



Article

Do Not Freak Me Out! The Impact of Lip Movement and Appearance on Knowledge Gain and Confidence

Amal Abdulrahman, Katherine Hopman and Deborah Richards * 

School of Computing, Faculty of Science and Engineering, Macquarie University,
Macquarie Park, NSW 2109, Australia; amal.abdulrahman@mq.edu.au (A.A.);
katherine.hopman@students.mq.edu.au (K.H.)

* Correspondence: deborah.richards@mq.edu.au

Abstract: Virtual agents (VAs) have been used effectively for psychoeducation. However, getting the VA's design right is critical to ensure the user experience does not become a barrier to receiving and responding to the intended message. The study reported in this paper seeks to help first-year psychology students to develop knowledge and confidence to recommend emotion regulation strategies. In previous work, we received negative feedback concerning the VA's lip-syncing, including creepiness and visual overload, in the case of stroke patients. We seek to test the impact of the removal of lip-syncing on the perception of the VA and its ability to achieve its intended outcomes, also considering the influence of the visual features of the avatar. We conducted a 2 (lip-sync/no lip-sync) × 2 (human-like/cartoon-like) experimental design and measured participants' perception of the VA in terms of eeriness, user experience, knowledge gain and participants' confidence to practice their knowledge. While participants showed a tendency to prefer the cartoon look over the human look and the absence of lip-syncing over its presence, all groups reported no significant increase in knowledge but significant increases in confidence in their knowledge and ability to recommend the learnt strategies to others, concluding that realism and lip-syncing did not influence the intended outcomes. Thus, in future designs, we will allow the user to switch off the lip-sync function if they prefer. Further, our findings suggest that lip-syncing should not be a standard animation included with VAs, as is currently the case.

Keywords: virtual coach; lip-sync; realism; behaviour change



Citation: Abdulrahman, A.; Hopman, K.; Richards, D. Do Not Freak Me Out! The Impact of Lip Movement and Appearance on Knowledge Gain and Confidence. *Multimodal Technol. Interact.* **2024**, *8*, 22. <https://doi.org/10.3390/mti8030022>

Received: 5 February 2024
Revised: 23 February 2024
Accepted: 28 February 2024
Published: 5 March 2024

Received: 5 February 2024
Revised: 23 February 2024
Accepted: 28 February 2024
Published: 5 March 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Virtual agents (VAs) are designed to mimic the natural human–human interaction for diverse purposes such as entertainment [1], education [2] and healthcare [3]. In many of these domains, VAs are increasingly being used to elicit engagement, assist users to complete activities and/or to persuade users to change their behaviours [4]. To maximise the effectiveness of VAs, a comprehensive understanding of VA design factors that optimise these outcomes is critical.

Investigating user experience during interaction with VAs has been an active topic for decades due to its importance in achieving the desired outcome of the interaction as well as in sustaining the intention to use VAs in the future [5]. User experience depends considerably on the VA's design/modalities (i.e., the VA's appearance and behaviour), which require multidisciplinary efforts such as psychology [6] and artificial intelligence [7].

One VA design feature that has received significant attention is agent realism. Increased use of the photo-realism of agents indicates a general acceptance that greater realism improves user experience; however, recent studies (e.g., [8]) have indicated that increased perceived humanness correlates with increased perceived eeriness, providing further support for the uncanny valley effect introduced by Mori [9]. While some researchers have focused on studying the effect of a single modality such as realism [8], others have

suggested the congruence between design aspects (e.g., verbal and non-verbal cues [10]) as being more important and significantly influencing user experience [11].

Research examining the congruence of the VA's appearance (human vs. animation) and voice (human vs. synthetic) [10,12] identified that users found a human face with a synthetic voice or a humanoid robot with a human voice caused significantly higher eeriness than when the face and voice were matched. Abdulrahman and Richards [10] concluded that although users reported higher eeriness after interacting with a VA with a synthetic voice compared to the same VA with a human voice, they achieved better outcomes (i.e., scored higher intention to change their behaviour) after interacting with the VA with a synthetic voice. This suggests that outcomes were related to the congruence of verbal and non-verbal features. In contrast, other studies failed to prove the importance of congruency on the desired outcome [13,14], which was explained by the dependency on the context such as education [15] and entertainment [16].

Besides appearance, the incongruence of facial expressions, including expressing emotions and lip movements, with the heard voice could lead to misunderstanding the message and impose a negative feeling towards the speaker [17,18]. Less accurate lip-syncing with the voice of animated characters has been found to engender freakiness and disturbance in users [19]. While researchers have investigated the effect of the various VAs modalities including voice (synthetic versus human) and facial expressiveness [20] and the incongruence between the two [18], limited research has examined the effect of the presence or absence of lip-syncing (i.e., lip movement versus no lip movement) on user experience and user outcomes (e.g., learning, self-efficacy or health outcomes). An exception to this is the work [21], which identified that an alignment between animation and lip-syncing in a VR environment (i.e., the conditions of animation and lip-syncing and no animation, no lip-syncing) leads to slightly better knowledge retention than when conditions are not aligned, prompting the authors to suggest that greater knowledge retention is potentially related to a lack of distraction in the no animation, no lip-syncing condition.

Despite the above findings, lip-sync is a basic animation included in the design of embodied conversational agents (ECAs). McDonnell et al. [22] demonstrated that users can accept motion artifacts in the design of VAs if they are more cartoon-like than realistic-looking (human-like) in appearance. As the lip-sync technique that is commonly used in the research is synthetic to mimic the natural lip shapes (e.g., [23]), we are interested in exploring the relevance of VA realism or naturalism on the user's perception of removing lip-syncing by testing the presence of lip-syncing with both a more realistic-looking VA versus a cartoon-like VA. Hence, in this paper, we investigate the impact of VA's realism and lip-syncing in the context of assisting psychology students to gain knowledge and confidence in their future roles as practicing therapists in the knowledge domain of emotion regulation. Section 2 of the paper examines the literature in the field, while Section 3 describes the study methodology. Sections 4 and 5 detail the study results and discussion, respectively, and Section 6 provides the study conclusions.

2. Literature Review

2.1. VA Design Features

There is encouraging evidence of the potential use of VAs in education; however, there remains a limited understanding of the specific VA design features that facilitate learning and promote competence and confidence/self-efficacy. These often are the intended outcomes from the VA–user interaction in an education setting. It has been suggested that learning and confidence can be increased through the provision of positive learning experiences, which can include instructional strategies and collaborative activities that promote engagement [24]. Arguably, VAs could play a pivotal role in this space; however, to achieve this outcome, there is a need for a greater understanding of their design features that influence instructional and interactive user experience.

Agent design and its impact on the formation of impressions during the VA–human interaction have been a topic of ongoing research and debate [25]. Prior studies investigated

design features such as realism [8], voice [26], personality [27], etc., with findings varying across the domains and at times being contradictory [25].

