12-12-2022

Exploring the Impact of Inclusive PCA Design on Perceived Competence, Trust and Diversity

Ricarda Schlimbach
Technische Universität Braunschweig, r.schlimbach@tu-bs.de

Susanne Robra-Bissantz
Technische Universität Braunschweig, s.robra-bissantz@tu-bs.de

Exploring the Impact of Inclusive PCA Design on Perceived Competence, Trust, and Diversity Awareness

Ricarda Schlimbach
TU Braunschweig
r.schlimbach@tu-bs.de

Susanne Robra-Bissantz
TU Braunschweig
s.robra-bissantz@tu-bs.de

ABSTRACT

Pedagogical Conversational Agents (PCAs) conquer academia as learning facilitators. Due to user heterogeneity and need for more inclusion in education, inclusive PCA design becomes relevant, but still remains understudied. Our contribution thus investigates the effects of inclusive PCA design on competence, trust, and diversity awareness in a between-subjects experiment with two contrastingly designed prototypes (inclusive and non-inclusive PCA) tested among 106 German university students. As expected by social desirability, the results show that 81.5% of the probands consider an inclusive design important. However, at the same time, the inclusive chatbot is highly significantly rated as less competent. In contrast, we did not measure a significant effect regarding trust, but a highly significant, strongly positive effect on diversity awareness. We interpret these results with the help of the qualitative information provided by the respondents and discuss arising implications for inclusive HCI design.

Keywords

Inclusion, Inclusive Design, Diversity, Pedagogical Conversational Agent, Chatbot.

INTRODUCTION

The heterogeneity of higher education students continues to rise (Lie et al., 2021), e.g., thanks to increasing student mobility, growing internationalization and migration. Growing global inclusion efforts attempt to ensure that everyone benefits from equal opportunities in learning and result in greater diversity in academia (UNESCO, 2015; Ydo, 2020), while also creating awareness for social inclusion (Lie et al., 2021). The socially inclusive design of learning artifacts – by representing characteristics of minorities and fostering social accessibility - becomes therefore a crucial societal task (Guenaga et al., 2012). Some conversational agents already target inclusiveness - for example, chatbots developed to assist refugees for fast integration (Caforio et al., 2021), digital assistants for people with neurodevelopmental disorders (Catania et al., 2019), or pedagogical conversational agents (PCAs) (Hobert & Meyer von Wolff, 2019) that promote the social integration of international students in their host country (Heo & Lee, 2019; Schlimbach, Khosrawi-Rad, et al., 2022). Gupta and Chen (2022) propose features to support disadvantaged learners, as their recent interview study with 215 undergraduates revealed that students see potential in using PCAs to support inclusive learning environments. Lempcke et al. (2020) have proven in an experiment that human-like designed chatbots positively impact perceived inclusiveness and might make *"education more readily accessible, both concerning affordability and geographical coverage"* (p.12).

Despite these findings, socially adaptive PCAs are scarce to find in research and practice (Schlimbach, Rinn, et al., 2022). Besides, research on the socially-inclusive design of learning artifacts to represent diversity in education is underrepresented, as existing studies rather limit their focus to the homogeneous marginalized group they target. However, according to Caforio et al. (2021), PCAs do hold the very potential to contribute to overall social inclusion by representing diversity – a resource that still remains untapped. We anticipate that socially inclusive PCAs potentially create awareness for diversity and might have an impact on other variables such as perceived competence or trust in the digital tutor. With that in mind, we aim to investigate the effect of an inclusive PCA design in a between-subjects experiment on students' perception: First, we measure perceived *trust* towards the PCA because lack of trust has been identified in the literature as a critical factor for the failure of chatbots in practice (Janssen et al., 2021) and because it is important for strong mentor-learner relationships (van Maele et al., 2014). Second, we examine perceived *competence* of the PCA because it correlates with continuous learning success (Sultan & Shafi, 2014). Third, we analyze *diversity awareness* towards the PCA as a measure of consciousness that positively influences learners' acceptance of minorities such as foreign cultures (Al-Obaydi, 2019; Holoien, 2013) essential for inclusiveness in education (Lie et al., 2021). Overall, we address the following research question (RQ):

RQ: To what extent does an inclusive PCA influence user perception of (a) perceived competence, (b) trust, and (c) diversity awareness?

Understanding inclusive PCA design and how it leads to effective and cognitive responses is of high practical relevance - as PCAs could thus potentially contribute as easily scalable mediators to higher educational and social inclusion.

