Virtual Tutor Personality in Computer Assisted Language Learning

Johanna Dobbriner[1][0000−0002−2129−3653], Cathy Ennis[1][0000−0002−1274−5347], and Robert Ross[1][0000−0001−7088−273X]

School of Computer Science, Technological University Dublin,
Dublin 7, Ireland, D07 ADY7

Abstract. The use of intelligent virtual agents in language learning has increased in recent years. Studies into several aspects of personalisation aiming to increase user engagement are an ongoing research topic with avatar personality being one such aspect. As a step towards our development of intelligent virtual avatars, we present two of our initial experiments to explore differences in user interaction with two contrasting avatar personalities – P1: open-minded, friendly and sociable and P2: closed-off, curt and distant. Each user interacted with a single personality in a video-call setting and gave feedback on the interaction. Our expectations, that P1 would be rated more enjoyable and induce participants to talk more, were only partially confirmed. While P1 did induce longer conversations in the participants, we found that interactions with both personalities were enjoyed and that user perception of P1 and P2 differed, but less than intended. Several possible causes for these results are discussed, and we outline impacts for follow on intelligent system design.

Keywords: Computer Assisted Language Learning · Virtual Tutor · Virtual Human · Wizard-of-Oz · Personality · Big Five Personality Model.

1 Introduction

Learning a foreign language is often difficult for several reasons, the availability of conversational partners to practice spoken interaction being one. Ideally, a language learner would immerse themselves in the language by spending some time in a foreign country and interacting with native speakers regularly. As this opportunity is not open to every student, one alternative for conversational practice is Computer Assisted Language Learning (CALL), specifically, virtual tutors: They are constantly available, financially more accessible, have infinite patience and can minimise the student's anxiety or embarrassment. A number of these systems for conversational practice have already been examined, embedded in video games, on their own or as part of a larger CALL system [3, 1, 2].

For these CALL applications, the automated tutor is usually embodied in some way, for example as a virtual avatar, to keep the student focused on the learning activities. In order to increase user engagement with a virtual avatar,

many avenues are being explored, among them customisation of physical characteristics or the use of gestures and facial expressions. Anything that distinguishes a virtual character could be useful. As such, assigning specific personality traits to a virtual tutor, expressed through personal preferences, appearance and behaviour, and perhaps tailored to the student's own personality, may increase engagement. To investigate this hypothesis, we first need to find out whether there are any observable variations in the interaction and feedback when students are confronted with different tutor personalities. To this end, we designed an interactive experiment with two opposing tutor personalities to explore that question and a separate survey to validate the two personalities we designed.

Fig. 1. Avatar expressing different personalities – P1 (left) and P2 (right)

2   Background

The use of virtual humans and conversational agents or chatbots for language learning has gained popularity in recent years, both separately and combined.

Chatbots are increasingly used in areas of day-to-day life, e.g. as personal assistants or in customer service. So far, they are often text-based as speech integration is still an area of active research [5]. General difficulties are synthesising natural sounding speech with appropriate intonation, pauses and the use of fillers (Hm, Ah,...), as well as keeping response delays appropriate for real-time conversation. The context of language learning can present an additional stumbling block as the speech of the person being taught is expected to contain errors in pronunciation or grammar as well as an accent from the speaker's native language. Furthermore, while dialogue systems are commonly categorised as either open-domain chatbots whose main purpose is to keep a conversation alive for as long as possible or task-based systems with a specific aim to fulfill, conversational practice in language learning requires elements of both: Interaction should take some time to give the student the opportunity to practice, but it also needs to be domain-specific so that the conversation makes sense and serves a purpose to the student.

To embody an interaction partner, virtual humans or avatars are often used in a CALL system. These can range from static images all the way to animated 3D characters, possibly even embedded within a VR environment. Computer games, for instance, can incorporate gamification in the language learning process, which is done frequently in CALL for student motivation and engagement. Some recent

examples of CALL chatbots and/or virtual avatars being used commercially or in scientific research are Mondly [15], VILLAGE [16] and CILLE [4].

Personalities can be described through numerous models in psychology. One widely used model to categorise different personalities by broad behavioural traits is the *OCEAN* model. This empirical model identifies five dimensions – openness, conscientousness, extroversion, agreeableness and neuroticism – that describe a person's character [6].

