Sharing Secrets with Robots?

Bhakta, R. , Savin-Baden, M. and Tombs, G.

Published version deposited in CURVE January 2014

Original citation & hyperlink:

Bhakta, R., Savin-Baden, M. & Tombs, G. (2014). Sharing Secrets with Robots?. In Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications 2014 (pp. 2295-2301). Chesapeake, VA: Association for the Advancement of Computing in Education (AACE).

http://www.editlib.org/p/147797/

CURVE is the Institutional Repository for Coventry University

http://curve.coventry.ac.uk/open

Sharing Secrets with Robots?

Roy Bhakta
Coventry University, Coventry, UK
r.bhakta@coventry.ac.uk

Maggi Savin-Baden
Coventry University, Coventry, UK
m.savinbaden@coventry.ac.uk

Gemma Tombs
Coventry University, Coventry, UK
aa8943@coventry.ac.uk

Abstract: This paper presents initial findings from a large-scale study that evaluated levels of student disclosure on sensitive topics. Four different conditions of survey delivery were applied and follow up interviews were undertaken. Non-parametric tests were used due to the data not satisfying the assumptions of parametric statistical tests; Wilcoxon Signed Ranks tests were conducted to examine the differences further. Preliminary data suggest that students disclosed additional information to the chatbot on more sensitive topics when the length of engagement was increased, but that this effect could be negated by the inclusion of the depth of engagement questions. Such findings suggest that the sensitivity of the student-chatbot conversation is critical in determining the influence of the chatbot, and that particular care should be taken when designing contextually-relevant 'icebreaker' questions.

Introduction

As internet access in the home, school, and workplace, has increased in recent years (Office for National Statistics, 2012), daily activities have increasingly become mediated through online environments. In education specifically, the rapidly growing popularity of the massive open online course (MOOC) would seem to suggest that the importance of online technologies is only likely to increase.

One technology with increasing relevance in the educational sector is the chatbot or pedagogical agent, defined as a character on a computer screen with embodied lifelike behaviours such as speech, emotions, locomotion, gestures, and movements of the head, the eye, or other parts of the body (Dehn & Van Mulken, 2000). Chatbots have been used in commercial and educational environments to dispense advice or support on a 24/7 basis, but have also been used to deliver educational scenarios in virtual worlds (e.g. Savin-Baden et al., 2009) as well as in 2D environments (Moreno & Flowerday, 2006). Such technologies would seem to offer potential for learning via MOOCs, as well as in other, campus-based, settings, as they help to support and even improve the level of interactive learning on a programme (Yanghee Kim & Wei, 2011). It therefore seems prudent to pause and take stock of what is currently known as well as what is assumed about the use of these technologies in educational settings.

This paper reports on the preliminary findings of a large-scale study into the use of chatbots on sensitive topics in educational settings. It begins by outlining the theoretical and conceptual basis for the study, derived from a detailed exploration of current research and literature on chatbots. It then outlines the methodology and procedures employed for this mixed-method study, and presents preliminary quantitative data. Finally, conclusions are drawn based upon the current data analyses, and recommendations made for educationalists, researchers, and developers who are, or may choose to use chatbots in educational settings.

Literature review

In the 60 years since Alan Turing began work on the Turing Test (1950) to evaluate the realism of chatbots (also known as virtual assistants, pedagogical agents, or HCI agents,), they have become an established, although niche, educational technology. To date, research into the realism of chatbots has formed both the greater part and the basis of use much of these technologies. However, little research has examined issues of disclosure and honesty, which were the focus of this study.

Research into chatbot realism has suggested that anthromorphic (human-like) chatbot appearances and voices are seen as typically desirable by the user base (Baylor, 2011), providing the basis for most chatbot

designs. Factors such as realistic eye-gaze and movement have been found to contribute significantly to the perceived quality of chatbot-user interactions (Garau et al., 2003), and might also improve student learning (Dunsworth & Atkinson, 2007). Other research has focused more closely upon conversational realism, for example. Morrissey and Kirakowski (2013), identified seven themes influencing users' perceptions of realistic chatbot-user interactions. These themes were: maintenance of themed discussions, responding to specific questions, responding to social cues, using appropriate linguistic register, greetings and personality, giving conversational cues, and inappropriate utterances and damage control. In relation to audible realism, Veletsianos (2009) found that verbal expressiveness in chatbots (pauses when speaking, emphasis on certain words, and improved enunciation) improved student learning retention. Students who interacted with the chatbot with increased expressiveness also rated the chatbot's ability to interact as higher than those who interacted with a standard chatbot. Other studies have demonstrated that computer-synthesized voices were perceived less favourably than human voices and might potentially reduce the learning potential of chatbots (Clark & Mayer, 2008). When computer-synthesized voices did not appear to match anthromorphic appearances, split-attention effect occurred in which students experienced a higher cognitive load due to competing demands for their attention, distracting from the learning task at hand (Garau et al., 2003). Split-attention effect, as well as reducing the potential for student learning, can also reduce the believability and perceived trustworthiness of the agent (Demeure, Niewiadomski, & Pelachaud, 2011).

