*Exploring AI in Education: A Quantitative Study of a Service-Oriented University Chatbot*

Mascha-Lea Fersch, University of Applied Sciences Ansbach, Germany
Sigurd Schacht, University of Applied Sciences Ansbach, Germany
Betiel Woldai, University of Applied Sciences Ansbach, Germany

The Paris Conference on Education 2023
Official Conference Proceedings

**Abstract**
The following paper presents the evaluation of an artificially intelligent assistant system (DIAS) with a service-oriented chatbot as a central communication element. The conversational AI (Artificial Intelligence) is supposed to increase information transparency in higher education environments and thus support students, teachers, and administrative staff. The exploratory study had two objectives: first, we intended to find out about the usability and utility of the DIAS chatbot using the CUQ (Chatbot Usability Questionnaire) score and benchmark the results against other conversational agents. Secondly, we were interested in possible effects among the different variables of interest, which could contribute to further theory development of chatbots in education. The results show that the DIAS chatbot scored above average, and can support students in finding relevant information, particularly if they use the assistant frequently. Positive aspects included the intuitive use, a welcoming persona (expressed in design & language) and easy navigation. The negative feedback showed potential for improvement particularly in content quality and handling dialogue mistakes, which is a general shortcoming of conversational AI at this development stage. The results can be used as a guidance for future research and theory building. However, they must be considered carefully due to several study limitations.

Keywords: Digital Learning, Service-Oriented Chatbot, Study Assistant, Educational Chatbot, AI

iafor

The International Academic Forum
www.iafor.org

## Introduction

Artificial intelligence (AI) can help to address various challenges in education such as an increasingly unstructured supply of information and educational resources (Schurz et al., 2021), and complement traditional learning formats to promote important future skills like media and information competencies (Mebis, 2018). The DIAS project at Ansbach University of Applied Sciences created a digital assistant, who steers information in a targeted manner among other components. In addition to an AI-based chatbot as the central communication element, the virtual guidance system also includes planning, study analysis, and motivation applications. DIAS offers students added value through a robust information platform and the opportunity to effectively plan and pursue their studies, while simultaneously creating more resources for individual support by relieving the burden of email traffic for the administration. While the first stage of the project evaluation was dedicated to finding out why and how digital assistants can be successful in educational settings (Fersch et al., 2022), the main objective of this study was to explore user experiences with the DIAS chatbot among a larger sample group (n=103) and benchmark the results against other conversational agents. Specifically, we wanted to learn about students' usability issues when using the DIAS chatbot to optimize the current chatbot version continually. In addition to usability aspects, we were interested in the perceived utility of the chatbot for students based on the project's objective to give students better control over their study situation and reduce the frustration and time spent gathering information. Furthermore, we intend to explore and structure the results to detect possible effects, which might contribute to the theoretical development of the research field. In the following we will present a theoretical background to chatbots in education, before explaining the DIAS chatbot component in detail. Consequently, we will highlight the methodology of this study, present and discuss the results and finally explain limitations and potential for future research.

## Theoretical Background

### Definitions and Classifications

Chatbots are "digital systems that can be interacted with entirely through natural language via text or voice interfaces. They are intended to automate conversations by simulating a human conversation partner and can be integrated into software such as online platforms, digital assistants, or be interfaced through messaging services" (Wollny et al., 2021, p.2). Conversational AI in the educational sector can provide interactive learning, ranging from knowledge tests to encouragement, learning advice, and reminder functions. A particular benefit of digital assistants in educational environments is the personalized learning experience, as they can individually adapt to the student's way and speed of learning (Clarizia et al., 2021). As a 24-hour support service, chatbots can facilitate academic information flow and meet the students' needs anytime (Alexander et al., 2019). DIAS can be classified as a "service-oriented" (Pérez et al., 2020) chatbot and according to the framework of Wollny et al. (2021) assumes an assisting pedagogical role, intending to increase the efficiency of education through answering FAQs, hence making information fast and easily available for students and teachers. Unlike teaching-oriented chatbots, service chatbots do not impart subject-specific knowledge, e.g. on language learning (Pérez et al., 2020).

**Evaluation of Educational Chatbots**

Literature suggests different quality criteria for educational assistants such as humanity, affect and accessibility (Radziwill & Benton, 2017; Smutny & Schreiberova, 2020), as well as learning success, acceptance, motivation, usability, technical correctness, psychological factors and further beneficial factors (Hobert & Meyer von Wolff, 2019) among other factors. Pérez-Marín (2021) defined a classification framework containing pedagogical and social aspects for chatbot design. Shawar and Atwell (2007) suggest choosing the evaluation criteria and methodology based on the user needs and the specific goal of the application, not necessarily on established standards.

The successful use of conversational AI in an educational context has been empirically proven in prior research (Pérez et al., 2020). Most publications concern teaching-oriented chatbots, of which we will present a selection in the following. In an evaluation of 47 educational chatbots, conversations with each were conducted and reviewed based on the aspects of teaching, humanity, affect and accessibility. The authors identified ten chatbots, which performed best in the beforementioned criteria. However, they also mentioned that chatbot development is still in the initial stage and cannot fully meet expectations (Smutny & Schreiberova, 2020). Verleger and Pembridge (2018) evaluated a chatbot interface for a computer programming course by reviewing and categorizing answers in the system's database, followed by a qualitative focus group study. Their results indicate the need for better development of the chatbot database before launch, as many students ceased using it due to the low response rate. Using conversational AI in language teaching can increase course performance compared to traditional teaching models. Among other aspects, study participants praised the intelligent tutor's easy usability and flexibility. (Vázquez-Cano et al., 2021). Also in mentoring processes, intelligent, digital assistants can be applied successfully. Thus, students who received feedback and guidance in reading and writing tasks during their self-study time reported a positive perception of the chatbot assistance (Neumann et al., 2021). Benotti et al. (2017) evaluated their chatbot programming platform with two observational studies based on the student's engagement such as participation, task completion and self-reported interest. Diachenko et al. (2019) report positive feedback from students and teachers who participated in an automated university course including a teaching-oriented chatbot.

