

Virtual Avatars as Children Companions : For a VR-based Educational Platform: How Should They Look Like?

Elsa Thiaville, Jean-Marie Normand, Joe Kenny, Anthony Ventresque

▶ **To cite this version:**

Elsa Thiaville, Jean-Marie Normand, Joe Kenny, Anthony Ventresque. Virtual Avatars as Children Companions : For a VR-based Educational Platform: How Should They Look Like?. ICAT-EGVE 2020 - International Conference on Artificial Reality and Telexistence and Eurographics Symposium on Virtual Environments, Dec 2020, Orlando, Florida, United States. pp.1-9. hal-03068405

HAL Id: hal-03068405
https://hal.science/hal-03068405

Submitted on 15 Dec 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

Virtual Avatars as Children Companions For a VR-based Educational Platform: How Should They Look Like?

Elsa Thiaville[1], Jean-Marie Normand[2], Joe Kenny[3] and Anthony Ventresque[1]

[1]SFI Lero & School of Computer Science, University College Dublin, Ireland
[2]École Centrale de Nantes, AAU UMR CNRS 1563, Inria Hybrid, France; [3]Zeeko, Ireland

Abstract

Virtual Reality (VR) has the potential of becoming a game changer in education, with studies showing that VR can lead to better quality of and access to education. One area that is promising, especially for young children, is the use of Virtual Companions that act as teaching assistants and support the learners' educational journey in the virtual environment. However, as it is the case in real life, the appearance of the virtual companions can be critical for the learning experience. This paper studies the impact of the age, gender and general appearance (human- or robot-like) of virtual companions on 9-12 year old children. Our results over two experiments (n=24 and n=13) tend to show that children have a bigger sense of Spatial Presence, Engagement and Ecological Validity when interacting with a human-like Virtual Companion of the Same Age and of a Different Gender.

1. Introduction

Good education is one of the most important goals for parents and society, with large amounts of resources spent on teaching and learning, and numerous experts working on designing and implementing better education, both in terms of quality and access. Technology has the potential to become a game changer for education: there are many resources students can access online, applications are being developed to support or complement education, etc. Recently, the development of consumer grade Virtual Reality (VR) technologies also lead to an increase of interest towards *virtual learning* where students are immersed in a Virtual Environment (VE) to learn, through, e.g., experiential learning which has proven to have a huge potential [JTMT09]. However getting the best of VR in education requires to get the embodiment right as the virtual world can have a profound impact on the quality of learning. While this question is well documented in the gaming context [Lan11], where, for instance, the question of the correlation between characters' appearance and gaming outcomes has been studied, to the best of our knowledge this has not been explored in the educational context. This question is critical in the education domain though - as are many dimensions associated with the learning experience - for what's at stake: better (or worse) learning outcomes. In the real world, the question of the impact of the appearance of teachers and teaching assistants is disputed (see for instance [MM05, MD09]) but there seems to be an impact of various elements on academic interests and outcomes based on personas/items in the education environment seems pretty clear (see for instance Cheryan et al. [CPDS09] which shows that objects can prevent or increase women's interest in computer science). Now, if the question of the appearance of teachers and teaching assistants is relevant for education and knowing that VR gives us the possibility to create and animate characters as much as we like - then it is of great importance to offer designers of virtual companions and teaching assistants some recommendations on how to create their characters. It is especially important as a large proportion of these virtual companions are going to interact with children, for which virtual worlds can be intimidating and need a calming figure that they can look up to when they have questions or do not feel well.

This paper investigates the question of the impact on children of the appearance of Virtual Companions (VCs) and in particular their age, gender and whether they are human- or robot-like. We perform two experiments ($n = 24$ and $n = 13$) on 9-12 years old children and evaluate their Spatial Presence (SP), Engagement, Ecological Validity (EV) and Negative Effects (NE); we also use objective metrics (time to complete, number of errors, etc.) to assess their performance in the VE. We observe that elementary school children (again, 9-12 years old) have a higher sense of presence and of ecological validity as well as show a better engagement when paired with learning companions of the same age and of a different gender. Moreover, we find that children's performance are comparable when they interact with a human-like or a robotic companion. The main contribution of this paper is to address this question of VC's appearance in virtual learning environment and we think that our experimental setup and conclusions will be of particular interest for the academic community and for education professional involved in the design and implementation of VR based educational contents containing virtual companions.

The remainder of this paper is organised as follows: section 2 presents the related work; section 3 introduces our research hy-

potheses and the VE used in our experiment); section 4 details and analyses our experiment where we study the impact of a VC's age and gender on school children; section 5 then analyses the impact of the type (human- or robot-like) of a VC on school children; section 6 discusses our findings and section 7 concludes this paper.

2. Related work

When it comes to Virtual Companions used in VR, they should be visually present in the VE (as it will increase motivation) and not only heard by the user. In some situations, it can be better to have multiple VCs, each with a different functionality, to help segment the information, e.g., "in a learning system motivational support is best kept separate from instructional information" [Bay09].

Moreover, VCs can be seen as social models and should be adapted to the user, as mentioned by Baylor [Bay09]: "for younger students, female agents may be more powerful role/social models overall, perhaps because of both parental influences and the fact that most schoolteachers are female". Challenging existing stereotypes (e.g., in engineering and STEM fields) with a VC has proven to be effective in learning for middle-school students [PBDRK09]. Of course this is context dependent, but overall people in VR tend to behave according to stereotypes, especially when embodied in virtual avatars, see e.g., [SS14]. In addition, in VR, users' behavior can change depending on the body that represents them: this effect has been coined by Yee et al. [YB07] as the Proteus Effect. Zanbaka et al. [AZGH06] suggest that gender stereotypes occurring in real life, also prevail in VR. Indeed, the authors showed that seems that when it comes to real speakers, participants are more convinced by the opposite gender, and that virtual speakers have a persuasion force similar to the one of real speakers.

