data fields have a number formatter	1000.8	(Apple, 2024)
P2.7			(Alharoon et al., 2021)
P3.1	Comments are nonjudgmental and descriptive rather than evaluative (focus on description rather than judgment).	Examples are provided assist in understanding	(Camarata & Slieman, 2020)
P3.2	Attribute positive feedback to internal causes and give it in the second person (you).	“You worked hard to explain the material well using relevant sources.”	(Camarata & Slieman, 2020)

P3.3	Expresses inclusivity, able to meet needs of the users independently form their gender, well-being, age, etc.		(Cheng, 2020)(Lowenthal & Dunlap, 2020) (Borsci et al., 2021)
P3.4	Explicitly use appreciation, validation, and normalization of mistakes.		(Cheng, 2020), (Lowenthal & Dunlap, 2020)
P3.5	Provide feedback to close performance gaps.		(Cheng, 2021)
P3.6	Provides a relevant and appropriate contribution to learner's needs.		(Borsci et al., 2021)
P3.7	Interaction enjoyment. The chatbot is perceived as enjoyable and engaging to operate with		(Borsci et al., 2021)
P4.1	Comments provide a good balance of positive and negative feedback.	You include a thought provoking topic, but it seems to me that it needs more elaboration with examples.	(Cheng, 2020)
P4.2	Provides learners with multiple opportunities to track their learning progress with timely feedback.		QM
P4.3	Give negative information in the first person (I) and then shift to third person (s/he), or shift from a statement to a question that frames the problem objectively.	"I thought I understood the organization of the material from the lecture, but then I was not sure . . ."	(Cheng, 2020)
P4.4	Offer specific suggestions that model appropriate behavior.	"Have you considered trying . . .? How do you think that would work?"	(Cheng, 2020)

P4.5	Use external data to inform feedback.	Standards from the IQA	(Cheng, 2021)
P4.6	Able to appropriately recognize the mood of the user from the conversation and to respond accordingly.		(Borsci et al., 2021)
P4.7	Messages with purpose clear without ambiguity		(Borsci et al., 2021)
T1.1	Fonts easily adjust to user setting sizes		WCAG 2.2
T1.2	There is a high-level of contrast in colors		(Alharoon et al., 2021; Nielsen & Molich, 1990; WCAG 2.2, 2024;)
T1.3	Buttons or areas to click are large enough for users to accurately select them.		(Nielsen & Molich, 1990)

T1.4	The user interface is recognizable and alerts the user to the required action.	The chat area should reflect a text message UI, search is a magnify glass.	(Nielsen & Molich, 1990)
T1.5	The technology is compatibile with current and future user browsers and devices		WCAG 2.2
T1.6	Ensure all functionality is available from a keyboard		WCAG 2.2
T1.7	There are no iteractions that rely entirely on color to convey information		WCAG 2.2
T2.1	Offers users a solution, like a shortcut that can solve the error immediately.		(Nielsen & Molich, 1990)
T2.2	Prevents errors from occurring.	If a step is missed, a message requires the user to fix a problem before they contine.	(Nielsen & Molich, 1990)
T2.3	Support Undo and Redo.		(Nielsen & Molich, 1990)
T2.4	The system provides instant feedback.		(Nielsen & Molich, 1990)
T2.5	The system reacts within a reasonable time.		(Nielsen & Molich, 1990)
T2.6	The interaction with the chatbot is perceived as free from errors.	messages appear send, appear to be received.	(Borsci et al., 2021)
T2.7	There are clear ways to exit the current interaction	Cancel buttons	(Nielsen & Molich, 1990)
T3.1	Items with similar functions are the same color, shape, and size, orientation.		(Yablonski, 2020)
T3.2	Navigation is clear.		(Nielsen & Molich, 1990)
T3.3	Multistep processes show users which step they are on, steps they've completed, steps ahead.		(Nielsen & Molich, 1990)

T3.4	Information required to use the design is visible.	field labels or menu items	(Nielsen & Molich, 1990)
T3.5	Minimize choices when response times are critical to increase decision time.		(Yablonski, 2020)
T3.6	The user should not have to remember information from one part of the interface to another.		(Nielsen & Molich, 1990)
T3.7	Icons are standard across the system		OSCQR
T4.1	Users are not surprised by any interaction.		(Nielsen & Molich, 1990)
T4.2	Help is available in context.	step-by-step wizards or additional information is available	(Apple, 2024)
T4.3	Explain the action or task the control initiates by first using a verb Add or remove a language from the list.”	Restore default settings, Add or remove a item from the list	(Apple, 2024)
T4.4	Offer in context help, tips, or explanations or help that are easy to understand.		(Apple, 2024)
T4.5	In context help or explanations are brief, and a maximum of 60 to 75 characters.		(Apple, 2024)
T4.6	In context help is action-oriented to describe what the feature does and how to use it.		(Yablonski, 2020)
T4.7	Tasks take as long as expected by the users		(Yablonski, 2020)
C1.1	Title or sentence case is the same for page titles, headers, buttons, and links.		(Apple, 2024)

