Learning by Building Chatbot: A System Usability Study and Teachers' Views about the Educational Uses of Chatbots

Author[1] [.....], Author[2] [.....]

[1]University .....
[2] University
Email1, Email2

Abstract. This article describes an experiment to investigate the usability of the "Learning by Building chatbot" environment and the views of in-service teachers about the potential educational uses of chatbots. Chatbots are text or voice-based conversational interfaces that use natural language to simulate human conversations serving as virtual assistants to users. They have been used in a variety of areas such as e-commerce and healthcare while their use in education is relatively new. Chatbots have the potential to offer new dynamic forms of interactions, but there is often a burden to users to deploy and maintain their own chatbots. Our research group has proposed a web-based environment, "Learning by Building Chatbot", that alleviates this burden by offering educators and students without any programming knowledge the opportunity to deploy and maintain their own chatbots for learning and training. The environment is a block-based, visual editing environment of creating RiveScript-powered chatbots for learning and training, without needing to know any RiveScript. Twelve in-service teachers registered in a postgraduate program in technology-enhanced learning used this environment to create their own chatbots. Afterwards, they completed the System Usability Scale (SUS) questionnaire. Results have shown that the visualized editing environment for building chatbot is marginally acceptable and fair to use with the SUS score 57.50. Participants found the system ease to use and felt quite confident to use it. Its functions are well integrated while there are a few inconsistencies in the current version. Moreover, participants self-reported their views about the educational uses of chatbots. They agreed that chatbots, if used appropriately, can be a valuable educational tool because they can automate administrative and teaching tasks and can be supportive and engaging for students. However, since chatbot development is challenging, teacher should be well supported to integrate chatbots in the educational practice if they intend to do so.

Keywords: Language Learning, NLTK, AWS, Natural Language Processing, Learning Materials

1 Introduction

Conversational User Interfaces (CUI) or chatbots are software applications that use natural languages to conduct online conversations with users via text or speech, serving as virtual assistants (Luo et al. 2019). Chatbots offer the capacity to humans to use their natural language to communicate rather than syntax specific computer commands, opening up possibilities for more dynamic and accessible forms of interaction. Latest developments in Natural Language Processing (NLP), Natural Language Understanding (NLU) and Artificial Intelligence (AI), combined with the increased computing power available helped chatbots to become more mainstream.

There is an increased use of chatbots in a variety of areas such as e-commerce, retail services, customer support, healthcare and lately in education. However, studies related to chatbots in education are still in an early stage (Hwang & Chang, 2021) and the potential of using chatbots in education have not yet fully explored (Winkler & Soellner, 2018). A recent literature review by Wollny et al. (2021) categorized chatbots in education into three categories: learning chatbots, assisting chatbots, and mentoring chatbots. Chatbots can support student learning, e.g., they can assist in language learning (Kim, 2019) or programming (Daud et al., 2019), they can assist students by providing administrative assistance, e.g., answer general questions about a course, timetable, assignment deadlines, etc. (Sandoval, 2018), they can mentor students by providing scaffolding and recommendations, e.g., through adaptive formative quiz feedback (Vijayakumar, Höhn, & Schommer, 2019).

To date most chatbot applications in education are offered as commercial off-the-shelf products. Users have limited capacity to create, modify and maintain their own chatbots unless they have a programming background. The current project is aiming to develop a platform that teachers can use to create their own chatbots or have their students to create their chatbots for learning and training without prior programming knowledge and skills.

Based on Hassenzahl (2005) software attributes can be “hedonic” and “pragmatic”. “Pragmatic attributes emphasize the fulfilment of individuals’ behavioral goals, hedonic attributes emphasize individuals’ psychological well-being”. Usability is a pragmatic attribute that refers to the fulfilment of users’ functional goals and therefore it is important to be measured (Hassenzahl, 2013). Usability is defined in ISO 9241-11:2018 (Section 3.1.1) as ‘the extent to which a system, product or service can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use’.