Fewer empirical studies have been conducted on service-oriented chatbots in the educational field. Dibitonto et al. (2018) researched requirements for a possible future university chatbot through a low-level dialog system on the Facebook Messenger platform, LiSA, which served as a survey tool and a presentation/foretaste of the potential future university chatbot. Questions about the student's expectations and satisfaction with LiSA were also evaluated. The majority of users were satisfied or indifferent with the chatbot itself and found the conversation experience (of the survey) to be interesting and enjoyable. Students mostly wished for information, especially notifications on lessons, events, and other campus activities. The authors also highlighted the importance of a chatbot personality to react to potential loops and dialogue mistakes, which might trigger rude answers from users. Another study found that administrative effort can be reduced, and matriculation is more likely to be finished on time when using an AI-based assistant with personalized messaging (Page & Gehlbach, 2017). In an approach to evaluate a candidate service chatbot, Santoso et al. (2018) tested the system by measuring the number of correct answers to requests by the researchers. Galko et al. (2018) tested standard UX metrics such as error rate and post-session testing with a relatively small sample group of five people for a service chatbot to facilitate the university

application process. They compared the results to the traditional application form and reported an improvement in user experience. The chatbot Jill Watson was developed to answer common FAQs in computer science courses. During the two-year usage of the AI assistant the research team mainly reported positive feedback on the chatbot. However, they also expressed ethical issues of the conversational agent especially when addressing underrepresented females in the male-dominated course (Eicher et al., 2018). A focus group on the DIAS chatbot in April 2022 showed that students particularly appreciate a vivid persona, appealing design, accurate, guided, direct answering, and optional push-messaging (Fersch et al., 2022). Educational chatbots are currently in a stage in which research is growing and expectations of the technology are high. However, there is still little practical experience (Smutny & Schreiberova, 2020; Wollny et al., 2021). While most of the research on educational chatbots deals with design questions concerning e.g. personality and pipeline, fewer studies have been published on the actual utility of chatbots in education (Wollny et al., 2021). Furthermore, most studies evaluating educational chatbots were conducted with an insignificant sampling population (Okonkwo & Ade-Ibijola, 2021).

**DIAS Chatbot Development**

A conversational AI represents the communication/information component of the DIAS system. In the first development phase, teaching and administrative staff collected quality-assured standard answers, which were integrated into the conversational AI in a rule-based approach. In the second phase, the chatbot will be trained for AI-based answer generation to deal with individual (non-standard) questions based on a self-uploading knowledge base. The chatbot's design, character traits, and conversational tone were based on a pre-developed persona, which aims to create a more personalized conversation experience (Braun & Alt, 2020). As an additional feature the DIAS chatbot also offers scientifically based learning advice in the following categories: concentration, productivity, and memorization e.g. "When compiling your notes, you should place the most important learning content at the beginning and end. The information positioned there can (usually) be better retained by your memory." The chatbot can currently be accessed at all times on the university's website, and will be integrated into messaging apps for more convenient use.
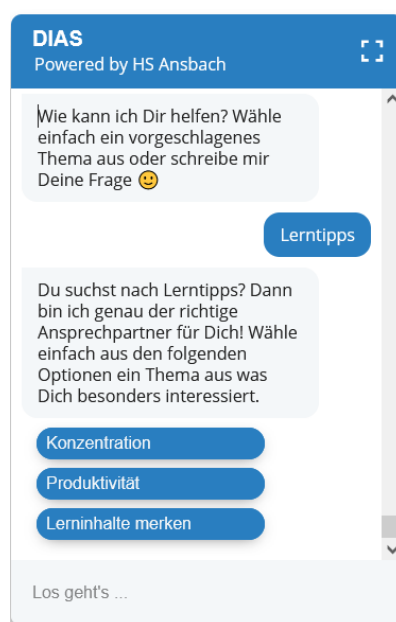


**Figure 1:** DIAS chatbot conversation

**Methodology**

**Research Design**

We used a questionnaire consisting of questionnaire instructions, content questions, open questions, and statistical information (see appendix table 1). For the content questions we decided to use an established psychometric scale, to ensure higher reliability and validity than single items and to make the study results comparable with similar studies. In the educational sector we could identify only one similar, quantitative study from Neumann et al. (2021), which used the System Usability Scale (SUS) for evaluation. However, as Larbi (2022, p.3) states – the SUS is "not recommended for usability testing of conversation-driven systems since they exploit other design principles". Two other studies from non-educational backgrounds were found to have developed and validated scales to assess the quality of conversational agents (Borsci et al., 2021; Holmes et al. 2019). For this study we decided to use the Chatbot Usability Questionnaire from Holmes et al. (2019), since it has been previously applied in other studies (Larbi et al., 2021; Larbi et al., 2022), thus providing a benchmark for comparison of results. The translation of the English scale was done with the forward and backward translation technique (Toma et al., 2017). Table 1 in the appendix lists the German translation of the Chatbot usability questionnaire. The questionnaire consists of 16 items with eight positive and eight negative statements and is rated on a 5-point Likert scale (from strongly disagree to strongly agree). The CUQ can be compared to the SUS as scores are likewise calculated out of 160 and normalized to 100, with an average benchmark score of 68 (Holmes et al., 2019; Larbi et al., 2022). Apart from the CUQ questionnaire we added two questions on how DIAS facilitates academic information flow (easier and faster access to information) and questions on age, semester, occupation, frequency of use and interaction duration.