C1.2	Content is relevant to the page, action, or image.		(Nielsen & Molich, 1990)
C1.3	Labels are consistent for navigation, buttons, or actions.	Next or Continue	(Nielsen & Molich, 1990)
C1.4	Screens are void of Empty States	A new email account has no emails, an cloud storage account has no files or folders	(Podmajersky, 2019)
C1.5	There is a visual consistency, same stroke size, text, icon style, colors, navigation, and layout.		(Apple, 2024)
C1.6	Only one or two typefaces are used. Emphasis is done by adjusting the font weight, size, and color.		(Apple, 2024)
C1.7	Uses first or second person.	You vs My	(Apple, 2024)
C2.1	Follows a grid layout, with equal space between rows and columns.		(Apple, 2024; Podmajersky, 2019)
C2.2	There is symmetrical balance, or asymmetrical balance.		(Yablonski, 2020)
C2.3	Layout follows a visual hierarchy that follows reading order, of left-right and top to bottom. Important information is at the top. Golden Ratio for font sizes.		(Apple, 2024; Nielsen & Molich, 1990; Yablonski, 2020)
C2.4	Group functions, like navigation or steps demonstrate visual continuance.		(Yablonski, 2020)
C2.5	Objects that belong together follow the Law of Common Region, and have a similar background or border, complex lists or choices are separated by lines to create and ease in finding items.		(Yablonski, 2020)
C2.6	Colors follow color theory and are used sparingly.		(Yablonski, 2020) (Alharoon et al., 2021). White, 2011)

C2.7	Use alignment to ease visual scanning and to communicate organization and hierarchy.		(Apple, 2024; Podmajersky, 2019)
C3.1	Refers to people directly such as you or your	avoids terms user, users, and is free of gender terms such as Mr. Mrs.	(Apple, 2024)
C3.2	Uses a neutral approachable tone.		(Apple, 2024)
C3.3	Uses an active voice for instructions, labels, and help to guide users from one step to the next		(Apple, 2024; Podmajersky, 2019)
C3.4	Language does not refer to specific senses	avoid see, hear, look	(Apple, 2024)
C3.5	Example or scenarios that depict people use diverse names.		(Apple, 2024)
C3.6	Example or scenarios that depict people are inclusive and bias free.		(Apple, 2024)
C3.7	Images of people depict a diversity		(Apple, 2024)
C4.1	Users can understand words without having to look up a definition.		(Nielsen & Molich, 1990)
C4.2	There are no idioms or colloquial expressions.	a dime a dozen, call it a day, break a leg	(Apple, 2024)
C4.3	Directions are clear on how to start.		(Nielsen & Molich, 1990)
C4.4	Directions follow a “To do X, do Y” action statements.	“To get started, create an account”	(Podmajersky, 2019)

C4.5	Images, icons, and concepts are clear to all users.	Search is a magnify glass	(Nielsen & Molich, 1990; OSCQR)
C4.6	Text fields show examples of the information to be entered.	name@example.com, Your Name,	(Apple, 2024; Podmajersky, 2019)
C4.7	Buttons are short one- or two-word labels that describe the result of selecting the button.	Reply, Next, Submit	(Apple, 2024; Podmajersky, 2019)

APPENDIX B

Heuristic Findings

Heuristic	Code	Notes	Location	Participant Affected	Ease of fix 1-easy, 2-unknown/maybe, 3-difficult	Target Audience
Consistent	C1.1	The case is different in simulations. Unarchive this sim, Permanently Delete	Admin Screen		1	All
Consistent	C1.1	The hierarchy is off and not consistent across this page	Home Screen		1	All
Consistent	C1.1	The button does not follow the sentence case to get started	Simulation Question		1	All
Consistent	C1.1	Main Menu is the name of something and, therefore, should be capitalized like other items. However, I would still favor a consistent menu.	User Screen		1	All
Consistent	C1.2	The pizzas do not align to the problem and caused confusion for multiple participants	Simulation	12345	1	All
Consistent	C1.3	Active? Archived? The question marks should be removed, there are no other questions and the item to select is not the question	Add a Simulation		1	All
Consistent	C1.3	Unarchive this sim vs archive this simulation should be consistent	Admin Screen		1	All
Consistent	C1.3	Labels do not follow a consistent pattern; some show click “here”	All		1	All
Consistent	C1.3	The end chat is not apparent.	Simulation	1345	1	All
Consistent	C1.3	the navigation is completely gone inside the simulation	Simulation		1	All
Consistent	C1.3	The button is different than others in the system	Simulation Question		1	All
Consistent	C1.3	Participants struggled to keep track as to where they were in the admin pages	Admin Screen	123	1	Faculty
Consistent	C1.4	User menu is an empty state but does provide a message	Add a Simulation		2	Faculty