In education, some studies of VA visual design identified that a highly human-like appearance failed to impact confidence or self-efficacy relative to more iconic renders [28], but in others, students exhibited increased knowledge when agents were depicted with more realism [29]. A systematic review of user interaction with agents concluded that agent appearance appears vital in some contexts but not in others and recommended that there is a need to explore the effect of appearance design from a more nuanced perspective, taking into account the respective context and/or task [25].

Noting the above, in this paper, we focus on the examination of the impact of a VA's features: realism and lip-syncing in the domain of emotion regulation education. Hence, our first hypothesis states the following:

Hypothesis 1 (H1). *There is a difference between the realistic-looking VA and the cartoon-like VA regarding (a) eeriness perception and (b) user-agent interaction experience.*

In addition to realism, this paper also examines the impact of lip-syncing on user outcomes. Research examining asynchronous lip-syncing, particularly in cartoon or virtual characters, was found to lead to negative user impressions [30]. In our previous research, we created different VAs to persuade children and families to follow treatment advice, change stigmatised attitudes to eating disorders and teach trauma survivors to regulate their emotions. While these studies have shown significant improvements in the desired outcome (i.e., adherence, health outcomes, attitudes and self-efficacy), we received negative feedback from users related to the lip-syncing animations, using terms such as freaky and creepy. In addition, Richards, Miranda Maciel and Janssen's [31] evaluation of a VA designed to help stroke survivors take charge of their recovery found that some stroke survivors, such as those with aphasia, were overwhelmed by the multiple audio and visual elements. In particular, in addition to the VA using audio to speak the dialogue, the dialogue was provided in text on the screen and the mouth of the avatar moved to match the audio (lip-syncing). Consequently, we opted to contribute to the area of VA design by understanding what the impact is, if any, of lip-syncing.

Hypothesis 2 (H2). *There is a difference between VAs with and without lip-sync in terms of (a) eeriness perception and (b) user-agent interaction experience.*

2.2. Confidence and Knowledge

Academic performance has been shown to be positively associated with the level of confidence one has in their ability to execute a course of action or attain a specific performance outcome [32]. This has implications for learning, as it suggests that confidence should be an education target alongside the building of knowledge and competence [33]. In the field of healthcare education, this is particularly important, as confidence is recognised as an important characteristic of the healthcare workforce and integral to patient experience [34].

According to Bandura [35], people develop confidence or self-efficacy from four sources: mastery experiences, vicarious experiences, social persuasion and emotional states. There is a small but growing body of evidence indicating that confidence (self-efficacy) can be successfully promoted in digital learning environments as a source of persuasion, including mental health apps [36], online self-management programs [37] and chatbot interfaces [38]. In addition, a recent study has demonstrated the efficacy of a VA in increasing a person's confidence in returning to work following injury [39].

Confidence is also a critical component of consultation skills. For psychologists, low perceived self-efficacy (i.e., confidence in their ability to perform the behaviour) in using consultation skills has been found as a barrier to engaging in consultation activities [40].

Noting this, there is a need to understand whether technology can be used in learning environments to assist novice psychologists in developing consultant self-efficacy.

Therefore, a key goal of this paper is to contribute to the field of VA design by understanding the design features that impact first-year psychology students' confidence in their knowledge and ability to impart that knowledge to others. Thus, we created four conditions/groups: (1) realistic VA with lip-syncing; (2) realistic VA without lip-syncing; (3) cartoon-like VA with lip-syncing and (4) cartoon-like VA without lip-syncing. This leads to our final hypothesis:

Hypothesis 3 (H3). *There is no between-groups difference in (a) knowledge gain, (b) confidence in their knowledge or (c) confidence to recommend/explain it to others.*

3. Methodology

3.1. Study Design

We conducted an online experiment approved by our Human Research Ethics Committee, utilising a 2×2 between-subjects design. The independent variables manipulated were the VA's appearance (human-like vs. cartoon-like) and lip-syncing (with vs. without). As our key goal was to validate the types of characters we were currently using in a range of different studies which had been made with Adobe FUSE, we selected one of the FUSE characters to which we had received comments of creepiness (Adobe Fuse software is no longer possible due to its discontinuation in 2020). However, previously designed agents can be edited using the animation software Adobe Mixamo (<https://www.mixamo.com/>, accessed on 1 October 2023) We refer to this model as realistic, acknowledging that the model is not photo-realistic but rather has human-like features. We then used Ready Player Me (<https://readyplayer.me/>, accessed on 1 October 2023) to transform our original Erica into a cartoon-like Erica. The Fuse platform uses high-resolution textures and advanced shaders to create life-like skin, hair and clothing, while Ready Player Me avatars have exaggerated features (e.g., bigger eyes) and simplified textures. The study design is illustrated in Figure 1. The following sections explain the design of the study, recruitment, questionnaires utilised to test the hypotheses and the VA's design.

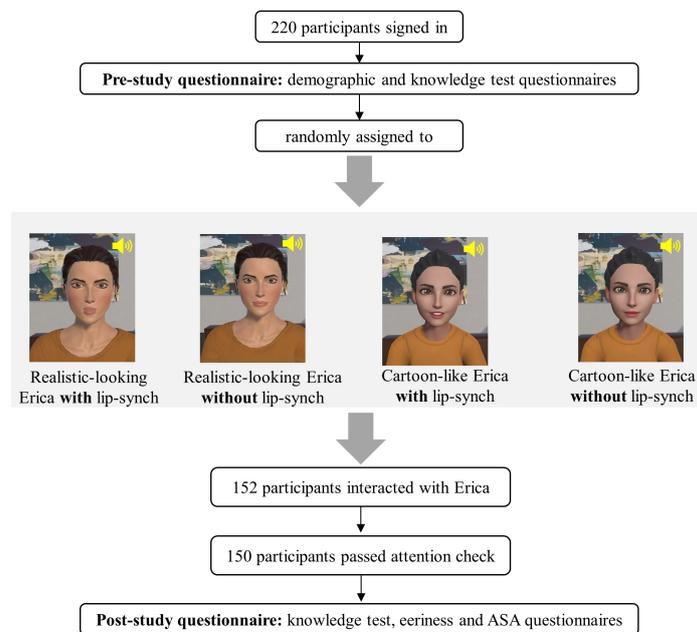


Figure 1. Study design. To visually demonstrate the concept, snippets of Erica were captured during interactions with users. Instances of Erica with/without the lip-sync feature depict moving/closed mouths. The microphone icon is consistently displayed across all versions of Erica to denote the presence of voice functionality.

3.2. Recruitment

Participants recruited from the university psychology pool were randomly assigned to one of the four groups. This pool comprised first-year psychology students who could receive course credit for research participation. The average age of this cohort was typically 21.7 (s.d. 6.747) years [41], comprising around 75% females from a range of cultural backgrounds. Students could select this study among several alternative studies to complete their course requirements. Informed written consent was gathered prior to the participants' involvement in the study, and participants could withdraw penalty-free at any stage. We designed the study to be completed within 30 min.