The influence of personality on different aspects of language learning has been studied extensively. However, more studies can be found focussing on the student's personality [9, 11] than the teacher's character traits. In the few publications available on teacher personality, the scarcity of available literature is explicitly noted [8]. The influence of personality on the strategies teachers employed when learning a language themselves is analysed, finding e.g. extroverted learners employ most types of strategies more frequently than introverts, particularly sociocultural interactive strategies and metastrategies like foreign language media consumption [17].

In virtual characters on the other hand, various studies can be found discussing personality, for instance the expression of personality traits in virtual humans [12] or the impact of a virtual avatar's personality on user engagement and perception [18]. Still, on the aspect of tutor personality in dialogue based CALL, no extensive research has been conducted.

3   Experiment Design

For our experiment, we chose English as the foreign language, due to the wealth of available speech synthesisers and recognisers as well as it being an easy language for which to find participants.

To investigate whether the interaction with and perception of the virtual tutor by language learners would change with different personality traits in the tutor, we designed the main experiment so that the participant would talk to a virtual avatar for a few minutes and then rate the avatar personality and give feedback on the conversation itself. We used an expressive avatar [14], animated in javascript with an Irish English, female voice [1] whose facial expression (see Figure 1), speech rate and emphasis could be controlled on-the-fly by the researcher.

Set up as a Wizard of Oz Study [10], we simulated an automated, speech-based dialogue system with a distinctive personality by creating a dialogue script to follow with alternative paths depending on how users reacted. All anticipated utterances in the script were then recorded as individual video clips. During an experiment run, the researcher could select and play appropriate clips to reply to the participant or to further the conversation. The dialogue script included an introduction and the conversational topics of hobbies, travel and animals.

---

[1] https://www.cereproc.com/en/node/1155

We designed two personalities to differ along three of the OCEAN model's five dimensions. As our context is language teaching, we chose to focus on Agreeableness, Extroversion and Openness as these can be expressed best in a dialogue setting. Due to the avatar's limited range of facial expressions and a lack of emotionally charged situations in the interaction, we left out Neuroticism. Conscientousness was also not included as a personality trait, as it did not apply to the topics discussed and would be better expressed in actions rather than pure conversation.

Table 1. Both avatar personality designs listed by personality dimension and supporting audio-visual characteristics

	Personality 1 (P1)	Personality 2 (P2)
Dimension		
Extroversion	– Loves to talk – Shares information about herself without prompting – Leads conversation	– Gives short answers – Only offers personal information when prompted – Verbally expresses discomfort at being among people
Agreeableness	– Empathises with the participant – Always friendly – Tries to understand dialogue partner's point of view – Patient with the participant	– Set in her own opinions, insistent on not changing them – Will express her opinions if they differ from the dialogue partner – Formal and curt wording, lacking most filler words
Openness	– Curious about other people's ideas and experiences – Ready to try new things – Interested in opinions and preferences different from her own	– Appears closed off – Dislikes new experiences, especially travel – Not interested in new places or experiences
Audio-Visual Features	– Smiling facial expression – Looks straight at the viewer – Emphatic lip movements – Fast talking speed	– Sad facial expression – Head slightly tilted, not looking directly at viewer – Little emphasis in lip movements – Slow talking speed

As we only worked with one avatar, both characters have a few things in common: They have the same physical characteristics, are female and named Saoirse. The different personalities were designed to represent two extremes expressed by creating dialogue scripts where the avatar exhibited specific personality traits. In addition, posture, facial expression and speech characteristics were fitted to support the script (see Figure 1). As detailed in Table 1, Personality 1 (P1) represents the higher end of our 3 scales, being open, friendly and sociable, while Personality 2 (P2) exhibits low scores along the 3 dimensions, behaving in a more closed off, curt and distant manner.

Personality Verification Survey In order to verify the two personalities we had designed, we ran an online survey where participants were shown a video

recording of each avatar in random order and asked to rate the avatar's personality on a standardised personality test [7] directly after the video. We also included a final question to determine which personality they preferred and why. For this survey we only invited native English speakers to ensure no participant overlap with the main experiment and the focus being on the personality without the added cognitive load of language learning.