In terms of collaboration with a VC, Yoon et al. [YKL*19] showed that in a remote collaborative situation in Augmented Reality (AR), participants preferred, in terms of social presence (i.e., the sense of being together), to interact with VCs represented as a whole-bodied avatar over just a upper body or only head & hands. They also showed that the preferred visual appearance of the VC (realistic or cartoonish) depended on the context of the situation (e.g., friendly conversation or work-related). Others studies focused on the attention (i.e., the percentage of time spent visually focusing on the VC) received by the agent. Walker et al. [WSR19] compared the effects of a small-sized VC compared to a VC of the same size as the user. Equal-sized VCs had significantly more influence on the user than small ones, who also received significantly less attention. Ennis et al. [EHO15] measured (via eye-tracking) the attention the user would put on a VC's body parts. Results show that users mostly stare at the VC's torso regardless of its motion and gender.

Schools recently started using interactive platforms for learning, which are thought to increase students' motivation by adding a VC that emotionally reacts to their answers [TA12] or by attracting their curiosity with auditory embodiment [DOC03]. Teachers can also present information in a more organized way [KHAL*09], especially in the context of specific needs, like autistic children. Nevertheless, all of these platforms are screen-based. Shin [Shi18] focused on the differences between video-based and VR-based content in the context of storytelling to stimulate empathy and embodied experience. Participants were grouped by personality traits (low or high empathy) and results suggest that more immersive techniques may induce more presence, which is positively correlated to embodiment. However, each user has its own understanding of a story and the empathy generated highly depends on personality: "In other words, VR developers propose immersion but users process it, based on their own preferences and needs". For movement learning, it appears that the learning is significantly better when virtual reality is used than when it is 2D platform [BPN*08, PBHJ*06].However, there is a shortcoming regarding VR-learning: most VR Head Mounted Displays (HMDs) are designed for 13+ years-old. The exact effects of VR on a child's development being still unclear [ON17]. Moreover, unlike videos, VR stages situations that are seen from a first-person point of view and it has been shown that children are more malleable to this kind of stimulus [GQ08]. Indeed, elementary school (6-7 years-old) children are more likely to create false memories after the use of VR than after 2D imagery [SB09]. Note that for younger children, both methods may cause false memory acquisition. As a consequence, one should be careful when manipulating VR with children.

3. User Experiments

We designed two experiments to evaluate the impact of a VC's visual appearance in a virtual learning context, a Virtual Environment where children had three tasks to perform. Based on our literature review, we chose to focus first on the age and gender of the Virtual Companion before studying differences when using a human-like or a robotic avatar. Before presenting the VE and the metrics used in both experiments, we detail our research hypotheses.

3.1. Research Hypotheses

In [AZGH06], Zanbaka et al. found that adult participants were more persuaded by a VC (displayed on a screen) representing the opposite gender. Considering this, our first hypothesis is that:

H1-a: Participants will have a stronger sense of Spatial Presence and Engagement when paired with a VC of the opposite gender.
H1-b: Participants will perform better (i.e., quicker and with less errors) when paired with a VC of the opposite gender.

Moreover, Baylor [Bay09] states that "while providing a social model from the same in-group as the user is generally advantageous, there are certain contexts where the opposite may be better". This leads us to study the impact of a VC that (i) "looks like you" or (ii) is different from you has in an educational context with children. We define a VC that "looks like you" to be of the same age, the same ethnic group and the same gender (the one with which the child identifies most). In this study, we decided to focus only on age and gender since both variables could be more easily split into a "same as me" and "different from me" than ethnic groups (which would give more possibilities in differences and would be prone to stereotype activation). Considering ethnic groups would also require to add participants. In a different study concerning the design of motivational agent, Baylor [Bay11] mentions that it is better to provide with an "an attractive agent that resembles them with respect to gender, ethnicity/race, age, and perceived competency in the domain". Considering this, we completed our hypothesis with:

H2-a: Participants will have a stronger sense of SP and Engagement when paired with a VC of the same age.

H2-b: Participants will perform better when paired with a VC of the same age.

On the other hand, in their study, Zanbaka et al. [AZGH06], showed that the degree of realism of the VC had no impact on the degree persuasion of the user. As this finding is in contradiction with that of Baylor [Bay11], we wanted to study the impact of the type of VC on school children. We thus hypothesized that:

H3-a: Participants will have the same feeling of Spatial Presence and Engagement with a human-like and a non-human-like VC.

H3-b: Participants will exhibit comparable performance (in terms of objective metrics) when interacting with a human-like VC or with a non-human-like VC.

3.2. Experimental Protocol and Apparatus

The experimental protocol, the apparatus, the Virtual Environment, the tasks and the measures were similar in both experiments. Therefore, we detail them before focusing on the two experiments.

Both experiments were conducted with an Oculus Quest HMD. The Oculus Quest has a resolution of 1440×1600 pixels per eye (2880×1600 total resolution), displayed at 72Hz, has a field of view (FOV) of $\sim 100^{\circ}$ and weighs 571g. The Quest was chosen for multiple reasons: it is wireless so participants can walk freely; it provides controllers that allow for easy interaction with the VE as well as headphones which allowed our participants to hear and interact vocally with the VC. In order for the VC to understand and reply to the participant we used a Wizard-of-Oz approach: the experimenter (located in the same room as the child, thus hearing the vocal interaction) typed the sentences that the virtual companion would speak on a computer connected to the same remote server as the Oculus Quest. Those typed sentences were transformed into speech in real-time by the Microsoft Cognitive Services for Unity [Mic20a]. Since only the HMD and the controllers are tracked, our participants did not have a full virtual body in the VE but only saw a representation of their virtual hands (see Figure 1a). Argelaguet et al. [AHTL16] investigated the sense of agency and ownership with different representations of a virtual hand (ball, skeleton hand and realistic arm) and found that the sense of agency was stronger for realistic hands but that the sense of ownership was stronger with the real arm. Our hand is a middle ground between their representations: it has the shape of a hand but no arm. This choice is also supported by the work of Lugrin et al. [LEK*18] where no significant difference was found over body ownership, immersion, emotional and cognitive involvement and over perceived control between the different conditions: (i) seeing only the controllers, (ii) seeing hands, (iii) seeing hands, forearms with a visible body.