**Participants & Procedure**

Participants were recruited directly through the chatbot on the university's homepage, as well as through social media announcements and printed handouts to students. The survey link was made accessible after users had interacted with the chatbot. As an incentive, participants could register for a lottery with several prizes, raffled among all subscribers. The survey was available for five weeks in October and November 2022.

**Data Analysis**

The exploratory data analysis was executed with the software SPSS. CUQ scores were calculated using the CUQ Excel Sheet provided by Ulster University (Holmes & Bond). For the thematic analysis of the open-ended answers the software F4Analyse was used. The data were coded, and grouped into categories, before being reviewed and adapted by other team members to ensure objectivity. Outliers, redundancies, and inconsistencies in the data set were analyzed using several procedures. Missing values in the data set were marked as missing.

**Results**

In total 113 completed and uncompleted answers were returned, of which ten were empty and therefore had to be deleted. Valid data was finally available for a subset ranging from 103 to 93 participants across the variables of interest.

Of the 85 respondents who indicated an occupation, 90,6% were in their bachelor studies, one person was an employee of the university and 8,2% were in their master studies. Most participants (45,2%) were in their first term at university and between 18 and 24 years (83%).

Most participants (70,9%) used the chatbot less than once a month. However, at the time of the survey the semester had just started and many of the participants, being in their first term, used the chatbot for the first time. 18% use the chatbot once a month, 9,7% once a week and 1% every day. In 78,6% of the cases the chatbot interaction took less than 5 minutes, and the rest of the participants interacted between 5-10 minutes with DIAS.

91,4% of respondents partly agree, agree or strongly agree that the chatbot delivers useful and informative responses. Of those 65,1% think that the chatbot enables them a faster access to useful information, and 78,3% perceive it easier to access information with the chatbot.

**Exploratory Analysis**

The histogram of the CUQ score shows a slightly skewed distribution (see figure 3). The Kolmogorov-Smirnov test was significant for all variables except the CUQ score, indicating a deviation from normal distribution for most of the data ($p < .05$). Outliers detected in the boxplots (figure 2) were not removed, as they were not considered to be measurement or data entry errors (Field, 2009).
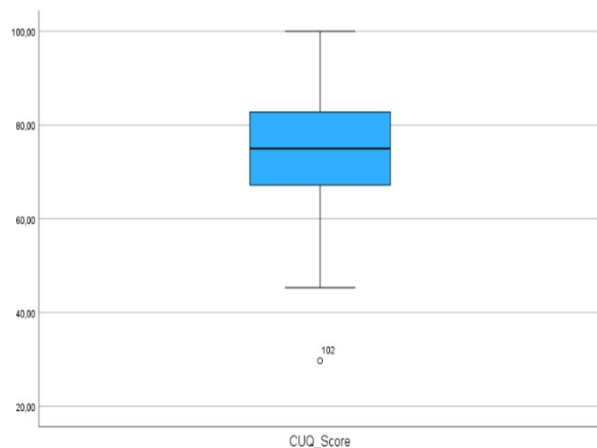


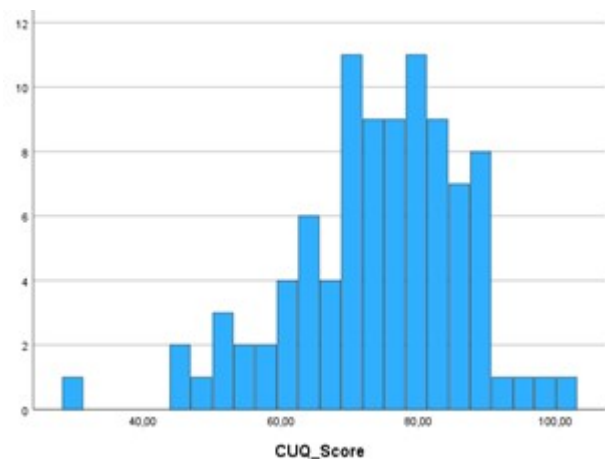**Figure 2:** Boxplot of the CUQ score      **Figure 3:** Histogram of the CUQ score

Based on the CUQ Score Calculation (Holmes & Bond), we analyzed the data of the 93 participants, who filled in the relevant questionnaire part. The analysis resulted in a mean CUQ score of 73,37, which is above the benchmark value of 68 (Holmes et al., 2019). Comparing the CUQ score of 73,37 with the benchmark of 68 in a one-sample t-test, the difference was strongly significant ($p<.001$). Compared to the usability study of a Social-Media Chatbot by Larbi et al. (2022) the DIAS chatbot scored 16 points higher. This difference was also significant ($p<.001$). The lowest score was 29,7, and the highest score was 100 (the highest possible score), with a median of 75.