The notion of trustworthiness has been identified as particularly important in sensitive situations, and essential in the formation of an emotional connection between a user and a chatbot (Savin-Baden, Tombs, Burden, & Wood, 2013). Further, the competence of the chatbot has been deemed important in student learning outcomes and perceptions of the chatbot (Kim, Baylor, & PALS Group, 2006). This potential formation of an emotional connection between the user and the chatbot is also affected by the concept of social presence, in which users might feel 'present' in an interaction with a chatbot. Work by Kim and Baylor focused upon the importance of facilitating social presence, suggesting that 'learners miss empathetic social encouragements and caring' (2006, p. 588). Thus whilst split-attention effect might have an impact upon users' experiences of social presence, alongside a multitude of other factors (Childs, 2010), so too might impersonal dialogue and interactions. This should be considered in relation to the work of Wheeless and Grotz (1977) who have argued that the disclosure of sensitive information requires the formation of a trust relationship. More recently, Corritore, Kracher and Wiedenbeck have suggested that websites can be the objects of trust, in which trust is defined as 'an attitude of confident expectation that one's vulnerabilities will not be exploited' (2003, p. 70). When disclosing sensitive information, the vulnerabilities and potential risk associated with trust – even of a chatbot – are that much higher.

The preceding literature review has highlighted the important role of the *user* in the interaction, an often neglected area of research (Lee, 2010). The following section presents preliminary findings from a study designed to explore the ways in which topical and contextual details of a chatbot-user interaction can influence experiences.

Methodology

This study aimed to explore the effect of using a chatbot on levels of information disclosure. It was funded by a large funding body as part of a larger-scale study into the potential influence of chatbots. Research work was undertaken by Coventry University, and development work was undertaken by Daden Ltd. Using a mixed methods design, students were asked to discuss and answer a Student Life survey about student lifestyle issues (finances, alcohol, plagiarism, drugs and sexual health). The Student Life survey was designed to increase levels of sensitivity over time and across subjects, meaning that finances, alcohol, and plagiarism, were perceived as less sensitive, whilst drugs and sexual health were perceived as more sensitive topics. Four different survey conditions were applied:

1) A non-chatbot survey delivered in one session
2) An identical chatbot-based survey delivered in one session (short-term chatbot)
3) An identical chatbot-based survey delivered across five sessions (long-term chatbot, exploring length of engagement)
4) A chatbot-based survey delivered across five sessions and including additional 'icebreaker' questions at the start of each session (long-term chatbot with icebreakers, exploring depth of engagement)

All students were asked to complete the Student Life survey twice, both with and without a chatbot. Thus three groups were formed, all completing survey condition 1, and either condition 2, 3, or 4. The outcome variable was the level of disclosure measured by words typed in response to each survey. 117 participants were recruited from both undergraduate and postgraduate courses across a variety of disciplines and provided

complete data through completion of both the chatbot and non-chatbot versions of the questionnaire (see Table 1).

Group	Male	Female	Mean Age (Standard Deviation)
Non-chatbot and short term chatbot (standard deviation)	5	28	20.15 (3.50)
Non-chatbot and long term chatbot (standard deviation)	6	32	22.79 (7.5)
Non-chatbot and long term chatbot with icebreakers (standard deviation)	6	40	20.72 (5.36)
Overall (standard deviation)	17	100	21.23 (5.8)