Before starting the experiment, children were explained that they would be immersed together with a VC, in a VE. They were told the VC was here to help them accomplish the different tasks. They were presented the three tasks they would have to perform and told they could ask (orally) the VC for help whenever they wanted or needed to. Prior to the study, participants spent 10 minutes inside the HMD with the Oculus tutorial to ensure that they know how to use the controllers. This was followed by a small break during which they were asked if they felt well, were told that the experiment was about to begin and that they could stop at any time without justification. Then, they had to put on the HMD again, and the experiment started. They spent 10 minutes in the VE with a VC as a learning companion. Overall, the experiment lasted for 40 minutes (20 minutes inside the HMD, 15 minutes for the questionnaire and 5 minutes for breaks and gearing-up time). When entering the environment, the VC greeted the student: it waived at them, verbally welcomed them and reminded them they could ask for help at anytime. During the whole experiment, the VC was making sure that the student was not struggling, by asking whether or not they needed help. It would provide verbal cues such as an indication to look at a specific spot, the signification of a word or the location of a card if needed. If the student was staking more time than a predefined time for a task, it would provide help.

3.3. The Virtual Learning Environment

Both experiments took place in a $4m^2$ room as shown in Figure 1c. This room contained all the material necessary for the participant to carry out the learning tasks. Six paintings hung on the walls of the room and on the South wall, 10 numbered tiles were used by the participants to indicate how many paintings are in the room.A wooden table was positioned in the North-West corner of the room. Ten objects lied on the table: two bananas, an apple, a pear, a cucumber, a zucchini, an onion, a garlic clove, some binoculars and a camera. Next to the table, 5 purple cubes hung on the wall, each of them having a label printed above them (see Figure 1b). On the room's floor lied 12 cards used for a memory card game (see subsection 3.4). There were 6 pairs of cards created specifically for this experiment (using the patterns presented in Figure 2), the back of the cards (a red pattern) were originally presented to participants (see Figure 1b). To select the cards, participants had to step on a card to select it (it turned green) and then press a button to flip it. If the pair was found, both cards stayed with the pattern visible otherwise they returned to their original state (back visible).

3.4. Learning Tasks

Evaluating the impact of the appearance of a VC in a learning environment is a complex task. Indeed, to do so, we would have to ensure that all students have a similar level of knowledge in the subject concerned. This implies either:

- using very simple tasks, the VC would not have any impact since the pupils would not need to interact with it;
- having access to a larger number of pupils to flatten the differences among them;
- choosing a subject in all pupils have the same level, that is difficult enough so that the VC could be of help and that leaves room for significant and measurable progress.

After discussing with a psychologist specialized in children learning, we decided to rely on playful tasks that nevertheless prove challenging and require an intellectual effort from the pupils. In a similar way, the total duration of the experiment was set to 10 minutes in accordance to school's regulations and to respect children's attention time. Therefore, we designed three tasks, each with its own purpose, performed in the following order: a counting task, a fetching task and a memory task. The first task allowed the participant to familiarize with the VE and the VC. The second task



(a) (b) (c) (d)

Figure 1: Participants experienced a first-person perspective of (a) the Virtual Environment and their virtual hands; (b) the Virtual Companion. (c) Top view of the learning Virtual Environment. (d) Participants had to fetch objects on the table (right) and put them in purple boxes (left).

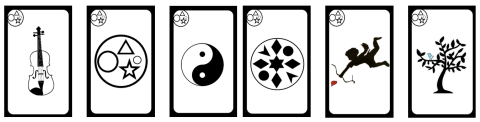


Figure 2: The 6 patterns used in the Memory task.

aimed at studying the willingness to ask a VC for help while the third task aimed at studying their ability to concentrate to solve the game. During the whole experiment we want the VC to be seen as a teaching companion, someone that would provide help on a given subject: the VC would not be the one giving the tasks. Hence, at the beginning of each task, textual indications appeared directly in the HMD to let participants know what they had to do. However, when help was needed the avatar would walk closer to the student without entering its personal space (as shown in Figure 1b, using a minimal distance of 0.7m). At the beginning, in the middle and at the end of a task, the avatar spoke to ensure that everything was going well and encourage dialogue.

The counting task The first task was designed to allow participants to familiarize with the VE, the presence of the VC and with the interaction mechanisms (grabbing objects with the controllers and moving around in the virtual room, achieved by actually walking in the real environment). In this task, participants had to count the number of paintings displayed in the room. To answer this question, participants had to grab a green sphere, hanging next to the numbered tiles and put it in the tile which number corresponded to the actual number of paintings (6) in the room. When they put the sphere in the correct tile, the tile turned green, otherwise it turned red. The task ended when the participant managed to put the sphere in the correct tile.

The fetching task. The second task is a fetching task, participants had to find 5 objects in the room: binoculars, a camera, two bananas, an onion (*alium cepa*) and a cucumber (*cucumis sativus*), see Figure 1d. Participants had to grab the object with the controller and put it in the corresponding purple bin. The bin turned green when it contained the correct object and red otherwise. No order was imposed to find the objects, and participants could ask the VC for help at any time. We used the onion's and the cucumber's scientific names to complexify the task and to "encourage" children to ask the VC for help.

The memory task. The third task is a classical memory card game, which participants played alone. This is an educational game, as opposed to "recreational" games, so it fits in our context. The goal is to find as many pairs as possible with or without the help of the VC, in the remaining time (i.e., 10 minutes minus the time taken by participants to achieve the two previous tasks). As a consequence, participants who performed better in the two first tasks had more time to find pairs. In this memory task, 12 cards are laid in 3 lines of 4 with, at first, only their back visible to participants (see Figure 1b). Participants could pick pairs of cards by turning first one and then the other. If both cards had the same face, participants scored a point and could go on to finding a new pair. Otherwise, both cards would return to being not visible and participants could select a new pair.

3.5. Metrics

In order to evaluate the impact of the virtual companion on the learning tasks, we used two types of metrics: (i) objective metrics, i.e., the outcomes of our three learning tasks (e.g., number of success, time taken to perform a task, etc.) and (ii) questionnaires to measure participants' engagement and social presence.