RESEARCH BACKGROUND

Pedagogical Conversational Agents as Social Actors

Conceptual ideas for PCAs in their role as learning companions emerged decades ago and have shifted from automated tutoring assistants to social actors with close social ties to their users (Krämer et al., 2011). Nass & Moon found in several experiments that people use more

social categories such as gender or roles when interacting with IT artifacts that exhibit human-like characteristics or behaviors. According to the theory that computers are social actors (CASA), humans exhibit human-like and social behaviors toward machines despite knowing that they are non-human (Nass et al., 1994; Nass & Moon, 2000), which reinforces anthropomorphic PCA design (e.g., Feine et al., 2019; Seymour et al., 2021). Studies prove that the inclusion of so-called "social cues", i.e., human identity and integrated (non-)verbal human behavioral patterns of PCAs (Seeger et al., 2018) contributes to the promotion of users' social behavior (Feine et al., 2019). Our study also echoes the finding that establishing *common ground* (Clark & Brennan, 1991) leads to positive perceptions of PCAs. A human-like avatar can enable the experience of *social presence* and promote learning, as due to the *persona effect* (Lester et al., 1997), the presence of a human character in an interactive learning environment can have a strong positive effect on the perceived learning experience and success. Schmid et al. (2022) and Araujo (2018) demonstrate a positive correlation between human-like design and perceived agent competence. Schroeder and Schroeder (2018) emphasize that human-like design leads to increased trustworthiness. Moreover, human-like agents favor feelings of enjoyment and likability (Qiu & Benbasat, 2009). At the same time, they also enforce socially desirable responses (Schuetzler et al., 2018).

Inclusive PCA Design

However, the referenced studies do not explicitly consider anthropomorphic design for the learning context, nor do they incorporate diversity of social clues with the goal of inclusion. Using adapted social cues (Seeger et al., 2018), PCAs could be intentionally designed to promote inclusiveness in the artifact (Lembcke et al., 2020), thus influencing human perception in the real world (Feine et al., 2019). For example, the *human identity* of the PCA is embodied by its name (Cowell & Stanney, 2005), gender (Schuetzler et al., 2018), and avatar representation (Gong, 2008). Emotions expressed through emojis (Hu et al., 2018) and visual design (Gupta & Chen, 2022) describe *nonverbal cues*, whereas language style (Gnewuch et al., 2018), self-references (Schuetzler et al., 2018), and a personal introduction (Cafaro et al., 2016) circumscribe the category of *verbal cues*. Each of these elements can potentially pick up underrepresented characteristics to foster inclusiveness (Holoien, 2013; Ydo, 2020).

RESEARCH DESIGN

Research Model and Hypothesis

To investigate the relationship between the inclusive design of a PCA and perceived *competence*, *trust,* and *diversity awareness*, we design two contrasting PCAs with respect to the manipulated variable *"inclusiveness"*. We hypothesize expected effects below.

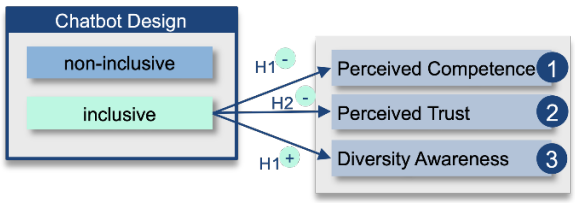

Figure 1. Hypotheses

Perceived Competence

A person is perceived as competent in the execution of an activity by effectively integrating accumulated knowledge, skills and experience (Fiske et al., 2007) for the situation at hand (Schmid et al., 2022). Our experiment measures the perceived domain expertise of the PCAs tutoring learning strategies. Since both PCAs teach identical content and have the same subject matter knowledge and response repertoire, their expertise does not differ. However, according to studies, individuals with disabilities are perceived as less competent (Rohmer & Louvet, 2018). In addition, due to *affinity bias*, people tend to rate people better if they are similar to them (Trainer et al., 2020). Given that our inclusive PCA embodies many underrepresented characteristics we hypothesize in accordance with CASA theory (Nass et al., 1994) that the biased perception will also transfer to PCA interactions.

H1*: The inclusive design of the PCA has a negative impact on perceived competence.*

Perceived Trust

Trust is a basic attitude of a person towards third parties or institutions, which can also deviate situationally. To avoid uncanniness and a negative perception of a PCA and potentially the entire learning experience, trust towards the PCA is particularly relevant (Seymour et al., 2021; Van Maele et al., 2014). Trust is composed of dimensions of credibility and benevolence (Gefen & Straub, 1997; Mcknight et al., 2011). High perceived trustworthiness is a prerequisite to establish social relationships between the PCA and its learner (Nißen et al., 2021). The greater the trust between tutor and student, and thus between PCA and learner, the more successful the educational intervention (Schulte-Pelkum et al., 2014). It has already been shown that anthropomorphic design, AI-supported interaction, and longer-term use, promote trust-building between user and PCA (Nißen et al., 2021; Wald et al., 2021; Zierau et al., 2020). Hereby the quality of its interpretation of requests and advice (Følstad et al., 2018) is particularly crucial. Kunkel et al. (2018) showed that even subtle visual cues showing similarity to the user have a positive effect on user trust towards the information system. Because the inclusive PCA embodies underrepresented features, it exhibits less similarity to the experiments' probands on average than its non-inclusive counterpart.

