

Analytical Steps for the Validation of a Natural User Interface

Madlen Wuttke^(✉), Sabine Völkel, Peter Ohler, and Nicholas H. Müller

Media-Psychology, Faculty of Humanities, Chemnitz University of Technology,
Thüringer Weg 11, 09126 Chemnitz, Germany
{madlen.wuttke, sabine.voelkel, peter.ohler,
nicholas.mueller}@phil.tu-chemnitz.de

Abstract. Pedagogical Agents are primarily researched regarding their depiction or appearance on screen, rather than their capabilities to react to a learner in front of it. As it has been previously reported, we developed an agent system, which is based on an electronic educational instance for a learning module to incorporate information from the environment as well as non-verbal reactions of a user. The system is then capable of proactively reacting to this additional information, establishing a natural user interface as it might be expected by any user. In order to validate our approach, the steps for our experimental setup are presented and discussed.

Keywords: Pedagogical · Agent · Proactive · Validation

1 Introduction

Within this paper, the authors refer to their validation efforts of the previously described concept [5]. A user study with 160 participants was conducted with four different variations of learner support aspects. To check for the persona effect, as stated by Lester et al. [9], the electronic educational instance (Fig. 1. EEI [6] and Enhanced capabilities [5].) is used in conjunction with a depicted pedagogical agent or without. The persona effect formulates that there is a per se beneficial effect of including a depicted agent onscreen. This might be true due to the expectancy of social adequate behavior [8], which is why the experimental setup displayed the agent onscreen and had variations where the agent's voice was audible but there was no depiction visible on the screen.

The development of our agent has been deeply influenced by the research activities of Reeves and Nass [8] who postulated a user's expectancies of a social adequate behavior when interacting with a machine. Our process of creating the electronic educational instance has been previously published [5–7, 12, 13]. Therefore, we will only present a limited overview of our previous development and focus on the validation aspects.

Ever since Lester et al. established the persona effect [9] research regarding pedagogical agents is heavily focused on the outward appearance of an agent. For example, the various depictions of an agent as well as their embodiment [2] have been a focus of research as has been their outward appearance, including their manners and behavior [3, 14–16]. Research regarding their appearance also includes their vocalization of learning material and

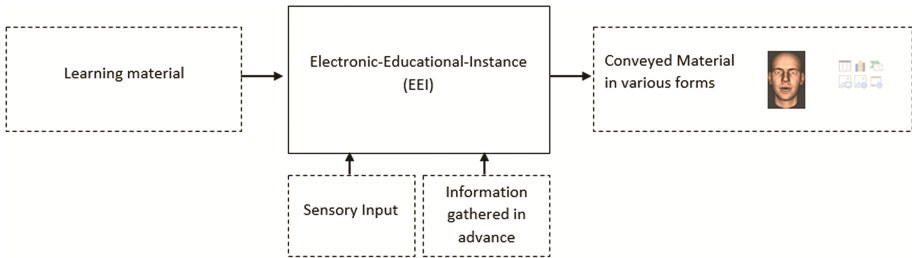


Fig. 1. EEI [6] and Enhanced capabilities [5]

their conversational behavior in general [4, 17, 18]. Heidig and Clarebout [1] published an overview of various aspects of pedagogical agents and their possible benefits for electronic learning.

These aspects of exploration all focus on one dimension of the user-agent relationship, namely having an agent appear more lifelike or to behave and communicate most helpfully. Within our approach, we argue for a shift to include additional input channels for the agent to get a more thorough grasp of the learning situation and about the context in which a learner is situated.

Based on our research, the possibilities for a user to interact with a learning system are still limited to traditional input channels like keyboard and mouse interfaces. Although, as Krämer [19] pointed out, the raw processing power of computer systems is already on a level, which would allow the implementation of much more natural ways of communication. Even though recent advances in the form of speech and gesture recognition has opened up new ways of interaction, this still has not advanced to the implementation of system-based reactions to a user's behavior in front of the screen or to environmental cues related to the learner. As it has been previously established by conceptual research [5] the integration of advanced input channels for computer and web based trainings should provide a measurable benefit during the conveyance of information, simply by being able to pause a training program if a student is looking away from the screen or there is too much noise in the environment. By implementing an electronic educational instance [6, 7] it is possible to enhance a wide array of possible variations of learning software. The idea behind the electronic educational software is to enable the computer system to behave in accordance with the media equation theory as postulated by Reeves and Nass [8]. As they stated, humans tend to subconsciously expect technical devices to behave as another human being would. Including the expectation of transporting non-verbal cues. An agent-system capable of identifying such cues would be immensely more useful when conveying learning material due to its capability to factor real time information into the presentation. Within our system, this explicitly means environmental cues like the noise level of the learning environment as well as the focused gaze of the learner in front of the computer. For the presented experimental validation, the attention onto the screen is key in having a thorough understanding of the learning material and the capability of actively utilizing learned knowledge.