Consistent	C1.4	If no simulations are listed there is nothing there, the add a new sim helps but otherwise appears confusing, participant who did not have an empty state clicked on the simulation to edit it instead of adding a new simulation	Admin Screen	2	2	Faculty
Consistent	C1.4	Text sizes do not demonstrate consistency	All		1	All
Consistent	C1.4	Empty state where sims should show, add a simulation there to start	Home Screen		2	All
Consistent	C1.4	When starting in a human-human chat there is no way to know if someone is there due to empty chat. Having the typing notification would help	Simulation	123	2	All
Consistent	C1.4	Empty state before user is added	User Screen		2	Faculty
Consistent	C1.4	Participants were unclear if they should start typing in the human to human ineraction because they student takes more time to start a conversation, leaving an empty state.	Simulation	123	3	No Facilitato r Present
Consistent	C1.4	Participants were not sure if someone was responding back to them in human-human interaction, a see typing feature like iMessage was mentioned by two participants	Simulation	12	3	No Facilitato r Present
Consistent	C1.4	Participants were unaware of where to add a simulation, the participant who did not arrive to a blank screen immedietly edited a simulation instead of adding one, showing the empty state is confusing.	admin Screen	123	2	Faculty
Consistent	C1.5	Active is the same color as Permanently Delete	Admin Screen		1	Faculty
Consistent	C1.5	Stroke is heavier on KV, which may cause confusion that it is more of an image and not an interactive element	Knowledge Visualizatio n	12345	1	All
Consistent	C1.5	The green header is not present on the simulation area	Simulation Question		2	All
Consistent	C1.5	There was a lack of consistency in how to return to a page in the admin screen section	Admin Screen	123	2	Faculty
Consistent	C1.7	Does not use my or your	Home Screen		1	All

Consistent	C1.7	The button says I'm ready, other areas say your	Simulation Question		1	All
Aesthetics	C2.1	Does not follow a grid layout, the header has considerably larger margins than other areas of the model	Home Screen		1	All
Aesthetics	C2.1	The lack of space between the KV ties it more closely to the problem.	Knowledge Visualizatio n	12345	1	All
Aesthetics	C2.2	Most items are right justified	All		2	All
Aesthetics	C2.2	KV is close to the problem	Knowledge Visualizatio n	12345	1	All
Aesthetics	C2.2	The text stretches across the page and can be more difficult to read	Simulation Question		2	All
Aesthetics	C2.3	The form entry title is the same font size and weight as the entry	Add a Simulation		1	All
Aesthetics	C2.3	Hierarchy could be improved to influence easier navigation	All		1	All
Aesthetics	C2.3	Fonts follow the following numbers: 1rem, 2rem, 3.5rem	All		1	All
Aesthetics	C2.4	The student role and bot component must align but there is no guidance to do so	Add a Simulation		1	Faculty
Aesthetics	C2.4	There is a lack of knowing if items are grouped together or interact with each other there is no way to know if this is the final step, lacks continuance.	Add a Simulation		1	Faculty
Aesthetics	C2.4	There are no visual steps	Add a Simulation		2	Faculty

Aesthetics	C2.4	There are no breadcrumbs to go back to the previous screen, no continuance steps to add users, no way to sort simulations, and new simulations show in the middle of the list due to naming convention.	Admin Screen		1	Faculty
Aesthetics	C2.4	There is no order of continuance.	Home Screen		2	All
Aesthetics	C2.4	Showing continuence by connecting steps allows users to see where they are and how to go back, multiple participants got lost as to where they were in the admin section	Admin Screen	123	2	Faculty
Aesthetics	C2.5	Using a common region, adding a simple box, or adding a horizontal line between the intro component and the visual component will help to visually identify that they belong together	Add a Simulation		1	Faculty
Aesthetics	C2.5	Menu items are together but part of the sentence; therefore, it is unclear to go there for most actions.	Home Screen	12345	1	all
Aesthetics	C2.5	Grouping the chat with the KV may help to make it clearer that they can be used together	Knowledge Visualizatio n	12345	2	All
Aesthetics	C2.5	There is no way to easily see the list of users' simulations	User Screen		3	Faculty
Aesthetics	C2.5	Participants felt like the chat and KV were two separate parts, they did not see them as connected elements, adding a box behind them can make it clearer that they work together	Simulation	12345	2	No Facilitato r Present
Aesthetics	C2.5	Connecting the users and simulation more together visually may help participants see where and if users are added to a simulation	Add a Simulation	123	2	Faculty
Aesthetics	C2.6	Red generally refers to stop; marking active items using red can be confusing; consider green. Red/green colorblindness is the most common.	Admin Screen		1	All