Table 1. Age and gender of participants involved in the study

A website was designed by Daden Ltd. to allow students to access an online questionnaire which comprised three demographics question and 24 items relating to the student lifestyle topics of finances, alcohol, plagiarism, drugs and sexual health. The website presented the questions first and then provided some information regarding that topic, before moving onto the next topic. When participants were using the chatbot version of the questionnaire, they were given an option to pick from 10 chatbots which ranged in age, gender, ethnicity and appearance. Chatbot voices were chosen to match appearances, but were computer-synthesized. The figure below provides an example of the website when completing a chatbot-based survey.
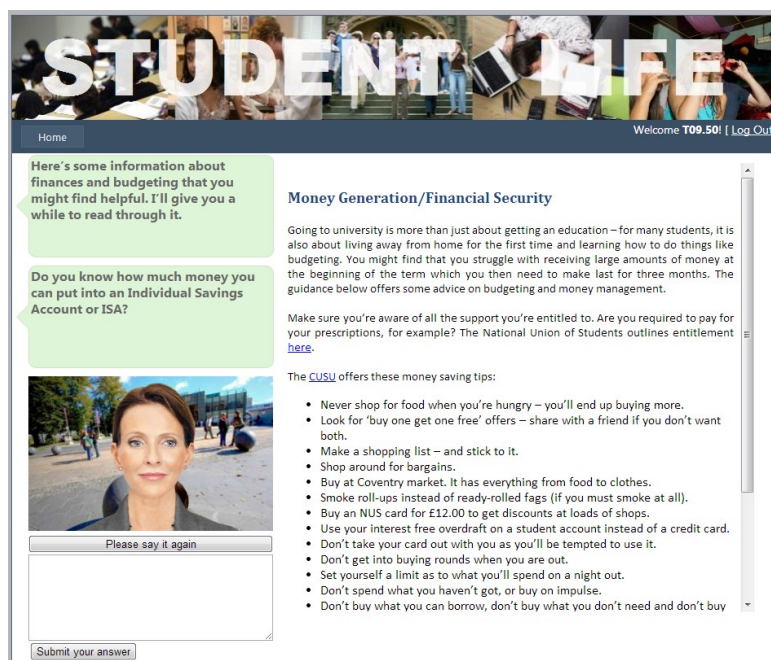

Figure 1. Example of a chatbot survey on the Student Life website

There was a gap of between two and three weeks between completing the chatbot and non-chatbot surveys. The order in which the three groups completed the chatbot and non-chatbot surveys was randomised so as to reduce order effects. After completing the chatbot-based survey, students then completed a short follow-up survey based at Bristol Online Surveys. This asked students to choose three words to describe their chatbot; the list is shown in figure 5, although students could also add their own terms.

Ethical approval was sought from both the funder's Ethics Committee and Coventry University Ethics Committee. Participation was voluntary, although incentives of £10 Amazon vouchers were provided in thanks for student participation. Participants were also made aware that they could withdraw from the study at any time up to a month after the last questionnaire had been completed and that all data would be anonymised, confidential, and securely stored.

Results

Preliminary results from this study suggest that different topics across the Student Life survey were perceived and responded to differently, and that the length of engagement with the chatbot survey had a significant effect. Data suggest that, overall, individuals were more likely to talk about sensitive topics such as drugs or sex when talking with a chatbot compared to the non-chatbot survey. In contrast, individuals showed a tendency to talk less about topics such as money, drink and plagiarism when talking to a chatbot. A significant increase in information disclosure to the chatbot was observed in Group 2, which compared the long-term chatbot survey to the non-chatbot survey; significance was not reached in the other two groups. Additionally, participants in the group comparing the long-term chatbot with icebreaker questions spoke significantly less about finances and alcohol in the chatbot condition. The following section outlines these findings and the methods by which they were reached.

Non-parametric tests were used due to the data not satisfying the assumptions of parametric statistical tests. Data from individual groups were found to deviate from a normal distribution in terms of both skewness and kurtosis. Transformation of the data was attempted; however, doing so did not rectify the problems of normality.

	Finances	Alcohol	Plagiarism	Drugs	Sex	Overall
Group 1 *Compares short-term chatbot condition to non-chatbot survey condition*						
Participant *n* =	33	33	33	33	33	33
No Chatbot	31	27	37	62	19	184
Chatbot	35	27	39	67	18	195
Group 2 *Compares long-term chatbot condition to non-chatbot survey condition*						
Participant *n* =	38	37	37	37	31	31
No Chatbot	42	36	43	**72**	**16**	220
Chatbot	34	32	48	**80**	**22**	242
Group 3 *Compares long-term chatbot with icebreaker questions condition to non-chatbot survey condition*						
Participant *n* =	46	46	46	44	38	38
No Chatbot	**38**	**34**	41.5	60	20	216
Chatbot	**28**	**26**	42	69	17.5	197

Table 2. Median word length of the answers given by participants in response to online Student Life surveys

Figures 2 and 3 (below) illustrate the differences between chatbot and non-chatbot responses in the two long-term chatbot groups.

