The Impact of the Narrator's Gender on Multimedia Learning

by

Maryssa Lefevre

A Thesis Presented in Partial Fulfillment
of the Requirements for the Degree
Master of Science

Approved November 2019 by the
Graduate Supervisory Committee:

Scotty Craig, Chair
Rod Roscoe
Russ Branaghan

ARIZONA STATE UNIVERSITY

December 2019

ABSTRACT

The utilization of multimedia videos has increasingly become more popular, especially in the field of education. In order to facilitate learning it is important to create a natural interaction between the learner and the on-screen material. This study focused on improving the facilitation of the information within a multimedia learning video by focusing on the gender and quality of computer-synthesized voices. Using a randomized pretest - posttest design the study looked at how the gender of the narrator affected a person's ability to learn and implement a new task. Narration was performed by a male and female, classic and modern synthesized voices to determine if there were gender effects across both generations of voices. The participants' learned knowledge was assessed through a multiple-choice assessment and a word to image matching transfer assessment. Results showed no significant results. Future studies should consider a more reliable knowledge assessment and utilize and larger sample size.

TABLE OF CONTENTS

LIST OF TABLES

Table Page

CHAPTER 1

INTRODUCTION

Society has reached a faster rate of technologic advancements and a higher degree of innovation than ever before, making this is a very exciting time for development. The most recent generation nicknamed ‘Generation Z’ by the media, is the first generation that will have technology readily available at their fingertips just shortly after birth. The incorporation of technology into every stage of life opens the door for new markets within the industry. An emerging market of interest is e-Learning technologies. Clark and Mayer (2016) define e-learning as instruction delivered through an electronic device that is intended to promote learning amongst users. There are several ways to implement e-learning, the focus of this paper will be on the use of multimedia instruction. Multimedia instruction is defined as a form of teaching that presents information in two or more perceptual modalities- visual, auditory, taste, or touch (Mayer, 2017). Combining different forms of information helps distribute the information processing across multiple senses rather than overloading a single channel. For the purposes of this study, the visual and auditory modalities with be utilized with emphasis on the auditory modality.

Current research focuses mainly on utilizing a pedagogical agent, which is defined as a computer-generated character that is utilized to assist the user in learning the material at hand (Craig, Gholson, & Driscoll, 2002). One problem that arises is not all multimedia environments will have the option to utilize a pedagogical agent. This makes it important to solely test the effects of different features in the narrator’s voice within the video. Specifically, this study will be focusing on how the gender of the narrator effects the learner’s ability to grasp and implements a new concept. After reviewing current

research, and a recent metanalysis, it has been shown that the current research has not properly investigated how the voice influences learning (Schroeder et al. 2013). The current research does not properly represent the male and female genders for the narrator/pedagogical agent or with participants. This study sought to help bridge the gap of voice and gender research to equally represent the two genders for both the narrator and well as the participants. This study will also test the voice effect determined by Mayer, Sabko and Mautone (2003). This study attempted to contribute to the research on how to properly assign a narrator to a user to increase the level of learning they achieve through a natural interaction.

With multimedia learning becoming increasingly utilized across several different platforms, it is important to ensure that it is designed in a manner that naturally stimulates learning for the users. With the minimal research on narration and gender, this study was intended to bridge the research gap to answer the question of how the quality and gender of a computer-generated narrator voice impacts the learner's perception of the narrator and their ability to understand and apply new information. The goal was to perform a replication study of the voice effect while also analyzing the effects of gender.

CHAPTER 2

LITERATURE REVIEW

2.1 MULTIMEDIA LEARNING

By analyzing previous research, Mayer and colleagues have determined many goals of multimedia learning. There are two that will be focused on for this study: (a) to design multimedia instructional messages in ways that reduce cognitive load, and (b) to increase the learner's interest (Mayer, Fennell, Farmer, & Campbell, 2004). Mayer has identified the multimedia effect and the personalization effect as two important concepts that aid in meeting these goals. The multimedia effect refers to the theory that students learn more in depth through multimedia with a presentation of words and pictures rather than with words alone (Mayer, 2003 a). This is utilized when a presentation uses on-screen text to highlight important information and a narrator to do a more in-depth explanation of the information. The learner will understand the information better if there are key points shown on the screen while a voice is reading the text aloud and elaborating on them. This activates the learners two channels of information processing, auditory and visual, to interpret the given information. The personalization effect builds on top of the multimedia effect to state that students learn better from multimedia instruction when words are presented in conversational style rather than a formal descriptive style (Mayer, 2003 a). Participants learn better with more relaxed social settings than in formal social settings. To incorporate this concept, it is better to use simpler language. These concepts make it apparent that the method in which information is delivered in is just as important as the information itself.

According to Reeves and Nass (1996) multimedia experiences that people participate in mimic real-life experiences they encounter in their everyday lives. They refer to this idea as the media equation. People are social creatures and seek social aspects in any interaction they encounter. This is easily observable when studying interactions people have with each other and animals. Recently, this has also been shown to be true of the interactions people have with technology. To demonstrate the media equation, Reeves and Nass (1996) conducted a study in which participants were asked to rate the performance of the computer after doing an activity. They found that the participants gave more positive reports of the computer's performance if the computer asked the participants questions of itself versus if a different computer asked the participants the same performance questions regarding the task facilitating computer. This shows that people follow the "social rule", in which they are polite to those who ask questions of themselves, with people as well as with computers (Reeves & Nass, 1996). People interact with computers just as they do people, causing a need for a natural experience for the user. A narrator's voice in multimedia instruction is a feature that can help create a natural interaction the user will have with the video.

Mayer and colleagues examined previous research and developed the social agency theory. This theory states that social cues that are in a multimedia video can activate social conversation schema within learners (Mayer, Sobko, & Mautone, 2003). When these social cues are activated between a multimedia video and a learner, there are basic human-to-human communication patterns that are present, and the learner acts as if they are in a conversation with the video. This sense of conversation leads the learner to engage in deep cognitive processing because they are interested in the information being

presented. Multimedia learning product designers want the learners to reach this state because this is how learners make sense of this information by combining it with information in their long-term memory (Mayer, Sobko, & Mautone, 2003). Mayer and colleague tested this theory by studying the effect a speaker's accent has within a multimedia video. In this study they found participants who watched a narrated video with a native-English speaker (standard-accent group) performed better on problem solving tests than those who watched a video with Russian accent (foreign-accent group) narration. The data revealed the participants were able to relate better with the standard-accent voice, and therefore engaged at a deeper level. This supports the idea of the media equation of mimicking a real-life situation (Reeves & Nass, 1996).

2. 2 NARRATION

When using a narrator within multimedia, it is important to look at variances in the voice. During the early 2000s there were two studies conducted that looked at the influence of a computer-generated voice versus a human voice (Atkinson, Mayer, & Merrill, 2005; Mayer, Sabko, & Mautone, 2003). In both studies it was found that the human voice was more effective in facilitating learning than the computer-generated voice, this is referred to as the voice effect.