Objective metrics. We first report a measure over the whole experiment: the number of times participants interacted with the Virtual Companion ($\#_{VC}$). Moreover, we recorded some task-specific measures. For the first task, we report: $\#_{T1}$: the number of trials before finding the correct number of paintings. For the second task, we manually report $Strat_{T2}$: the strategy used by participants to find the correct objects: Help when participants asked the VC for help or $T\&E$ when participants used a trial and error strategy (i.e., putting all the objects one by one until they found the correct one). Finally, for the third task, we report: (i) $\#_{pairs}$: the number of pairs found by participants; (ii) T_{T3}: the time spend in the third task. This last measure served as a global estimation of how participants performed overall, indeed as the experiment lasted 10 minutes, the longer participants tried task 3, the quicker they solved tasks 1 and 2. **Questionnaire.** The choice of the questionnaire was carried out after a careful discussion with a psychologist specialized in chil-

dren. The questionnaire had to be: phrased in a simple way and short so that children did not got bored. We therefore chose to rely on the ITC-SOPI questionnaire [LFDK01] that allowed us to assess SP, Engagement, EV and NE of our system. Other questionnaires could have been used to also try to evaluate e.g., social presence but using multiple questionnaires was advocated against by the psychologist, thus we used only the ITC-SOPI. The questionnaire was composed of 38 questions in total, each one rated using a 5-point Likert scale where 1 corresponded to a strong disagreement. Children could ask the experimenter for help with the questions if they had trouble understanding them. According to the ITC-SOPI guidelines, 4 dimensions emerge out of the 38 questions, each one having a score between 1 and 5 computed as follows:

- Spatial Presence: mean value of 19 questions;
- Engagement: mean value of 13 questions;
- Ecological Validity: mean value of 5 questions;
- Negative Effects: mean value of 6 questions.

To evaluate how the participants felt about the VC, we looked at questions Q_{23} (*"I had the sensation that the characters were aware of me"*) and Q_{35} (*" I had the sensation that parts of the displayed environment (e.g. characters or objects) were responding to me."*). After the questionnaire, the informal discussion allowed us to get the children overall feeling about the experiment as well as some remarks they could have.

4. Experiment 1: Influence of the Virtual Companion's visual appearance

This experiment was designed to address our research hypotheses **H1** and **H2** on the impact of the visual appearance (age and gender) of the Virtual Companion on learning children. To do so, we designed a within-group experiment ($n = 24$) with two independent variables (age and gender of the VC) where participants were randomly assigned to one of four groups:

- Same Age Same Gender (SASG): the VC is of the Same Age (i.e., a child) and of the Same Gender as the participant ($n = 5$);
- Same Age Different Gender (SADG): the VC is of the Same Age and of a Different Gender ($n = 6$);
- Different Age Same Gender (DASG): the VC is of a Different Age (i.e., an adult) and of the Same Gender ($n = 6$);
- Different Age Different Gender (DADG): the VC is of a Different Age and of a Different Gender ($n = 7$).

Participants were immersed in the VE along with the VC corresponding to their group (see subsection 4.1) and had to perform the tasks described in subsection 3.4.

4.1. Human-Like Avatars

In the first experiment, the VC is human-like. We designed four avatars (one for each condition, see Figure 3) using MakeHuman [Mak20] and Mixamo [Ado20] was used to fully rig and animate them (each possessed two animations: idle and walking). Here, we vary the age and sex of the avatar, but the skin, hair and eye colors do not vary. According to Baylor's work [Bay11], it is preferable to use an avatar that resembles the participant in terms of ethnicity, among other things. In their study on the prediction of pigmentation of individuals according to their genotype, Walsh et al. [WLW*13] conducted a survey in several countries, including the one where our experiment took place. According to their surveys, 77% of the population in this country has light eyes (blue, green or heterochromic). As for hair, 51% of those polled had dark brown or black hair and 32% had dark blond hair. Therefore, our avatars were designed with dark hair and light eyes (Figure 3). Avatars' voices were generated using Microsoft Speech synthesizers [Mic20b].



Figure 3: Human-like avatars. From left to right: adult male, adult female, child male and child female.

4.2. Participants

For the first experiment, participants were initially 27 elementary-school students, but due to technical issues during the experiment, we had to remove 3 of them from the statistical analysis. The remaining 24 participants (12 boys and 12 girls) were aged between 11 and 12 (M=11.46 SD=0.51). The school was a mixed school and parents filled out consent forms before the beginning of the experiment. Children did not know the purpose of the experiment. All of them had used a VR equipment at least once before.

4.3. Results

Data was analyzed using one-way ANOVA with $\alpha = 0.05$ as significance level. We studied the influence of the type of virtual companion (SASG, SADG, DASG or DADG) on Spatial Presence, Engagement, Ecological Validity, Negative Effects, the number of trials for Task 1, the strategy used by participants in Task 2 (help or trial and error), the number of pairs found in Task 3 (cf. Table 1).

4.3.1. Influence of the Virtual Companion's visual appearance

The statistical analysis showed an effect of the VC's visual appearance over SP ($F_{(3,20)} = 4.935; p = 0.01$) and over Engagement ($F_{(3,20)} = 4.917; p = 0.010$). No influence of the avatar was found on EV ($F_{(3,20)} = 1.951; p = 0.154$) nor on NE ($F_{(3,20)} = 0.551; p = 0.653$). An effect of the type of VC over T_{T3} ($F_{(3,20)} = 3.427; p = 0.037$) was also found.

Regarding objective metrics, no effect of the VC's visual appearance was found on $\#_{T1}$ ($F_{(3,20)} = 0.217; p = 0.884$), on $\#_{VC}$ ($F_{(3,20)} = 1.473; p = 0.252$) nor on $\#_{pairs}$ ($F_{(3,20)} = 0.498; p = 0.668$).Finally, no significant difference was found for Q_{23} ($F_{(3,20)=1.599;p=0.221}$) nor for Q_{35} ($F_{(3,20)} = 0.333; p = 0.802$).

Shapiro-Wilk tests for normality indicate that all our data, except EV, follow a normal distribution and Levene's tests confirmed the equality of variances. For EV, a Bartlett test confirmed equality of variance. Therefore, as post-hoc analysis, we carried out bidirectional Student's t-tests with a 0.05 significance level for the

Table 1: Results of Experiment 1.