The System Usability Scale (SUS) is validated questionnaire (Brooke, 1996) that can be used to assess the usability of any hardware, software system, device or service (Bangor, Kortum, and Miller 2008, 2009; Kortum and Bangor 2013; Lewis and Sauro 2009), and also it has wide acceptance and easy administration. The questionnaire consists of 10 items that are answered using a 5-step Likert scale ranging from “strongly disagree” to “strongly agree”, resulting in a single score between 0 and 100 (in 2.5 points increments) where higher scores indicate better usability. Scores[1] below ‘50’ is

[1] https://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html

unacceptable, over ‘50’ is OK, over ‘70’ is acceptable, and a score of ‘85’ can be considered ‘Excellent’.

Nielsen (2012) has found the minimum number of test users for a usability study is 5 and Hwang & Salvendy (2010) have the general rule 10±2 for optimal sample size and for major usability evaluations and meta-analysis results is 10. This study aims to investigate the perceived usability that in-service teachers have after they use the proposed “Learning by Building Chatbot” environment.

2 Learning by Building Chatbot

Our research team has proposed a block-based, visual editing environment based on RiveScript that enables users to create, view, test and manage their own chatbots (Removed, 2022). Fig. 1 shows a screenshot of the “Learning by Building Chatbot” environment.

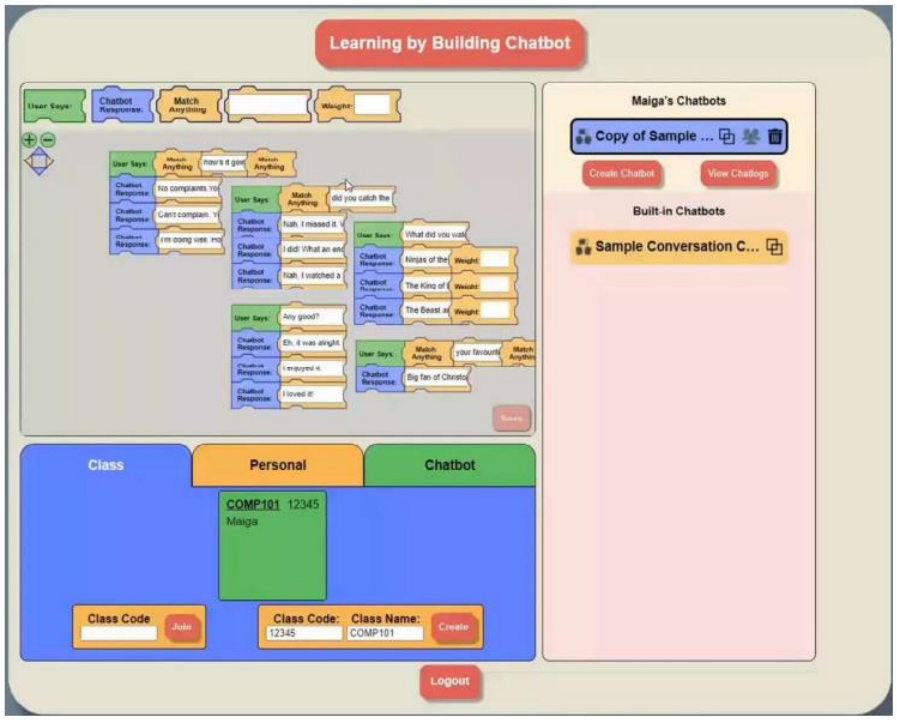

Fig. 1. The Visualized Editing Environment for Building Chatbot
(link is temporary removed for blind reviewing).

While RiveScript has a learning curve with distinct lexicons, formatting requirements and other nuances for the non-programmer, the “Learning by Building Chatbot” alleviates this burden for users. The editing environment handles the syntax, semantics,

and other idiosyncrasies of the RiveScript language. New chatbots can be easily built by dragging and snapping together their building blocks and typing inside the blocks the conversations to be carried out, as shown in Fig. 2.