Technology has advanced drastically over the past 15 years, calling for a replication study. Craig and Schroeder (2017; 2018) sought to reanalyze this effect by conducting two studies that examined text-to-speech generated voices. In both studies the researchers compared a human voice, a classic synthesized voice (Microsoft speech engine) and modern synthesized voice (Neospeech voice engine). However, in the first

study they used an on-screen character and in the second study they isolated the voices. When looking at performance the results produced from the studies revealed that participants who viewed the presentation with a human voice and a modern synthesized voice had better transfer test results than those with the classic synthesized voice (Craig & Schroder, 2017; Craig & Schroder, 2018). This shows that technology truly has come a long way. Learners may have better knowledge acquisition with the modern synthesized voice over the human voice because it produces a more consistently pitched and comprehensible narration than the human voice. Although these findings are interesting, they fail to assess the effects the gender of the virtual human has on the learner information processing system.

2.3 GENDER

In one study, the researchers sought to analyze the impact of peer-like and expert-like stereotypes on a learners' perceptions of the computer-generated agent, task-related attitudes, and learning achievement within a task (Tze Wei, Su-Mae, & Jayothisa 2013). Participants were randomly assigned to a computer lab that used a peer-like agent or an expert-like agent. They used female computer animated agents that varied on their appearance and their voice inflections. The results did not reveal a significant main effect on the learner's achievement. However, the results from this study did identify three main effects on the learners' attitude towards the agent: learners perceived the peer-like agent to be more friendly than the expert-like agent, female learners perceived agents as more knowledgeable than the male learners did, and female learners were more trusting of the information than male learners were. An additional t-test showed the female participants

were more trusting of the expert-like agent than to the peer-like agent. Overall, this supports the social agency theory of giving higher ratings to the peer-like agent based on relatability (Mayer, 2003 b). The findings are ambiguous for males because the study did not include male agents. This leaves the question of whether male learners would have been more receptive to the information if it was presented by a male agent based on relatability. Until male agents are included, one cannot explicitly say this to be true or false.

In a similar study Baylor and Kim performed two separate studies to look at agent realism and the agent's role in regard to facilitating learning (2004). They were interested in learning how these two factors are affected by gender and ethnicity**.** In the first study they utilized a 2x2x2 design to look at agent gender (Male vs. Female), agent ethnicity (Caucasian vs. African-American), and agent realism (realistic vs. cartoon-like). The results for this study revealed a significant main effect for agent gender on self-regulation and self-efficacy in which the presence of the male agent led to increased self-efficacy. When further analyzing the results, the researchers found that the participants perceived the male agents as more intelligent, interesting and useful than the female agents.

In the second study, the researchers utilized a 2x2x3 design in which they looked at agent gender (Male vs. Female), agent ethnicity (Caucasian vs. African American), and agent role (expert vs. motivator vs. mentor). The results for this study also revealed a significant main effect for agent gender on self-efficacy. Students who worked with the female agents showed higher self-efficacy beliefs than students who worked with the male agents. Contradictory to these results, further analysis shows that participants still perceived the male narrators to be more knowledgeable. Participants who were assigned

to the motivator and mentor agents showed higher self-efficacy beliefs than students who were assigned to an expert agent (Baylor & Kim, 2004). These results show that although men are perceived to be more knowledgeable, they inhibit a learner's level of self-efficacy. This is an important factor to consider when choosing gender of an agent or a narrator because this shows that the genders promote different levels of motivation to the user.

In a study performed by Rosenberg and colleagues, they focused on the race and gender of computer-simulated agent. The researchers hypothesized the participants would be more positively influenced by computer-based models that matched them in terms of their race and gender (Rosenberg, Plant, Droerr, & Baylor, 2010). Specifically, they looked at how race and gender their interest, self-efficacy, and stereotypes about engineering. Their study consisted of four test groups: white female, white male, black female, black male. The participants were solely female participants, African American and White. The results showed the majority of the effects were based on race. Participants responded positively to engineering if they interacted with an agent that matched their race. In regard to gender the Black participants responded better to the female Black agent and the male White agent. This shows they relate most to the black female agent but view the White male agent more as an expert (Chaiken, 1994; Debono, 1988; Hovland, 1951). However, the White participants only showed significant results for preference over the white agents; showing insignificant results for gender preference. For the White participants the two most influential agents were the White female and the White male agents. This could show that White females relate more to being White than to being female (Rosenberg et al., 2010). The fact that the participants in this study were

all female limits the generalizability of the results when considering how this would affect the male population.

Although it was reported that the main effects in Rosenberg and colleagues' (2010) study were due to race, it is still important to consider gender stereotypes when looking at the results. Both the White and Black participants gave higher ratings to the White male narrator than to the White female narrator. Corresponding with the social agency theory, the researchers hypothesized that the participants would give higher rating to the female agents in both races because this was perceived as the more relatable trait (Mayer, 2003 b; Rosenberg et al. 2010). However, society places a large importance on gender roles. This transpires across many situations that deal with family, occupational, and social roles (Rice & Barth, 2016; Hollingshead & Fraidin, 2003). Gender roles are used to gather information about a person's knowledge and skill level (Hollingshead & Fraidin, 2003). When looking at occupation, women have been shown to choose careers that are aligned with "family values" whereas men choose careers that are focused towards "individualistic goals" (Rice & Barth, 2016). This creates a ratio difference within different fields, and in retrospect creates a gender stereotypes within these fields. Rosenberg's et al. (2010) study was looking at women's perception of the Engineering field, which is primarily dominated by men and is therefore acknowledged as a masculine career path. Gender stereotypes could be a factor in explaining why the White male narrator had a higher rating than the White female narrator due to the fact that males are seen to be more knowledgeable within the field based on male to female employment ratio.

CHAPTER 3

OVERVIEW

Previous studies have revealed that learners adhere better to agents in which they can relate to (Mayer, 2003 b). This has been shown when analyzing voice accent, expert-like vs peer-like agents, and race within multimedia learning (Mayer, 2003 b; Rosenberg et al., 2010; Tze Wei et al., 2013). However, there has also been evidence that people tend to view men as more knowledgeable and expert-like (Rosenberg et al., 2010). There are two gaps within previous research: 1) There is no research that looks at only the voice of a narrator to analyze gender and 2) there has not been an equal representation of either male and female agents or learners to properly analyze the effects gender has on learning. The current study aims to fill these gaps by isolating the voice of the narrators and looking at the effects the gender and voice quality of a computer-generated narrator has on the participants' ability to grasp and implement a new concept. The prediction is that based on social agency theory, when participants are shown a multimedia video with a narrator that they can relate to they will be more interested in the information, have a deeper understanding, and therefore have better application results of the provided information in a given task. The current study looked at gender and voice clarity as the relatable factors by utilizing a male and female version of an older and newer generation of computer-synthesized voices. There were three sets of hypotheses for this study:

Hypothesis 1: Voice effect

Hypothesis 1A: When participants are shown a multimedia learning video narrated with a modern-synthesized voice, they will give the narrator a higher likeability rating.

Hypothesis 1B: When participants are shown a multimedia learning video narrated with a modern-synthesized voice, they will learn more and score higher on their application tests.

Hypothesis 1C: Participants who give a narrator with a modern-synthesized voice a higher likeability rating will perform better on their application tests.