H2*: The inclusive design of the PCA has a negative effect on perceived trust.*

Diversity Awareness

The third variable to be examined is *diversity awareness*, as the ability to perceive and appreciate heterogeneity (Al-Obaydi, 2019). The diversity concept includes in particular the aspects of age, gender, cultural origin, physical or mental disability, sexual orientation, and worldview (Holoien, 2013). All people should be considered equally and not be disadvantaged because of one of these aspects. Diversity awareness training programs (Colette et al., 2002) and diversity intervention in education (Keime et al., 2002), aim to increase diversity awareness by confronting underrepresented expressions. We expect that inclusively designed social cues are consciously perceived by the probands in the interaction and therefore positively influence diversity awareness.

H3*: The inclusive design of the PCA increases diversity awareness among users.*

Experimental Set-Up

The experiment comprises an in-between-subject design with 106 students of different disciplines of a German Technical University, recruited via the university's online campus and email lists in June 2022. For the experiment, we programmed two prototypical PCAs with help of *Google Dialogflow* in the form of text-based chatbots with an identical functional scope and the same learning content, but manipulated the variable *"inclusive design"*.

Manipulation in the PCA Design

We modeled the conceptualization of the inclusive (iPCA) and non-inclusive chatbot (niPCA) on Heo & Lee's (2019) approach, which decomposes into five phases: (1) concept definition, (2) persona definition, (3) stylistic features, (4) conversation design, and (5) prototyping and testing.

Our PCA teaches learning strategies. We defined the personas (2) as the basis for the PCAs to be created as tutors together with students by designing a persona for a stereotypical professor in Germany and a contrarian persona with underrepresented characteristics and interests. Building upon these personas resulted in the contrasting design (iPCA vs. niPCA) (3) of the two chatbots with the following variations as listed in **Table 1**. With the help of intents, we entered potential questions or answers from the users in *Google Dialogflow* (4), to which the chatbots should respond. Predefined answer buttons make the chatbot less error-prone and also control the interaction flow generated in the experiment for better comparability of the tested prototypes.

	Social Cues	iPCA	niPCA
Human Identity	Name	Aylin A.	Dr. F. Zweistein
	Gender	Female	Male
	Ethnicity	Afro-American	White
	Avatar Embodiment	Physical disability (wheelchair)	No discernible disability
Non-Verbal	Emotions	Emojis as visual support against language barrier	No use of Emojis
	Visual Design	Enlarged font, color contrasts	Standard font, black and white
Verbal Cues	Language Style	Simplified language with short main sentences	Technical language
	Inclusive Language	Gender-inclusive communication	Use of the generic masculine
	Self-Reference in Introduction	Aylin, your tutor with expertise in the subject area of learning strategies	Franz, your tutor with expertise in the subject area of learning strategies

Table 1. Overview of Manipulated Cues

In the prototyping and testing phase (5), we had a student seminar group test the PCAs in order to detect and solve technical errors and incomprehensibilities.

Conducting the Experiment

First, we randomly assigned the students (according to their birth month) to one of the two chatbots and gave interaction instructions. We then asked the experiment takers to complete a questionnaire created with *LimeSurvey.* To ensure direct comparability, we prepared identical surveys in German language for both chatbots. In addition to socio-demographic data (age, gender, education) as well as a manipulation and scenario check, we queried the subjects' perceptions of the three constructs *competence*, *trust*, and *diversity awareness* on a 7-point Likert scale. Complementarily, in the qualitative part of the survey we also embedded open questions regarding students' attitudes towards inclusion, queried their impression on the overall PCA design and asked for a self-assessment of minority characteristics they identify with as well as aspects that stood out positively or negatively in the interaction with the chatbot. For the statistical analysis we used the statistical software Jamovi version 2.2.5.0, while we manually clustered and analyzed the free text data in a peer-reviewed process.

RESULTS

Of the 106 probands, 54 participants tested the iPCA and the remaining 52 tested the niPCA. The manipulation check was successful, as on average the iPCA was rated as strongly inclusive (6.1 of 7) and the niPCA as rather non-inclusive (3.2 of 7). All subjects were able to place themselves in the scenario of interacting with a PCA tutoring different learning types and methods in order to acquire new knowledge. Because of the random assignment to one of the two groups, the socio-demographic characteristics of both groups (iPCA/niPCA) were almost congruent, as expected. While 17% of subjects in both groups had already tried a chatbot in a learning context, the remaining 83% were interacting with a PCA for the first time. 35.2 %/26.9 % of the students were studying for a master's degree, 59.2 %/61.5 % studied towards a bachelor's degree, and 5.6 %/11.5 % were currently in transition to become freshmen, with 85.2 %/84.6 % of participants aging 20-29. Among the participants 40.7 %/48.1 % were male, 57,5 %/50%

female, and 1.8%/1.9% diverse. The self-attributed diversity characteristics provided in the free text fields are also very comparable in both groups, including first generation university students (29.6 %/30.8 %), immigration (13 %/15.4 %), sexual orientation (7.4 %/5.8 %), (light) physical limitations (3.7 %/ 3.8 %), giftedness with an IQ > 129 (1.9 %/ 3.8 %), and isolated additional mentions such as religious affiliation, obesity, or left-handedness. Due to this strong congruence of both groups, the risk of a diversity bias between the groups is kept low.