**Figure 4:** Mean question scores of the positive items of the CUQ scale
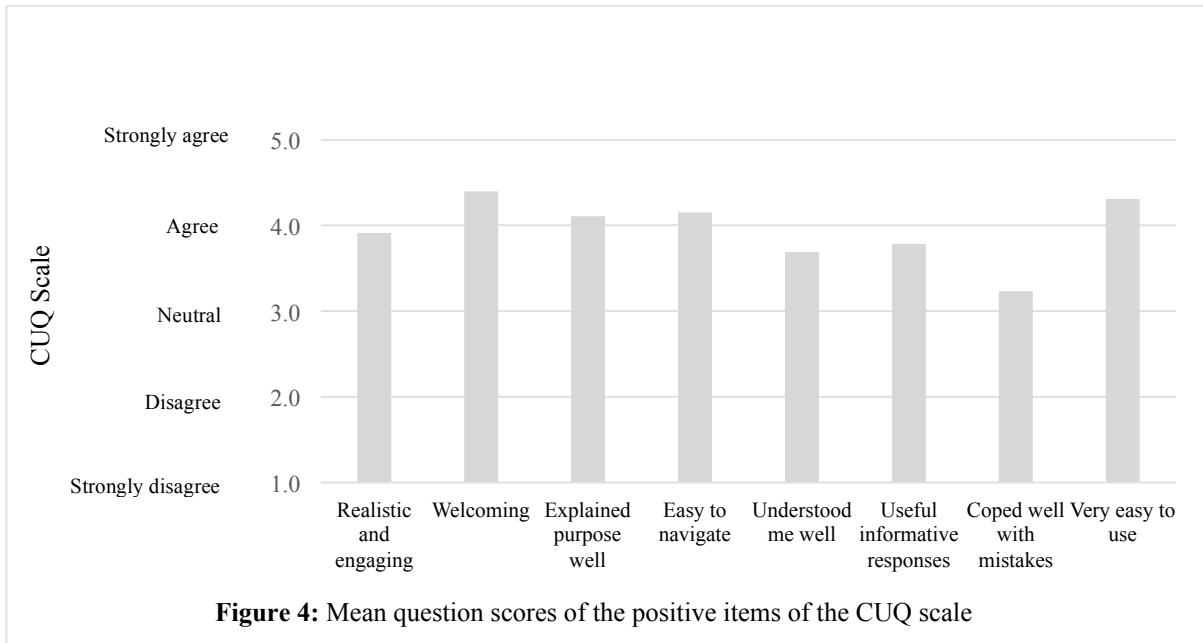
Among the positive aspects of the CUQ questionnaire the ease of use achieved the highest average ranking, followed by a welcoming way and an easy navigation (see figure 4). In general, questions relating to positive aspects all received similar ratings. Coping well with mistakes received the lowest average score of 3, 2.



**Figure 5:** Mean question scores of the negative items of the CUQ scale
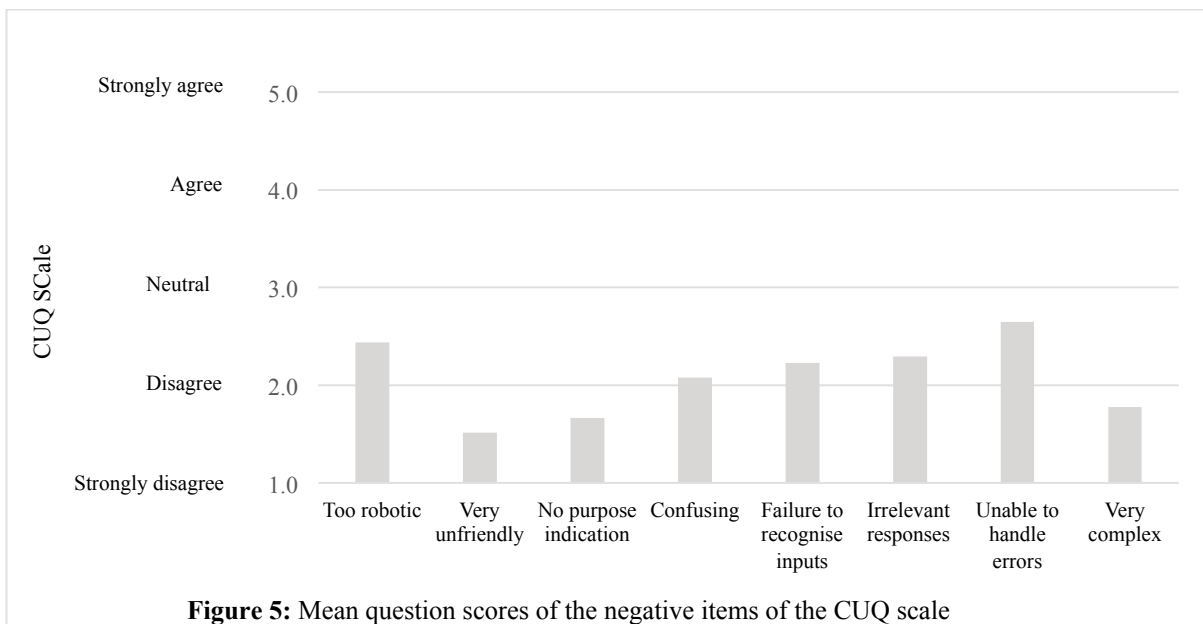
For the negative aspects (see figure 5) the questions whether DIAS can handle errors and whether the chatbot seemed too robotic received the highest average ranking, meaning that participants tended to agree more with these statements than with the other negative aspects. The handling of errors corresponds with the comparatively lower score for "coped well with mistakes" among the positive aspects. The questions concerning unfriendliness and complexity had the lowest average ranking, again corresponding with the opposite poles of positive aspects. As we cannot assume a normal distribution for all variables except the CUQ

score, and wanted to analyze ordinal data, correlations between the different variables were calculated using Kendall's tau (Field, 2009).