Hypothesis 2: Gender Match Effect

Hypothesis 2A: When participants are shown a multimedia learning video with a narrator of the same gender as themselves, they will give the narrator a higher likeability rating.

Hypothesis 2B: When participants are shown a multimedia video with a narrator of the same gender as themselves, they will learn more and score higher on their application tests.

Hypothesis 2C: Participants who give a narrator of the same gender a higher likeability rating will perform better on their application tests.

Hypothesis 3: Mismatch Gender Effects

Hypothesis 3A: Female participants who are shown a multimedia learning video with a male narrator will be less affected by gender roles and will give the narrator a higher likeability rating than the male participants who are shown a video with a female narrator.

Hypothesis 3B: Female participants who are shown a multimedia learning video with a male narrator will be less affected by gender roles and will

have higher application test results than the male participants who are shown a video with a female narrator.

Hypothesis 3C: Participants who give a narrator a higher likeability rating will perform better on their application tests.

The first prediction should coincide with Craig and Schroder's findings in that participants will perform better in the modern synthesized voices over the classic synthesized voices (2017; 2018). However, it will be interesting to see if there is a similar effect across the two different generations of computer-generated voices based on the gender matchings of the participants and the narrator. The results from the following two predictions will help contribute to the scarce research conducted on gender effects in multimedia environments.

CHAPTER 4

METHODS

This study implemented a randomized treatments design with a utilization of a pretest (Shadish, Cook, & Campbell, 2002). This was done through the randomization feature within Qualtrics survey system. The online Qualtrics systems was the tool used to implement all stages of the study after recruitment. Participants were randomized into one of four treatments: Male classic synthesized voice (Microsoft speech engine), Female classic synthesized voice (Microsoft speech engine), Male modern synthesized voice (NeoSpeech voice engine), and Female modern synthesized voice (NeoSpeech voice engine). Assessments for the study included an open-ended pretest question in which the participant was instructed to discuss their understanding of learnability before watching an instructional video, and two posttest knowledge assessments that include a 20 question multiple-choice test and three application transfer questions. A questionnaire was also included to measure the participant's perceptions of the narrator and their learning experience.

4.1 PARTICIPANTS

A convenient source of participants was recruited from Arizona State University's subject pool. This pool consists of intro to psychology students and intro to human systems engineering students who are required to participate in student research studies in order to receive credit for their class. Participants were required to speak English because the content of the video was only be presented in English. Each participant received .50 research credit as compensation for their participant in this study. Each participant was

randomly assigned to one of four voice groups. Due to the limited participants available there was no way to ensure an even amount of male and female participants would be recruited for the study. In total there were n = 52 male respondents, n = 16 female respondents, and n = 1 non-binary respondents. For purposes of this study the non-binary participant was excluded. Two other participants were excluded from the final data analysis because they answered an attention test question inaccurately. This question was randomly placed in the survey to asses if they were paying attention to the questions or simply marking answers. The final data set count consisted of n = 50 male participants and n = 16 female participants, equating to an overall of n = 66 participants. The heavy gender imbalance created an uneven participant gender representation, as seen in previous studies. The current study could not draw accurate conclusions in regard to participant gender.

There were demographic questions incorporated into the survey to evaluate the range of participants. Since Race/Ethnicity questions was an open-ended question, some of the answers were combined to account for typos and similar Ethnicities; for example, “Caucasian” was combined with “White”, and “East Asian” was combined with “Asian”. The majority of the participants were born in the year 2000 or later (n = 52 or 78%). The majority of participants describe themselves as being White (n = 37 or 56%), with the two next largest groups being Hispanics and Asians (n = 9 or 13%). The majority of the participants answered they have only received a high school diploma or equivalent (n = 38 or 57%). Based on these numbers, the majority of the students are first- or second-year white male college students. There were no significant differences in the participants race and gender across the four test groups.

4. 2 LEARNING MATERIALS

The purpose of this study was to isolate the voice of a narrator to test the effects the gender of a narrator's voice has on the learner's perception of the narrator and how it correlates to their ability to process and implement the information. To do this, four versions of a multimedia video were produced using VideoScribe. The material within the videos consists of information on how to design systems for learnability. Components of multimedia videos previously mentioned to stimulate higher levels of learning have been incorporated into the video (Mayer, 2003a; Mayer, 2003b). Refer to Appendix A for a screenshot of an overview of the final design of the video. The video is a simple design that is free of clutter, has images and examples, and has on screen text that will correspond with the narration (Mayer, 2003a; Mayer & Moreno, 2003). The on-screen material and the voice narration were consistent across all four videos to create a control within the study.

4.3 STUDY DESIGN

A 2x2x2 between-subjects study with a pre-test and a post-test was conducted. There were two levels to three independent variables. The single subject independent variable was the gender of the participants- male or female. The two experimental independent variables were the gender of the narrator's voice within the video- male or female, and the text-to-speech software used. The text-to-speech softwares that were utilized are the classic Microsoft speech engine and the modern NeoSpeech engine (Atkinson, Mayer & Merrill, 2005; Craig & Schroder, 2017). The two experimental variables created four test groups:

- Male modern speech engine, Paul (n = 17).
- Male classic speech engine, Sam (n = 18).
- Female modern speech engine, Kate (n = 17).
- Female classic speech engine, Mary (n = 14).

4.4 ASSESMENTS

In order to assess the participants' learned knowledge, it is important to assess the amount of information they know about the topic of learnability before given the instructional video. To do so, a pre-test what implemented before the participants were shown the video. The pre-test stated "**Learnability** is an important concept related to usability of a product. In your own words, what is **learnability**? Try to define the concept and give examples of "good learnability" and "poor learnability" in regard to product development. You will have 2 minutes to complete this question." The question was timed to ensure that participants could not perform an internet search to answer the question. This qualitative question was scored quantitatively based on key points discussed in the video. For each key point the participant mentioned in their response they received one point. Participants could receive up to five points: 1 point if they defined learnability as the ease of use of a product, 1 point if they stated the purpose or learnability, 1 point if they mentioned user research, 1 point if they defined any design principles, and 1 point if they mentioned learnability testing. All means for pre-test scores are presented in **Table 3** and **Table 4**. Scores were analyzed for reliability by running a Kappa analysis. This was done with the assistance of a second researcher. Both researchers coded ten of the responses separately. The codes were then combined into a

table to compare all five codes from each response to create one overall Kappa score. The Kappa analysis received an overall score of 0.64, meaning the scale is moderately reliable.

The two dependent variables are the participants' ability to apply the information as well as their perception of the narrator. They were given two different assessments. The first assessment consisted of questions about their perception of the narrator within the video. The Agent Persona Instrument (API), updated by Schroder, Romine, and Craig (2017), was used to assess the participants' perceptions of the narrator's persona by utilizing 29 specific questions and a 5-point Likert scale (1 = strongly disagree, 3 = neither agree nor disagree, 5 = strongly agree). An additional question was included to test whether the participants were paying attention or just marking answers. Two participants were removed from the analysis for not choosing "Strongly Agree" for the attention test question. The API questionnaire is displayed below in Appendix C. This model focuses on measuring how the participants views how well the narrator was able to facilitate learning, the narrators' level of credibility, how human-like they are, and how well they engaged the participant. Baylor and Ryu (2003) provide a clear definition of these measurements:

Facilitates Learning: how well the agent enables learning and reflection.