With 81.5% in total, the majority of respondents demand for inclusive PCA design in order to represent diversity and respect everyone. All implemented inclusive social cues (cf. p.3) were mentioned and positively highlighted by almost all participants in the iPCA free text fields with the exception of the increased background color contrast for better readability, which was not proactively noticed. For the niPCA, the students did not identify any inclusive features, except that Franz seemes accessible to all people and addresses different types of learners in terms of content provision. Several mentions even criticized the non-inclusive design of Dr. Franz Zweistein by making suggestions for a more inclusive design such as, "*Franz is the absolute stereotype of a German man. He could use gender-inclusive language, he could support people with impairments like blind people by additional functions, but he only acts as a conservative tutor and nothing else."*

In the statistical analysis, we aimed for quantitatively measuring the impact of the inclusive design on user perceptions. For this purpose, we included only the 97 questionnaires that had fully completed the quantitative section. First, we conducted an exploratory factor analysis to examine the factor loadings for each construct and its attributed items. **Table 2** provides an overview of the items with associated (α) and AVE, as well as factor loadings.

Construct	Items	Loadings
Competence **(α) = 0.859 AVE = 0.632**	The chatbot has a distinct knowledge.	0.907
	The chatbot is extremely well qualified.	0.914
	I feel that the chatbot could answer any question for me.	0.610
	The chatbot was able to give me professional advice.	0.534
Trust **(α) = 0.825 AVE = 0.572**	I am convinced that the chatbot has good intentions.	0.753
	The Chatbot seems very trustworthy.	0.716
	I had the feeling that the chatbot is honest.	0.891
	The trust in the accuracy of the information the chatbot gave me.	0.545
Diversity awareness **(α) = 0.859 AVE = 0.558**	~~The chatbot shows sincere interest towards me.~~	~~0.535~~
	The chatbot acts without prejudice	0.504
	The chatbot is concerned about my well-being.	0.806
	The chatbot is considerate of all users.	0.806
	I can identify with the chatbot.	0.642
	I feel that the chatbot is inclusive.	0.828

Table 2. Constructs with Measurements

We considered only items with factor loadings above the threshold of .50 (Gefen & Straub, 2005). Here, the constructs we targeted, *competence, trust,* and *diversity awareness*, were confirmed. We then examined the internal validity of each construct for Crohnbach's alpha (α). We had to remove one item of the diversity construct to meet the prerequisite according to Urbach and Ahlemann (2010) that (α) must have a minimum value of .80 with an average variance extracted (AVE) of at least .50. Since the item reliability statistic showed no further improvement for (α) or McDonald`s ω
if additional items were dropped, we retained the remaining items. We then calculated weighted sum scores to create a metric variable for each of the three constructs (DiStefano et al., 2009).

To perform a two-tailed t-test, we tested the constructs for normal distribution according to Shapiro- Wilk as well as homogeneity of variances according to Levene and, due to the violation of this precondition (variances were unequal), we decided to perform the Welch's - t- test. **Table 3** shows the calculated values.

		t-value	df	p		Effect Size
Competence	Welch's t	-2.567	91.6	0.012*	Cohen's d	-0.523
Trust	Welch's t	0.555	94.9	0.580	Cohen's d	0.113
Diversity awareness	Welch's t	6.255	89.4	< .001***	Cohen's d	1.275

Table 3. Welch's t-Test

The Welch's t-value is -2.567 for the construct competence, indicating a significant difference (p = 0.012) between the two groups iPCA and niPCA. Cohen's d with the effect size of - 0.523 indicates that the inclusive design of the PCA has a medium negative effect with respect to the variable competence. Thus, the null hypothesis that there is no difference between the means of both groups regarding perceived competence must be rejected. For the construct trust, the measured difference (p = 0.580) is not significant, so the null hypothesis that there is no difference between the means of both groups cannot be rejected at a 95 % confidence level. On the other hand, the test demonstrates at a probability level of p <0.001 and a t-value of 6.255 that the mean value of diversity awareness from the iPCA is highly significant above the corresponding one of the niPCA. Here, the inclusive design exhibits a high effect size (Cohen's d = 1.275). A look at the descriptive statistics of the median and arithmetic mean illustrates the above-mentioned effects also expressed in absolute values as shown in **Table 4**.