Main Experiment As in-person experiments with local participants and fixed technical equipment were not feasible at the time of writing due to the 2020/2021 Coronavirus pandemic, an experiment website [2] was built using primarily ReactJS, NodeJS and WebRTC with separate views for the researcher and the participant (Figure 2). On signing up for the experiment, a participant was randomly assigned an avatar personality to interact with. Once a participant had signed in, they could start the experiment in their own time and were presented with an initial survey to collect some meta information (native language, age group, ...). Next, in the main part of the experiment, the participant entered a video chat with the avatar. On the experimenter's side, the software OBS Studio
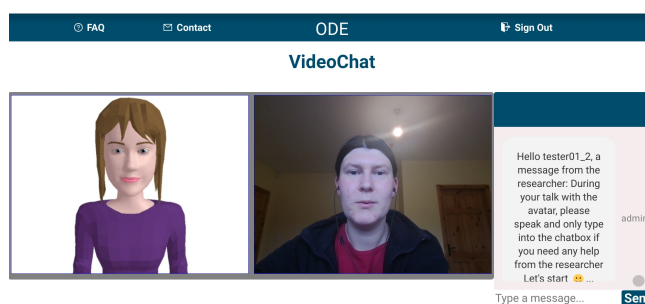

Fig. 2. Participant view during interaction with P1 avatar

[3] was used to assemble the video clips of each personality into an OBS scene and replay them as needed in the video call via virtual camera. Each video call was recorded via screen recorder and saved locally. After the video call, the participant was presented with another questionnaire rating the avatar's personality and giving feedback on the interaction.

To assess the avatar's perceived personality after the interaction, we decided to use adjectives symbolising the 3 personality dimensions, with 4 items per dimension, 2 positive, 2 negative each (see Table 2). Participants were asked to rate on a 5-point scale how well each description fit the avatar they had just talked to. Next to each adjective, a definition of the word was available in case the user had not encountered the word before.

The participants of this experiment were adult English learners, 18 years or older with normal hearing, as the main mode of interaction was spoken dialogue.

We expected P1 to be more pleasant and enjoyable to converse with, which would show itself in markedly positive user feedback and high scores on the

---

Table 2. Descriptors used in the avatar's personality evaluation by the participant

Dimension	High	Low
Extroversion	assertive	shy
	outgoing	distant
Agreeableness	patient	annoyed
	caring	demanding
Openness	curious	bored
	imaginative	conservative

avatar personality survey. In contrast, we anticipated low personality scores and for fewer participants to enjoy the interaction with P2. Regarding the dialogue, we also hypothesised that P1 freely sharing much of herself would animate participants to talk more, and therefore result in longer interactions than with P2.

4   Results

Personality Verification Survey In this survey, with only native English speakers, we recruited 33 participants, 67% female and 33% male, aged between 18 and 54 years with 33% in the 18-24 and 48% in the 25-34 age groups.
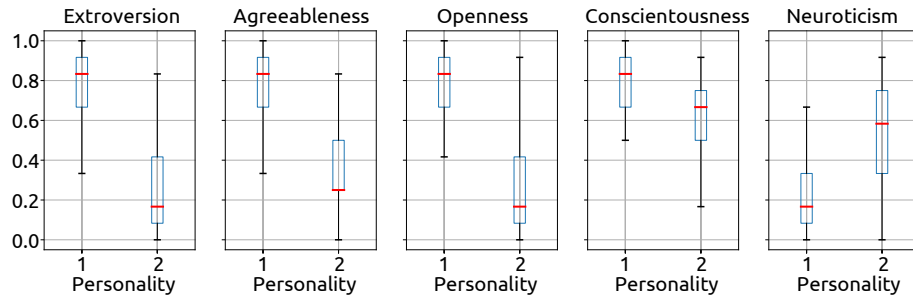
Fig. 3. Box plots of Avatar personality scores between avatar personalities in the Personality Verification Survey