3.3. Procedure

As shown in Figure 1, upon consent, participants first completed a pre-study survey covering demographics (Section 3.5.1) and assessing baseline knowledge (Section 3.5.2) on the emotion regulation topic to be reviewed with the VA and their confidence in the strategies and in recommending these strategies to others who need it as an indication of their readiness to practice their knowledge. Participants were then randomly assigned to interact with one of the four versions of Erica. We used the "distribute evenly" randomisation feature in the Qualtrics survey software (<https://www.qualtrics.com>, accessed on 22 February 2024) to ensure equal numbers in each group. We did not use stratified allocation to groups, as we were unable to control who selected our study.

After the interaction, the participants received the post-study questionnaire, which included baseline questionnaires as well as the eeriness questionnaire [42] and the Artificial Social Agent (ASA) questionnaire [43] to evaluate the user-VA interaction experience. The eeriness and ASA questions were randomised. and Qualtrics research software was used to design the questionnaires and collect data.

3.4. Dialogue

Increasing the confidence (i.e., self-efficacy) of psychologists in their consultation abilities includes improving one's consultation knowledge [28,40]. To this end, as an intervention, we designed Erica's dialogue to review the knowledge of interest, emotional regulation strategies, with the students in an interactive way.

The dialogue is structured as a state-based or tree-based dialogue where the flow of the dialogue progresses according to the user's choice of answers. Figure 2 illustrates part of the designed dialogue. When the current state is the agent's turn to speak and it is "Agent43", the agent will introduce two strategies for the user to select from. The dialogue engine then moves forward to the next state "User43" and waits for the user's choice before continuing the conversation.

In all conditions, Erica converses with participants via speech (digital voice), regardless of the presence of lip-syncing, and text dialogue (her utterances appear as text on the screen). The participant can converse with Erica by selecting response options from a predefined set of answers. We developed this estimated 10–12 min interactive experience using the UNITY3D Game engine (<https://unity.com/products/unity-engine>, accessed on 1 October 2023) Salsa LipSync (<https://crazyminnowstudio.com/unity-3d/lip-sync-salsa/>, accessed on 1 October 2023) and a custom dialogue engine was designed to control audio, text and state-based response branching.

3.5. Measures

3.5.1. Basic Demographic Data

Demographic data allow for the description of the study population. In this study, we asked the participants to report their cultural background and gender.

Current state	Utterances	Next state
Agent43:	I have two more strategies that can be useful to help with disengaging from excessive rumination or worrying. One is called positive refocusing and the other is planning and problem-solving.	User43
	Which one would you like to hear more about?	
User43:	- Positive refocusing - Planning and Problem Solving	Agent44 Agent46
Agent44	Positive refocusing involves learning to increase someone's positive emotions. Have you heard of this strategy before?	
User44:	- Yes, but I am not sure how it works - Yes, and I have heard about it before - No, can you please tell me more	Agent45b Agent45a Agent45b
Agent45a:	That's great. Maybe you can see whether the techniques I outline are the same as the ones you have heard about?	Agent45b
Agent45b:	There are a couple of ways that one can increase their positive emotions. One way is by paying attention to the positive events going on in one's life and celebrating their successes. The second is by actively increasing the positive events in their life. For example, making sure they engage in activities they enjoy.	Agent46
Agent46:	When a person finds that it is hard to stop worrying or ruminating about a particular thing, it can often be helpful to "take a break" and participate in an activity that they enjoy. This might be going for a walk outdoors or reading a good book. It is important to recognise that this type of positive refocusing is not the same as trying to suppress a worry or thought, or put a worry out of their head.	User45
	Does that make sense?	
User45:	- Yes - No - I'm not sure	Agent47a Agent47b Agent47c

Figure 2. A snippet from Erica's dialogue discussing positive refocusing based on the user's choice. In column 1 (current state), "Agent" indicates the following text will be displayed and uttered by Erica and "User" indicates the options provided to the user. The final column (next state) manages dialogue flow indicating the next state.

3.5.2. Knowledge Test

At baseline, participants were asked whether they had studied emotion regulation before and were then given 4 multiple-choice questions to test their knowledge of worrying, rumination, positive refocusing, and planning and problem-solving. With every question, participants were asked to rate their confidence in their response using a scale of 1—not at all confident to 7—completely confident. Using the same confidence scale, participants were then asked, "How would you rate your level of confidence in being able to recommend or explain how to implement each strategy to someone who needs to regulate their emotions?" These data were collected before and after interacting with Erica to measure if the participants gained knowledge or increased their confidence as a result of the interaction.

3.5.3. Eeriness Questionnaire

Building upon the widely used Godspeed questionnaire, Ho and MacDorman [42] developed a more comprehensive tool to assess VA anthropomorphism and other dimensions like eeriness to capture the "uncanny valley", the region where VAs appear unsettlingly human-like. It includes eight items (bipolar scales) measuring eeriness, ranging from negative impressions (−3) on the left to positive (+3) on the right, with a neutral midpoint (0). The items can be found in Figure 3. The items were presented in a random order to the participants to control for order bias.