Measures	SASG (n=5)		SADG (n=6)		DASG (n=6)		DADG (n=7)	
	Mean	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
$\#_{T1}$	1.3	0.8	1.2	0.4	1.2	0.4	0.6	0.9
$\#_{VC}$	2	2	1.16	1.6	2.33	1.21	0.6	0.89
$\#_{pairs}$	2.9	2.3	2.8	2.1	4.2	2.6	2.8	2.3
T_{T3} (s)	305.3	21.7	300.8	99.1	211.8	36.4	326.0	82.5
	Help	**T&E**	**Help**	**T&E**	**Help**	**T&E**	**Help**	**T&E**
$Strat_{T2}$	2	5	1	5	2	4	0	5
Quest.	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
SP	3.48	0.40	4.08	0.45	3.67	0.20	3.29	0.34
Eng.	3.73	0.33	4.18	0.22	3.85	0.33	3.4	50.47
EV	3.00	0.59	3.54	0.17	2.92	0.52	2.86	50.77
NE	2.55	0.87	2.67	1.05	2.00	0.82	2.67	1.44
Q_{23}	3.14	1.07	4.17	0.98	4.17	0.75	3.80	1.10
Q_{35}	3.71	0.76	3.67	1.03	4.00	0.89	3.40	1.34

questionnaires and the objective metrics. Results showed (see Figure 4a) that SP was rated significantly higher in the SADG condition (M=4.08, SD=0.45) than in the SASG condition (M=3.48, SD=0.40): $t(10.168) = -2.526; p = 0.030$. A significant difference was also found between SADG and DADG (M=3.29, SD=0.34): $t(8.948) = 3.29; p = 0.009$. No significant difference was found between SADG and DASG (M=3.67, SD=0.20). Moreover, no significant differences were found between SASG and DADG, between SASG and DASG nor between DADG and DASG.

In terms of Engagement (see Figure 4b), a significant difference was found between SADG (M=4.17, SD=0.21) and SASG (M=3.74, SD=0.33) with $t(10.392) = -2.909; p = 0.015$; as well as between SADG and DADG (M=3.40, SD=0.47): $t(5.381) = 3.406; p = 0.017$. No significant difference was found between SADG and DASG (M=3.85, SD=0.33). Besides, no significant differences were found between SASG and DADG, between SASG and DASG, nor between DADG and DASG.

Concerning the influence of the type of VC over T_{T3}, the statistical analysis showed significant differences between SASG and DASG: $t(9.156) = 5.172; p = 0.0005$; and between DADG and DASG: $t(5.294) = 2.869; p = 0.033$. However, unlike for SP and Engagement, SADG showed no difference with SASG, nor with DADG. No significant difference was observed between SADG and DASG, neither between SASG and DADG.

Overall these results showed significant differences between SADG and SASG for both SP and Engagement which partially supports **H1-a**. They also showed significant differences between SADG and DADG for both SP and Engagement which partially supports **H2-a**. In terms of performance, the results showed a significant difference between DASG and SASG for T_{T3}, in favor of DASG, which does not support **H2-b**. A significant difference between DASG and DADG (in favor of DASG) over T_{T3} has also been found and does not support **H1-b**.

4.3.2. Influence of the Virtual Companion's gender

A statistical analysis explored the effect of the gender of the VC: either of the Same Gender (SG) as the participant (i.e., SASG and DASG groups) or of a Different Gender (DG) (SADG and DADG).

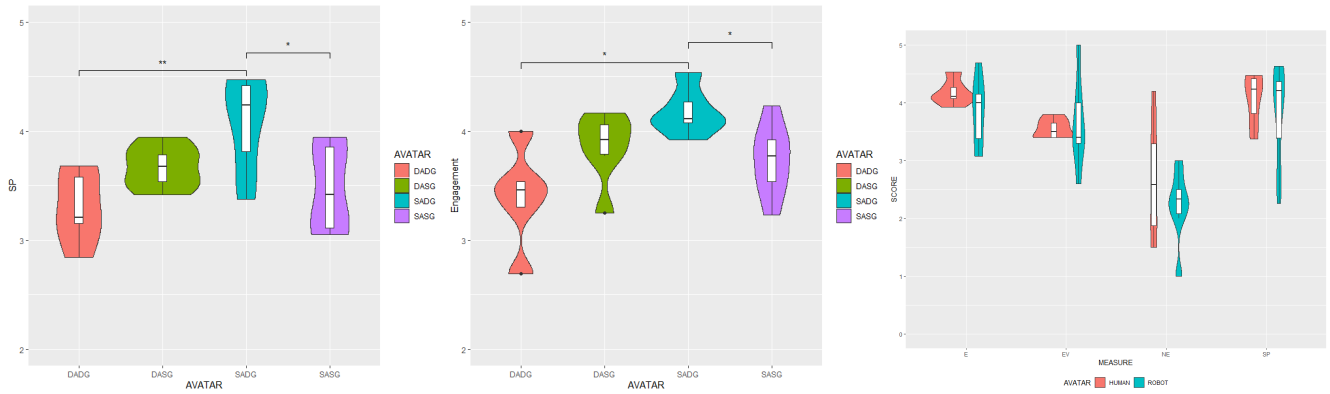
Shapiro-Wilk tests for normality indicate that our data follows a normal distribution except for EV and Levene's tests confirmed equality of variances. For EV, we conducted a Barlett test which also confirmed the equality of variances. As a consequence, bidirectional Student's t-tests were conducted between the SG and DG groups. No influence of the VC's gender was found on SP, on Engagement, on EV nor on NE. Concerning the time spent on each task, the analysis showed no influence of gender on T_{T3}. The statistical analysis also showed that gender had no influence over $\#_{T1}$, nor over $\#_{pairs}$. However, an influence of gender in favor of SG (M=2.08; SD=1.625) compared to DG (M=0.909; SD=1.300) was found over $\#_{VC}$ ($t(21.950) = 2.084; p = 0.049$). As there was no significant difference between SG and DG for most metrics, our hypotheses **H1-a** and **H1-b** (a positive influence of SG would be found over SP, Engagement and as well over the performance) are not supported when comparing only the genders. The significant difference reported over $\#_{VC}$ even goes against these hypotheses.