	Group	N	Mean	Median	SD	SE
Competence	iPCA	50	4.76	5.00	1.093	0.155
	niPCA	47	5.37	5.50	1.246	0.182
Trust	iPCA	50	5.80	6.00	0.835	0.118
	niPCA	47	5.71	5.75	0.808	0.118
Diversity awareness	iPCA	50	5.38	5.60	1.036	0.147
	niPCA	47	3.91	3.80	1.253	0.183

Table 4. Descriptive Statistics

In summary, H1 and H3 are confirmed, while H2 is not.

Qualitatively, the experiment participants stated that the iPCA seemed competent due to its fast, understandable communication and its extensive knowledge, but criticized its eloquence (compared to a human). For the niPCA, they associate professionalism with the doctorate. Both groups indicated that trust in the respective PCA was built by its human design, its self-reference, and the personalized assistance offered for learning. Some participants in the niPCA group also cited that they perceived the niPCA's

expertise to be higher than their own expertise, which, generated trust. Most iPCA users reported being consciously aware of the diversity characteristics of physical disability, an Afro-American female tutor, her foreign name, and gender-inclusive language during the interaction, while niPCA's users reported not thinking about diversity during the interaction and did not provide any diversity-related free-text responses.

DISCUSSION

In the present between-subject user experiment with 106 students from a German university, we demonstrated that the inclusive design of a PCA had a significant ($p = 0.012$) negative effect of medium effect size (Cohen's $d = - 0.523$) on perceived competence and a highly significant ($p < 0.001$) positive effect on subjects' diversity awareness with a high effect size (Cohen's $d = 1.275$). This result is surprising in that 81.5% of both groups were in favor of the inclusive design of PCAs to represent diversity and inclusion of minorities, but at the same time the inclusive design had a negative effect on competence, even though both PCAs had the identical professional competence in terms of instructed knowledge and offered features. Thus, our experiment reveals a discrepancy between the social desirability of inclusion and unconscious social bias towards inclusive design in Human-Computer-Interaction.

In the physical world, it has already been demonstrated that individuals with a physical or mental disability are more likely to be perceived as incompetent (Fiske et al., 2007; Rohmer & Louvet, 2018) even in contexts for which the respective limitation is irrelevant. Other studies show that perceptions of competence are sometimes subject to further biases such as gender bias (Oh et al., 2019) or affinity bias (Trainer et al., 2020). Our study provides initial evidence that, congruent with CASA and social response theory (Nass et al., 1994; Nass & Moon, 2000), these real-world biases can also be applied to virtual information systems such as PCAs. If this relationship is solidified by further studies, it will have far- reaching implications for the design of PCAs, other information systems and the metaverse on the long-term: On the one hand, it increases the responsibility to think about inclusive design and its implications right from the beginning as a core component of an ethical code. On the other hand, it also creates new opportunities. While positive discrimination in the real world always involves situationally disadvantaging real, non-minority people, PCAs offer the potential to exploit only its positive aspects in the virtual world. Visual cues alone can have significant effects on user perceptions (Kunkel et al., 2018), so simply representing diversity characteristics in avatar design could go some way toward conveying underrepresented role models via PCAs and thereby increase diversity awareness. In the practical implementation of PCAs, however, this potential is still hardly realized (Schlimbach, Rinn, et al., 2022).

While in reality, despite quotas and diversity promotion programs, breaking down disadvantaging structures is only gradually succeeding (WEF, 2022), PCAs, which are still in their infancy, offer precisely the opportunity to take these aspects into account from the outset. They could assist in situating knowledge to address discrimination, facilitate dialogic encounters between minorities, and contribute to the integration of experience and reflection on diversity (Lie et al., 2021) as a future mission of HCI design. However, initial studies show that even in chatbots stereotypical (gender) bias is reproduced (Feine et al., 2020) and the potentials of flexible adaptation to underrepresented characteristics (especially beyond cognition) are hardly considered (Schlimbach, Rinn, et al., 2022).

However, accelerated by the trend of increased digital teaching, discriminatory structures could be broken up much more quickly, adapted more flexibly, and implemented on a far-reaching scale; because in the digital space, inclusion can be better implemented and made tangible (Guenaga et al., 2012). Already our short-term experiment showed strong effects on diversity awareness thanks to inclusive design. Potentially, with a long-term effect still to be investigated. PCAs could thus also support inclusion in society as a whole, since diversity-sensitive learning is crucial for anchoring mutual respect and tolerance (Lie et al., 2021). To this end, it is necessary to mitigate the unconscious (social) bias of future users (Muchnik et al., 2013) in addition to the algorithmic bias (Yapo & Weiss, 2018), which has already been frequently addressed in HCI design. However, this requirement is particularly challenging because underrepresented characteristics, as well as perceptions of diversity, are themselves context-specific and diverge widely between cultures, application fields, and target groups (Koopmans & Schaeffer, 2017). Accordingly, the number and complexity of needed studies results much higher. For example, it appears to be interesting to repeat our experiment in a different cultural setting or with a more heterogeneous target group than similarly aged students from the same German university. In this respect, our results so far are severely limited.