Aesthetics	C2.6	Colors feel sporadic and are not carried though out the design	All		1	All
Aesthetics	C2.7	A heading could also be added above the student role and the bot if grouped together	Add a Simulation		1	Faculty
Aesthetics	C2.7	Add Users at the top would be more logical	Admin Screen		1	Faculty
Aesthetics	C2.7	The KV is aligned more with the problem than the chat, if the problem was across the top it could be clearer that the KV would more closely match the chat	Knowledge Visualizatio n	12345	2	All
Aesthetics	C2.7	Adding a heading to the intro component and visual component may help to identify that they belong together.	Simulation		2	All
Inclusive	C3.1	System calls learners users	Admin Screen		1	Faculty
Inclusive	C3.1	The AI-based CA defaults to Mrs. if no name is entered there is no feedback	Simulation	5	3	All
Inclusive	C3.1	The problem contains Mrs. Jones.	Simulation Question		1	All
Inclusive	C3.2	The lack of information about the feedback makes it feel non approachable	Simulation		2	All
Inclusive	C3.3	There are no labels on the KV to guide users	knowledge Visualizatio n	12345	2	All
Inclusive	C3.3	Participant was unable to clearly connect users to the simulation quickly	Add a Simulation	2	2	Faculty
Inclusive	C3.4	There is no guidance on what to do with the feedback	Feedback	345	2	No Facilitato r Present

Inclusive	C3.5	The names used in the scenario do not appear to be diverse	Simulation Question		1	All
Inclusive	C3.6	The AI-based CA calls the instructor Mrs. if no name was entered. One participant added, “how do you know who I am, I could be a Mr.”	Simulation	5	2	All
Clarity	C4.1	Menu items are confusing; they would need to be simplified.	Add a Simulation		1	Faculty
Clarity	C4.1	Time is not aligned to users' time zone, and therefore confusing.	Admin Screen		1	Faculty
Clarity	C4.1	While users may understand the meaning of words, it is unclear how their questioning style could improve	Feedback		2	No Facilitato r Present
Clarity	C4.1	Participants were very unclear what the meaning of the words in the add/edit a simulation meant, even with a list of what to select they all asked if they were clicking on the right thing.	Add a Simulation	123	2	Faculty
Clarity	C4.1	Teacher/Student was confusing to some.	User Screen	123	2	Faculty
Clarity	C4.2	unarchive can be changed to active to match the system	Add a Simulation		1	Faculty
Clarity	C4.2	The terms used in the menu are unclear to the average user. For the bot, if there is a more stable bot or preferred bot, it could be marked as such.	Add a Simulation		1	Faculty
Clarity	C4.2	? Mark next to the Active and Archived should be removed.	Add a Simulation		1	All
Clarity	C4.2	Unclear understanding of award, as it is the first thing that shows.	Home Screen		2	All
Clarity	C4.2	The button says I'm ready, lets get started	Simulation Question		1	All
Clarity	C4.3	Directions would help to guide the user.	Admin Screen		1	Faculty

Clarity	C4.3	Adding a button to add users may make more sense than a test, the test could be added to the add simulation menu *This could be my misunderstanding of needing to test the simulation directions, which would help to guide the user.	Admin Screen		2	Faculty
Clarity	C4.3	There are no directions on how to read the feedback	Feedback		2	No Facilitator Present
Clarity	C4.3	The empty state makes it unclear what should show there, perhaps your sim will show here when starting	Home Screen		3	All
Clarity	C4.3	There are no directions how to start	Simulation		1	No Facilitator Present
Clarity	C4.4	To get started create a new simulation.	Admin Screen		1	Faculty
Clarity	C4.4	No directions	Feedback		2	No Facilitator Present
Clarity	C4.4	There are no directions on how to actually start the chat	Simulation		1	No Facilitator Present
Clarity	C4.4	When first starting, you may not want to edit users but add them; perhaps change to To enroll learners, add participants	User Screen		1	Faculty
Clarity	C4.4	To assign a learner to a simulation, first add a user.	User Screen		1	No Facilitator Present
Clarity	C4.4	Participants felt hesitent to add in items, despite having them in front of them, one participant explained this but also went on to say she didn't feel like she would break anything if she selected the wrong thing.	Admin Screen	123	2	Faculty