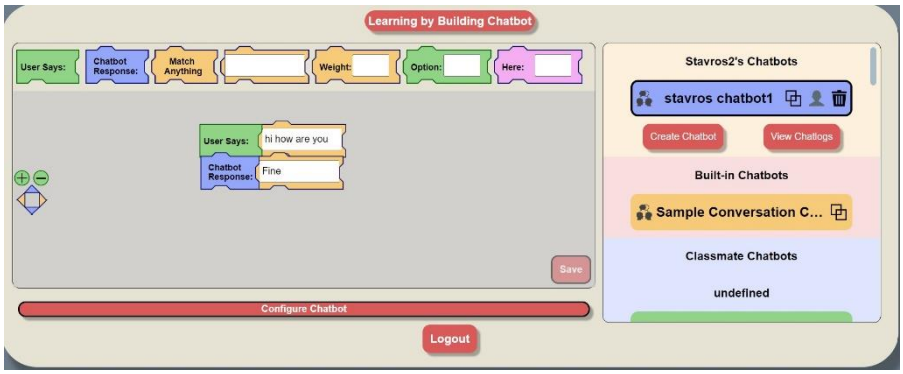

Fig. 2. Chatbot Creation.

Chatbots can be tested as Fig. 3 shows and can also be shared with the classmates and teacher. Therefore, the platform offers a co-creating feature that users can share the chatbot they designed to others and multiple people can work on different parts of a chatbot together.

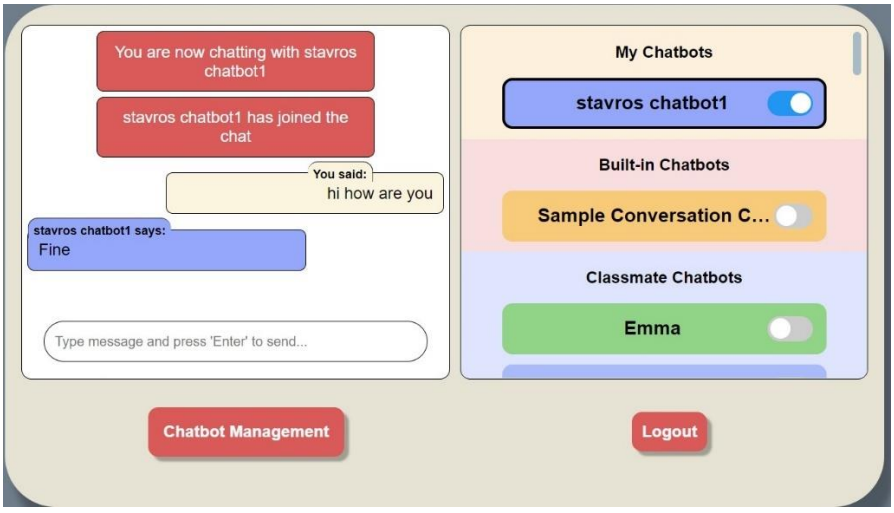

Fig. 3. Chatbot Test.

The platform allows the chatbots to be managed by their users and allows teachers to review student chatbot interactions from a single website without the need to self-host a chatbot or distribute a program. Moreover, the platform offers a few configuration options such which class the chatbot belongs to (Fig. 4), chatbot name, the default

chatbot language and reading/writing speed of the chatbot (Figure 5). While waiting for the chatbot's response, users can also see the real-time feedback "typing indicator" that shows the name of chatbot "is typing…". Both the speed and typing indicator make the chatbots more like users' human counterpart.