Credibility: the value of information provided by the agent.

Human-Like: these questions asses the genuineness of the narrator's nonverbal communication in regard to personality and emotional expression.

Engaging: how well the agent influences the participant to learn the information.

To create the test groups, a basic randomized design was used (Shadish, 2002). As previously mentioned, participants were assigned to one of four voice groups. The groups were mixed with a total of n= 50 males n = 16 females, equating to an overall of 66 participants within the study. To ensure the test groups are even in size, Qualtrics randomized option was used to evenly assign participants into a group.

The second assessment was used to test their abilities to apply the information. A knowledge assessment was created with questions in the form of multiple choice and word to image matching. The questions test their abilities to recall, comprehend, and transfer information. The recall portion had ten multiple choice questions that asked the participant to remember factual information. An example of this is "How do you measure learnability?" The comprehension questions also had ten multiple choice questions; however, these questions go a step further and ask the participant to infer which concept matches the given scenario. An example of this is "The video provided examples of a learnable and a non-learnable webpage design. What were the problematic features in the non-learnable design?" For the transfer portion, participants were asked three different questions about design principles. They were instructed to choose one of the three website wireframes presented that best represented the principle discussed from the video. A transfer question is displayed below in Appendix B. Having these three types of questions allowed the researchers to test for knowledge retention and applicability skills of the participants. Their results from this test were assessed quantitatively based on how many questions were answered correctly.

4.5 PROCEDURES

Participants were recruited through ASU's subject pool. The subject pool consisted of ASU students enrolled in PSY 101 or HSE 101. Students under the age of 18 were exempted from participating in the study. Those who volunteered to participate in the study were provided with a link to the survey created through Qualtrics survey website. Qualtrics was utilized to administer every stage of the study proceeding recruitment. Participants who described their gender as anything other than male or female were excluded from the final data analysis.

The instructional videos were produced using VideoScribe. The videos are a five to six-minute explanation of how to design systems for learnability. The time varied due to the older voices having a slower speed. The narration script was created by the researcher. The narration was made with a free online text-to-speech generator and NeoSpeech software. Camtasia was then used to record and edit the narration. The videos were then imbedded into the Qualtrics survey used to administer the study.

Participants used Qualtrics to sign the consent form and perform a sound check test to ensure they will be able to hear the video narration. Participants had to listen to a recording and type in the code provided. Participants were not able to move forward until they passed the sound check. Participants were then provided demographic questions and a pre-test question to answer. They were given instructions on how to perform the study as well as prompted to please pay attention as the information is presented in the video. Once the instructions were read, the participants were randomly placed into one of the four test groups determined by the video they were presented. After the video, they were given the API to assess their opinion of the narrator's relatability and level of knowledge.

Participants were then given the knowledge assessment to test their comprehension of the information presented in the video. They were allocated 30 minutes to complete the study to receive .50 research credits for their final class grades. The results from all tests were then analyzed individually and then compared to one another to see if there is a correlation between likeability of the narrator and knowledge acquisition.

CHAPTER 5

RESULTS

The current study originally had a randomized 2x2x2 pretest-posttest design. However, there was a mistake made during the data collection process which prevented a true random collection of the classic male voice condition. This led to uneven sample sizes across the conditions, which called for a Levene's test of homogeneity to be executed for all dependent measures to test for equal variances. The tests indicated that variances were not significantly different across all dependent measures, except the API subscale for Human-Likeness. This allowed for ANOVA analyses to be conducted with a standard significance level of $\alpha = .05$ for the majority of the measures. The participants' scores for the API Human-Like subscale received a Levene's test of homogeneity score of $p < 0.05$, which means the four test groups did not have equal error variances. To account for this, the scores for the Human-Like subscale were transformed into their ln scores. The purpose of the data transformation was to remove the assumption of normal group distribution. With the data transformation the participants' scores received a Levene's test of homogeneity score of $p = 0.096$. Mean scores for each condition and all dependent measures, including ln (Human-Like), are presented below in **Table 1**, **Table 2**, **Table 3**, and **Table 4**.

Table 1. Means and Standard Deviations for Male Participants' Ratings on the Subscales of the Agent Person Instrument Separated by Condition.

Voice Condition	N	Facilitates Learning *M* (SD)	Credibility *M* (SD)	ln (Human-Like) *M* (SD)	Engaging *M* (SD)
Male Modern	10	32.9 (10.071)	19.1 (4.886)	0.9841 (0.17251)	20.4 (6.552)
Male Classic	16	26.13 (7.136)	16.31 (2.6)	0.7587 (0.10897)	15 (4.899)
Female Modern	12	31.58 (5.368)	17 (2.374)	0.9423 (0.20140)	20.42 (5.664)
Female Classic	12	33.50 (11.453)	18.42 (3.988)	0.9203 (0.22856)	20.17 (9.054)

Note. Number of participants in each condition (N), mean sores (*M*), and standard deviations (SD)

Table 2. Means and Standard Deviations for Female Participants' Ratings on the Subscales of the Agent Person Instrument Separated by Condition.

Voice Condition	N	Facilitates Learning *M* (SD)	Credibility *M* (SD)	ln (Human-Like) *M* (SD)	Engaging *M* (SD)
Male Modern	7	31.86 (9.191)	17.57 (4.685)	0.9605 (0.25406)	24.29 (9.656)
Male Classic	2	26 (4.243)	15 (1.414)	0.7720 (0.10333)	15 (1.414)
Female Modern	5	31.6 (11.371)	20 (4.583)	0.8794 (0.15727)	17.8 (7.791)
Female Classic	2	26 (1.414)	18 (2.828)	0.9065 (0.29343)	17.5 (6.364)

Note. Number of participants in each condition (N), mean sores (*M*), and standard deviations (SD)

Table 3. Means and Standard Deviations for Male Participants' Scores on Pre-Test, Knowledge Assessment, and Transfer Assessment Separated by Condition.

Voice Condition	N	Pre-Test *M* (SD)	Knowledge Assessment *M* (SD)	Transfer Assessment *M* (SD)
Male Modern	10	0.70 (0.823)	13.50 (1.314)	2.2 (0.789)
Male Classic	16	0.88 (1.025)	13.92 (1.165)	2.25 (.577)
Female Modern	12	.92 (.996)	13.60 (1.506)	2.33 (.492)
Female Classic	12	1.5 (1.243)	13.44 (1.999)	1.75 (.866)

Note. Number of participants in each condition (N), mean sores (*M*), and standard deviations (SD)

Table 4. Means and Standard Deviations for Female Participants' Scores on Pre-Test, Knowledge Assessment, and Transfer Assessment Separated by Condition.