Clarity	C4.5	The items in the menus are not clear to users	Add a Simulation		1	No Facilitator Present
Clarity	C4.5	Active is in red, which signifies stop, and would change to green if defined by color.	Admin Screen		1	No Facilitator Present
Clarity	C4.5	Icons may be helpful in clueing in the user towards an action	All		2	All
Clarity	C4.5	users are unaware that they can interact with the KV, making it difficult to use	Knowledge Visualization		1	No Facilitator Present
Clarity	C4.5	Participants found it hard to find the home area	All	12345	2	All
Clarity	C4.6	The _New Simulation_ puts it in the center of the menu and is confusing	Add a Simulation		1	No Facilitator Present
Clarity	C4.7	Archive this simulation is long	Admin Screen		1	Faculty
Clarity	C4.7	Navigation menu is hard to understand	All		2	All
Clarity	C4.7	Start simulation button has 5 words, and 2 sentences	Simulation Question		1	All
Clarity	C4.7	Add New User can be shortened	user Screen		1	Faculty
Clarity	C4.7	Home link is confusing to participants	All	12345	2	All
Quality	P1.1	No directions are provided on how to add a simulation or that items must match, which choice is preferred, etc.	Add a Simulation	13	1	Faculty

Quality	P1.1	Participant 2 was asked to add a simulation, but a previous simulation was already in there, and she then edited a simulation rather than adding a simulation.	Admin Screen	2	2	Faculty
Quality	P1.1	Limited directions on how to add a simulation are also confusing. If a simulation is already there, a user may choose to just edit that simulation.	Admin Screen	123	2	Faculty
Quality	P1.1	There were brief directions to get started, but nothing that would explain the simulation to a learner before they entered the simulation. If there are multiple simulations, it may not be clear where to start depending on the naming of the simulations. Participants asked for confirmation before clicking on a simulation, further signaling confusion.	Home Screen	12345	1	All
Quality	P1.1	There are no details on what a learner should do after completing the simulation and reviewing the feedback.	Simulation	1234	1	All
Quality	P1.1	There are no details on what a learner should do after completing the simulation, or when they should end the chat.	Simulation	12345	1	No Facilitator Present
Quality	P1.1	The question references the Lesson 19 Problem set; this is not clear to anyone who is not following that curriculum	Simulation Question	3	1	No Facilitator Present
Quality	P1.1	Not immediately clear how to add a user; in context, help needed	User Screen	123	2	Faculty
Quality	P1.1	Participant attempted to speak to the screen	Simulation	2	2	Faculty

Quality	P1.2	It is unclear if having a certain number of probing questions means XYZ, etc.	Feedback	CHECK	1	No Facilitator Present
Quality	P1.2	There is no information or links to describe how a user will be evaluated or if they will be evaluated	Simulation Question	CHECK	2	No Facilitator Present
Quality	P1.3	There is no information available that guides the user by using graceful language and therefore this was not met on multiple areas of ACTS	Simulation Question - Feedback		2	No Facilitator Present
Quality	P1.4	There is not information that helps to orient the learners to the simulation, they do not know what to expect when they start the simulation. Participants asked clarification going into the simulation, asking if they just start to chat. The AI Bot helped to clarify this some vs. a human who may not have posted a message yet. One participant attempted to verbally speak to the screen because she was more familiar with using that type of simulation.	Simulation Question - Simulation	12345	1	No Facilitator Present
Quality	P1.4	Participant emphasized the need for clear expectations of what and how they were about teach, the participant was familiar with the stystem and previously rated preservice teachers preparing students to enter the simulation.	Simulation Question	1	2	No Facilitator Present

Quality	P1.5	There is no feedback on how to interpret the feedback, was their questioning method quality or not, it is unclear.	Feedback	345	1	All
Quality	P1.5	There is no information on how fast messages will come and how fast they must respond back, 1 participant who used the human-human expressed anxiety because she was not sure if the learner understood what she said and was typing back or confused and didn't know how to answer.	Simulation	12	1	No Facilitator Present
Quality	P1.5	There is no information that the learner is in a safe environment of if their teaching is private (using this for research will alter obviously alter the privacy aspect).	Simulation Question - Simulation		2	No Facilitator Present
Quality	P1.6	There is no information pertaining to how serious a learner should take the simulation, how fast they must respond, in turn, they also are not aware if they should refrain from looking at their phone and being cognitively present.	Simulation		1	No Facilitator Present
Quality	P1.7	there is no help documentation	Full System		2	No Facilitator Present

Visual Knowledge	P2.1	All five participants did not recognize the knowledge visualization at the start of the simulation, two participants did not use the KV during the simulation, one was unable to use it because the controls were blocked by the chat. Teo participants used it but did not understand if the learner could see it.	Knowledge Visualization	12345	2	No Facilitator Present
Visual Knowledge	P2.2	The numbers and slices are not immediately aligned to the problem, one participant attempted to teach to the 6 slices, thinking that was what the student was thinking.	Knowledge Visualization	145	1	All
Visual Knowledge	P2.5	The sliding steppers are slightly unclear on how to use them without prior instructions, one participant tried to click on the slices to change if they were selected or not, likely due to a familiar interaction with another software.	knowledge Visualization	135	2	All
Encouragement	P3.1	The feedback has no context that is text-based that could come across as judgmental, some participants questioned the meaning of the feedback but no one expressed that they felt judged. The script is highly helpful and meet the other portion of this criteria, participants who did not understand the categories could makes sense of the script	Feedback	345	2	No Facilitator Present
Encouragement	P3.2	If written feedback is added to the UI, it should be written in the second person and seem personalized	Feedback	345	3	No Facilitator Present