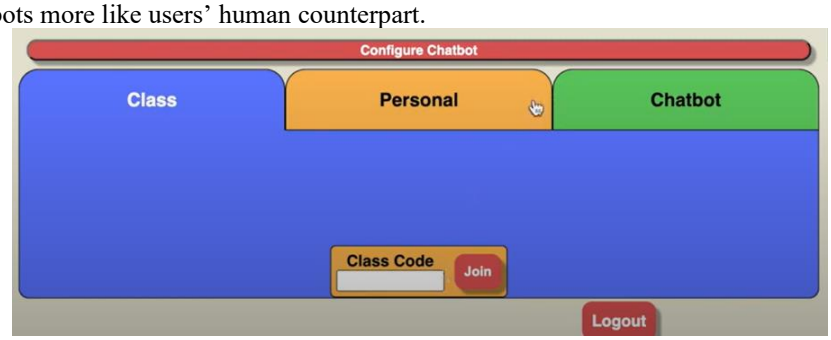

Fig. 4. Chatbots can be created for different classrooms.

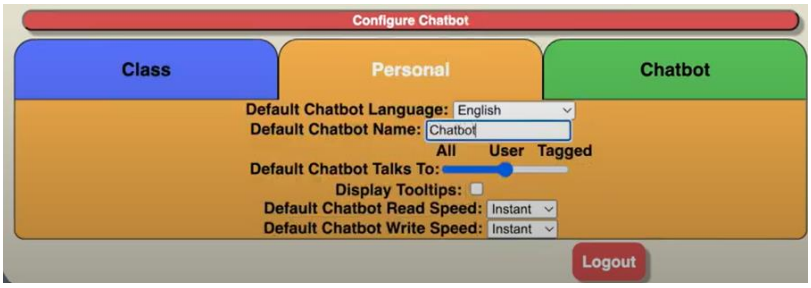

Fig. 5. Predefined configuration for chatbots created by the user.

3 Methodology

3.1 Participants and Procedures

The study took place in the context of the course "Technology Enhanced Teaching and Learning: Theory and Practice". This course is part of the Postgraduate Certificate in Digital Education within the MEd in Education Studies program at the University [details are omitted for review]. The aim of the course is to provide students with a theoretical background on technology enhanced learning and introduce them to a variety of educational technologies. Conversational Agents is one of these technologies. Twelve in-service teachers who were registered in the course participated in the study. The number of participants is within the usability study's general rule 10±2 for optimal sample size.

Once ethical approval was granted by the University [details are omitted for review] ethics committee, researchers invited students to participate in the study. Participation was voluntarily and anonymous. The study took place during fall semester 2022. After a dedicated guest lecture on the use of chatbots in education, students were introduced

to the “Learning by Building Chatbot” environment. Participants familiarised themselves with the environment, during an in-class session, by using it to build their own chatbots. They were asked to build simple chatbots that could be used to greet their students e.g., “Hi, how are you?”, “I am fine, thank you, and you?”, “I am well, thank you” or “Not bad”. Participants were then further instructed to continue using the environment and build very simple “Frequently Asked Questions” scenarios about their classes, e.g., timetable, due dates, marking criteria, assessment, etc. Finally, they have been asked to complete an online questionnaire made available to them through Qualtrics.

3.2 Instruments

The questionnaire had three parts. The first part included a series of socio-demographic and teaching related questions namely: gender, age, country of teaching, general digital skills level (basic, intermediate, advance), level and subject of teaching as well as previous experience with chatbots in education. The second part included the 10-item SUS questionnaire aiming to measure the usability of the “Learning by Building Chatbot” environment. The third part included the following open-ended questions aiming to increase our understanding on teachers’ views on the potential applications, benefits, and challenges of chatbots in education:

- Q1: What do you think about the (potential) use of the chatbots for teaching and learning purposes? Any particular areas of possible applications?
- Q2: How would you perceive the value of the chatbots in your own teaching? What do you think are the associated benefits and/or challenges.

The following two sections present the results following with the main conclusions and related discussions.

4 Results

Based on the responses to the first part with the socio-demographic and teaching related questions, participants were 6 female, 5 male, and 1 preferred not to say. Their distribution in terms of their age was 4 participants 21-30 years old, 2 participants 31-40 years old, 5 participants 41-50 years old and 1 participant 51-60 years old. In terms of their general digital skills level, 3 participants had basic digital skills (use of a basic range of software such as office; and devices such as computer, tablet), 8 participants had intermediate digital skills (use of a big variety of software such as, audio and video processing software; and devices such as smart interactive whiteboards), and 1 participant had advanced digital skills (use of highly innovative and complex digital and communication technologies, advanced skills such as programming, software and web development).