| Variable | Mean | s.d. | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|---|
| 1 Semester | 3,05 | 2,150 | 1 | | | | |
| 2 Age group | 2,18 | 0,416 | 0,077 | 1 | | | |
| 3 Frequency of Use | 1,42 | 0,748 | -,223[*] | -0,047 | 1 | | |
| 4 Interaction duration | 1,21 | 0,412 | -0,104 | -,244[*] | 0,093 | 1 | |
| 5 CUQ_Score | 73,37 | 12,605 | 0,030 | -0,071 | 0,086 | -0,089 | 1 |

*p < .05, **p<.01 two-tailed tests, for the correlation results of all CUQ items please consult the appendix

**Table 1:** Means, Standard Deviations and Correlations of Selected Variables Included in the Study

When looking at the correlations between the different variables (see table 1 and appendix table 2), there was a significant relationship between the frequency of use and the perceived usefulness of responses ($\tau$ = ,185, p < .05). Furthermore, the semester of the student participants was significantly related to the frequency of use ($\tau$ = -.223, p < .05). Interaction duration was related to a realistic and engaging perception of DIAS ($\tau$ = -.199, p < .05) and to the age group ($\tau$ = -.244, p < .05).

**Open-Ended Question Feedback**

The positive and negative feedback gained through open questions at the end of the questionnaire was classified into the following categories, partly based on the first qualitative study of the DIAS system (Fersch et al., 2022): general aspects, the benefit/utility of the chatbot, conversation management, design & language, further functions, and content (divided into content concerning study programs and examination/semester timetable). Furthermore, participants made several suggestions for possible new features of the chatbot.

The following table shows selected examples from the open-question section:

| Category / Subcategory | Example |
|---|---|
| **Positive Feedback** | |
| General | "nice idea, helpful" <br> "he is cute" |
| Utility | "the learning tips are really good" <br> "Helps to find certain things on the website quickly" <br> "support around the clock" |
| Conversation management | "easy to use" <br> "intuitive use, linking of websites" |
| Design & Language | "I particularly like the way of expression" <br> "You have the feeling that you are not chatting with a bot, but with a human being" <br> "The look is very appealing" |

| Negative Feedback | |
|---|---|
| Content - general | "If he does not find an answer to the search terms, then he should also clearly note this and offer suggestions"<br>"Provide even more information" |
| Content– study programme | "My degree programme […] was not recognised - not even when I selected it from the suggestions" |
| Content– examination/ semester timetable | "Couldn't answer all my questions or answered them incorrectly (e.g. when is the examination period in the winter semester)" |
| Conversation management | "Direct access to the respective module manual"<br>"At the beginning, perhaps another subdivision, in which area you are looking for something" |
| Design & Language | "I think the text is too long for a chat message"<br>"Gender[-neutral] language, feels foreign when chatting. I personally don't find it nice to read in "casual" environments either" |
| **Suggestions for new features** | "The possibility to search for specific professors, currently he does not answer to "Professor *name*""<br>"Be able to display images (e.g. contact cards/site plan) directly in the chat"<br>"Search function for rooms. e.g. where is room 50.1.2?" |

The answers to the open questions reflect the CUQ scores on their different dimensions. Participants appreciated the quick access to university-related information, the 24/7 support, the learning tips and the ease of use. Participants also highlighted an appealing design and the informal way of speaking adapted to the linguistic habitus of students. However, a major shortcoming of DIAS concerns content quality. Several participants mentioned that the chatbot did not recognize their answer or question in general or specifically concerning their study program and examination/semester dates. Apart from positive and negative feedback, there were also some valuable suggestions for new features such as a search function for rooms and professors/employees of the university and the ability to display images in the chat window.

**Discussion**

In comparison with both the benchmark of the CUQ score (Holmes et al., 2019) and a similar study by Larbi et al. (2022), the DIAS chatbot achieved significantly higher results, suggesting that the chatbot is perceived comparatively well by the sample group. In particular, the friendly and welcoming character of DIAS (expressed in design and language), intuitive use, and easy navigation were praised. This can also be attributed to the previous focus group discussion of the prototype version. Based on the results from this discussion, the beforementioned aspects could be improved continuously from an early development stage (Fersch et al., 2022).

Regarding utility, most respondents strongly agreed, agreed, or partly agreed that DIAS delivers useful, appropriate, and informative information. Of those the majority also thinks that the chatbot enables them a faster and easier access to relevant information. We assume these perceptions will increase over time, as the chatbot data set is further developed and enriched. This is also backed up by the open feedback where participants highlighted for example the "support around the clock" and the better overview of the partly crowded

information on the website through the chatbot. This positive feedback suggests that DIAS facilitates information flow and fulfills its role as a service-oriented chatbot, thus supporting the project's objective to give students better control over their study situation and reduce the frustration and time spent searching for information. At the same time, it also implies a relief for student service and study counseling, however, this has yet to be evaluated.