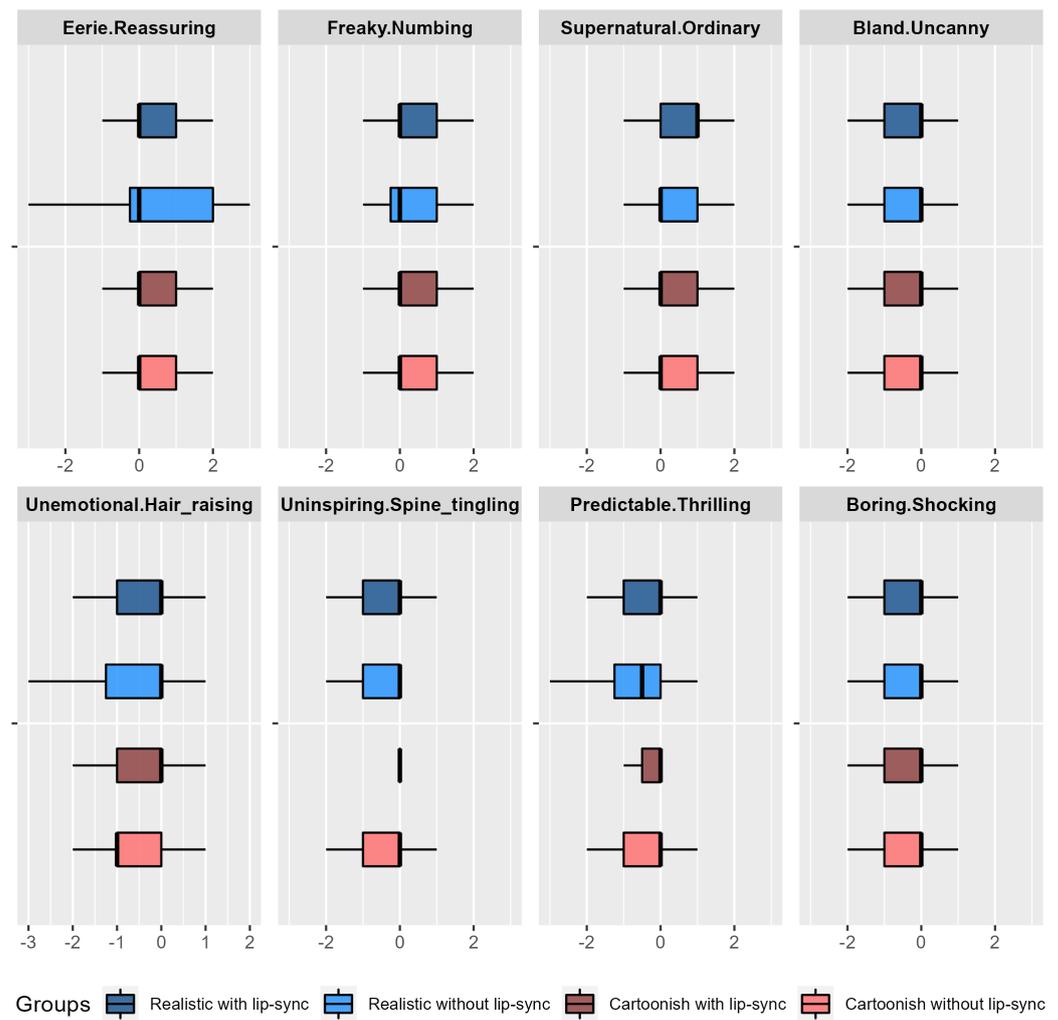


Figure 3. Eeriness items' scales. The box plots show the participants' scores on scales, with the left side representing a greater tendency towards a negative perception of the item. Negative and positive labels of the scales are presented at the top of each plot, respectively.

3.5.4. ASA—Short Questionnaire

The ASA questionnaire [43] evaluates VAs using the most common constructs/concepts endorsed by the VA community. The questionnaire includes 24 items, including constructs such as acceptance, likeability and performance. Participants are asked to rate their experience with Erica on a 7-point Likert scale from -3 as strongly disagree to $+3$ as strongly agree. Additionally, participants are provided with the option to choose “Not Applicable” when they deem a question irrelevant to the given context. As an example, human-like appearance is measured with the item “Erica has the appearance of a human” on the 7-point Likert scale. Participants receive the 24 items in a random order.

3.5.5. Logfile Data

We collected all participant responses to Erica and the duration of their interaction. For the purpose of the data analysis we undertook, we used the logfile data to determine whether participants had actually interacted with Erica or not. Only participants who completed their interaction with Erica were included in the data analysis. It is crucial to ensure participants reviewed the emotion regulation strategies with Erica. This way, any change in their knowledge or confidence can be confidently linked to their interaction with the agent.

3.6. Data Analysis

We adopted a Bayesian approach over traditional frequentist tests, driven by the advantages offered by Bayesian statistics in providing a more intuitive interpretation of uncertainty [44,45]. Unlike classical hypothesis testing, a key benefit of the Bayesian test is the ability to directly quantify and update uncertainty using probability distributions instead of relying solely on p-values from null hypothesis significance testing. Bayesian statistics provide credible intervals (CIs), which are conceptually more straightforward than frequentist confidence intervals and offer a more intuitive measure of uncertainty, especially for noisy, low-sample data where Bayesian testing reduces false positives without sacrificing discovery sensitivity [46].

We utilised the first aid Bayesian R package [47] to apply Bayesian analysis to estimate the posterior distribution of models' parameters for each hypothesis. We assessed the models by checking the posteriors and reporting the posteriors' distribution of the means and the associated credible intervals. A 95% credible interval represents a range of values where the reported mean is likely to fall. If the credible interval excludes zero, it indicates evidence against the null hypothesis (i.e., there is a between-group difference).

4. Results

4.1. Participants

Figure 1 summarises participant flow. A total of 220 participants were recruited, completed the pre-interaction questionnaires and were randomly assigned to one of the four versions of Erica. Although all participants completed the study with the post-interaction questionnaires, only 152 out of the 220 reached the end of the conversation with their assigned Erica. Those who did not complete a full interaction ($n = 68$) were deemed ineligible for analysis, resulting in unequal numbers of participants in each of the four conditions. Additionally, 2 out of the 152 participants failed the attention check, resulting in 150 participants being deemed eligible for the analysis. The distribution of the participants across the four groups, along with their gender, is presented in Table 1.

Approximately 29% of the participants identified themselves with an Oceania background, 16% with South East Asian, 22% with a mixed background, 10% with North African and Middle Eastern, 7% did not associate themselves with any cultural group, and the remaining participants represented different backgrounds.

Table 1. Number of participants and gender distribution among the four experimental groups.

Acronym	Setting	N	Female (Male, Non-Binary)
RL+LS	Realistic-looking with lip-sync	37	78% (22%, 0%)
RL-LS	Realistic-looking without lip-sync	35	77% (17%, 1%)
CL+LS	cartoon-like look with lip-sync	28	89% (11%, 0%)
CL-LS	cartoon-like look without lip-sync	50	66% (32%, 2%)

4.2. Eeriness

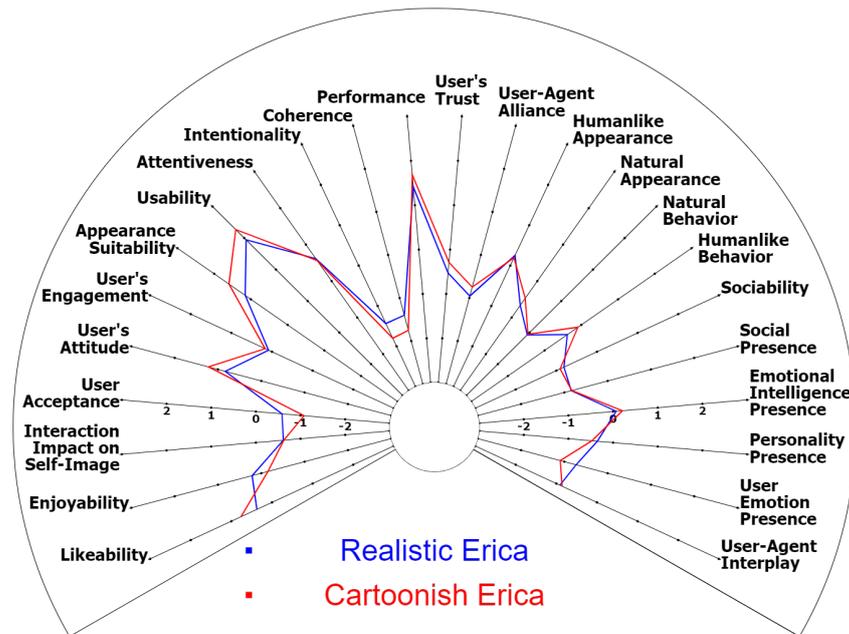
Bayesian analysis identified no significant between-group differences, including realism and lip-syncing differences on any scale of the eeriness questionnaire. Further, no interaction was identified between the look and the lip-syncing factors on the eeriness items. The eeriness scores of the four experimental groups are provided in Figure 3.