Most of the participants were teaching in primary education (5 participants) with the tertiary education (4 participants) and secondary education (2 participants) to follow,

while 1 participant described the education level they were teaching as other. Their teaching experience in years was varied with 5 participants to have 1-5 years teaching experience, 4 participants to have 6-10 years teaching experience and 3 participants to have more than 16 years teaching experience.

The distribution of the participants in terms of the subject they were teaching was 3 participants were teaching Science (Math, Physics, Chemistry, Biology), 1 participant was teaching Informatics/Technology, 1 participant was teaching Social Sciences, 2 participants were teaching Language/Literature while 5 participants reported their teaching subject as other (as these participants were teaching in primary schools). Finally, regarding the previous use of chatbots in class, only 1 out of the 12 participants said that they had used chatbots in their classes before. The respondents' socio-demographic and teaching related characteristics are presented in Table 1.

Table 1. Participants' socio-demographic and teaching related characteristics (N=12).

Gender		Age		Digital skills level		Teaching level		Teaching Experience (#in years)		Teaching Subject		Previous chatbot use in class	
Male	6	21-30	4	Basic	3	Primary	5	1-5	5	Science (Math, Physics, Chemistry, Biology)	3	Yes	1
Female	5	31-40	2	Intermediate	8	Secondary	2	6-10	4	Informatics/ Technology	1	No	11
Prefer to self-describe/Not say	1	41-50	5	Advanced	1	Tertiary	4	11-15	0	Social Sciences	1		
		51-60	1			Other	1	>16	3	Languages/ Literature	2		
										Other	5		

Regarding the results of the SUS questionnaire, a total of twelve responses were collected. While Brooke (1996) suggested that SUS is a unidimensional instrument with its questionnaire items better not considered individually, to highlight each one questionnaire item, Table 2 shows the responses on the individual questionnaire items along with the median, the mean and standard deviation.

Table 2. SUS questionnaire and statistics for each item.

	Strongly Disagree 1	2	3	4	Strongly Agree 5	Median	Mean	SD
1. I think that I would like to use this system frequently.	1	6	2	3	0	2	2.58	0.99
2. I found the system unnecessarily complex.	2	5	5	0	0	2	2.25	0.75
3. I thought the system was easy to use.	1	4	3	3	0	3	2.92	1.16

4. I think that I would need the support of a technical person to be able to use this system.	3	6	3	0	0	2	2.00	0.74
5. I found the various functions in this system were well integrated.	0	6	5	1	0	2.5	2.58	0.67
6. I thought there was too much inconsistency in this system.	0	6	5	1	0	2.5	2.58	0.67
7. I would imagine that most people would learn to use this system very quickly.	1	1	6	4	0	3	3.08	0.90
8. I found the system very cumbersome to use.	1	6	4	1	0	2	2.42	0.79
9. I felt very confident using the system.	1	2	4	5	0	3	3.08	0.99
10. I needed to learn a lot of things before I could get going with this system.	4	3	3	2	0	2	2.08	1.16

Based on (Brooke, 1996) the overall System Usability Score, representing the composite measure of the overall usability of the system was found 57.50, which is indicating the environment and its functionality at this moment is OK for using. Regarding the third part, participants' responses to the open-ended questions about their views on the potential applications, benefits, and challenges of chatbots in education highlighted the following.

The majority said that chatbots can be used for administrative purposes: "Provide administrative information about marking procedure, assignments, school policies", "Provide feedback to common student questions mostly related to admin stuff e.g., absences, marking policies etc.", "I think they could be good for answering basic admin questions such as due dates, class timetable".