Our experiment did not show a significant correlation between inclusive PCA design and user- generated trust. Since a longer-term interaction with the PCA is required to build trust (Nißen et al., 2021), we admit here a limitation of our study. Presumably, the short interaction time was not sufficient to measure a significant effect. The scientific long-term monitoring of a learning relationship with a PCA might decipher the emergence of trust and could thereby also analyze what effects trust has on long-term learning success with the PCA. However, such long-term studies with PCAs are overall highly underrepresented (Khosrawi-Rad et al., 2022; Schlimbach, Rinn, et al., 2022).

Furthermore, we deliberately chose to make inclusion holistic in our experiment by manipulating different social cues accordingly. This has also been shown in other studies, e.g., to measure social presence or anthropomorphism in experiments with chatbots (e.g.,

Araujo, 2018). Thus, we were able to measure (highly) significant effects due to inclusive design. At the same time, however, this approach limits us in that we could not analyze the effect sizes and intercorrelations of individual manipulated cues. We therefore propose to limit the manipulation to only one set of social cues at a time (human identity/(non)-verbal) in a second step and then to investigate the manipulative effect of isolated social cues (e.g., name or depicted physical disability on the avatar) in a further third step. This multi-step approach could help to better understand combined effects and define the threshold at which an inclusive design highlights significantly measurable effects on user responses. For practical implementation, sound design guidelines could then be derived to support designers and programmers in inclusive design, especially directives for PCA adaptation. Features might be designed adaptive or predetermined to specifically counteract unconscious social bias or offer choices for users to consciously customize them to better identify with their PCA. Extensive further research is needed in this regard to uncover culture- and context-specific differences and to measure the manipulation effects in a long-term study. Thereby, changing attitudes toward diversity over time shall be measured as well and might lead to new findings. Potentially, inclusive PCAs could thus not only support the needs of the individual learner, but even make a contribution to more (educational) inclusion for society as a whole.

ACKNOWLEDGMENTS

We would like to acknowledge *Henry Akmann* for programming the prototypes as well as the TUBS master students *Mareike Geers*, *Lena Kook, Lucas Lampka* & *Anika Stoppel* for their valuable input and continuous support in conducting this study.