Our personality design was verified, as participants consistently rated P1 significantly higher than P2 on the Extroversion, Agreeableness and Openness scales just as we had intended. Figure 3 shows the box plots of these scores for both avatars on all 5 personality scales. For better readability, we normalised the scores between 0 and 1, i.e. 1 on the Extroversion scale would signify extreme extroversion while 0 is extreme introversion. A T-test at a significance threshold of 0.05 confirms these results with Openness ($T(18) = 10.651$, $p = 1.857e - 15$) the most significant, followed by Extroversion ($T(18) = 9.704$, $p = 3.889e - 14$) and Agreeableness ($T(18) = 7.693$, $p = 1.620e - 10$). Even on the remaining two scales the differences were significant, more so regarding Neuroticism ($T(18) = -6.436$, $p = 3.001e - 8$) – P1 scored much lower with a median under 20% which translates to greater emotional stability compared to P2 – than Conscientousness ($T(18) = 3.810$, $p = 3.404e - 4$) where both personalities were rated relatively high but P1 still received significantly higher scores, also shown in Figure 3.

In terms of preference, participant responses also confirmed our expectations with only a single participant indicating they would rather talk to P2 and the remaining 97% stating a preference for P1.

Main Experiment Over the course of this pilot experiment, 18 participants completed the main study, 44% male and 56% female, aged 18 - 45 years, predominately under 35, who had been learning English between 7 and 22 years ($M = 12.5$, $SD = 4.33$). A majority of 78% reported some experience with virtual characters or environments and the remaining 22% noted no experience at all. At 83%, most participants spoke German as their native language with another 11% Italian and 5.6% (one person) Chinese. Due to the automated random assignment of experiment groups – 1 for P1, 2 for P2 – and the even number of participants, there were 9 participants per avatar personality.

Going forward, it must be stated that with a sample size as low as this, any statistics computed on the collected data cannot be very robust and all results are to be taken as indicative.
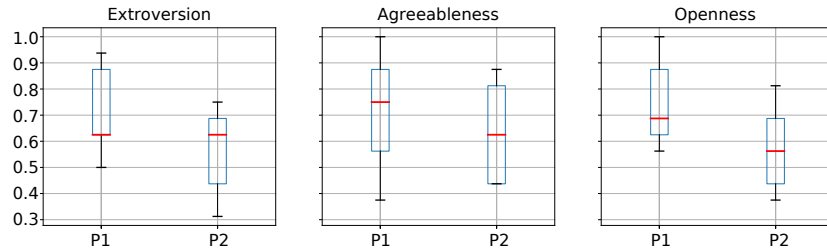

Fig. 4. Box plots of Avatar personality scores between avatar personalities for Extroversion (Avatar_E), Agreeableness (Avatar_A) and Openness (Avatar_O) in the main experiment

Avatar personality scores To answer the question, whether participants were generally able to tell apart the different personalities they interacted with, let us first look at the scores the avatar achieved in the 3 personality dimensions.

Figure 4 shows box plots of the 3 personality dimensions comparing both avatar personalities. At first glance, the plots look mostly as expected, with P1 generally scoring higher than P2, but a closer look reveals a few unexpected results.

While P1 generally achieved higher scores, the box plots for Agreeableness (Figure 4, middle) stretch over a larger interval and overlap more than they differ. For Extroversion, while the 25th and 75th percentile are higher in the first personality, the median is actually the same at 0.625 in both groups and P2 achieved scores to just under 0.7 which is surprising for an introverted character. The clearest distinction is found for Openness with far less overlap between the box plots (Figure 4, right).

A T-test at a significance threshold of 0.05 confirms these results as shown in Table 3, with Openness ($T(18) = 2.166$, $p = 0.046$) showing the only significant difference between P1 and P2, whereas Extroversion ($T(18) = 1.816$, $p = 0.088$) is marginally significant and may prove distinct with more participants. Participants still rated P1 as slightly more agreeable, but the mean comparison in

Table 3. p-value and t-statistic from Welch's T-test comparing the mean values between P1 and P2 for 18 participants

Measurement	t(18)	p value)
Avatar Extroversion	1.816	0.088
Avatar Agreeableness	1.099	0.288
Avatar Openness	2.166	0.046
Speaking Time Ratio	2.525	0.022
Number of Interruptions	- 0.622	0.543

Table 3 reveals no significant difference between the two groups. At a p value of 0.288, not even a trend is discernible, meaning that we failed to convey the intended difference in agreeableness.