Figure 2. Median number of words for the responses to sexual health and drug related topics in Group 2

Figure 3. Median number of words for the responses to financial and alcohol related topics in Group 3

Wilcoxon Signed Ranks tests were conducted to examine the differences further and revealed that only participants in the long term chatbot group responded with longer answers when posed by a chatbot than without the chatbot when discussing drugs ($z = 2.023$, $p < 0.05$, $r = 0.238$) and sexual health ($z = 2.952$, $p < 0.05$, $r = 0.375$). However, participants in the long term engagement group responded with shorter answers when posed by a chatbot than without the chatbot when discussing finances ($z = 1.956$, $p < 0.05$, $r = 0.202$) and alcohol ($z = 2.475$, $p < 0.05$, $r = 0.255$).

Participants were also asked to describe their experience of using the chatbot by selecting several words that best describe how they feel about the avatar that they had chosen (see Figure 4 below). Over 50 per cent had chosen their avatar based on the perception of it being Friendly, Professional and Approachable. However, less than 7 per cent perceived their chosen avatar to be Hip, Judgemental or Disconcerting.

Figure 4. Characteristics seen to be most important in choosing a chatbot

Discussion

The findings suggest that both the topic under discussion and the length of engagement with the chatbot are important in determining information disclosure levels. In this study, students who answered the Student Life survey over a period of two weeks (but did not answer the off-topic questions designed to increase depth of engagement) disclosed significantly more information to the chatbot when discussing sexual health and drug use, but not on the financial, plagiarism, and alcohol questions. No such significance was found in students who answered the survey designed to increase both length and depth of engagement, or in students who answered the chatbot survey in one one-off session, although similar trends can be observed.

The general trend towards longer responses on the topics of drugs and sexual health might perhaps be attributed to a current focus on issues of personal safety in the university context, which focuses particularly on drugs and sexual health. With regard to the long-term chatbot condition particularly, these findings would seem to suggest that a stronger emotional connection – leading to increased disclosure – might be forged in this condition. Interviews undertaken by Kim (2007) have yielded data indicating that chatbot characteristics of 'friendliness' are particularly important, as illustrated by our findings in Figure 4. Veletsianos and Miller (2008) have speculated that longer-term interactions with chatbots may have a positive effect on emotional engagement. Thus far, empirical studies of chatbots have neglected this area; these findings, although preliminary, reinforce the argument that examinations of longitudinal chatbot engagement are essential (Veletsianos & Russell, 2014).

Findings from this study also suggest that the increased disclosure in the long-term chatbot setting might be negated by the inclusion of the 'icebreaker' questions designed to increase the depth of interaction. This demonstrates that the inclusion of these questions certainly has an effect but perhaps not the desired one. We have identified two possible reasons for this finding. Firstly, Culley and Madhavan have cautioned that 'as the agent becomes increasingly morphologically similar to a human, it is also likely that operators will engage in correspondence bias more frequently by ascribing human motivations, reasoning abilities and capabilities to this non-human system' (2013, p. 578). Consequently, the inclusion of questions designed to facilitate increased engagement might have resulted in students feeling a sense of talking to a person and thus perhaps being less willing to disclose information. We derive this potential explanation from findings of an earlier study in which qualitative responses indicated that student willingness to disclose sensitive information to chatbots was attributed partially to chatbots being *almost* like a person (Savin-Baden et al., 2013). As the long-term chatbot was designed to create an emotional connection by asking off-topic questions, it may have shifted in student perceptions to seeming more humanistic and thus likely to seem 'judgemental' of responses to sensitive questions. These findings should be considered in relation to Lessler and O'Reilly's (1997) finding that self-administered surveys could yield more truthful responses than interview methods, particularly in sensitive disclosure situations. However, Hasler, Touchman, and Friedman (2013) have also found in a comparison of human interviewees with virtual world chatbots, that chatbots and human interviewees were equally successful in collecting information about participants' backgrounds. Chatbots would thus seem to offer and interstitial or liminal space , a space between human and technology, a space which might be perceived as 'safe' when disclosing sensitive information. It seems important, therefore, to pay particular attention to Culley and Madhavan's (2013) caution to be aware of the implications of increased avatar realism. Yet the lessening of student disclosure in this particular setting might also be attributed to split-attention effect and a lack of conversational realism. The chatbot in this study asked how students were, but was unable to respond when asked "How are you?" Chatbot ability to respond to conversational norms and engage in bridging topics has been identified as particularly important in users' perception of realistic conversations (Morrissey & Kirakowski, 2013), and this should also be considered in future studies. This study suggests that particular attention should be paid to design when seeking to facilitate increased depth of engagement, and that the concept of chatbot realism is particularly important in sensitive settings. Whilst this study focused on student lifestyle issues, we contend that the same might apply in counselling or healthcare educational situations, for example. Earlier studies on disclosure to chatbots have largely failed to consider the influence of the sensitivity of questions; our findings suggest that different topics even across the same study can yield different levels of disclosure.