Voice Condition	N	Pre-Test *M* (SD)	Knowledge Assessment *M* (SD)	Transfer Assessment *M* (SD)
Male Modern	7	.86 (1.215)	13.14 (2.911)	2 (1.155)
Male Classic	2	1.5 (2.121)	13.50 (.707)	1.5 (0.707)
Female Modern	5	1.6 (1.140)	12 (2.828)	1.2 (.447)
Female Classic	2	1 (1.414)	14 (0)	1.5 (0.707)

Note. Number of participants in each condition (N), mean sores (*M*), and standard deviations (SD)

5.1 PERCEPTIONS

The first research question looked at the role the narrator's voice has on the learners' perception of the learning experience conducted by the narrator. For this question the study looked at how the genders of the narrator and participant and the clarity of the narrator's voice may play a role. The purpose of these analyses were to see if participants would give the narrators different ratings based on the social agency theory (Mayer, Sobko, & Mautone, 2003). ANOVA analyses were used to look at the

differences between the four different narration groups on the subscales of the API: Facilitates Learning, Credibility, ln (Huma-Like), and Engagement. Means and standard deviations of participants' ratings of their perceptions of the narrators are shown in **Table 1** and **Table 2**.

A 2x2x2 factorial ANOVA that looked at gender of participants, gender of the voices, and voice quality was conducted on participants' rating for how the narrator facilitates learning. The analysis did not indicate a significant three-way interaction between participant gender, voice gender, and voice quality (F (1, 58) = 0.551, p = 0.461, n_{p2} = .009). The analysis also did not indicate any significant two-way interactions between participant gender—voice gender (F(1, 58) = 0.309, p = 0.581, n_{p2} = 0.005), participant gender—voice quality (F (1, 58) = 0.337, p = 0.564, n_{p2} = 0.006), or voice gender—voice quality (F(1,58) = 0.620, p = 0.434, n_{p2} = 0.011). Similarly, there were no significant main effects for participants gender (F (1, 58) = 0.579, p = 0.450, n_{p2} = 0.10), voice gender (F(1, 58) = 0.261, p = 0.612, n_{p2} = .004), or voice quality (F (1,58) = 0.337, p = 0.564, n_{p2} = 0.005).

A 2x2x2 factorial ANOVA that looked at gender of participants, gender of the voices, and voice quality was conducted on participants' rating for how credible the narrator was. The analysis did not indicate a significant three-way interaction between participant gender, voice gender, and voice quality (F (1, 58) = 0.558, p = 0.466, n_{p2} = .010). The analysis also did not indicate any significant two-way interactions between participant gender—voice gender (F(1, 58) = 1.310, p = 0.257 n_{p2} = 0.022), participant gender—voice quality (F (1, 58) = 0.456, p = 0.502 n_{p2} = 0.008), or voice gender—voice quality (F(1,58) = 1.016 p = 0.318, n_{p2} = 0.017). Similarly, there were no significant main

effects for participants gender (F (1, 58) = 0.003, p = 0.957, n_{p2} = 0.00), voice gender (F(1, 58) = 1.314, p = 0.256, n_{p2} = .022), or voice quality (F (1,58) = 1.572, p = 0.215, n_{p2} = 0.026).

A 2x2x2 factorial ANOVA that looked at gender of participants, gender of the voices, and voice quality was conducted on the ln of participants' rating for how human-like the narrator was. This analysis was conducted with the ln of the Human-like scores. The analysis did not indicate a significant three-way interaction between participant gender, voice gender, and voice quality (F (1, 58) = 0.002, p = 0.961, n_{p2} = 0.000). The analysis also did not indicate any significant two-way interactions between participant gender—voice gender (F(1, 58) = 0.003, p = 0.074, n_{p2} = 0.001), participant gender—voice quality (F (1, 58) = 0.125, p = 0.725, n_{p2} = 0.002), or voice gender—voice quality (F(1,58) = 2.968, p =0.090 , n_{p2} = 0.049). Similarly, there were no significant main effects for participants gender (F (1, 58) = 0.128, p = 0.722, n_{p2} = 0.002), voice gender (F(1, 58) = 0.507, p = 0.479, n_{p2} = 0.009), or voice quality (F (1,58) = 2.830, p = 0.098, n_{p2} = 0.047).

A 2x2x2 factorial ANOVA that looked at gender of participants, gender of the voices, and voice quality was conducted on participants' rating for how engaging the narrator was. The analysis did not indicate a significant three-way interaction between participant gender, voice gender, and voice quality (F (1, 58) = 0.179, p = 0.674, n_{p2} = 0.003). The analysis also did not indicate any significant two-way interactions between participant gender—voice gender (F(1, 58) = 1.025 , p = 0.316 n_{p2} = 0.017), participant gender—voice quality (F (1, 58) = 0.189, p = 0.666 n_{p2} = 0.003), or voice gender—voice quality (F(1,58) = 2.436 p = 0.124, n_{p2} = 0.040). Similarly, there were no significant main

effects for voice gender ($F(1, 58) = 0.017$, $p = 0.895$, $n_{p2} = 0.000$), participants gender ($F(1, 58) = 0.024$, $p = 0.878$, $n_{p2} = 0.000$), or voice quality ($F(1,58) = 2.830$, $p = 0.098$, $n_{p2} = 0.047$).

5.2 LEARNING MEASURES

The second research question regarded the impact the narrator's voice would have on learners' abilities to learn and implement a new concept. For this question the study looked at how the gender of the narrator and participant, and the clarity of the narrator's voice may play a role. ANCOVA analyses were used to examine the learned differences among groups based on their pretest and posttest scores. The pretest was scored on a five-point scale. To ensure reliability a Kappa analysis was performed. The final analysis received a Kappa score of .64, making the coding guide moderately reliable.

A 2x2x2 factorial ANCOVA that looked at the difference in the participants pre-test score and their knowledge assessment scores. Participants gender, gender of the voices, and voice quality were all factors assessed when analyzing participants' multiple-choice knowledge assessment scores. The analysis did indicate there was a single main effect for the pre-test ($F(1, 57) = 4.632$, $p = 0.036$, $n_{p2} = 0.075$). However, the analysis did not indicate a significant three-way interaction between participant gender, voice gender, and voice quality ($F(1, 57) = 0.606$, $p = 0.440$, $n_{p2} = 0.011$). The analysis also did not indicate any significant two-way interactions between participant gender—voice gender ($F(1, 57) = 0.104$, $p = 0.749$, $n_{p2} = 0.002$), participant gender—voice quality ($F(1, 57) = 1.098$, $p = 0.299$ $n_{p2} = 0.019$), or voice gender—voice quality ($F(1, 57) = 1.259$, $p = 0.267$, $n_{p2} = 0.022$). Similarly, there were no significant main effects for participants

gender (F (1, 57) = 0.941, p = 0.336, n_{p2} = 0.016), voice gender (F(1, 57) = 0.107, p = 0.744, n_{p2} = 0.002), or voice quality (F (1, 57) = 0.933, p = 0.338, n_{p2} = 0.016).