Audio Analysis Aside from the direct personality survey, we also analysed the audio track recorded from each participant, computing speaking duration, interruptions and pauses within a participant's turn as well as between turns.

The total duration of the participant's interaction with the avatar varied significantly between avatar personalities: While P1 interactions lasted an average of 12.15 minutes ($SD = 1.81$, $Min = 8.61$, $Max = 14.25$), P2 conversations were markedly shorter at 7.55 minutes on average ($SD = 1.54$, $Min = 5.77$, $Max = 10.26$). However, this marked difference was at least partly due to differences in the avatar script as the P1 script included overall longer utterances.

Since learners' speaking practice is the main goal of the application and to fairly measure how much the avatar induced the participant to talk, we compared the ratio of the the participant's total speaking time divided by the avatar's total speaking time for each conversation. The distributions of this ratio between the two groups are very different: P1 participants talked between 1.17 and 3.14 times as much as the avatar with the 25th percentile at 1.80 and the 75th percentile at 2.39, whereas participants in P2 range between talking just over half as much as the avatar's at 0.59 to 2.52 times as much, the latter of which is an outlier. A mean comparison of this measure (Table 3) has a significant result at a p value of 0.022, thus confirming our hypothesis that P1 would animate learners to talk more than P2.

Another useful measure are interruptions that inevitably occur during video chats due to network latency or misjudging when the other party is going to speak. An automated dialogue system especially may need to adjust to a user's individual response times and minimum pause duration for a specific speaker to adequately mimic natural conversations. During the experiment, the number of involuntary interruptions by the avatar occurred independent of the avatar personality due to connectivity variations and human error of the researcher controlling the avatar, but the frequency of interruptions may influence the participant's perception of the avatar's personality. As such, we counted the number of interruptions for each conversation and found no significant difference between the personalities (see Table 3).

The pauses in the conversation present a further aspect of analysis that may be of interest, particularly the participant's reaction time, i.e. the interval until the participant speaks after an utterance from the avatar and the participant's pauses while speaking. Analysis of participant response times across groups re-

veals no real difference between groups for both measures, the participant silences within a speaker's turn having a median around 1 second and the reaction times with a median of 2 seconds across groups.

Qualitative Feedback As a final part of the experiment, participant feedback was collected to determine enjoyment of the interaction, positive and negative aspects noted by the participants, their attentiveness during the dialogue and suggestions for further topics to discuss with the avatar.

All participants replied with *Yes* when asked whether they had enjoyed the interaction. When we gauged the participants' attentiveness by having them recall what they talked about, each participant recalled the main topics of conversation (hobbies, travel and pets).

In a question about positive and negative aspects of the avatar's conversation, frequent positive aspects of P1 were: patient (repeating words or phrases), interested, asking questions/introducing topics, kind and interesting to talk to. Negative aspects included: interruptions/impatience and topic changes without answering a user question. Showing too little emotion and talking too fast were also brought up, but only once. For P2, the main positive aspect was a good understanding of the user with matching responses. Unexpectedly, considering the intended lower agreeableness level, the avatar being nice, polite and having a sense of humor was also mentioned. A possible explanation might be that P2, while frequently disagreeing with the participant, was never openly rude, and polite disagreement may even be perceived as truer understanding than constant agreement with the learner. Frequent negative aspects were that the avatar was sometimes hard to understand and was stiff or unnatural to talk to. The avatar looking sad and not answering some questions was also sporadically noted.

Topic suggestions for further conversation with the avatar were numerous for both groups and several people explicitly stated they would like to talk to the avatar again. Notably for P2, the avatar's general sadness and fear was repeatedly suggested, demonstrating an urge to help the avatar develop a more positive attitude.

5   Discussion

While in our validation survey both personalities were clearly distinguishable, the results of the main experiment show less of a perceived difference between the two. Part of that may be explained by having native English speakers for the validation and language learners in the main experiment along with the different number of participants, the mode of delivery and different focus: In the survey, 33 participants watched a video with instructions to focus on personality and had a direct comparison of both avatars, whereas the main experiment was interactive, with 18 language learners more focused on understanding the avatar and formulating a response and they only saw one avatar.