4.3. User-Agent Interaction Experience

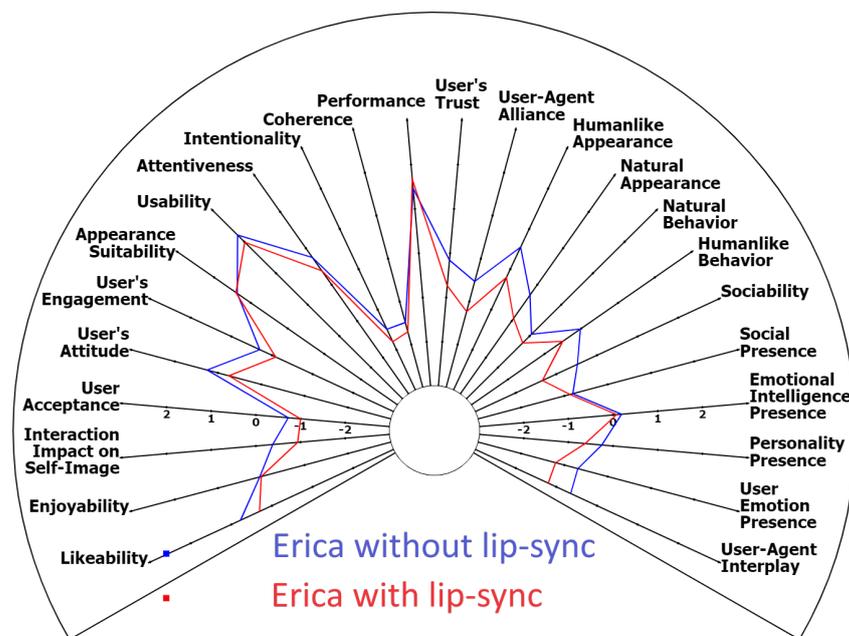
The participants' scores for the ASA items are presented in Figure 4a, which compares the look/realism settings, and Figure 4b, which compares the lip-syncing settings. The Bayesian independent t -test revealed only 2 of 28 items had significant differences between realistic and cartoon-like Erica, favouring the latter, in terms of (1) **usability** ($\Delta\mu = -0.32 [-0.63, -0.03]$) and (2) **attitude** ($\Delta\mu = -0.45 [-0.86, -0.04]$). The magnitude

of the means indicates the direction and strength of the observed difference between the two groups.

Further, the test revealed some differences between the lip-syncing (with vs. without lip-sync) groups: (1) **human-like appearance** ($\Delta\mu = 0.77 [1.3, 0.23]$), (2) **natural appearance** ($\Delta\mu = 0.66 [1.2, 0.17]$), (3) **sociability** ($\Delta\mu = 0.87 [1.4, 0.36]$), (4) **user-agent alliance** ($\Delta\mu = 0.72 [1.3, 0.19]$), (5) **user attitude** ($\Delta\mu = 0.50 [0.92, 0.091]$), (6) **interaction impact on self-image** ($\Delta\mu = 0.56 [1.1, 0.011]$), (7) **user emotion presence** ($\Delta\mu = 0.53 [1, 0.021]$) and (8) **user-agent interplay** ($\Delta\mu = 0.58 [1.1, 0.018]$), favouring Erica without lip-syncing.



(a)



(b)

Figure 4. ASA comparing Erica with the different settings. (a) Realistic-looking Erica vs. cartoon-like Erica. (b) Erica with lip-sync vs. Erica without lip-sync.

4.4. Knowledge Test

The mean scores of participants in the knowledge test were calculated pre- and post-interaction. Table 2 provides a detailed breakdown of these statistics for the four experimental groups, where knowledge mean values are within the 0 to 1 range and confidence scores are within the 1 to 7 range. Despite Erica's design, there was no notable change in the participants' knowledge gain. Nevertheless, a statistically significant increase in their confidence levels was observed. This increase extended to their confidence in recommending the strategies to others. No significant difference was found when comparing knowledge gain or confidence between the different experimental groups, stratified based on Erica's look, lip-syncing or into the four groups.

Table 2. Knowledge test statistics before and after interacting with Erica. The bold font in the last column indicates a significant change with a 95% confidence interval. The acronyms of the settings are listed in Table 1.

	Setting	Before Interaction		After Interaction		$\mu_1 - \mu_2$ [95% CI]
		μ_1	σ_1	μ_2	σ_2	
RL+LS	knowledge	0.84	0.23	0.83	0.23	−0.01 [−0.10, 0.07]
	confidence	4.98	1.00	5.93	1.00	0.95 [0.66, 1.20]
	recommending	4.26	0.99	5.11	0.98	0.85 [0.56, 1.20]
RL-LS	knowledge	0.76	0.26	0.77	0.25	0.01 [−0.08, 0.10]
	confidence	5.09	1.03	5.70	0.95	0.58 [0.35, 0.79]
	recommending	4.46	0.96	5.22	1.27	0.75 [0.47, 1.00]
CL+LS	knowledge	0.82	0.24	0.83	0.25	3.7×10^{-7} [− 9.7×10^{-5} , 9.8×10^{-5}]
	confidence	4.81	1.03	5.75	1.11	0.93 [0.57, 1.30]
	recommending	4.39	0.87	5.37	1.20	0.94 [0.55, 1.30]
CL-LS	knowledge	0.83	0.22	0.87	0.18	0.04 [−0.030, 0.11]
	confidence	4.89	0.79	5.81	0.86	0.93 [0.71, 1.10]
	recommending	4.32	0.67	5.33	1.06	0.99 [0.74, 1.20]

5. Discussion

The results revealed that, on average, participants showed a neutral impression towards the four VAs (Erica versions) on the eight scales of eeriness, as shown in Figure 3. This is further reflected in the analysis, where no significant between-group differences were captured in eeriness perception based on the appearance of VAs or the presence of lip-syncing. Previously, on the uncanny valley curve introduced in [9], the findings in [48] grouped the agents into six categories based on their appearance and the eeriness they cause. Animated agents, including cartoon-like and human-like agents, were found to be in the same category, which supports the idea that the eeriness source is not solely derived from the appearance but results from a blend of various design features (e.g., voice, animation and look) [10,12] or the application context [49]. This is supported by the feedback of a participant in our study who stated, "I was surprised by the progress of technology. This makes me feel a little creepy". This suggests that the students might not be fully prepared for such technology, and hence, the eeriness is not caused by a specific feature of the technology and context. As a future direction, it would be interesting to measure the students' acceptance of technology and to explore its influence on outcomes.