2 Real-World Applications

Implementing a proactive functionality to any technology promises broad new forms of human computer interactions in general. Due to the EEI and its enhanced capabilities as being a standalone component, it is possible to implement our sensory concept into pretty much any form of technology. For example, Smart TV systems, which would pause and resume a movie or a sports game until the attention is reverted back onto the screen or the audience remains within the visible reach of the television screen. Another form would be the implementation into the modern aspects of mobile learning scenarios, in which a system would react to the users surrounding and would be able to proactively offer another form of material conveyance, e.g. an audio transcript while waiting at a bus stop and a combined audio-video demonstration once seated and commuting [20].

Given a real-world learning scenario, a human teacher would be able to react to a student. Would there be a noise disturbance in the environment of the classroom, a teacher would wait for the disturbance to go away or reaches a low enough level to be sure that audible information would be understandable. In addition, if one or more students divert their gaze away from what is shown by the teacher, the presentation would be paused until attention has been reverted to the topic at hand. Additionally, if a teacher would detect confusion amongst students, a different approach would be used to look at a specific topic from another perspective and thus possibly have the learning material conveyed more easily.

3 Experimental Validation

Therefore, we enhanced the electronic educational instance [8] to be able to detect a users' gaze. A commonly used webcam is able to detect the eyes of a user and therefore infers that the learning material is actively consumed (Fig. 2. The experimental setup with webcam and eye-tracking-camera.). Once the user is looking away from the screen and the software notices a deviation of the gaze away from the monitor, the presentation is paused and only resumed, once the gaze and therefore the attention is reverted back to the learning material.

Test subjects were asked to study an already established web-based training in use at the institute of media research from 2006 to 2012. Within the WBT, the basic functions of the Adobe Dreamweaver Suite are explained. Three basic ideas were to be tested during the experimental run.

- Does the proactive function lead to a better understanding of the learning material once the learner's attention is diverted?
- Does the depiction of the agent (persona effect) have a measurable impact on the understanding of the learning material?
- Is it necessary to include an intervention of the proactive functionality in order to allow a learner to repeat a session or to continue from the point of diverted attention?

To ensure that our experiment is capable of testing for various empirical aspects (e.g. the persona effect [9]) six groups of participants were tested (Table 1).