Considering the frequency of use and aspects of utility, a positive relationship exists between the perception of useful, informative responses and the frequency of use. This might be due to the fact that the more often a user interacts with DIAS, his or her ability to ask questions in a way that DIAS will understand them increases, and consequently the chatbot delivers better or more valuable answers. Furthermore, the frequency of use is lower in higher semesters, which seems logical as the service-oriented chatbot can more likely provide support to students who have just started at university.

Despite the positive feedback, there is still some improvement potential, which can mostly be deducted from the open-ended questions. Particularly, content quality can be further improved, as well as the handling of input that DIAS does not recognize or expect. Regarding navigation, design & language the open feedback was divided: while some participants mentioned the answer text to be too much, others commented that it has just the right scope. Direct access to relevant websites was partly appreciated and partly criticized, as some participants appreciated the links to different websites, others wished for more links. Participants also made valuable suggestions for additional features such as an in-chat search for employees/ professors and rooms. Like other publications (see chapter on the theoretical background), our study reflects the general shortcomings of conversational agents at this stage of development, particularly concerning the reaction to dialogue mistakes (Dibitonto et al. 2018).

**Limitations and Future Research**

The study underlies some restrictions which also offer potential for future research. For once, we relied on an ad hoc sample, and the characteristics of the target population were not exactly reflected in the sample group. Therefore, our results do not allow any conclusions to be drawn about the population as a whole and only serve as a critical assessment of correlations in the sample and as a basis for future studies with probability sampling methods.

In addition, there was no formal validation of the German CUQ questionnaire version. Although forward and backward translation can ensure a high degree of translation equivalence, a formal validation would further improve internal validity. The English version has been validated according to the authors (Holmes and Bond), however, with a relatively small number of participants (n=26). For an extensively validated instrument, future research might be advised to refer to the studies by Borsci et al., (2021 & 2022).

Furthermore, while our study also included some items on the utility of the chatbot, it focused on the chatbot's usability. To our knowledge there is no established, validated questionnaire which comprehensively surveys the utility of chatbots in an educational context. However, utility can give more information about whether a system actually provides necessary features, instead of whether these features are usable (Johannessen & Hornbæk, 2014). A possible utility questionnaire should be specifically adapted to the educational context and consider the difference between service-oriented and teaching-oriented chatbots, an aspect which should usability questionnaires should reflect too. Another suggestion would be to

consider ethical dimensions in a potential new questionnaire. A qualitative conversational agent should also respond to a diverse audience and be free from any biases towards minorities, particularly in an educational context.

In total, future research can use this exploratory study to establish and test new theories on using service-oriented chatbots in education. As a practical implication of the study, the DIAS system will be further improved, particularly regarding a better reaction to mistakes or misunderstandings. Furthermore, the research team intends to repeat the survey in due time to find out how the CUQ score, as well as the frequency and length of interactions with the conversational agent will develop and whether these parameters will increase over time. The suggestions for additional features will be tested and if possible, implemented with other ideas from the initial focus group on the chatbot prototype (Fersch et al., 2022).

**Conclusion**

The main objective of this study was to explore user experiences with the DIAS chatbot, benchmark the results against those of other conversational agents and detect possible effects, which might contribute to the theoretical development of the research field. We can conclude that DIAS received an above-average ranking in comparison with the benchmark and with another chatbot. Most participants perceived the answers to be informative and useful. Positive aspects included the intuitive use, a welcoming persona (expressed in design & language) and easy navigation. The negative feedback showed potential for improvement particularly in content quality and handling dialogue mistakes. In addition, we found that the chatbot is more frequently used by first-year students in the sample group and the perceived usefulness of responses increases as the frequency of use increases. With these findings we hope to have contributed to the still scarce empirical literature on service-oriented educational chatbots and to a further development of the field of applied AI in teaching and learning.

# Appendix

| | English Questionnaire Version | German Translation |
|---|---|---|
| **Positive aspects** | | |
| Q1 | The chatbot's personality was realistic and engaging. | Die Persönlichkeit des Chatbots war realistisch und einnehmend. |
| Q3 | The chatbot was welcoming during initial setup. | Der Chatbot war einladend bei Gesprächsbeginn. |
| Q5 | The chatbot explained its scope and purpose well. | Der Chatbot hat seinen Umfang und seinen Zweck gut erklärt. |
| Q7 | The chatbot was easy to navigate. | Der Chatbot war einfach zu navigieren. |
| Q9 | The chatbot understood me well. | Der Chatbot hat mich gut verstanden. |
| Q11 | Chatbot responses were useful, appropriate, and informative. | Die Antworten des Chatbots waren hilfreich, angemessen und informativ. |
| Q13 | The chatbot coped well with any errors or mistakes. | Der Chatbot kam gut mit fehlerhaften Eingaben oder Irrtümern zurecht. |
| Q15 | The chatbot was very easy to use. | Der Chatbot war sehr einfach zu bedienen. |
| **Negative aspects** | | |
| Q2 | The chatbot seemed too robotic. | Der Chatbot wirkte zu roboterhaft. |
| Q4 | The chatbot seemed too roboti | Der Chatbot wirkte sehr unfreundlich. |
| Q6 | The chatbot failed to recognize a lot of my inputs | Der Chatbot gab keinen Hinweis zu seinem Zweck. |
| Q8 | It would be easy to get confused when using the chatbot. | Die Nutzung des Chatbots könnte leicht zu Verwirrung führen. |
| Q10 | The chatbot failed to recognize a lot of my inputs. | Der Chatbot hat viele meiner Eingaben nicht erkannt. |
| Q11 | Chatbot responses were not relevant. | Die Antworten des Chatbots waren nicht relevant für mich. |
| Q13 | The chatbot seemed unable to handle any errors. | Der Chatbot schien nicht in der Lage zu sein, mit fehlerhaften Eingaben umzugehen. |
| Q15 | The chatbot was very complex. | Der Chatbot war sehr komplex. |