Further, in our investigation, the analysis failed to detect any interaction between appearance and lip-syncing. To draw a more comprehensive conclusion, further research with systematic variation in the levels of appearance and accuracy of lip-syncing is essential. However, it is not a straightforward matter to test the relationship to provide definitive results, due to differences in context and the purpose or role of the VAs. For example, the study [30] explored the perception of the uncanny valley in human-like VAs exhibiting different emotions and using different mouth shapes to create different magnitudes of mouth

movements, and it found that different VAs' emotions being displayed and participant gender both impacted the results. Also, despite having 20 combinations, none of their treatments included no lip movement.

As a further point concerning eeriness, it is possible that human-like Erica was not realistic enough to produce a strong uncanny valley effect. However, we were not trying to induce the uncanny valley effect. As explained in the Introduction, we had received comments from users of our health-related conversational agents (which includes Erica) concerning character freakiness, particularly related to lip-syncing. Before deciding to give users the option to switch off lip-syncing, we wanted to check the impact of disabling lip-syncing on our intended outcomes while also exploring whether a cartoon-like character would influence user experience.

The examination of the user-agent interaction experience indicated that the cartoon-like Erica was perceived significantly higher in terms of usability and user attitude (i.e., a favourable evaluation toward the interaction with the agent). Thus, Hypothesis 1 was not supported, as there were no significant differences in eeriness perception between groups, H1(a), and only 2 of the 24 dimensions of the user experience with Erica were different, H1(b).

While the examination of lip-syncing effects on eeriness perception did not yield any significant impact (neutral scores on all eeriness items), it did uncover a noteworthy influence on various dimensions of the user-agent interaction experience. The inclusion of lip-syncing might introduce elements that disrupt the overall perception of Erica, creating a divergence from the qualities associated with human-likeness, natural interaction and positive user experiences. Concerning the education context, this observation aligns with the proposition put forth in [50] that the design of pedagogical agents should incorporate only one feature, suggesting including voice without visual features, as the latter may introduce cognitive load and consequently have a detrimental effect on students' experiences. This is because humans naturally anticipate reactive listening, which includes facial and postural mirroring alongside speech content, rather than mere motor mimicry. This could clarify the withdrawal of a participant who provided feedback to cartoon-like with lip-syncing Erica, stating *"Large eyes is a feature exploited by many to invoke a feeling of cuteness to the viewers. Due to this, I found 'Erica' to be nothing but a nuisance"*. Unfortunately, a limitation of this study is that we did not measure the cognitive load aspects. Hence, we can conclude that the second hypothesis, H2, is partially supported. While the presence of lip-syncing did not impact the eeriness perception, H2(a), it negatively influenced the user-agent interaction experience in many dimensions, H2(b). This latter finding is consistent with [51], which found that including human speech with a real human appearance provided a positive effect, but a negative effect occurred when artificial speech was added to a VA.

The assessment of knowledge gain and confidence levels yielded insightful results and partially supports H3. While no substantial change in knowledge was evident, supporting H3(a), participants displayed a notable increase in confidence, both in their understanding and in recommending the strategies, following their interaction with Erica, which contradicts H3(b) and H3(c). Interestingly, no significant differences were observed among the various designs of Erica. The lack of change in knowledge gain may be attributed to the fact that only 60% of the participants reported not having studied the topic before. These findings align with the barriers highlighted in the literature, suggesting that novice psychologists may possess the necessary consultation knowledge but encounter challenges related to confidence in its practical application [40]. Given the absence of a significant correlation between eeriness, user experience and changes in knowledge or confidence, we posit the viability of employing Erica as a persuasive technology to motivate novice psychologists to enhance and apply their knowledge with others, regardless of whether lip-syncing is included or not. This supports our intention to deploy our VAs with the option to switch off lip-syncing according to the preference and needs of the individual.

6. Conclusions

The study presented in this paper explores and contributes to our understanding of the perception of VAs and their impact on behaviour change, focusing on two design aspects: eeriness perception and user-agent interaction experience. The absence of significant between-group differences suggests that appearance and lip-syncing did not distinctly affect eeriness perception. The study also uncovered that while lip-syncing did not significantly impact eeriness perception, it did negatively influence the user-agent interaction experience in various dimensions, indicating that the inclusion of lip-syncing may disrupt the overall perception of the VA. This is noteworthy because lip-syncing is a common VA animation, perhaps due to an assumption that it is expected by the user and/or that it is beneficial to the interaction experience.

The findings of this study have implications for both theory and practice relating to the design and use of conversational agents. From a theoretical perspective, our work confirms the importance of congruence and realism with respect to virtual agents, i.e., while lip movement is normal and expected in humans, lip movement and realistic appearance is not required in virtual humans. In fact, as reported by others, eeriness and uncanny valley are associated with a high level of realism. These findings go against a trend to increase realism both in appearance and lip movement in virtual agents and virtual reality models in general. From a practical perspective, while users have preferences regarding appearance and lip movement, these preferences do not necessarily impact the intended outcomes and benefits to humans. These findings suggest that the developers of virtual agents should focus more on the intended benefits of the virtual agent for the human, rather than on measurements such as believability, naturalness or liking. Specifically, VA developers should reconsider whether lip movement is included and/or whether they should allow the user to choose to switch it on or off.

Participants did not experience a substantial change in knowledge but exhibited increased confidence following interaction with the VA. This study highlights the importance of considering factors beyond knowledge acquisition, such as confidence, in evaluating the effectiveness of VA interactions in the education context.

Further research is warranted, specifically with diverse variations in the design features of VA, encompassing aspects such as appearance and voice, in combination with lip-syncing, in the area of education and behaviour change. Related to the findings reported in this study, in our upcoming study using realistic Erica, we are providing the option for users to turn off lip movement. That study will involve up to six interactions over a 3-week period, and we will turn lip-syncing on at the start of each session. We intend to use these data to determine how often people prefer to not have lip movement and whether they care enough to turn it off and to keep turning it off and to identify if there are any profiles or patterns in who chooses or when they choose to turn off lip movement.

Author Contributions: Conceptualization, A.A., K.H. and D.R.; methodology, A.A., K.H. and D.R.; software, A.A., K.H. and D.R.; validation, A.A., K.H. and D.R.; formal analysis, A.A.; investigation, A.A., K.H. and D.R.; resources, A.A., K.H. and D.R.; data curation, A.A.; writing—original draft preparation, A.A. and D.R.; writing—review and editing, A.A., K.H. and D.R.; visualization, A.A.; supervision, D.R.; project administration, D.R.; funding acquisition, K.H. and D.R. All authors have read and agreed to the published version of the manuscript.

Funding: This work is partially supported by Australian Research Council: DP20010213 and the Digital Health CRC Limited (“DHCRC”). DHCRC is funded under the Australian Commonwealth’s Cooperative Research Centres (CRC) Program.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and approved by the Human Research Ethics Committee of Macquarie University (reference code 520231587553549 and date of approval-10/10/2023).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Data is governed by a restricted access clause within the DHCRC Project Agreement. Requests to access data presented in this study should be directed to the corresponding author. The data are not publicly available due to privacy policies of the DHCRC Industry partner.