A 2x2x2 factorial ANCOVA that looked at the difference in the participants pre-test score and their transfer assessment scores gender. Participants' gender, gender of the voices, and voice quality were all factors assessed when analyzing participants' transfer assessment scores. The analysis did not indicate there was a main effect for the pre-test (F (1, 57) = 1.634, p = 0.206, n_{p2} = 0.028). The analysis did not indicate a significant three-way interaction between participant gender, voice gender, and voice quality (F (1, 57) = 2.862, p = 0.096, n_{p2} = 0.048). The analysis also did not indicate any significant two-way interactions between participant gender—voice gender (F(1, 57) = 0.150 , p = 0.700 n_{p2} = 0.003), participant gender—voice quality (F (1, 57) = 0.190, p = 0.664, n_{p2} = 0.003), or voice gender—voice quality (F(1, 57) = 0.075, p = 0.785, n_{p2} = 0.001). Similarly, there were no significant main effects for voice gender (F (1, 57) = 1.838, p = 0.181, n_{p2} = 0.031), or voice quality (F (1, 57) = 0.754, p = 0.389, n_{p2} = 0.013). However, there were significant main effects for participants gender (F (1, 57) = 6.619, p = 0.013, n_{p2} = 0.104).

CHAPTER 6

DISCUSSION

The current study presented a multimedia learning video with a common Human Factors concept. There were four different videos created that used four different voices to examine how the gender and clarity of the narrator's voice affected the learners' perception of the narrator and their ability to learn and implement a new concept. Gender of the participants was also used as a variable in hopes of seeing an effect based on the social agency theory. The study had three main predictions with three sub-predictions for each, equating to nine predictions in total.

6.1 VOICE EFFECT

When participants are shown a multimedia learning video narrated with a modern-synthesized voice, they will give the narrator a higher likeability rating.

Craig and Schroder have performed two studies that look at how the voice of a narrator play a role on the perception of the narrator and the participants ability to learn new information (2017; 2018). Both studies utilized a human voice and the same voices from the female classic and modern synthesized voice conditions as the current study. Their first study utilized an on-screen agent to help facilitate the presentation of information. In this study they found participants gave significantly lower ratings in the classic voice condition than in the modern voice condition for the Facilitates Learning and Credibility subscale. However, there was no significance found for the two groups for the Human-Like and Engaging subscales (Craig & Schroder, 2017).

In Craig and Schroder's second study they did not use an on-screen voice, solely isolating the narrator's voice. Their findings were consistent with their first study in that they did not find a significant difference between the two computer-generated voices for the Human-Like and Engagement subscales. However, the findings for the study also found no significance difference in the scores for the Facilitates Learning and Credibility subscales (Craig & Schroder, 2018).

In a study performed by Mayer and colleagues looked at a human voice in comparison to a computer-generated voice. Participants gave a higher likeability rating to the human-voice over the computer-generated voice (Mayer, Sabko, & Mautone, 2003). Due to low effect sizes of this study it is difficult to report definitive findings. However, for the Human-Like and Engagement subscales, the findings from the current study are consistent with the findings from previous research. This shows that regardless of the narrator's or the participant's gender, the participants will not view a computer-generated voice as human-like or engaging. The findings from this study were in agreement with Craig and Schroder's second study for the facilitated learning and credibility subscales. It seems as though the presence of a real human cannot be replaced with the current technology. In order to authentically engage a person, it is best to use a human voice with an on-screen character. Mayer explains that the difference between the ratings of the narrators is due to the computer-generated voice seeming "less dynamic, less attractive, and less superior" than a human voice (Mayer, Sabko, & Mautone, 2003).

When participants are shown a multimedia learning video narrated with a modern-synthesized voice, they will learn more and score higher on their application tests.

Previous research that has looked at learning effects of narrator voices has not been as consistent as the research on narrator likeability. Craig and Schroder's studies did show some learning effects; however, they were not constant results across the two studies. Participants in their first study performed significantly better on the transfer test from the modern synthesized voice than the human voice and classic synthesized voice (Craig & Schroder, 2017). In their second study, participants in the modern-synthesized voice condition performed better on the multiple-choice section than participants in the classic-synthesized voice (Craig & Schroder, 2018). However, studies performed by Mayer and Atkinson do show consistent learning effects across voice conditions. Both studies report significant difference in the voice groups for retention and transfer tests (Atkinson, Mayer, & Merrill, 2005; Mayer, Sabko, & Mautone, 2003).

The current study did not show any significant knowledge assessment results across voice groups, creating an inconsistency in the overall research. However, with the finding of the previous research, it can be concluded that this study does not report reliable results. This largely due to uneven group sizes and a possibly unreliable assessment. Previous research suggests that there is high importance to still look at the learning effects of a voice. Mayer claims a human voice creates less cognitive load, whereas Craig claims technology has advanced enough to where there is no difference in cognitive load (Mayer, Sabko, & Mautone, 2003; Craig & Schroder, 2018). Based on all the research there is not sufficient evidence to support Mayer's findings on learnings effects. In order to draw true conclusions from Mayer's research effects it is important to

perform replication studies. These inconsistencies could be due to different sample groups collected by the different studies. However, if the findings from Mayer's studies are true then they should be generalizable to the whole population.

6.2 GENDER EFFECTS

There were two separate sets of predictions made at the start of the study for gender that predicted gender effects and mismatch gender effects. However due to the little research previously performed and the non-significant findings for gender and mismatch gender effects from the current study, the findings for these predictions will be discussed together.

When participants are shown a multimedia learning video with a narrator of the same gender as themselves, they will give the narrator a higher likeability rating; Female participants who are shown a multimedia learning video with a male narrator will be less affected by gender roles and will give the narrator a higher likeability rating than the male participants who are shown a video with a female narrator.

The majority of previous gender research on multimedia learning has focused on how the gender of an on-screen agent affects a person's perception of the agent. The past literature has drawn multiple conclusions on this topic, making it necessary for more research to be conducted. The overall consensus has been that female narrators are more beneficial than male narrators to promote self-efficacy because they are women are viewed as less knowledgeable than men (Baylor & Kim, 2004; Rosenberg, Plant, Doerr, & Baylor, 2010). The results of this study are inconsistent with findings of participants

perceptions of the narrators. There was no significant difference in the narrators' API. This could largely be due to uneven and small sample sizes prevented the study from having a representative sample to draw significant conclusions from, creating a Type II error.

When participants are shown a multimedia video with a narrator of the same gender as themselves, they will learn more and score higher on their application tests; Female participants who are shown a multimedia learning video with a male narrator will be less affected by gender roles and will have higher application test results than the male participants who are shown a video with a female narrator.

The results from this study show that the gender of a narrator does not play an effect on a person's ability to learn new information. Looking at previous research, there has also not been any evidence to show that the gender of the narrator has a direct effect on a person's ability to learn new information (Baylor & Kim, 2004). When this prediction was made it was based more off the social agency theory. When further analyzing the research, it seems as though social agency theory relates more to the type of voice being used- human voice vs. computer-generated voice. Although gender does promote relatability, it does not enhance performance. Research on the voice effect suggests that learning effects were affected by cognitive load (Mayer, Sabko, & Mautone, 2003). The gender of a speaker does not affect that because there are not any noticeable variations in tones, pitches, speed, etc that would affect how a person processes the information.

6.3 LIKEABILITY VS. TEST PERFORMANCE

Participants who give a narrator a higher likeability rating will perform better on their application tests.

This prediction was made for all three hypothesis groups. Due to similarity in findings, this single prediction across all three prediction groups will be discussed together.