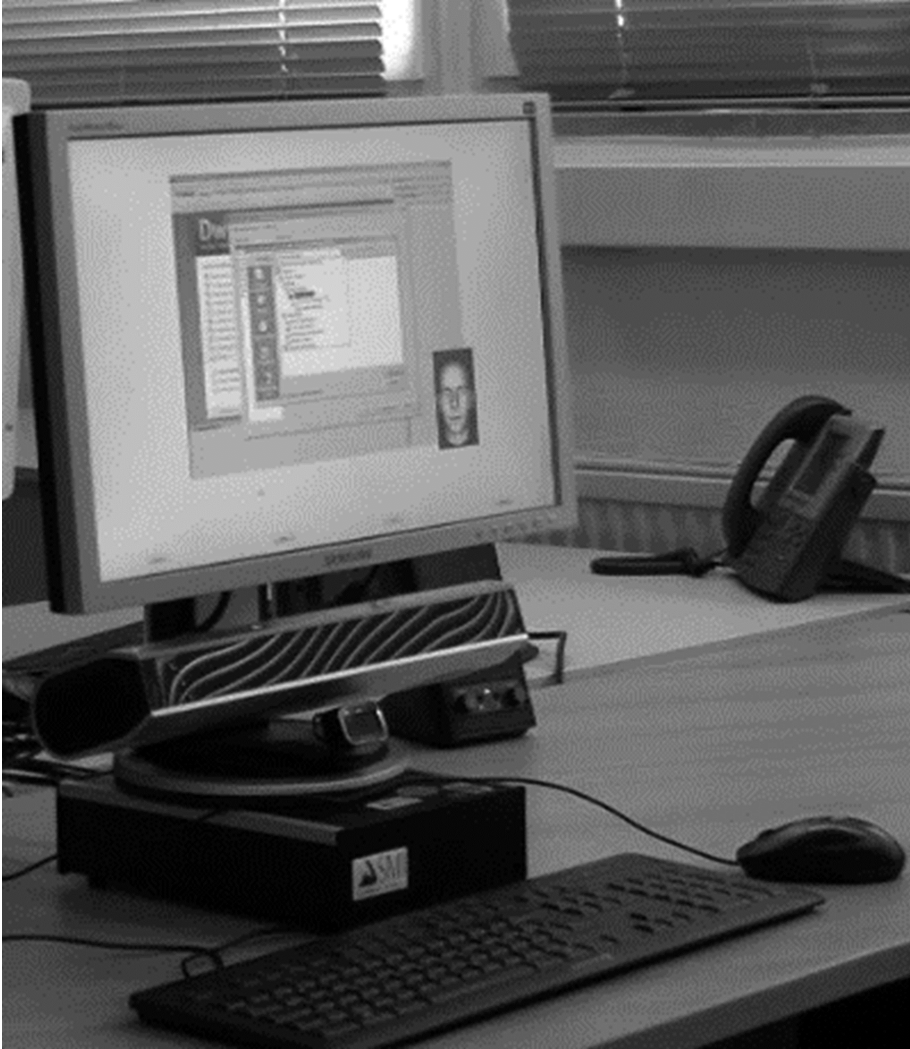


Fig. 2. The experimental setup with webcam and eye-tracking-camera

Table 1. Experimental groups

Group 1	Depicted agent, with proactivity
Group 2	Depicted agent, without proactivity
Group 3	Audio-only agent, with proactivity
Group 4	Audio-only agent, without proactivity
Group 5	Audio-only agent, with proactivity, no intervention message
Group 6	Depicted agent, with proactivity, no intervention message

In order to create a lifelike agent onscreen, we used the FaceShift software, which, using a Microsoft Kinect camera, allows the immediate capture of mimic and voiced mouth movements onto a 3D head model (Fig. 3. Faceshift capture of speaker).

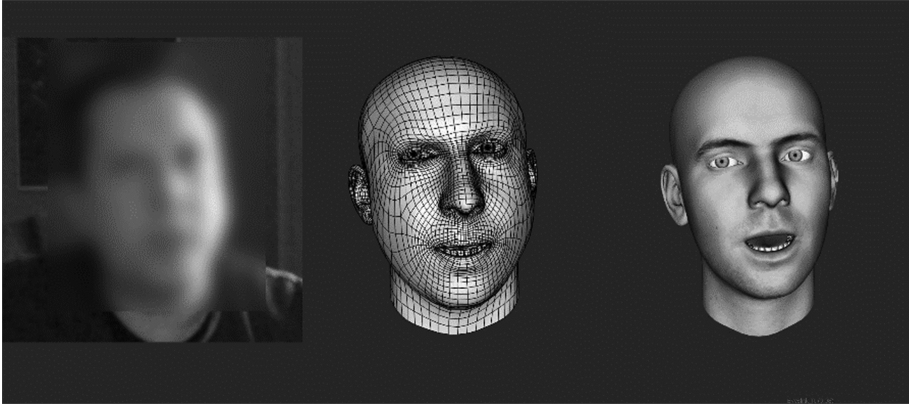


Fig. 3. Faceshift capture of speaker

The recorded re-readings of the information material were recorded and afterwards synchronized with the original web based training by using Adobe Premiere (Fig. 4. Synchronization of WBT and character animation).