Aside from that, certain implications of the online interaction likely also affected specific personality dimensions: The overall high extroversion scores for P2 could be due to the avatar being perceived as more assertive than intended

in its aim to keep the conversation alive. For participants who tend to give short replies and not ask many questions themselves, the avatar's continued questions may be recognised as outgoing, leading to a higher extroversion score.

Agreeableness, on the other hand seems to be highly varied independent of avatar personality, and generally rather high, which is further reinforced in the feedback survey where the P1 is repeatedly noted as kind or friendly – as intended – but P2 is unexpectedly deemed nice and polite by two participants. In the script, agreeableness was incorporated implicitly by a generally more informal speaking style for P1 but not in explicit statements, which would likely be more easily picked up by native speakers. Additionally, the avatar's involuntary interruptions certainly had an impact on perceived patience that was also noted explicitly as a negative in the P1 feedback. Countering that, the willingness to repeat utterances which was present in both avatars as this is a learner's environment, would be perceived as higher patience. Both interruptions and repetitions therefore likely affected the agreeableness score as noise across personalities. Participants' direct comments also indicated that P2 was perceived as sad and depressed rather than disagreeable.

With the avatar's questions in both personalities, the openness score, particularly curiosity was always rated at least neutral or higher and thus elevated the score across groups. However, the explicitly conservative statements of P2, along with more in-depth questions and requests for recommendations in P1 appear to be sufficient to be perceived as distinct from each other.

With regards to the speaking time ratio, P1 appears to encourage the participant to speak more than P2 does, as expected if the participant is mirroring the avatar's conversational style [13]. Encouraging the user to talk more is one aim of a language learning application, so P1 may be more suited to that task. However, advanced learners might be shown P2 with the task to encourage the avatar to speak more.

Regarding participant feedback, the universal enjoyment of interacting with the avatar was group-independent and is likely at least in part due to the novelty of the experience in contrast to conventional teaching. Gathering feedback and recording student progress over an extended period of time would be required for more valid results. However, the positive and negative aspects of the personalities will help us improve them in further iterations of this project.

As a final aspect of this study, it should be noted that our participants were all volunteers and therefore bringing a high baseline of openness and agreeableness to the interaction. If used e.g. as an extension to traditional language teaching in a classroom or even just in a larger group of learners, the acceptance of and engagement with a virtual language tutor like this would likely differ.

6   Conclusions and Future Work

In this pilot study we designed and verified two different personalities for a virtual language tutor, exploring through an interactive experiment whether participants interacted differently with it depending on the tutor's personal-

ity. While the personality differences appeared distinct in our validation survey, in our main study we found that only one out of three personality dimensions, Openness, was perceived significantly differently between both groups. Extroversion, a second dimension came close to the significance threshold and the third dimension, Agreeableness was widely distributed with relatively high scores in both groups. However, the participant's speaking time relative to the avatar as a fourth measure, was significantly higher in the personality designed to be more pleasant overall, thus matching our expectations, and qualitative feedback for the application itself was encouraging.

While this initial study has limitations, we see it as an important step in validating our approach to data-driven customisable assistive agents. Such agents require the synthesis of a number of aspects of artificial intelligence, machine learning, and cognitive science, but in so doing provide us a very real and beneficial application domain for intelligent systems development. In current research we are now beginning to put our findings to work in prototype development. With respect to the specific future work building on our activities here, we will focus on making the personality differences more apparent during the interaction, automating and extending the study by building a chatbot for each personality and possibly adding more personalities as well as exploring measures to automatically adapt to the student.

Acknowledgments

References

1. Cheng, A., Yang, L., Andersen, E.: Teaching language and culture with a virtual reality game. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. pp. 541–549 (2017). https://doi.org/10.1145/3025453.3025857
2. Collins, N., Vaughan, B., Cullen, C., Gardner, K.: Gaeltechvr: Measuring the impact of an immersive virtual environment to promote situated identity in irish language learning. Journal For Virtual Worlds Research **12**(3) (2019). https://doi.org/10.4101/jvwr.v12i3.7356
3. Dalton, G., Devitt, A.: Gaeilge gaming: Assessing how games can help children to learn irish. International Journal of Game-Based Learning (IJGBL) **6**(4), 22–38 (2016). https://doi.org/10.4018/IJGBL.2016100102
4. Divekar*, R.R., Drozdal*, J., Chabot*, S., Zhou, Y., Su, H., Chen, Y., Zhu, H., Hendler, J.A., Braasch, J.: Foreign language acquisition via artificial intelligence and extended reality: design and evaluation. Computer Assisted Language Learning **0**(0), 1–29 (2021). https://doi.org/10.1080/09588221.2021.1879162