Encouragement	P3.3	If written feedback is added it should be inclusive	Feedback	345	3	No Facilitator Present
Encouragement	P3.4	Validate and normalize the feedback	Feedback	345	3	No Facilitator Present
Encouragement	P3.5	Participants were not clear on how they could improve their learning based on the feedback provided.	Feedback	345	2	No Facilitator Present
Encouragement	P3.6	Feedback is appropriate if the learners are aware of the meaning	Feedback	345	1	No Facilitator Present
Encouragement	P3.7	One participant became frustrated working with the AI-based CA, and commented about it having no emotions.	Simulation	4	3	No Facilitator Present

Supportive	P4.1	No feedback is given, but there is a sense of positive and negative because it shows a mix of questions and other types of messaging. However, the all green bars appear to make it all positive.	Feedback	345	1	All
Supportive	P4.2	There was no way for a learner to see their progress from one simulation to the next in one screen, learners should not have to remember information from one screen to the next in order to compare.	Feedback	CHECK	3	No Facilitator Present
Supportive	P4.3	Feedback is not written out	Feedback		3	No Facilitator Present
Supportive	P4.4	No written feedback was provided to offer suggestion.	Feedback	345	2	No Facilitator Present
Affective	S1.1	Participant expressed that the AI-based CA had very little emotion	Simulation	4	3	No Facilitator Present

Affective	S1.2	The chatbot never laughed, joked, or teased	Simulation		3	All
Affective	S1.2	Participant expressed that the AI-based CA had very little emotion	Simulation	4	3	No Facilitator Present
Affective	S1.3	Participant expressed that the AI-based CA had very little emotion	Simulation	4	3	No Facilitator Present
Affective	S1.4	The chatbot never made any emojis	Simulation		3	All
Affective	S1.7	Empathy was often followed by a leading statement, changing the conversation to the bot teaching. For example, after saying I made a mistake, the bot responded "That's okay! We can figure it out together. How do we find out how much Mrs. Jones gave to Gary?"	Simulation		3	No Facilitator Present
Interactive	S2.6	Participant felt backed into a corner and could not change the route of the AI-CA	Simulation	4	3	No Facilitator Present
Cohesive	S3.2	If the instructor does not provide a name, the bot defaults to saying only "Mrs." which can be considered biased, especially since the majority of preservice teachers are likely not married women.	Simulation	5	3	All
Cohesive	S3.5	The bot can easily switch to the teacher, especially if the user continues to say "ok."	Simulation		3	No Facilitator Present
Cohesive	S3.5	Participant felt backed into a corner and could not change the route of the AI-CA	Simulation	4	3	No Facilitator Present

Cohesive	S3.5	Particpant gave up because the AI-CA was not responding accordingly	Simulation	4	3	No Facilitator Present
Cohesive	S3.6	Participant felt backed into a corner and could not change the route of the AI-CA	Simulation	4	3	No Facilitator Present
Cohesive	S3.7	The bot was working on a problem about fractions but seemed to only talk about slices of pizza; the participant felt backed into a corner at this point and abruptly ended the chat.	Simulation	4	3	No Facilitator Present
Cohesive	S3.7	The bot did a good job at following past conversations. I randomly mentioned that I needed to take my dog out. Later, I said, “Guess what my dog needs to do again?” and the bot responded, “Your dog needs to go out again?”	Simulation			All
Cohesive	S3.7	Particpant gave up because the AI-CA was not responding accordingly	Simulation	4	3	No Facilitator Present
Adaptable	T1.1	The chat covers the knowledge visualization when the screen is not properly sized	Knowledge Visualization		2	All
Adaptable	T1.2	Update Simulation Settings button and cancel button do not have proper contrast	Add a Simulation		1	All
Adaptable	T1.2	Update Simulation Settings button and cancel button do not have proper contrast; the green used is very light and not compliant, but the border is, but would recommend increasing the border or increasing the value of the green.	Admin Screen		1	All