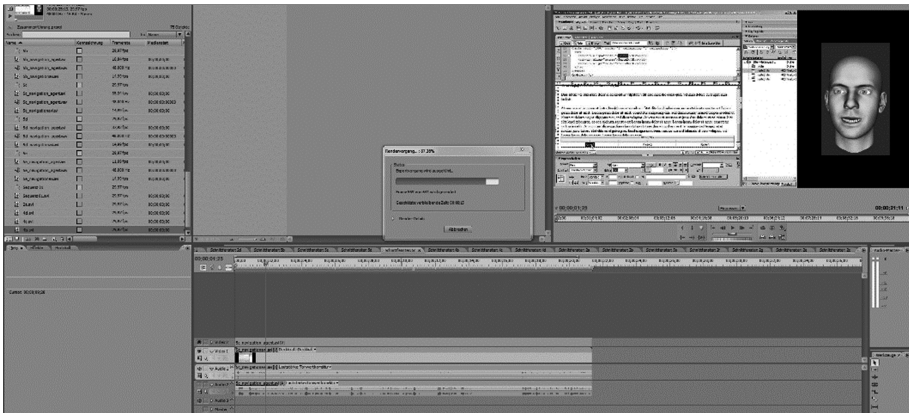


Fig. 4. Synchronization of WBT and character animation

The displayed Agent interaction onscreen differs only marginally from group to group in order to test for aspects of the persona effect, the benefits of proactivity and the intervention of the agent system (Fig. 5. Displayed variations onscreen). Main differences are the depiction of the agent itself and the display of interaction buttons once the presentation stopped during the experiment due to an averted gaze.

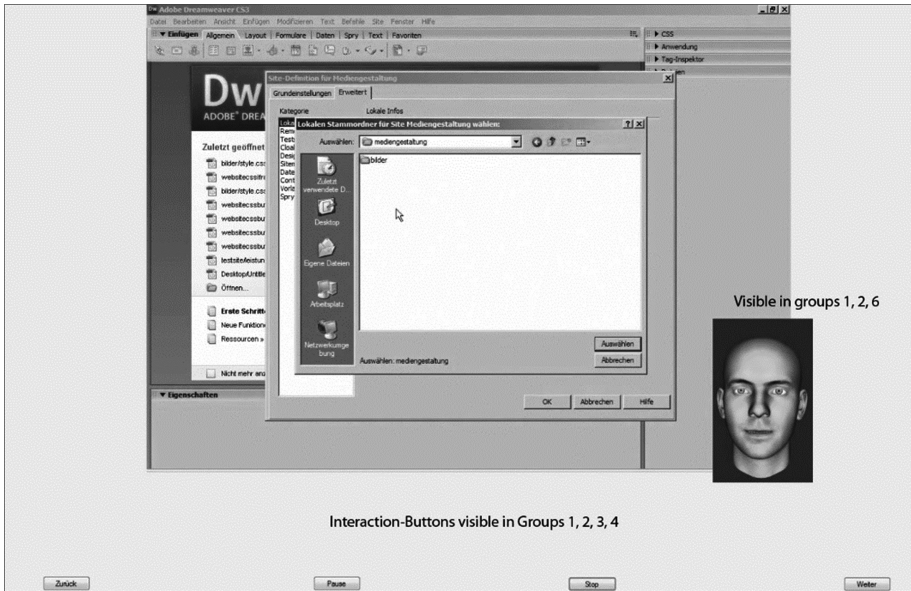


Fig. 5. Displayed variations onscreen

Before the experimental variations started, the general knowledge of participants regarding websites, html code and the Dreamweaver software is gathered via an electronic LimeSurvey questionnaire. Similar questions were asked once the experiment was over to check for increased knowledge due to the web-based training and to expose missing knowledge during the diverted attention.

To ensure any validation could be traced back to the manipulations, the participants were asked to perform a second task while studying the material. At two standardized segments of the presentation, at a critical information conveyance, a monitor to the left of the participants flashed and a speaker played an alarm. Participants were given a password before the experiment and told to enter it, once something happens on the screen to their left.

4 Conclusion

Within this paper, we presented the steps for the validation of our implemented electronic educational instance (EEI). We argue to broaden the scope of pedagogical agent research to include various other aspects when deciding how to better enable an information conveyance using electronic learning media. Especially, not to focus any longer on instructional design methods and new forms of information segmentation, but to include new and readily available input channels into electronic learning systems.