Acknowledgments: Thanks to Meredith Porte for technical assistance and to all participants involved in the study.

Conflicts of Interest: K.H. has received a research scholarship through the Digital Health CRC Limited and is an employee of Insurance Australia Group working within the Compulsory Third Party Insurance Business Unit. The remaining authors have no conflicts of interest to declare.

References

1. Yuan, X.; Chee, Y.S. Design and evaluation of Elva: An embodied tour guide in an interactive virtual art gallery. *Comput. Animat. Virtual Worlds* **2005**, *16*, 109–119. [[CrossRef](#)]
2. Aljameel, S.S.; O’Shea, J.D.; Crockett, K.A.; Latham, A.; Kaleem, M. Development of an Arabic conversational intelligent tutoring system for education of children with ASD. In Proceedings of the 2017 IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA), Annecy, France, 26–28 June 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 24–29.
3. Provoost, S.; Lau, H.M.; Ruwaard, J.; Riper, H. Embodied Conversational Agents in Clinical Psychology: A Scoping Review. *J. Med. Internet Res.* **2017**, *19*, e151. [[CrossRef](#)]
4. Ter Stal, S.; Kramer, L.L.; Tabak, M.; op den Akker, H.; Hermens, H. Design features of embodied conversational agents in ehealth: A literature review. *Int. J. Hum.-Comput. Stud.* **2020**, *138*, 102409. [[CrossRef](#)]
5. Loveys, K.; Sebaratnam, G.; Sagar, M.; Broadbent, E. The effect of design features on relationship quality with embodied conversational agents: A systematic review. *Int. J. Soc. Robot.* **2020**, *12*, 1293–1312. [[CrossRef](#)]
6. Ruhland, K.; Peters, C.E.; Andrist, S.; Badler, J.B.; Badler, N.I.; Gleicher, M.; Mutlu, B.; McDonnell, R. A Review of Eye Gaze in Virtual Agents, Social Robotics and HCI: Behaviour Generation, User Interaction and Perception. *Comput. Graph. Forum* **2015**, *34*, 299–326. [[CrossRef](#)]
7. Gan, Q.; Liu, Z.; Liu, T.; Zhao, Y.; Chai, Y. Design and user experience analysis of AR intelligent virtual agents on smartphones. *Cogn. Syst. Res.* **2023**, *78*, 33–47. [[CrossRef](#)]
8. Thaler, M.; Schlögl, S.; Groth, A. Agent vs. avatar: Comparing embodied conversational agents concerning characteristics of the uncanny valley. In Proceedings of the 2020 IEEE International Conference on Human-Machine Systems (ICHMS), Rome, Italy, 7–9 September 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 1–6.
9. Mori, M. the uncanny valley. *Energy* **1970**, *7*, 33–35. [[CrossRef](#)]
10. Abdulrahman, A.; Richards, D.; Bilgin, A.A. A Comparison of Human and Machine-Generated Voice. In Proceedings of the 25th ACM Symposium on Virtual Reality Software and Technology, New York, NY, USA, 12–15 November 2019. [[CrossRef](#)]
11. Isbister, K.; Nass, C. Consistency of Personality in Interactive Characters: Verbal Cues, Non-Verbal Cues, and User Characteristics. *Int. J. Hum.-Comput. Stud.* **2000**, *53*, 251–267. [[CrossRef](#)]
12. Mitchell, W.J.; Szerszen Sr, K.A.; Lu, A.S.; Schermerhorn, P.W.; Scheutz, M.; MacDorman, K.F.J.i.P. A mismatch in the human realism of face and voice produces an uncanny valley. *i-Perception* **2011**, *2*, 10–12. [[CrossRef](#)]
13. Zanbaka, C.; Goolkasian, P.; Hodges, L. Can a virtual cat persuade you?: The role of gender and realism in speaker persuasiveness. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, New York, NY, USA, 22 April 2006. [[CrossRef](#)]
14. Lee, E.J. The more humanlike, the better? How speech type and users’ cognitive style affect social responses to computers. *Comput. Hum. Behav.* **2010**, *26*, 665–672. [[CrossRef](#)]
15. Dickerson, R.; Johnsen, K.; Raji, A.; Lok, B.; Stevens, A.; Bernard, T.; Lind, D.S. Virtual patients: Assessment of synthesized versus recorded speech. *Stud. Health Technol. Inform.* **2006**, *119*, 114–119.
16. Torre, I.; Latupeirissa, A.B.; McGinn, C. How context shapes the appropriateness of a robot’s voice. In Proceedings of the 2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN), Naples, Italy, 31 August–4 September 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 215–222. [[CrossRef](#)]
17. Bahrack, L.E.; Hollich, G. Intermodal Perception. In *Reference Module in Neuroscience and Biobehavioral Psychology*; Elsevier: Amsterdam, The Netherlands, 2017. [[CrossRef](#)]
18. Torre, I.; Carrigan, E.; Domijan, K.; McDonnell, R.; Harte, N. The Effect of Audio-Visual Smiles on Social Influence in a Cooperative Human-Agent Interaction Task. *ACM Trans. Comput. Hum. Interact.* **2021**, *28*, 1–38. [[CrossRef](#)]
19. Tinwell, A.; Grimshaw-Aagaard, M.; Williams, A. Uncanny behaviour in survival horror games. *J. Gaming Virtual Worlds* **2010**, *2*, 3–25. [[CrossRef](#)]
20. Milcent, A.S.; Kadri, A.; Richir, S. Using facial expressiveness of a virtual agent to induce empathy in users. *Int. J. Hum. Comput. Interact.* **2022**, *38*, 240–252. [[CrossRef](#)]
21. Peixoto, B.; Melo, M.; Cabral, L.; Bessa, M. Evaluation of animation and lip-sync of avatars, and user interaction in immersive virtual reality learning environments. In Proceedings of the 2021 International Conference on Graphics and Interaction (ICGI), Porto, Portugal, 4–5 November 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 1–7.