Adaptable	T1.2	Blue links do not meet WCAG standards, Blue on white is at 3.98:1 and the blue on green is at 3.20:1 must be 4:5	All		1	All
Adaptable	T1.2	Orange color is not compliant color is a 2.53 and needs to be at a 4.5 for contrast	Knowledge Visualization		1	All
Adaptable	T1.3	Participants could not clearly see where the home button was	All	12345	1	All
Adaptable	T1.4	The choices for Active or Archived uses a checkbox	Add a Simulation		1	Faculty
Adaptable	T1.4	There is no defined navigation	All	12345	1	All
Adaptable	T1.4	The stepper slider is not clear to users.	Knowledge Visualization	12345	2	All
Adaptable	T1.4	If chatting with a human and they do not immediately message, it makes it unclear for the user to know what to do first	Simulation	123	2	All
Adaptable	T1.4	After reading the feedback it is unclear what to do, there is no way to exit the system	Simulation	12345	1	All
Adaptable	T1.4	Participant tried to click on the pizza instead of using the sliders	Simulation	1	2	All
Adaptable	T1.4	Participants mentioned having a white board would be easier or was what they were used to using	Simulation	12345	3	No Facilitator Present

Adaptable	T1.5	There is no logo at the top or a home icon to clue a user to go home, even with the word home written out, participants did not know how to go back to the main screen	Feedback	12345	1	All
Adaptable	T1.5	The knowledge visualization becomes hidden when the screen is narrow.	Knowledge Visualization	1	2	All
Adaptable	T1.5	In human-human chats there is no way to see that the other learner is typing, one participant said it did not operate like Apple's iMessage	Simulation	12	2	All
Adaptable	T1.6	The slider can not be operated by the computer, this shows as an accessibility problem in Firefox developer tools.	Knowledge Visualization		3	For users who would need to have keyboard only access to the simulation.
Adaptable	T1.7	The name of the website and home button do not appear as links because they look like normal non clickable fonts	All	12345	1	All
Adaptable	T1.7	Links rely on color to know they are links	All		1	All

Dependab le	T2.1	It may not be apparent how to know why a simulation is not showing in the list because it is archived	All		1	Faculty
Dependab le	T2.1	There are no links to learn more about the feedback.	Feedback	345	2	All
Dependab le	T2.1	Participant stated that when she made an error adding a user the screen did not jump down or highlight where the error was	Add a Simulation	3	2	Faculty
Dependab le	T2.2	When clicking "Add a new simulation" it is expected to go a new screen where a simulation is added in.	Admin Screen		2	Faculty
Dependab le	T2.2	There is no way to exit a simulation to start over.	Simulation	4	1	All
Dependab le	T2.2	When adding users, if a user navigates away before hitting save there is no warning that changes will be lost, and the user has to go into the simulation or back to the users to try and figure out the problem.	User Screen	123	2	Faculty
Dependab le	T2.2	Participants were not alerted that the KV was updated, a learner who is aware that they need to click on that may not know that it changed	Simulation	12345	3	All
Dependab le	T2.3	There is no undo function if you add a simulation twice, you have to archive and then delete	Admin Screen		2	Faculty
Dependab le	T2.3	There is no way to undo a sent message in the simulation.	Simulation	4	3	All

Dependab le	T2.4	There is no way to know if user updated the image, with the exception of one user who was clued and prompted by the researcher, no user knew to click on view image.	Simulation	12345	2	All
Dependab le	T2.4	The end chat link is slow and makes it unclear if it is working.	Simulation	12345	2	All
Dependab le	T2.4	Adding a user error, a participant added a username with a space and did not immediately see the warning message.	user Screen	3	2	Faculty
Dependab le	T2.4	When selecting “save changes” on the user menu nothing appears to happen.	user Screen	123	2	Faculty
Dependab le	T2.5	The chat during human-human feels slower waiting for the person to type, possibly because you can't see they are typing.	Simulation	12	2	All
Dependab le	T2.5	The end chat link is slow and makes it unclear if it is working.	Simulation	12345	2	All
Dependab le	T2.6	There is not verification that messages went through, users are unsure if they sent.	Simulation	12	2	All

Dependable	T2.6	Participants were unclear if someone was tryping, thinking, or if they saw their message	Simulation	123	3	No Facilitator Present
Dependable	T2.7	the home button is not clear and hidden by just having text next to the name, however, is in a familiar spot.	All	12345	1	All
Dependable	T2.7	It is unclear if after reading the feedback how a user should exit the simulation	Simulation	12345	1	All
Dependable	T2.7	There is no way to exit a simulation to start over. Some users did not see where to exit right away.	Simulation	345	1	All
Intuitive	T3.1	Some links are black others blue, some are bolded others are the same color.	All	123	1	All
Intuitive	T3.1	The all green bars appear to make it all positive feedback	Feedback	345	1	All
Intuitive	T3.2	There is no navigation menu or common ways of navigating (hamburger menu)	All	12345	2	All