Certain limitations of this study need to be taken into account. Firstly, we have only measured disclosure levels, i.e. number of words disclosed in the chatbot and non-chatbot settings. From that, we are unable to derive the quality and truthfulness of the chatbot-student interaction. Future work will focus upon the content of these interactions and consider qualitative comparisons of the chatbot and non-chatbot responses. Secondly, there was a high female bias in the sample, meaning that conclusions cannot be drawn based upon gender.

Conclusion

Honesty and truth telling thus remains complex issues to manage in a networked society and the amount of truthful information divulged was dependent on how well the participant engaged with the pedagogical agent. Tourangeau and Smith (1996) suggested that the reasons participants provide different answers to the same questions under different data collection methodologies are a function of three issues –

- The degree of perceived privacy;
- The legitimacy it confers;
- The cognitive burden it imposes on the respondents.

The use of chatbots in both commercial and educational settings has the potential to disrupt the ways in which we interact in online settings, as "robots [become] part of our relational futures" (Turkle, 2010, p. 4); it is from this platform that we suggest there is a greater need to understand the ways in which individuals relate and disclose information to chatbots. As pedagogical agent technologies are increasingly integrated into commercial and educational arenas, it seems likely that they will transfer to mobile as well as blended learning settings.

References

Baylor, A. L. (2011). The design of motivational agents and avatars. *Educational Technology Research and Development*, *59*(2), 291–300. doi:10.1007/s11423-011-9196-3

Childs, M. (2010, November). *Learners' Experience of Presence in Virtual Worlds* (Doctoral). University of Warwick.

Clark, R., & Mayer, R. E. (2008). *E-learning and the science of instruction* (2nd ed.). San Francisco: Jossey-Bass.

Corritore, C. L., Kracher, B., & Wiedenbeck, S. (2003). On-line trust: concepts, evolving themes, a model. *International Journal of Human-Computer Studies*, *58*(6), 737–758. doi:10.1016/S1071-5819(03)00041-7

Culley, K. E., & Madhavan, P. (2013). A note of caution regarding anthropomorphism in HCI agents. *Computers in Human Behavior*, *29*(3), 577–579. doi:10.1016/j.chb.2012.11.023

Dehn, D. M., & Van Mulken, S. (2000). The impact of animated interface agents: a review of empirical research. *International Journal of Human-Computer Studies*, *52*(1), 1–22. doi:10.1006/ijhc.1999.0325

Demeure, V., Niewiadomski, R., & Pelachaud, C. (2011). How is believability of a virtual agent related to warmth, competence, personification, and embodiment? *Presence: Teleoperators and Virtual Environments*, *20*(5), 431–448.

Dunsworth, Q., & Atkinson, R. K. (2007). Fostering multimedia learning of science: Exploring the role of an animated agent's image. *Computers & Education*, *49*(3), 677–690. doi:10.1016/j.compedu.2005.11.010

Garau, M., Slater, M., Vinayagamoorthy, V., Brogni, A., Steed, A., & Sasse, M. A. (2003). The impact of avatar realism and eye gaze control on perceived quality of communication in a shared immersive virtual environment (p. 529). ACM Press. doi:10.1145/642611.642703

Hasler, B. S., Tuchman, P., & Friedman, D. (2013). Virtual research assistants: Replacing human interviewers by automated avatars in virtual worlds. *Computers in Human Behavior*, *29*(4), 1608–1616. doi:10.1016/j.chb.2013.01.004

Kim, Y. (2007). Desirable characteristics of learning companions. *International Journal of Artificial Intelligence in Education*, *17*(4), 371–388. Retrieved from http://ijaied.org/pub/1110/file/1110_Kim07.pdf