Previous research that looked at narrator likeability and test performance does not directly discuss the relation of the two factors. However, when looking at the results conclusions can be drawn. Overall, it has been found that groups who gave a narrator a high likeability rating have higher application results (Craig & Scroder 2017; Craig & Schroder 2018; Mayer, Sabko, & Mautone, 2003). Although not all the research shows statistically significant results, there is a clear pattern that has been developed. In the current study participants in the classic male voice received the lowest ratings on all four API subscales but had the highest assessment results. This is inconsistent with the findings from previous research. Sample sizes from the groups were not equal which may have played a role on the inconsistencies. The male classic voice condition had the highest male participants and the lowest female participants.

6.4 LIMITATIONS

As a new researcher, this study did not go without limitations. There were multiple obstacles that occurred during data collection due to limited time and resources available. There are many ways that this study can be improved to accurately asses the effects a narrator's voice has on a person's ability to learn and implement a new concept.

The goal of this study was to evenly represent male and female narrators and participants. However, this was not achieved for both groups. The ASU subject-pool that was used consist of students in predominantly engineering degrees because it utilizes students from the engineering campus. This means that there is a large sample of males then females available for recruitment. During data collection there was a glitch found in the survey flow that did not allow for true randomization. The video for the male classic synthesized voice was not working properly. Participants would start the survey answer the demographic questions and then the pre-test question and would watch the entire video. However, the system would not allow the participant to move forward. This led to participants exiting the survey and restarting it until the survey gave them a video with a different voice so they could move on. To correct the group sizes, the researcher adjusted the group count accordingly to represent the true number of survey collections for each group at that current time. Based on this adjustment, Qualtrics collected data for the male classic synthesized voice until that group was equal with the rest. Once the groups were equal, Qualtrics returned to randomly assigning participants to a test group.

Due to the glitch in the survey, many of the participants had to restart the survey to move forward. The issue with this is these participants were now exposed to the information in the multimedia learning video. This is problematic because they now knew what information to discuss in the pre-test, creating a priming effect. When reviewing the participants' surveys there were several answers that were almost verbatim to the information discussed in the video. This skews the data and does not show a true representation of what these participants knew about learnability before watching the video.

Lastly, this study would benefit from a more reliable knowledge assessment. The scope of this study was very large. It required the researcher to design the study, create a multimedia learning video, and develop a knowledge assessment that accurately assess the respondents' abilities to recall and transfer information. Based on the responses received it seems as though the knowledge assessment was not created to accurately test their recall skills. Many of the questions had answers such as "D. All of the above" or "D. Both A and B are correct". However, these were not necessarily the correct answer choices. This could be an issue with the comprehension of the questions themselves or just plain negligence of the participant. Either way, to produce the most accurate knowledge assessment scores it would be best to utilize a tool that already holds reliability, such as the lighting video and assessment used in Mayer's and Craig's studies (Mayer, 2003b; Craig & Schroder, 2017).

CHAPTER 7

CONCLUSION

The current study attempted to answer the question of how the gender and voice quality of a computer-generated narrator affects a person's ability to learn and apply a new concept. The study predicted that there would be three main effects: voice effect, gender match effects, and gender mismatch effects. There was no significance found across all dependent variables, which is inconsistent with much of the past research on participants' perception of the narrator. In regard to learning, the research is consistent in that gender does not create significant differences in test results.

There were several issues that arose during data collection that could have skewed the data. This study called for 160 participants in order to achieve a high enough power analysis, however there were only 66 responses analyzed. It is important to have a high enough effect size in order to draw accurate conclusions from the data. It would be beneficial to replicate this study with a larger sample size. Although learning effects may not be observed, there may be significance in the participants' perceptions of the narrator and their levels of self-efficacy. By looking at these factors combined there could be interesting results in how self-efficacy effects their ability to learn new information.

As the world shifts to a more technology-based system, there is an increased need to understand how to best design multimedia learning videos to promote the most natural learning environment possible. Even if there is no relation to the narrator's voice and learning, there has been shown evidence that there is a relation with their perception of the narrator. If a participant exhibits some amount of joy in the learning material, then they will be more interested in continuing on with their education.

REFERENCES

Atkinson, R. K., Mayer, R. E., & Merrill, M. M. (2005). Fostering social agency in multimedia learning: Examining the impact of an animated agent's voice. Contemporary Educational Psychology, 30(1), 117-139. doi:10.1016/j.cedpsych.2004.07.001

Baylor, A. L., & Kim, Y. (2004). Pedagogical Agent Design: The Impact of Agent Realism, Gender, Ethnicity, and Instructional Role. Intelligent Tutoring Systems Lecture Notes in Computer Science, 592-603. doi:10.1007/978-3-540-30139-4_56

Baylor, Amy & Ryu, Jeeheon. (2003). The API (Agent Persona Instrument) for Assessing Pedagogical Agent Persona.

Clark, R. C., & Mayer, R. E. (2016). e-Learning and the science of instruction (4th ed.). Hoboken, NJ: Wiley.

Chaiken, S., and D. Maheswaran. (1994). Heuristic processing can bias systematic processing: Effects of source credibility, argument ambiguity, and task importance on attitude judgment. Journal of Personality and Social Psychology 66 (3): 460–73.

Craig, S. D., Gholson, B., & Driscoll, D. M. (2002). Animated pedagogical agents in multimedia educational environments: Effects of agent properties, picture features and redundancy. Journal of Educational Psychology, 94(2), 428-434. doi:10.1037//0022-0663.94.2.428

Craig, S. D., & Schroeder, N. L. (2017). Reconsidering the voice effect when learning from a virtual human. Computers & Education, 114, 193-205. doi:10.1016/j.compedu.2017.07.003

Craig, S. D., & Schroeder, N. L. (2018). Text-to-Speech Software and Learning: Investigating the Relevancy of the Voice Effect. *Journal of Educational Computing Research,* 073563311880287. doi:10.1177/0735633118802877

Debono, K.G., and R.J. Harnish. (1988). Source expertise, source attractiveness, and the processing of persuasive information—a functional- approach. Journal of Personality and Social Psychology 55 (4): 541–46.

Hollingshead, A. B., & Fraidin, S. N. (2003). Gender stereotypes and assumptions about expertise in transactive memory. Journal of Experimental Social Psychology, 39(4), 355-363. doi:10.1016/s0022-1031(02)00549-8

Hovland, C.I., and W. Weiss. (1951). The influence of source credibility on communication effectiveness. Public Opinion Quarterly 15: 635–50.

Mayer, R. E. (a)(2003). The promise of multimedia learning: Using the same instructional design methods across different media. *Learning and Instruction, 13*(2), 125-139. doi:10.1016/s0959-4752(02)00016-6

Mayer, R. E. (b) (2003). Principles of Multimedia Learning Based on Social Cues : Personalization, Voice, and Image Principles. *The Cambridge Handbook of Multimedia Learning,* 201-212. doi:10.1017/cbo9780511816819.014

Mayer, R. E., Fennell, S., Farmer, L., & Campbell, J. (2004). A Personalization Effect in Multimedia Learning: Students Learn Better When Words Are in Conversational Style Rather Than Formal Style. Journal of Educational Psychology, 96(2), 389-395.