**Table 1:** Chatbot Usability Questionnaire (Holmes, 2019) including the German translation

| Variables | Mean | s.d. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 11 | 12 | 13 | 14 | 15 | 16 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 Realistic and engaging | 3,91 | 0,816 | 1 | | | | | | | | | | | | | | |
| 2 Too robotic | 2,44 | 0,972 | -,326** | 1 | | | | | | | | | | | | | |
| 3 Welcoming | 4,40 | 0,739 | ,328** | -,183* | 1 | | | | | | | | | | | | |
| 4 Very unfriendly | 1,52 | 1,274 | 0,053 | 0,084 | -0,042 | 1 | | | | | | | | | | | |
| 5 Explained purpose well | 4,11 | 0,714 | 0,116 | -0,102 | ,236* | 0,038 | 1 | | | | | | | | | | |
| 6 No purpose indication | 1,67 | 0,948 | -,204* | 0,078 | -,242* | 0,144 | -,309** | 1 | | | | | | | | | |
| 7 Easy to navigate | 4,15 | 0,884 | 0,101 | -0,112 | 0,145 | 0,064 | ,239* | -0,178 | 1 | | | | | | | | |
| 8 Confusing | 2,08 | 1,003 | -,216* | 0,146 | -,249** | 0,018 | -,266** | ,345** | -,524** | 1 | | | | | | | |

# References

Alexander, B., Ashford-Rowe, K., Barajas-Murphy, N., Dobbin, G., Knott, J., McCormack, M., Pomerantz, J., Seilhamer, R., & Weber, N. (2019). *Educause Horizon Report: 2019 higher education edition*. EDUCAUSE.

Benotti, L., Martinez, M. E. & Schapachnik, F. (2018). A Tool for Introducing Computer Science with Automatic Formative Assessment. *IEEE Transactions on Learning Technologies*, *11*(2), 179–192. https://doi.org/10.1109/tlt.2017.2682084

Borsci, S., Malizia, A., Schmettow, M., Van Der Velde, F., Tariverdiyeva, G., Balaji, D. & Chamberlain, A. (2021). The Chatbot Usability Scale: the Design and Pilot of a Usability Scale for Interaction with AI-Based Conversational Agents. *Personal and Ubiquitous Computing*, *26*(1), 95–119. https://doi.org/10.1007/s00779-021-01582-9

Borsci, S., Schmettow, M., Malizia, A., Chamberlain, A. & Van Der Velde, F. (2022). A confirmatory factorial analysis of the Chatbot Usability Scale: a multilanguage validation. *Personal and Ubiquitous Computing*. https://doi.org/10.1007/s00779-022-01690-0

Braun, M., & Alt, F. (2020). Identifying personality dimensions for characters of digital agents. *Character Computing*, 123–137. https://doi.org/10.1007/978-3-030-15954-2_8

Clarizia, F., Colace, F., Lombardi, M., Pascale, F., & Santaniello, D. (2018). Chatbot: An Education Support System for Student. In A. Castiglione, F. Pop, M. Ficco & F. Palmieri (Eds.), *Cyberspace Safety and Security: 10th International Symposium*, CSS (pp.291-302), Amalfi, Springer.

Diachenko, A. V., Morgunov, B. P., Melnyk, T., Kravchenko, O. & Zubchenko, L. V. (2019). The Use of Innovative Pedagogical Technologies for Automation of the Specialists' Professional Training. *International journal of higher education*. https://doi.org/10.5430/ijhe.v8n6p288

Dibitonto, M., Leszczynska, K., Tazzi, F. & Medaglia, C. M. (2018). Chatbot in a Campus Environment: Design of LiSA, a Virtual Assistant to Help Students in Their University Life. *Springer International Publishing eBooks*, 103–116. https://doi.org/10.1007/978-3-319-91250-9_9

Eicher, B., Polepeddi, L. & Goel, A. (2018). Jill Watson Doesn't Care if You're Pregnant. *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*. https://doi.org/10.1145/3278721.3278760

Field, A. (2009). *Discovering Statistics Using SPSS*. (3rd ed.). SAGE Publications.

Fersch, M. L., Schacht, S., Woldai, B., Kätzel, C. & Henne, S. (2022). Digital Learning Assistants in Higher Education Environments: A Qualitative Focus Group Study. *BCE Official Conference Proceedings*. https://doi.org/10.22492/issn.2435-9467.2022.28

Galko, L., Porubän, J. & Senko, J. (2018). Improving the User Experience of Electronic University Enrollment. *International Conference on Emerging eLearning Technologies and Applications*. https://doi.org/10.1109/iceta.2018.8572054

Hobert, S., & Meyer von Wolff, R. (2019). Say Hello to Your New Automated Tutor – A Structured Literature Review on Pedagogical Conversational Agents. *Wirtschaftsinformatik und Angewandte Informatik*, 301–314. https://aisel.aisnet.org/cgi/viewcontent.cgi?article=1199&context=wi2019