22. McDonnell, R.; Breidt, M.; Bühlhoff, H.H. Render me real? investigating the effect of render style on the perception of animated virtual humans. *ACM Trans. Graph.* **2012**, *31*, 1–11. [[CrossRef](#)]
23. Basori, A.H.; Ali, I.R. Emotion expression of avatar through eye behaviors, lip synchronization and MPEG4 in virtual reality based on Xface toolkit: Present and future. *Procedia-Soc. Behav. Sci.* **2013**, *97*, 700–706. [[CrossRef](#)]
24. Pesonen, H.; Leinonen, J.; Haaranen, L.; Hellas, A. Exploring the Interplay of Achievement Goals, Self-Efficacy, Prior Experience and Course Achievement. In Proceedings of the 2023 Conference on United Kingdom & Ireland Computing Education Research, Swansea, Wales, UK, 7–8 September 2023; pp. 1–7. [[CrossRef](#)]
25. Elshan, E.; Zierau, N.; Engel, C.; Janson, A.; Leimeister, J.M. Understanding the design elements affecting user acceptance of intelligent agents: Past, present and future. *Inf. Syst. Front.* **2022**, *24*, 699–730. [[CrossRef](#)]
26. Im, H.; Sung, B.; Lee, G.; Kok, K.Q.X. Let voice assistants sound like a machine: Voice and task type effects on perceived fluency, competence, and consumer attitude. *Comput. Hum. Behav.* **2023**, *145*, 107791. [[CrossRef](#)]
27. Zhou, M.X.; Mark, G.; Li, J.; Yang, H. Trusting virtual agents: The effect of personality. *ACM Trans. Interact. Intell. Syst. (TiiS)* **2019**, *9*, 1–36. [[CrossRef](#)]
28. Baylor, A.L.; Kim, Y. Pedagogical agent design: The impact of agent realism, gender, ethnicity, and instructional role. In *Proceedings of the International Conference on Intelligent Tutoring Systems*; Springer: Berlin/Heidelberg, Germany, 2004; pp. 592–603.
29. Salehi, V.; Nia, F.T. Effect of levels of realism in mobile-based pedagogical agents on health e-learning. *Future Med. Educ. J.* **2019**, *9*, 40–45.
30. Tinwell, A.; Grimshaw, M.; Williams, A. Uncanny speech. In *Game Sound Technology and Player Interaction: Concepts and Developments*; IGI Global: Hershey, PA, USA, 2011; pp. 213–234.
31. Richards, D.; Miranda Maciel, P.S.; Janssen, H. The Co-Design of an Embodied Conversational Agent to Help Stroke Survivors Manage Their Recovery. *Robotics* **2023**, *12*, 120. [[CrossRef](#)]
32. Phan, N.T.T.; Chen, C.H. Taiwanese Engineering Students' Self-Efficacy and Academic Performance. *Arab. World Engl. J.* **2022**.
33. Lucero, K.S.; Chen, P. What do reinforcement and confidence have to do with it? A systematic pathway analysis of knowledge, competence, confidence, and intention to change. *J. Eur. CME* **2020**, *9*, 1834759. [[CrossRef](#)] [[PubMed](#)]
34. Owens, K.M.; Keller, S. Exploring workforce confidence and patient experiences: A quantitative analysis. *Patient Exp. J.* **2018**, *5*, 97–105. [[CrossRef](#)]
35. Bandura, A. Self-efficacy: Toward a unifying theory of behavioral change. *Psychol. Rev.* **1977**, *84*, 191. [[CrossRef](#)]
36. Bakker, D.; Kazantzis, N.; Rickwood, D.; Rickard, N. A randomized controlled trial of three smartphone apps for enhancing public mental health. *Behav. Res. Ther.* **2018**, *109*, 75–83. [[CrossRef](#)] [[PubMed](#)]
37. Farley, H. Promoting self-efficacy in patients with chronic disease beyond traditional education: A literature review. *Nurs. Open* **2020**, *7*, 30–41. [[CrossRef](#)]
38. Chang, C.Y.; Hwang, G.J.; Gau, M.L. Promoting students' learning achievement and self-efficacy: A mobile chatbot approach for nursing training. *Br. J. Educ. Technol.* **2022**, *53*, 171–188. [[CrossRef](#)]
39. Brinsley, J.; Singh, B.; Maher, C.A. A digital lifestyle program for psychological distress, wellbeing and return-to-work: A proof-of-concept study. *Arch. Phys. Med. Rehabil.* **2023**, *104*, 1903–1912. [[CrossRef](#)]
40. Meaghan C. Guiney, Abigail Harris, A.Z.; Cancelli, A. School Psychologists' Sense of Self-Efficacy for Consultation. *J. Educ. Psychol. Consult.* **2014**, *24*, 28–54. [[CrossRef](#)]
41. Hopman, K.; Richards, D.; Norberg, M.M. A Digital Coach to Promote Emotion Regulation Skills. *Multimodal Technol. Interact.* **2023**, *7*, 57. [[CrossRef](#)]
42. Ho, C.C.; MacDorman, K.F. Revisiting the uncanny valley theory: Developing and validating an alternative to the Godspeed indices. *Comput. Hum. Behav.* **2010**, *26*, 1508–1518. [[CrossRef](#)]
43. Fitrianie, S.; Bruijnes, M.; Li, F.; Abdulrahman, A.; Brinkman, W.P. The Artificial-Social-Agent Questionnaire: Establishing the Long and Short Questionnaire Versions. In Proceedings of the 22nd ACM International Conference on Intelligent Virtual Agents, IVA '22, New York, NY, USA, 6–9 September 2022. [[CrossRef](#)]
44. Sidebotham, D.; Barlow, C.J.; Martin, J.; Jones, P.M. Interpreting frequentist hypothesis tests: Insights from Bayesian inference. *Can. J. Anesth./J. Can. d'anesthésie* **2023**, *70*, 1560–1575. [[CrossRef](#)]
45. Wagenmakers, E.J.; Marsman, M.; Jamil, T.; Ly, A.; Verhagen, J.; Love, J.; Selker, R.; Gronau, Q.F.; Šmíra, M.; Epskamp, S.; et al. Bayesian inference for psychology. Part I: Theoretical advantages and practical ramifications. *Psychon. Bull. Rev.* **2018**, *25*, 35–57. [[CrossRef](#)]
46. Benavoli, A.; Corani, G.; Demšar, J.; Zaffalon, M. Time for a Change: A Tutorial for Comparing Multiple Classifiers through Bayesian Analysis. *J. Mach. Learn. Res.* **2017**, *18*, 2653–2688.
47. Bååth, R. Bayesian first aid: A package that implements Bayesian alternatives to the classical *.test functions in R. *Proc. useR* **2014**, *2014*, 2.
48. Mathur, M.B.; Reichling, D.B. Navigating a social world with robot partners: A quantitative cartography of the Uncanny Valley. *Cognition* **2016**, *146*, 22–32. [[CrossRef](#)]
49. Sharma, M.; Vemuri, K. Accepting human-like avatars in social and professional roles. *ACM Trans. Hum.-Robot Interact. (THRI)* **2022**, *11*, 1–19. [[CrossRef](#)]

-
50. Schroeder, N.L.; Adesope, O.O. A Systematic Review of Pedagogical Agents' Persona, Motivation, and Cognitive Load Implications for Learners. *J. Res. Technol. Educ.* **2014**, *46*, 229–251. [[CrossRef](#)]
 51. Gurung, N.; Grant, J.B.; Hearth, D. The Uncanny Effect of Speech: The Impact of Appearance and Speaking on Impression Formation in Human–Robot Interactions. *Int. J. Soc. Robot.* **2023**, 1–16. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.