Even nowadays, most modern e-learning software is limited to mouse and keyboard inputs and learning success is mostly defined by having a learning success test implemented at the end of a chapter. However, the information whether or not a learner is

able and capable to understand certain information is available long before this end point of the conveyance – it is visible due to non-verbal cues of averted attention. Due to the proposed EEI module, any existing learning software could be rapidly upgraded to notice any diversion from the screen and due to ubiquitous webcam-in-screen-frame designs of modern notebooks, tablets and 2-in-1 computers; there are not even monetary investments necessary.

Preliminary analysis of our validation shows a clear proof of the obvious: it is beneficial to the information conveyance, if a user is actually looking at the screen and attentive when said information is delivered.

Furthermore, our next steps for implementation will include a cognitive workload assessment based on recorded pupil dilations [10]. Due to this, we will be able to know in real-time whenever any given material is too hard to understand for any given user in front of the learning system. Therefore, the agent system is capable to proactively change the conveyance of the material, to repeat certain aspects until cognitive load levels [21, 22] normalize or to change the mode of presentation. This would allow for a unique learner calibration of learning material, which right now is, if at all, performed by questions about preliminary knowledge beforehand.

In addition, we are about to implement an auditory sensor array which checks for environmental noise levels and decides whether or not the speaker output-level is suitable for the current situation or if it would be beneficial to raise the amplitude. If noise levels were too disruptive, the presentation would be stopped proactively entirely while the system waits for the surrounding circumstances to normalize.

Moreover, beyond these readily available modes, we are currently working on implementing a facial action coding reconnaissance module. This FACS [11] would present another non-verbal cue about the learner's inner state as the cognitive workload module. Due to certain facial muscle activations, it is possible to detect basic emotions on a user's face and therefore compute an emotional valence level. As long as this would be in a neutral or positive domain, the information conveyance might be all right or even entertaining. However, once the emotional valence level threshold is crossed into a negative sphere, the system should be able to incorporate this information. Further closed eyelids might indicate boredom or at least sleepiness and is therefore not beneficial for a learning success. Therefore, the system would be able to know when a unique learner is not able to follow the conveyance of information and suggest either a break or to choose another variation of conveyance by, for example, switching to an easier but more time-consuming variant.

Due to these non-verbal detection modules, it would be possible to detect specific instances where additional learner support would be beneficial instead of the standard time for the detection of faulty conveyance – during the learner success test at the end of a computer or web-based training program.

Although we discuss all of these non-verbal cue detection modules inside a learning setting, there are numerous possible applications, which would benefit from a proactive sensory phalanx. Smart home systems and self-driving cars would benefit from knowing a user's valence state, as would streaming-services when recommending movies or tv-shows. The cognitive load level could be used to detect a pilot's overextension in aviation and deviations from established gaze patterns in air-traffic-controllers might indicate a problem.

Therefore, although this paper is merely outlining the implementation aspects of our validation, we showed what steps are necessary to implement a proactive agent system and we will continue to include said non-verbal detection modules into every-day-life human-computer-interactions.