Mayer, R. E., & Moreno, R. (2003). Nine Ways to Reduce Cognitive Load in Multimedia Learning. Educational Psychologist, 38(1), 43-52. doi:10.1207/s15326985ep3801_6

Mayer, R. (2017). Using multimedia for e-learning. *Journal of Computer Assisted Learning, 33*(5), 403-423. doi:10.1111/jcal.12197

Mayer, R. E., Sobko, K., & Mautone, P. D. (2003). Social cues in multimedia learning: Role of speaker's voice. Journal of Educational Psychology, 95(2), 419-425. doi:10.1037/0022-0663.95.2.419

Paolacci, G., & Chandler, J. (2014). Inside the Turk: Understanding mechanical turk as a participant pool. Current Directions in Psychological Science, 23, 184e188.

Reeves, B., & Nass, C. (1996). The media equation. New York: Cambridge University Press.

Rice, L., & Barth, J. M. (2016). A Tale of Two Gender Roles: The Effects of Implicit and Explicit Gender Role Traditionalism and Occupational Stereotype on Hiring Decisions. Gender Issues, 34(1), 86-102. doi:10.1007/s12147-016-9175-4

Rosenberg-Kima, R. B., Plant, E. A., Doerr, C. E., & Baylor, A. L. (2010). The Influence of Computer-based Model's Race and Gender on Female Students' Attitudes and Beliefs Towards Engineering. *Journal of Engineering Education, 99*(1), 35-44. doi:10.1002/j.2168-9830.2010.tb01040.x

Schroeder, N. L., Adesope, O. O., & Gilbert, R. B. (2013). How Effective are Pedagogical Agents for Learning? A Meta-Analytic Review. Journal of Educational Computing Research, 49(1), 1-39. doi:10.2190/ec.49.1.a

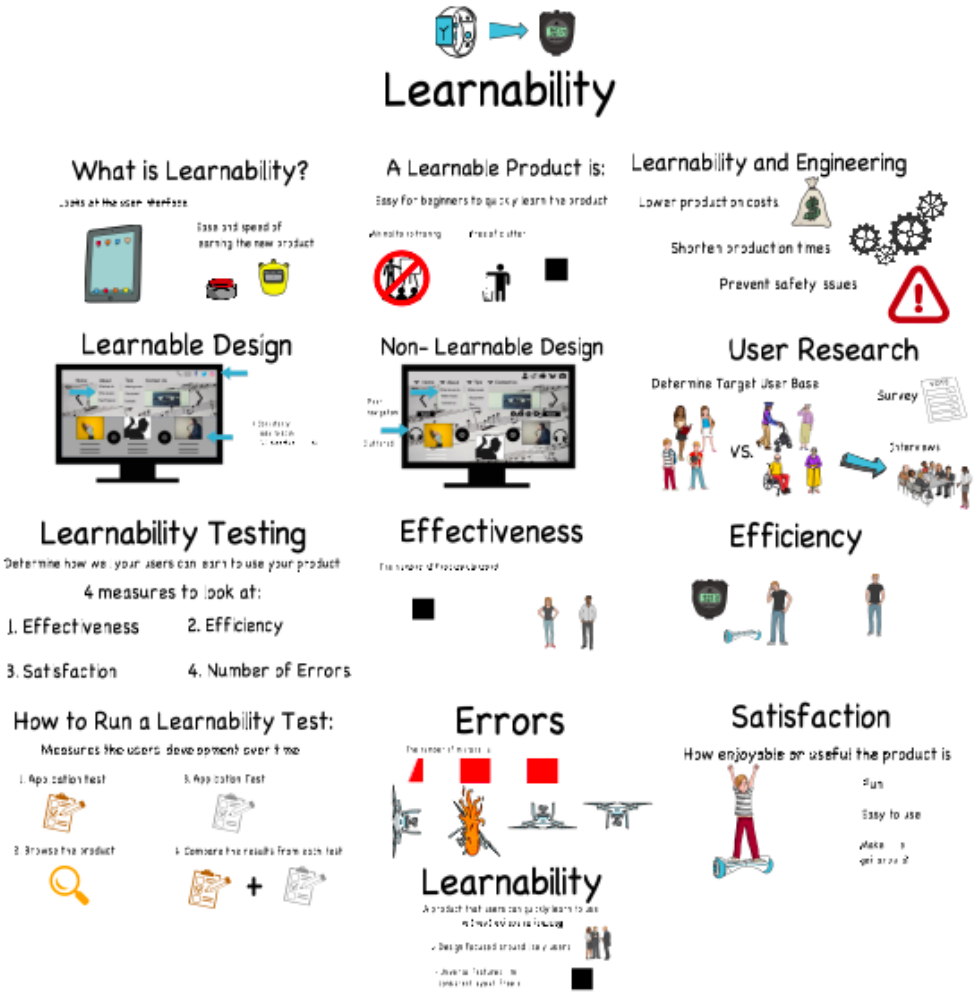
Schroeder, N. L., Romine, W. L., & Craig, S. D. (2017). Measuring pedagogical agent persona and the influence of agent persona on learning. Computers & Education, 109, 176-186. doi:10.1016/j.compedu.2017.02.015

Shadish, W., Cook, T., & Campbell, D. (2002). Experimental and Quasi-Experimental Designs for Generalized Causal Inference. Belmont, CA: Wadsworth Cengage Learning.

Tze Wei, L., Su-Mae, T., & Jayothisa, C. (2013). The Effects of Peer-Like and Expert-Like Pedagogical Agents on Learners' Agent Perceptions, Task-Related Attitudes, and Learning Achievement. *Journal Of Educational Technology & Society*, *16*(4), 275-286.

APPENDIX A

MULTIMEDIA LEARNING VIDEO OVERVIEW



This is an overview of the information that was presented in the multimedia learning video created for this study.

APPENDIX B

APPLICATION QUESTION

25.

Which image is an example of a design that is **free of clutter**?

This is an example of the application questions created to tests the participants' ability to apply the new information.

APPENDIX C

AGENT PERSON INSTRUMENT (API)

Agent Persona Instrument (API) : Final Instrument

All items should be presented with a 5-point Likert scale, ranging from 1=Strongly disagree to 5=Strongly agree.

Facilitating Learning (10 items)

The narrator led me to think more deeply about the presentation.
The narrator made the instruction interesting.
The narrator encouraged me to reflect what I was learning.
The narrator kept my attention.
The narrator communicated the main idea clearly.
The narrator helped me to concentrate on the presentation.
The narrator focused me on the relevant information.
The narrator helped me learn the material.
The narrator was good at teaching.
The narrator was easy to learn from.

Credible (5 items)

The narrator was knowledgeable.
The narrator was intelligent.
The narrator was useful.
The narrator was helpful.
The narrator was an effective teacher.

Human-like (4 items)

The narrator has a personality.
The narrator's emotion was natural.
The narrator was human-like.
The narrator showed emotion.

Engaging (5 Items)

The narrator was engaging.
The narrator was enthusiastic.
The narrator was entertaining.
The narrator was motivating.
The narrator was easy to connect with.

This is the updated API used to test participants' perception of the overall likeability. 44