Holmes, S. & Bond, R. *The Chatbot Usability Questionnaire (CUQ)*. Ulster University. https://www.ulster.ac.uk/research/topic/computer-science/artificial-intelligence/projects/cuq

Holmes, S., Moorhead, A., Bond, R., Zheng, H., Coates, V. & McTear, M. F. (2019). Usability testing of a healthcare chatbot: Can we use conventional methods to assess conversational user interfaces? *European Conference on Cognitive Ergonomics*. https://doi.org/10.1145/3335082.3335094

Johannessen, G. H. J. & Hornbæk, K. (2014). Must evaluation methods be about usability? Devising and assessing the utility inspection method. *Behaviour & Information Technology*, *33*(2), 195–206. https://doi.org/10.1080/0144929x.2012.751708

Larbi, D., Denecke, K. & Gabarron, E. (2022). Usability Testing of a Social Media Chatbot for Increasing Physical Activity Behavior. *Journal of Personalized Medicine*, *12*(5), 828. https://doi.org/10.3390/jpm12050828

Larbi, D., Gabarron, E. & Denecke, K. (2021). Social Media Chatbot for Increasing Physical Activity: Usability Study. *Studies in health technology and informatics*. https://doi.org/10.3233/shti210604

Mebis (2018) *„E-Learning" und „Blended Learning"*. Mebis – Landesmedienzentrum Bayern. Retrieved June 22, 2022, from https://www.mebis.bayern.de/p/26271

Neumann, A. T., Arndt, T., Köbis, L., Meissner, R., Martin, A., de Lange, P., Pengel, N., Klamma, R. & Wollersheim, H. W. (2021). Chatbots as a Tool to Scale Mentoring Processes: Individually Supporting Self-Study in Higher Education. *Frontiers in Artificial Intelligence*, *4*. https://doi.org/10.3389/frai.2021.668220

Okonkwo, C. W. & Ade-Ibijola, A. (2021). Chatbots applications in education: A systematic review. *Computers & Education: Artificial Intelligence*, *2*, 100033. https://doi.org/10.1016/j.caeai.2021.100033

Page, L. C. & Gehlbach, H. (2017). How an Artificially Intelligent Virtual Assistant Helps Students Navigate the Road to College. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2940297

Pérez, J. Q., Daradoumis, T. & Puig, J. M. M. (2020). Rediscovering the use of chatbots in education: A systematic literature review. *Computer Applications in Engineering Education*, *28*(6), 1549–1565. https://doi.org/10.1002/cae.22326

Pérez-Marín, D. (2021). A Review of the Practical Applications of Pedagogic Conversational Agents to Be Used in School and University Classrooms. *Digital*, *1*(1), 18–33. https://doi.org/10.3390/digital1010002

Radziwill, N. & Benton. M.C. (2017). Evaluating Quality of Chatbots and Intelligent Conversational Agents. *arXiv: Computers and Society*. https://arxiv.org/pdf/1704.04579

Santoso, H. A., Winarsih, N. A. S., Mulyanto, E., Saraswati, G. W., Sukmana, S. E., Rustad, S., Rohman, M. S., Nugraha, A. & Firdausillah, F. (2018). Dinus Intelligent Assistance (DINA) Chatbot for University Admission Services. *2018 International Seminar on Application for Technology of Information and Communication*. https://doi.org/10.1109/isemantic.2018.8549797

Schurz, K., Schrumpf, J., Weber, J., Lübcke, M., Seyfeli, F. & Wannemacher K. (2021). Towards a User-Focused Development of A Digital Study Assistant Through a Mixed Methods Design, In D.G., Sampson, D. Ifenthale, I. Pedro, M.L. Mascia, (Eds.), *18th International Conference on Cognition and Exploratory Learning in Digital Age (CELDA 2021)*. IADIS Press. https://doi.org/10.33965/celda2021_202108l006

Shawar, B. A. & Atwell, E. (2007). Different measurements metrics to evaluate a chatbot system. *North American Chapter of the Association for Computational Linguistics*. https://doi.org/10.3115/1556328.1556341

Smutny, P. & Schreiberova, P. (2020). Chatbots for learning: A review of educational chatbots for the Facebook Messenger. *Computers & Education*, *151*, 103862. https://doi.org/10.1016/j.compedu.2020.103862

Toma, G., Guetterman, T. C., Yaqub, T. M., Talaat, N. & Fetters, M. D. (2017). A systematic approach for accurate translation of instruments: Experience with translating the Connor–Davidson Resilience Scale into Arabic. *Methodological innovations*, *10*(3), 205979911774140. https://doi.org/10.1177/2059799117741406

Vázquez-Cano, E., Mengual-Andrés, S. & López-Meneses, E. (2021). Chatbot to improve learning punctuation in Spanish and to enhance open and flexible learning environments. *International Journal of Educational Technology in Higher Education*, *18*(1). https://doi.org/10.1186/s41239-021-00269-8

Verleger, M. & Pembridge, J. J. (2018). A Pilot Study Integrating an AI-driven Chatbot in an Introductory Programming Course. *Frontiers in Education Conference*. https://doi.org/10.1109/fie.2018.8659282

Wollny, S., Schneider, J., Di Mitri, D., Weidlich, J., Rittberger, M. & Drachsler, H. (2021). Are We There Yet? - A Systematic Literature Review on Chatbots in Education. *Frontiers in Artificial Intelligence*, *4*. https://doi.org/10.3389/frai.2021.654924

**Contact email**: mascha-lea.fersch@hs-ansbach.de