References

1. Heidig, S., Clarebout, G.: Do pedagogical agents make a difference to student motivation and learning? *Educ. Res. Rev.* **6**, 27–54 (2011)
2. Lusk, M.M., Atkinson, R.K.: Animated pedagogical agents: does their degree of embodiment impact learning from static or animated worked examples? *Appl. Cogn. Psychol.* **21**, 747–764 (2007)
3. Wang, N., Johnson, W.L., Mayer, R.E., Risso, P., Shaw, E., Collins, H.: The politeness effect: pedagogical agents and learning outcomes. *Int. J. Hum.-Comput. Stud.* **66**, 98–112 (2008)
4. Veletsianos, G.: How do learners respond to pedagogical agents that deliver social-oriented non-task messages? Impact on student learning, perceptions, and experiences. *Comput. Hum. Behav.* **28**, 275–283 (2012)
5. Wuttke, M., Heidt, M., Rosenthal, P., Ohler, P., Müller, N.H.: Proactive functions of a pedagogical agent – steps for implementing a social catalyst function. In: Zaphiris, P., Ioannou, A. (eds.) *Learning and Collaboration Technologies, LCT 2016*. LNCS, vol. 9753, pp. 573–580. Springer, Cham (2016). doi:[10.1007/978-3-319-39483-1_52](https://doi.org/10.1007/978-3-319-39483-1_52)
6. Wuttke, M.: Pro-active pedagogical agents. In: Fakultät für Informatik (ed.) *Proceedings of International Summer Workshop Computer Science*, pp. 59–62, July 2013
7. Wuttke, M., Heidt, M.: Beyond presentation - employing proactive intelligent agents as social catalysts. In: Kurosu, M. (ed.) *HCI 2014*. LNCS, vol. 8511, pp. 182–190. Springer, Cham (2014). doi:[10.1007/978-3-319-07230-2_18](https://doi.org/10.1007/978-3-319-07230-2_18)
8. Reeves, B., Nass, C.: *The Media Equation. How People Treat Computers, Televisions, and New Media Like Real People and Places*. Cambridge University Press, New York (1996)
9. Lester, J.C., Converse, S.A., Kahler, S.E., Barlow, S.T., Stone, B.A., Bhogal, R.S.: The persona effect: affective impact of animated pedagogical agents. In: Pemberton, S. (ed.) *Human Factors in Computing Systems: CHI 1997 Conference Proceedings*, pp. 359–366. ACM Press, New York (1997)
10. Rosch, J.L., Vogel-Walcutt, J.J.: A review of eye-tracking applications as tools for training. *Cogn. Technol. Work* **15**(3), 313–327 (2013)
11. Ekman, P., Friesen, W.V., Hager, J.C.: *Facial Action Coding System - The Manual*. Research Nexus Division of Network Information Research Corporation, Salt Lake City (2002)
12. Wuttke, M., Martin, K.-U.: Natural forms of communication and adaptive behaviour in human-computer-interaction. In: Kurosu, M. (ed.) *HCI 2014*. LNCS, vol. 8511, pp. 641–647. Springer, Cham (2014). doi:[10.1007/978-3-319-07230-2_61](https://doi.org/10.1007/978-3-319-07230-2_61)
13. Wuttke, M., Belentschikow, V., Müller, N.H.: Storytelling as a means to transfer knowledge via narration – a scenario for a narrating pedagogical agent. *i-com* **14**(2), 155–160 (2015). doi: [10.1515/icom-2015-0034](https://doi.org/10.1515/icom-2015-0034)
14. Graesser, A.C., Person, N.K., Harter, D., Group, T.R.: Teaching tactics and dialog in AutoTutor. *Int. J. Artif. Intell. Educ.* **12**, 257–279 (2001)
15. Kim, Y., Baylor, A.L., Shen, E.: Pedagogical agents as learning companions: the impact of agent emotion and gender. *J. Comput. Assist. Learn.* **23**, 220–234 (2007)
16. Chan, T.W., Chou, C.Y.: Exploring the design of computer supports for reciprocal tutoring systems. *Int. J. Artif. Intell. Educ.* **8**, 1–29 (1997)

17. Domagk, S.: Do pedagogical agents facilitate learner motivation and learning outcomes? The role of the appeal of agent's appearance and voice. *J. Media Psychol.* **22**(2), 82–95 (2010)
18. Nass, C., Isbister, K., Lee, E.J.: Truth is Beauty, Researching Embodied Conversational Agents: Embodied Conversational Agents, pp. 374–402. The MIT Press, Cambridge (2000)
19. Krämer, N.C.: Soziale Wirkungen virtueller Helfer. Gestaltung und Evaluation von Mensch-Computer-Interaktion. Kohlhammer, Stuttgart (2008)
20. Martin, K.-U., Wuttke, M., Hardt, W.: Sensor based interaction mechanisms in mobile learning. In: Zaphiris, P., Ioannou, A. (eds.) *Learning and Collaboration Technologies. Technology-Rich Environments for Learning and Collaboration, LCT 2014. LNCS*, vol. 8524, pp. 165–172. Springer, Cham (2014). doi:[10.1007/978-3-319-07485-6_17](https://doi.org/10.1007/978-3-319-07485-6_17)
21. Chandler, P., Sweller, J.: Cognitive load theory and the format of instruction. *Cogn. Instr.* **8**(4), 293–332 (1991)
22. Sweller, J., Chandler, P.: Evidence for cognitive load theory. *Cogn. Instr.* **8**(4), 351–362 (1991)