4.3.3. Influence of the Virtual Companion's age

We also explored the effect of the age of the VC: either of the Same Age (SA) as the participant (i.e., SASG and SADG groups) or of a Different Age (DA) (i.e., DASG and DADG groups). Shapiro-Wilk tests for normality indicate that our data except for EV follows a normal distribution and Levene's tests confirmed the equality of variances. For EV, we conducted a Bartlett test which also confirmed equality of variance. Bidirectional Student's t-tests were conducted on the results between the two groups: SA and DA. No influence of the virtual companion's age was found on SP, on Engagement, on EV nor on NE. Concerning the time spent on each task, the analysis showed no influence of age on T_{T3}. The statistical analysis also showed that age had no influence over $\#_{T1}$, over $\#_{pairs}$, nor over $\#_{VC}$. When grouping VCs by their age only i.e., SA and DA, results showed no significant difference for any of the metrics. This supports neither **H2-a** nor **H2-b**.

4.4. Discussion

This first experiment aimed at evaluating the impact of the age and gender of a VC serving as companion for children in a learning environment. Results show that in terms of SP, Engagement and EV the SADG Virtual Companion gave the best results overall. When grouping the four VCs either by age or by gender, results show that neither **H1** nor **H2** is supported but show a partial support for **H1-a** and **H2-b**. Concerning the overall feeling about the VC, while there is no significant difference between the four groups, SADG and DASG have the highest average score for Q_{23}, and DASG the highest for Q_{35} which is further evidence that SADG and DASG are the most suitable candidates. Regarding objective metrics, the only significant differences were in favor of DASG in terms of performance (number of pairs found and overall speed to achieve the three tasks) but participants did not interact more with the VC in this condition. We further studied the impact of the VC's age and



(a) (b) (c)

Figure 4: Combined violin and box plots for (a) Spatial Presence (SP) in Experiment 1; (b) Engagement in Experiment 1 and (c) Engagement (E), Ecological Validity (EV) and SP in Experiment 2.

gender. We did not find any impact of the age of the VC on any of the metrics. Regarding gender, we found a trend towards a slight increase of interaction with the VC when it is of the same gender as the child. Nevertheless, it should be noted that in both conditions the number of interactions with the VC is relatively low. As a conclusion, there is no clear advantage for a single condition over all the others, but in the subjective evaluation the SADG condition performed better in three dimensions: SP, Engagement and EV. It should be noted that both SADG and DADG have the highest score for NE. However, NE represents the overall feeling of dizziness, eyestrain, etc. and should not depend on the VC.

5. Experiment 2: Influence of the Virtual Companion's type

This experiment aimed at evaluating the impact of the Virtual Companion's type on our metrics, addressing **H3**. We assumed that children would have similar interactions with both the human-like VC and the non-human-like VC. Benefiting from our first experiment, we chose to rely on a humanoid VC but that does not look like a human, therefore we decided to use a robotic virtual companion. Moreover, since results from our first experiment showed that SADG VCs offered the best results regarding subjective metrics, we chose a within-subject design with two experimental groups:

- Human: the VC was a human-like avatar (see subsection 4.1) of SADG as the participant ($n = 6$);
- Robot: the VC was a robotic avatar (see subsection 5.1) of SADG as the participant ($n = 7$).

5.1. Robotic Avatars

We used a Unity Asset [Uni20] to implement our robotic VC. We manually modified the original gray robot to obtain two "gendered" version of the robot (see Figure 5): the pink robot corresponds to the female version whereas the blue/grey one corresponds to the male version. In Europe and the United States, a common stereotype is to associate pink with the feminine and blue with the masculine [Kar11, LD11, INT19]. This is why we have chosen to differentiate the appearance of robots according to this same principle. For both VCs, the voices used (female and male) were identical to those of the first experiment. The robot VC was approximately 1 meter tall, smaller than the participants whose average height was ∼1m40.



Figure 5: Robots used in the second experiment: both are child-like, one being identified as a boy (left: blue/grey robot), the other being identified as a girl (right: pink robot).

5.2. Participants

Fourteen (14) participants took part in this second experiment, but one participant was removed from analysis due to a technical problem during the experiment. The remaining 13 participants (6 girls and 7 boys) were aged from 9 to 13 (M=11.46, SD=0.77). As previously, they came from a mixed-school, parental consent was obtained beforehand and they did not know the purpose of the experiment. All of them had used a VR equipment before.

5.3. Results

Shapiro-Wilk tests for normality indicate that all our data except for SP follows a normal distribution and Levene's tests confirmed the equality of variances. For SP, we conducted a Bartlett test which also confirmed equality of variance. Therefore, as post-hoc analysis we carried out bidirectional Student's t-tests with a 0.05 significance level for the questionnaires and the objective measures (see Figure 4c). No significant difference was found for SP ($t(9.16) = 0.72; p = 0.49$) between the Robot group (M=3.8;

SD=0.88) and the Human group (M=4.08; SD=0.45). No significant difference was found in terms of Engagement ($t(7.78) = 1.43; p = 0.19$) between the Robot group (M=3.83; SD=0.59) and the Human group (M=4.17; SD=0.21). No significant difference was found for EV ($t(6.68) = -0.38; p = 0.72$) for NE ($t(7.91) = 0.93; p = 0.38$), nor for Q_{23} ($t(10.94) = 0.73; p = 0.48$) or for Q_{35} ($t(10,97) = -0.52; p = 0.62$). In terms of direct measures, the statistical analysis showed that the type of the VC had no influence on T_{T3} ($t(10.23) = -0.05; p = 0.963$), neither on $\#_{T1}$ ($t(5) = 1; p = 0.36$) nor on $\#_{pairs}$ ($t(10.93) = -1,05; p = 0.315$). Those results support both **H3-a** and **H3-b**, as no significant difference were found between the Human and the Robot groups.

Table 2: Results of Experiment 2.