REFERENCES

1. Al-Obaydi, L. (2019). Cultural Diversity, Awareness and Teaching: A Study in an EFL Context. *Journal of Asia TEFL*, *16*, 987–995.
2. Araujo, T. (2018). Living up to the chatbot hype: The influence of anthropomorphic design cues and communicative agency framing on conversational agent and company perceptions. *Computers in Human Behavior*, *85*, 183–189.
3. Cafaro, A., Vilhjálmsson, H. H., & Bickmore, T. (2016). First impressions in human-agent virtual encounters. *ACM Transactions on Computer-Human Interaction*, *23*(4), 1–40.
4. Caforio, A., Pollini, A., Filograna, A. S., & Passani, A. (2021). Design issues in Human-centered AI for Marginalized People. *ITAIS 2021 Proceedings*.
5. Catania, F., Nardo, N. D., Garzotto, F., & Occhiuto, D. (2019). Emoty: An Emotionally Sensitive Conversational Agent for People with Neurodevelopmental Disorders. *HICSS-52 Proceedings*.
6. Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In *Perspectives on socially shared cognition*, pp. 127–149.
7. Colette, C., van Vuuren, L., & Delene, V. (2002). The effectiveness of a diversity awareness training programme. *South African Journal of Industrial Psychology*, *28*.
8. Cowell, A. J., & Stanney, K. M. (2005). Manipulation of non-verbal interaction style and demographic embodiment to increase anthropomorphic computer character credibility. *International Journal of Human-Computer Studies*, *62*(2), 281–306.
9. DiStefano, C., Zhu, M., & Mîndrilã, D. (2009). Understanding and Using Factor Scores: Considerations for the Applied Researcher. *Practical Assessment, Research and Evaluation*, *14*(20).
10. Feine, J., Gnewuch, U., Morana, S., & Maedche, A. (2019). A Taxonomy of Social Cues for Conversational Agents. *International Journal of HC Studies*, *132*, 138–161.
11. Feine, J., Gnewuch, U., Morana, S., & Maedche, A. (2020). *Gender Bias in Chatbot Design*, pp. 79–93.
12. Fiske, S. T., Cuddy, A. J. C., & Glick, P. (2007). Universal dimensions of social cognition: Warmth and competence. *Trends in Cognitive Sciences*, *11*(2), 77–83.
13. Følstad, A., Nordheim, C., & Bjørkli, C. (2018). *What Makes Users Trust a Chatbot for Customer Service? An Exploratory Interview Study*, *INSCI 2018 Proceedings*, 194–208.
14. Gefen, D., & Straub, D. W. (1997). Gender Differences in the Perception and Use of E-Mail: An Extension to the Technology Acceptance Model. *MIS Quarterly*, *21*(4), 389.
15. Gefen, D., & Straub, D. W. (2005). A Practical Guide to Factorial Validity Using PLS-Graph: Tutorial and Annotated Example. *CAIS*, *16*(5).
16. Gnewuch, U., Morana, S., Adam, M., & Maedche, A. (2018, June 1). *Faster Is Not Always Better: Understanding the Effect of Dynamic Response Delays in HCI.*
17. Gong, L. (2008). How social is social responses to computers? The function of the degree of anthropomorphism in computer representations. *Computers in Human Behavior*, *24*(4), 1494–1509.
18. Guenaga, M., Menchaca, I., Romero Yesa, S., & Eguíluz, A. (2012). A Tool to Evaluate the Level of Inclusion of Digital Learning Objects. *Procedia Computer Science*, *14*, 148–154.
19. Gupta, S., & Chen, Y. (2022). Supporting Inclusive Learning Using Chatbots? A Chatbot-Led Interview Study. *Journal of Information Systems Education*, *33*(1), 98–108.
20. Heo, J., & Lee, J. (2019). CiSA: An Inclusive Chatbot Service for International Students and Academics. In C. Stephanidis: *HCI International 2019 – Late Breaking Papers*, 153–167.
21. Hobert, S., & Meyer von Wolff, R. (2019). Say Hello to Your New Automated Tutor – A Structured Literature Review on Pedagogical Conversational Agents. *Wirtschaftsinformatik 2019 Proceedings*.
22. Holoien, D. S. (2013). Do Differences Make a Difference? The Effects of Diversity on Learning, Intergroup Outcomes, and Civic Engagement. *Princetion University Reports*, 18.
23. Hu, T., Xu, A., Liu, Z., You, Q., Guo, Y., Sinha, V., Luo, J., & Akkiraju, R. (2018). *Touch Your Heart: A Tone-aware Chatbot for Customer Care on Social Media*. 1–12.
24. Janssen, A., Grützner, L., & Breitner, M. H. (2021). *Why do Chatbots fail? A Critical Success Factors Analysis*, 18.
25. Keime, S., Landes, M., Rickertsen, G., & Wescott, N. (2002). *Promoting Cultural Awareness and the Acceptance of Diversity through the Implementation of Cross-Cultural Activities*.
26. Khosrawi-Rad, B., Rinn, H., Schlimbach, R., Gebbing, P., Yang, X., Lattemann, C., Markgraf, D., & Robra-Bissantz, S. (2022, June 18). Conversational Agents in Education – A Systematic Literature Review. *ECIS 2022 Proceedings*.
27. Koopmans, R., & Schaeffer, M. (2017). *Statistical and Perceived Diversity and Their Impacts on Neighborhood Social Cohesion in Germany, France and the Netherlands*.