Kim, Y,, & Baylor, A. L. (2006). A Social-Cognitive Framework for Pedagogical Agents as Learning Companions. *Educational Technology Research and Development*, *54*(6), 569–596. doi:10.1007/s11423-006-0637-3

Kim, Y., Baylor, A. L., & PALS Group. (2006). Pedagogical Agents as Learning Companions: The Role of Agent Competency and Type of Interaction. *Educational Technology Research and Development*, *54*(3), 223–243. doi:10.1007/s11423-006-8805-z

Kim, Y., & Wei, Q. (2011). The impact of learner attributes and learner choice in an agent-based environment. *Computers & Education*, *56*(2), 505–514. doi:10.1016/j.compedu.2010.09.016

Lee, E.-J. (2010). The more humanlike, the better? How speech type and users' cognitive style affect social responses to computers. *Computers in Human Behavior*, *26*(4), 665–672. doi:10.1016/j.chb.2010.01.003

Lessler, J. T., & O'Reilly, J. M. (1997). Mode of interview and reporting of sensitive issues: Design and implementation of audio computer assisted self-interviewing. In L. Harrison & A. Hughes (Eds.), *The validity of self-reported drug use: Improving the accuracy of survey measurements* (pp. 366–382). Rockville, MD: National Institute of Drug Abuse.

Moreno, R., & Flowerday, T. (2006). Students' choice of animated pedagogical agents in science learning: A test of the similarity-attraction hypothesis on gender and ethnicity. *Contemporary Educational Psychology*, *31*(2), 186–207. doi:10.1016/j.cedpsych.2005.05.002

Morrissey, K., & Kirakowski, J. (2013). "Realness" in chatbots: Establishing quantifiable criteria. In M. Kurosu (Ed.), *Human-Computer Interaction: Interaction Modalities and Techniques* (Vol. 8007, pp. 87–96). Berlin: Springer-Verlag.

Office for National Statistics. (2012). *Internet Access - Households and Individuals, 2012* (Statistical Bulletin). London: Office for National Statistics. Retrieved from http://www.ons.gov.uk/ons/dcp171778_275775.pdf

Savin-Baden, M., Hilton, S., Kavia, S., Linsey, T., Morris, D., Poulton, T., … Younis, A. (2009). *PREVIEW project final report*. JISC. Retrieved from https://curve.coventry.ac.uk/cu/items/38cd8bee-ae1d-d50c-e9c0-16a1dec639757a/1/PREVIEW%20Final%20Report.pdf

Savin-Baden, M., Tombs, G., Burden, D., & Wood, C. (2013). "It's Almost like Talking to a Person'. *International Journal of Mobile and Blended Learning*, *5*(2), 78–93. doi:10.4018/jmbl.2013040105

Tourangeau, R., & Smith, T. W. (1996). Asking sensitive questions: the impact of data collection mode, question format, and question context. *Public Opinion Quarterly*, *60*, 275–304.

Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, *49*, 433–460.

Turkle, S. (2010). In good company? On the threshold of robotic companions. In Y. Wilks (Ed.), *Close engagements with artificial companions: key social, psychological, ethical and design issues* (pp. 3–10). Amsterdamn ; Philadelphia, PA: John Benjamins Pub. Company.

Veletsianos, G. (2009). The Impact and Implications of Virtual Character Expressiveness on Learning and Agent-Learner Interactions. *Journal of Computer Assisted Learning*, *25*(4), 345–357. doi:10.1111/j.1365-2729.2009.00317.x

Veletsianos, G., & Miller, C. (2008). Conversing with pedagogical agents: A phenomenological exploration of interacting with digital entities. *British Journal of Educational Technology*, *39*(6), 969–986. doi:10.1111/j.1467-8535.2007.00797.x

Veletsianos, G., & Russell, G. (2014). Pedagogical Agents. In M. Spector, D. Merrill, J. Elen, & M. J. Bishop (Eds.), *Handbook of Research on Educational Communications and Technology* (4th ed., pp. 759–769). New York: Springer Academic. Retrieved from http://www.veletsianos.com/wp-content/uploads/2013/08/pedagogical_agents_synthesis_veletsianos_handbook.pdf

Wheeless, L., & Grotz, J. (1977). The measurement of trust and its relationship to self-disclosure. *Human Communication Research*, *3*(3), 250–257. doi:10.1111/j.1468-2958.1977.tb00523.x

Acknowledgements