Measures	**Human** (n=6)		**Robot** (n=7)	
	Mean	**SD**	**Mean**	**SD**
$\#_{T1}$	1.2	0.4	1.0	0.
$\#_{pairs}$	2.8	2.1	4.1	2.3
T_{T3} (s)	300.8	99.1	303.3	89.0
	Help	**T&E**	**Help**	**T&E**
$Strat_{T2}$	1	5	0	7
Quest.	**Mean**	**SD**	**Mean**	**SD**
SP	4.08	0.45	3.80	0.89
Eng.	4.18	0.21	3.83	0.59
EV	3.54	0.17	3.66	0.77
NE	2.67	1.05	2.21	0.63
Q_{23}	4.17	0.98	3.71	1.25
Q_{35}	3.66	1.03	4.00	1.29

5.4. Discussion

In this second experiment, we focus on the differences between two humanoid VCs: Human and Robot. No significant differences are found for SP, Engagement, EV or NE, which supports **H3-a**. Nevertheless, one can see that the overall the Human condition showed better scores in most dimensions of the ITC-SOPI questionnaire: highest scores for SP and Engagement but lowest score for EV and highest score for NE. Regarding the direct measures, no significant difference was found on any of them, which supports **H3-b**.

6. General Discussion

Overall, our results show that children show a preference towards interacting in a playful environment with a Virtual Companion of the Same Age Different Gender but, when comparing a VC with a human appearance with one that looks like a robot, our results did not show any significant difference. It should be noted that in both experiments, the children interacted much less with the VC that what we expected. During the informal feedback, most of them said that even if the VC was really helpful, it was weird. Some were reticent to interact with the VC since, given the experimental set-up, they could hear the experimenter typing the sentences. Others would rather figure the tasks on their own than ask the VC; this reaction could be accentuated by the playful nature of the tasks and it may not be the case for educational tasks. Moreover, our results are in contradiction with the literature [AZGH06, Bay11] concerning the visual appearance of the VC, this could be explained by the fact that participants in our study were children, unlike previous work.
As any user experiment, this study has some limitations, the first obvious one concerns the participants. We asked them not to discuss it among themselves until all students in the class have had the opportunity to participate in the experiment, but we have no guarantee that this has been respected. The first experiment was carried out over two and a half consecutive days, with the pupils leaving the classroom to complete the tutorial, the experiment and then the questionnaire. The second experiment took place over two consecutive afternoons. Then, concerning the type of the VCs, we focused only on humanoid VCs (human or robot). Further studies may be needed to study the impact of other humanoid representations (cartoons, puppets, etc.) or non-humanoid VCs such as animals. Furthermore, during the distribution in the different groups, we did not ask the participants about their perception of the avatar: what gender and age did they associate the avatar with? In addition, the situation described in this paper could easily be reproduced in real life. It could be interesting to look at differences between the situation in the VE with the VC and one happening in real life. Another concern lies in the use of the Wizard-of-Oz approach: while it allows for a more natural interaction in terms of content of the discussion, it adds a delay in the VCs' speech. This may have had an impact on the interaction between the VC and the children. Finally, we limited ourselves to playful but non academic activities. Thus, the results we have obtained in terms of presence and engagement could be different by changing the nature of the proposed activities.

7. Conclusion

In this paper, we studied the impact of age, gender and type of a Virtual Companion that engaged with a child in educational yet playful tasks in a Virtual Environment. To do so, we designed a first experiment in which 24 children aged between 9 and 12 were randomly assigned in one of 4 groups where the VC was either of Same Age Same Gender, Same Age Different Gender, Different Age Same Gender, Different Age Different Gender with the participant. While no clear winner arise in terms of direct measures or questionnaires, the SADG group showed the best results and seemed the best candidate. Then, to evaluate the impact of the type, we decided to use a human-like or a robot VC of the SADG as that of the participant. Here as well, results of a study involving 13 children between 9 and 12, do not exhibit a clear winner, but the human-like VC obtained the best scores in terms of questionnaires. Overall, our results tend to show that children have a higher sense of Spatial Presence, Engagement and Ecological Validity when interacting with a human-like Virtual Companion of the Same Age and of a Different Gender. More experiments are nevertheless required to confirm these results, especially using more academic tasks.

Acknowledgments

This work was supported with the financial support of the Science Foundation Ireland grant 13/RC/2094.

References

[Ado20] ADOBE: Mixamo, 2020. Last accessed 31 August 2020. URL: https://www.mixamo.com/. 5

[AHTL16] ARGELAGUET F., HOYET L., TRICO M., LÉCUYER A.: The role of interaction in virtual embodiment: Effects of the virtual hand representation. In *2016 IEEE Virtual Reality* (March 2016), pp. 3–10. 3

[AZGH06] A. ZANBAKA C., GOOLKASIAN P., HODGES L.: Can a virtual cat persuade you? The role of gender and realism in speaker persuasiveness. In *Conference on Human Factors in Computing Systems - Proceedings* (01 2006), vol. 2, pp. 1153–1162. 2, 3, 8

[Bay09] BAYLOR A. L.: Promoting motivation with virtual agents and avatars: role of visual presence and appearance. *Philosophical Transactions of the Royal Society B: Biological Sciences 364*, 1535 (2009), 3559–3565. 2

[Bay11] BAYLOR A. L.: The design of motivational agents and avatars. *Educational Technology Research and Development 59*, 2 (2011), 291–300. 2, 3, 5, 8

[BPN*08] BAILENSON J., PATEL K., NIELSEN A., BAJSCY R., JUNG S.-H., KURILLO G.: The effect of interactivity on learning physical actions in virtual reality. *Media Psychology 11*, 3 (2008), 354–376. 2

[CPDS09] CHERYAN S., PLAUT V. C., DAVIES P. G., STEELE C. M.: Ambient belonging: how stereotypical cues impact gender participation in computer science. *Journal of Personality and Social Psychology 97*, 6 (2009), 1045. 1

[DOC03] DARVES C., OVIATT S., COULSTON R.: The impact of auditory embodiment on animated character design. In *Proceedings of the International Joint Conference on Autonomous Agents and Multi-Agent Systems, Embodied Agents Workshop* (05 2003). 2

[EHO15] ENNIS C., HOYET L., O'SULLIVAN C.: Eye-tracktive: Measuring Attention to Body Parts when Judging Human Motions. In *EG 2015 - Short Papers* (2015), Bickel B., Ritschel T., (Eds.), The Eurographics Association. doi:10.2312/egsh.20151009. 2