28. Krämer, N., Eimler, S., Rosenthal-von der Pütten, A. M., & Payr, S. (2011). Theory of Companions: What Can Theoretical Models Contribute to Applications and Understanding of Human-Robot Interaction? *Applied Artificial Intelligence*, *25*, 474–502.
29. Kunkel, J., Donkers, T., Barbu, C.-M., & Ziegler, J. (2018). Trust-Related Effects of Expertise and Similarity Cues in Human-Generated Recommendations. *Proceedings of the Conference on Theory-Informed User Modeling for Tailoring and Personalizing Interfaces*. HUMANIZE18, Tokyo, Japan.
30. Lembcke, T.-B., Diederich, S., & Brendel, A. B. (2020). Supporting Design Thinking Through Creative and Inclusive Education Facilitation: The Case of Anthropomorphic Conversational Agents for Persona Building. ECIS 2020.
31. Lester, J. C., Converse, S. A., Kahler, S. E., Barlow, S. T., Stone, B. A., & Bhogal, R. S. (1997). The persona effect: Affective impact of animated pedagogical agents. *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*, 359–366.
32. Lie, R., Boogaard, B., & Witteveen, L. (2021). *Towards design principles for diversity sensitive learning*, 134–146.
33. Mcknight, D. H., Carter, M., Thatcher, J. B., & Clay, P. F. (2011). Trust in a specific technology: An investigation of its components and measures. *ACM Transactions on Management Information Systems*, *2*(2), 12:1-12:25.
34. Muchnik, L., Aral, S., & Taylor, S. (2013). Social Influence Bias: A Randomized Experiment. *Science (New York, N.Y.)*, *341*, 647–651.
35. Nass, C., & Moon, Y. (2000). Machines and Mindlessness: Social Responses to Computers. *Journal of Social Issues*, *56*(1), 81–103.
36. Nass, C., Steuer, J., & Siminoff, E. (1994). *Computer are social actors*. 204. https://doi.org/10.1145/259963.260288
37. Nißen, M., Selimi, D., Janssen, A., Cardona, D., Breitner, M., Kowatsch, T., & Wangenheim, F. (2021). See you soon again, chatbot? A design taxonomy to characterize user-chatbot relationships with different time horizons. *Computers in Human Behavior*, *127*(2).
38. Oh, D., Buck, E. A., & Todorov, A. (2019). Revealing Hidden Gender Biases in Competence Impressions of Faces. *Psychological Science*, *30*(1), 65–79.
39. Qiu, L., & Benbasat, I. (2009). Evaluating anthropomorphic product recommendation agents: A social relationship perspective to designing information systems. *Journal of Management Information Systems*, *25*(4), 145–182.
40. Rohmer, O., & Louvet, E. (2018). Implicit stereotyping against people with disability. *Group Processes & Intergroup Relations*, *21*(1), 127–140.
41. Schlimbach, R., Khosrawi-Rad, B., & Robra-Bissantz, S. (2022). *Deriving Design Knowledge for eLearning Companions to Support International Students*. Bled 2022.
42. Schlimbach, R., Rinn, H., Markgraf, D., & Robra-Bissantz, S. (2022). A Literature Review on Pedagogical Conversational Agent Adaptation. *PACIS 2022 Proceedings*.
43. Schmid, D., Staehelin, D., Bucher, A., Dolata, M., & Schwabe, G. (2022). Does Social Presence Increase Perceived Competence? Evaluating Conversational Agents in Advice Giving through a Video-Based Survey. *26*, *ACM Human Computer Interaction* (6).
44. Schroeder, J., & Schroeder, M. (2018). Trusting in Machines: How Mode of Interaction Affects Willingness to Share Personal Information with Machines. HICSS *Proceedings.*
45. Schuetzler, R. M., Giboney, J. S., Grimes, G. M., & Nunamaker, J. F. (2018). The influence of conversational agent embodiment and conversational relevance on socially desirable responding. *Decision Support Systems*, *114*, 94–102.
46. Schulte-Pelkum, J., Schweer, M. K. W., & Pollak, B. (2014). Dyadic trust relations between teachers and students. An empirical study about conditions and effects of perceived trustworthiness in the classroom from a differential perspective, 70–86.
47. Seeger, A.-M., Pfeiffer, J., & Heinzl, A. (2018). Designing Anthropomorphic Conversational Agents: Development and Empirical Evaluation of a Design Framework. *ICIS 2018.*
48. Seymour, M., Yuan, L., Dennis, A., & Riemer, K. (2021). Have We Crossed the Uncanny Valley? Understanding Affinity, Trustworthiness, and Preference for Realistic Digital Humans in Immersive Environments. *Journal of the Association for Information Systems*, *22*(3).
49. Sultan, S., & Shafi, M. (2014). Impact of Perceived Teachers' Competence on Students' Performance: Evidence for Mediating/ Moderating Role of Class Environment. *Journal on Educational Psychology*, *8*(1), 10–18.
50. Trainer, T., Taylor, J., & Stanton, C. (2020, November 18). *Choosing the Best Robot for the Job: Affinity Bias in Human-Robot Interaction*.
51. UNESCO. (2015). *Education 2030: Incheon Declaration and Framework for Action for the implementation of Sustainable Development*, https://unesdoc.unesco.org/ark:/48223/pf0000245656.
52. Urbach, N., & Ahlemann, F. (2010). Structural equation modeling in information systems research using Partial Least Squares. *Journal of Information Technology Theory and Application*, *11*.
53. Van Maele, D., Van Houtte, M., & Forsyth, P. (2014). Introduction: Trust as a Matter of Equity and Excellence in Education. In *Trust and School Life: The Role of Trust for Learning, Teaching, Leading, and Bridging*.
54. Wald, R., Heijselaar, E., & Bosse, T. (2021). Make your own: The Potential of Chatbot Customization for the Development of User Trust. *Adjunct Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization*, 382–387.
55. WEF. (2022). *Global Gender Gap Report 2022*. World Economic Forum. https://www.weforum.org/reports/global-gender-gap-report-2022/
56. Yapo, A., & Weiss, J. (2018). *Ethical Implications of Bias in Machine Learning*.https://doi.org/10.24251/HICSS.2018.668
57. Ydo, Y. (2020). Inclusive education: Global priority, collective responsibility. *PROSPECTS*, *49*(3), 97–101.
58. Zierau, N., Engel, C., Söllner, M., & Leimeister, J. M. (2020). Trust in Smart Personal Assistants: A SLR and Development of a Research Agenda. Wirtschaftsinformatik *Proceedings.*