5. Fryer, L., Coniam, D., Carpenter, R., Lăpușneanu, D.: Bots for language learning now: Current and future directions. Language Learning & Technology **24**(2), 8—-22 (2020). https://doi.org/10125/44719

6. Goldberg, L.R.: An alternative "description of personality": the big-five factor structure. Journal of personality and social psychology **59**(6), 1216 (1990). https://doi.org/10.1037/0022-3514.59.6.1216

7. Gosling, S.D., Rentfrow, P.J., Swann, W.B.: A very brief measure of the big-five personality domains. Journal of Research in Personality **37**(6), 504–528 (2003). https://doi.org/https://doi.org/10.1016/S0092-6566(03)00046-1

8. Göncz, L.: Teacher personality: a review of psychological research and guidelines for a more comprehensive theory in educational psychology. Open Review of Educational Research **4**(1), 75–95 (2017). https://doi.org/10.1080/23265507.2017.1339572

9. Kanero, J., Oranç, C., Koşkulu, S., Kumkale, G.T., Göksun, T., Küntay, A.C.: Are tutor robots for everyone? the influence of attitudes, anxiety, and personality on robot-led language learning. International Journal of Social Robotics pp. 1–16 (2021). https://doi.org/10.1007/s12369-021-00789-3

10. Kelley, J.F.: An iterative design methodology for user-friendly natural language office information applications. ACM Transactions on Information Systems (TOIS) **2**(1), 26–41 (1984). https://doi.org/10.1145/357417.357420

11. Liang, H.Y., Kelsen, B.: Influence of personality and motivation on oral presentation performance. Journal of psycholinguistic research **47**(4), 755–776 (2018). https://doi.org/10.1007/s10936-017-9551-6

12. Nunnari, F., Héloir, A.: Generation of virtual characters from personality traits. In: Beskow, J., Peters, C.E., Castellano, G., O'Sullivan, C., Leite, I., Kopp, S. (eds.) Intelligent Virtual Agents - 17th International Conference, IVA 2017, Stockholm, Sweden, August 27-30, 2017, Proceedings. Lecture Notes in Computer Science, vol. 10498, pp. 301–314. Springer (2017). https://doi.org/10.1007/978-3-319-67401-8_39

13. Sinclair, A.J., Ferreira, R., Gašević, D., Lucas, C.G., Lopez, A.: I wanna talk like you: Speaker adaptation to dialogue style in l2 practice conversation. In: International Conference on Artificial Intelligence in Education. pp. 257–262. Springer (2019). https://doi.org/10.1007/978-3-030-23207-8_48

14. Sloan, J., Maguire, D., Carson-Berndsen, J.: Emotional response language education for mobile devices. In: 22nd International Conference on Human-Computer Interaction with Mobile Devices and Services. MobileHCI '20, Association for Computing Machinery, New York, NY, USA (2020). https://doi.org/10.1145/3406324.3417603

15. Studios, A.: Mondly – learn languages online for free (2021), https://www.mondly.com/, accessed on 27/09/2021

16. Wang, Y., Petrina, S., Feng, F.: VILLAGE - virtual immersive language learning and gaming environment: Immersion and presence. Br. J. Educ. Technol. **48**(2), 431–450 (2017). https://doi.org/10.1111/bjet.12388

17. Zhou, C., Intaraprasert, C.: Language learning strategies employed by chinese english-major pre-service teachers in relation to gender and personality types. English Language Teaching **8**(1), 155–169 (2015). https://doi.org/10.5539/elt.v8n1p155

18. Zibrek, K., Kokkinara, E., McDonnell, R.: The effect of realistic appearance of virtual characters in immersive environments - does the character's personality play a role? IEEE Trans. Vis. Comput. Graph. **24**(4), 1681–1690 (2018). https://doi.org/10.1109/TVCG.2018.2794638