[GQ08] GOODMAN G. S., QUAS J. A.: Repeated interviews and children's memory: It's more than just how many. *Current Directions in Psychological Science 17*, 6 (2008), 386–390. 2

[INT19] ISHII K., NUMAZAKI M., TADO'OKA Y.: The effect of pink/blue clothing on implicit and explicit gender-related self-cognition and attitudes among men. *Japanese Psychological Research 61*, 2 (2019), 123–132. 7

[JTMT09] JARMON L., TRAPHAGAN T., MAYRATH M., TRIVEDI A.: Virtual world teaching, experiential learning, and assessment: An interdisciplinary communication course in second life. *Computers & Education 53*, 1 (2009), 169–182. 1

[Kar11] KARNIOL R.: The color of children's gender stereotypes. *Sex roles 65*, 1-2 (2011), 119–132. 7

[KHAL*09] KONSTANTINIDIS E. I., HITOGLOU-ANTONIADOU M., LUNESKI A., BAMIDIS P. D., NIKOLAIDOU M. M.: Using Affective Avatars and Rich Multimedia Content for Education of Children with Autism. In *Proceedings of the 2nd International Conference on Pervasive Technologies Related to Assistive Environments* (2009), PETRA '09, ACM, pp. 58:1–58:6. 2

[Lan11] LANKOSKI P.: Player character engagement in computer games. *Games and Culture 6*, 4 (2011), 291–311. 1

[LD11] LOBUE V., DELOACHE J. S.: Pretty in pink: The early development of gender-stereotyped colour preferences. *British Journal of Developmental Psychology 29*, 3 (2011), 656–667. 7

[LEK*18] LUGRIN J., ERTL M., KROP P., KLÜPFEL R., STIERSTORFER S., WEISZ B., RÜCK M., SCHMITT J., SCHMIDT N., LATOSCHIK M. E.: Any "Body" There? Avatar Visibility Effects in a Virtual Reality Game. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces* (2018), pp. 17–24. 3

[LFDK01] LESSITER J., FREEMAN J., DAVIDOFF J. B., KEOGH E.: A cross-media presence questionnaire: The itc-sense of presence inventory. *Presence: Teleoperators and Virtual Environments 10*, 3 (June 2001), 282–297. URL: http://research.gold.ac.uk/483/. 5

[Mak20] MAKEHUMAN: MakeHuman Community, 2020. Last accessed 31 August 2020. URL: http://www.makehumancommunity.org. 5

[MD09] MARTIN A. J., DOWSON M.: Interpersonal relationships, motivation, engagement, and achievement: Yields for theory, current issues, and educational practice. *Review of Educational Research 79*, 1 (2009), 327–365. 1

[Mic20a] MICROSOFT: Microsoft Cognitive Services for Unity, 2020. Last accessed 31 August 2020. URL: https://docs.microsoft.com/en-ie/azure/cognitive-services/speech-service/language-support. 3

[Mic20b] MICROSOFT: Speech Synthesizer, 2020. Last accessed 31 August 2020. URL: https://docs.microsoft.com/en-ie/dotnet/api/system.speech.synthesis.speechsynthesizer?view=netframework-4.8. 5

[MM05] MARTIN A., MARSH H.: Motivating Boys and Motivating Girls: Does Teacher Gender Really Make a Difference? *Australian Journal of Education 49*, 3 (2005), 320–334. 1

[ON17] O.BAILEY J., N.BAILENSON J.: Considering virtual reality in children's lives. *Journal of Children and Media* (2017). 2

[PBDRK09] PLANT E. A., BAYLOR A. L., DOERR C. E., ROSENBERG-KIMA R. B.: Changing middle-school students' attitudes and performance regarding engineering with computer-based social models. *Computers & Education 53*, 2 (2009), 209–215. 2

[PBHJ*06] PATEL K., BAILENSON J. N., HACK-JUNG S., DIANKOV R., BAJCSY R.: The effects of fully immersive virtual reality on the learning of physical tasks. In *Proceedings of the 9th Annual International Workshop on Presence, Ohio, USA* (2006), pp. 87–94. 2

[SB09] SEGOVIA K. Y., BAILENSON J. N.: Virtually true: Children's acquisition of false memories in virtual reality. *Media Psychology 12*, 4 (2009), 371–393. 2

[Shi18] SHIN D.: Empathy and embodied experience in virtual environment: To what extent can virtual reality stimulate empathy and embodied experience? *Computers in Human Behavior 78* (2018), 64 – 73. URL: http://www.sciencedirect.com/science/article/pii/S0747563217305381. 2

[SS14] SLATER M., SANCHEZ-VIVES M. V.: Transcending the Self in Immersive Virtual Reality. *Computer 47*, 7 (2014), 24–30. 2

[TA12] THENG Y.-L., AUNG P.: Investigating effects of avatars on primary school children's affective responses to learning. *Journal on Multimodal User Interfaces 5*, 1 (Mar 2012), 45–52. 2

[Uni20] UNITY: Asset Store, 2020. Last accessed 31 August 2020. URL: https://assetstore.unity.com/packages/3d/characters/robots/cute-robot-137652. 7

[WLW*13] WALSH S., LIU F., WOLLSTEIN A., KOVATSI L., RALF A., KOSINIAK-KAMYSZ A., BRANICKI W., KAYSER M.: The hirisplex system for simultaneous prediction of hair and eye colour from dna. *Forensic Science International: Genetics 7*, 1 (2013), 98–115. 5

[WSR19] WALKER M. E., SZAFIR D., RAE I.: The Influence of Size in Augmented Reality Telepresence Avatars. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* (March 2019), pp. 538–546. 2

[YB07] YEE N., BAILENSON J.: The proteus effect: The effect of transformed self-representation on behavior. *Human communication research 33*, 3 (2007), 271–290. 2

[YKL*19] YOON B., KIM H., LEE G. A., BILLINGHURST M., WOO W.: The effect of avatar appearance on social presence in an augmented reality remote collaboration. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* (March 2019), pp. 547–556. 2