Intuitive	T3.2	The navigation currently changes from screen to screen so there is no familiar place to look	All	12345	2	All
Intuitive	T3.3	It is not clear if you should add a simulation or users first, there are no steps	Admin Screen	123	1	Faculty
Intuitive	T3.3	A clear understanding of simulation process, users will enter simulation, see a question, and then receive feedback.	Simulation		1	Faculty
Intuitive	T3.4	There is no help in context to get help if needed.	All	12345	1	All
Intuitive	T3.4	There is no additional information for feedback to make it clear as to the meaning	Feedback	12345	1	No Facilitator Present
Intuitive	T3.4	The sliders do not have labels	Knowledge Visualization	12345	1	All
Intuitive	T3.4	There is not information that helps to orient the learners to the simulation, they do not know what to expect when they start the simulation.	Simulation Question - Simulation	12345	1	No Facilitator Present
Intuitive	T3.4	Screen covered the KV first time Particpant went though	Simulation	1	2	No Facilitator Present

Intuitive	T3.5	The choices for adding a new simulation are overwhelming to many users	Add a Simulation	123	1	Faculty
Intuitive	T3.5	Having the Show KV in every text is overwhelming and it could cause a learner to need to click on it every time a new chat message is sent	Simulation	12345	3	All
Intuitive	T3.6	User/simulation menus are separate; you can not click to add users from the simulation created, so users must remember what they just called the simulation. Participants only had one simulation to use so they likely didn't run into this problem.	Add a Simulation		2	Faculty
Intuitive	T3.6	The users have no way to go back to the full simulation description in the simulation or a way to see it again	Simulation	1	2	No Facilitato r Present
Intuitive	T3.7	There are no icons used throughout the system.	All		1	All
Predictabl e	T4.1	It is unclear if items need to match from the menu; in turn, if the simulation does not work, it may take a number of trial and error attempts to make it work.	Add a Simulation	123	1	Faculty
Predictabl e	T4.1	If items do not match, the test does not seem to alert the user to the problem.	Add a Simulation	123	1	Faculty
Predictabl e	T4.1	I expected to be taken to an add simulation screen instead of adding one to the list. The list order is confusing, as it is alphabetical instead of by date, so a new sim will show up in the middle of the list.	Admin Screen		2	Faculty
Predictabl e	T4.1	Save changes does not appear to do anything, possibly adding a confirmation message or changes saved.	user Screen	123	2	Faculty
Predictabl e	T4.1	Users tried to click on the pizza, and were surprised it wasn't working.	Simulation	1	3	All

Predictable	T4.2	There is no help available in context	All	12345	2	All
Predictable	T4.3	There is no help available in context, but when added should be actionful	All	12345	1	All
Predictable	T4.4	There is no help available in context, but help should be in plain languate and tips	All	12345	2	All
Predictable	T4.5	There is no help available in context, but should be brief if added	All	12345	2	All
Predictable	T4.6	There is no help available in context, but if added it should be describe what the feature does and how to use it.	All	12345	2	All
Predictable	T4.7	users are unaware if the other person is typing in human-human interaction	Simulation	123	2	All
Predictable	T4.7	feedback took a long time load after ending simulation	Simulation	12345	1	All

APPENDIX C

Questionnaire

ACTS Research User Experience Survey

To understand the usability of ACTS, we are asking users to complete a walkthrough of the system and complete the following feedback survey. You may review the survey before using the system and can fill it out as you progress through the system or complete after completing the system walkthrough.

Please rate the following questions from Strongly Agree to Strongly Disagree.

	Strongly Agree			Neutral			Strongly Disagree
1. The AI chatbot generated text that demonstrates emotional expressions.	○	○	○	○	○	○	○
2. There was a high level of interactivity between myself and the AI chatbot.	○	○	○	○	○	○	○
3. The AI chatbot demonstrated a genuine interest in the conversation.	○	○	○	○	○	○	○
4. The AI chatbot demonstrated clear, accurate, and contextually appropriate communication.	○	○	○	○	○	○	○
5. The system demonstrated clear directions, rules, and evaluation criteria.	○	○	○	○	○	○	○
6. The system demonstrated instructional images that are clear and easy to use.	○	○	○	○	○	○	○
7. The feedback was encouraging and non-judgmental.	○	○	○	○	○	○	○
8. The feedback was appropriate and supported my learning.	○	○	○	○	○	○	○
9. I knew how to navigate and use the system.	○	○	○	○	○	○	○
10. I controlled the system and could avoid or recover from errors.	○	○	○	○	○	○	○
11. I was able to use the system without any training or help.	○	○	○	○	○	○	○
12. I felt confident in predicting what would occur before interacting with any element in the system.	○	○	○	○	○	○	○
13. The navigation, instructions, and error messages are consistent throughout the system.	○	○	○	○	○	○	○
14. The overall appearance of the system is aesthetically pleasing.	○	○	○	○	○	○	○
15. The voice and tone are consistent and inclusive throughout the system.	○	○	○	○	○	○	○
16. The content is clear and easy to understand throughout the system.	○	○	○	○	○	○	○