Most participants said that chatbots can also be used for teaching and learning purposes: "It would be used to test factual knowledge", "It is good for digital literacy development" and "Good for comprehension in literacy", "I think it can be used in teaching languages and math" or "One use could be building a revision tool, whereby you essentially create a digital glossary on a series or unit of lessons".

One participant mentioned that chatbots are appealing for young generations because they are interactive: "Young people would enjoy the interactivity chatbot generate". One participant said that chatbots can be used to train preservice teachers: "Provide teaching scenarios simulating common classroom dialogues that student teachers can respond and practice with" and another one said that they can be used for communication with the parents: "In primary it could be good for communication with parents".

Five participants agreed that chatbots could be used to automate tasks that are performed frequently to "answering a series of most asked questions" and integrate teachers' feedback "in areas that students need most clarification".

The majority of participants agreed that the use of chatbots can be beneficial for students as "they can help learners control their learning and be creative". One said that they can enhance student engagement as "Chatbots can be engaging for pupils" and another one said that they "Encourage independence during tasks and more technology awareness for pupils." Most participants agreed that chatbots have "Lots of benefits, save time, extra support" and seven participants said that chatbots be useful not only in teaching but in assessment as well: "It is a useful learning and assessment tool."

Participants seemed to agree that "Any activity where knowledge must be stored for easy access would suit the use of a chatbot." However, they found that it is difficult for chatbots to "mimic human logic and empathy."

From the teachers' point of view, two participants mentioned that chatbots, once developed they can be very useful for teachers because they can "facilitate processes" and can "free up administration time". However, participants agreed that it is "Challenging for teachers to build their own chatbots." One teacher was skeptical about the generalized use of chatbots in education – "I do not believe these types of systems will ever be fully accepted by the teaching professionals" … "Until such times as the subtle variations in human behavior (variations in spelling, addition of grammar symbols, etc.) can be automated without programming." A participant emphasized that "The time potentially required to construct a worthwhile system that could be used in place of an actual person" is a real challenge. Unless "…pre-made chatbots can be produced to specific coursework then their potential use would be much greater." A participant also said, "The time required to build a chatbot (particularly by those whose digital skills are limited) to a level where it can be deemed usable and interactive could prevent implementation by teaching professionals". Another one suggested a way that teachers could be motivated to use chatbots if they receive appropriate training – "Chatbot use can be demonstrated to teaching professionals in their area of expertise."

In conclusion, despite the aforementioned challenges, participants tended to agree that chatbots can be useful if used appropriately – "I can imagine a myriad of uses for a system such as this if used correctly."

5 Discussion

The result found for the SUS of the "Learning by Building chatbot" is below 68 which is considered as average. However, Lewis and Sauro (2009) did a sensitivity data analysis with 19 datasets, and they came out "the mean of our Overall SUS data was 62.1, with a 99.9% confidence interval ranging from 58.3 to 65.9" instead of earlier Bangor et al. (2008) distribution 70.1 (with a 99.9% confidence interval ranging from 68.7 to 71.5). Our result **57.50** is falling a little bit outside of 58.3 (i.e., the 99.9% interval range).

Moreover, as Sauro (2018) reported, acceptable score corresponds to roughly above 70, **marginally acceptable to 50-70**, and unacceptable to below 50 (Bangor et al., 2008). The report also categorized "promoters" are close to 81, **"passives" are between 53 and 81**, and "detractors" are associated with 53 and below. In respect to the use of the adjectives including "Good," "OK," and "Poor", "excellent" is associated with 85, "good" is above 71, **"OK" (or "Fair") is above 51**. Finally with respect to grades, A+ is 84.1 to 100, A is 80.8 to 84, A- is 78.9 to 80.7, B+ is 77.2 to 78.8, B is 74.1 to 77.1, B- is 72.6 to 74, C+ is 71.1 to 72.5, C is 65 to 71, C- is 62.7 to 64.9, and D is 51.7 to 62.6. Therefore, we can safely say that the visualized editing environment for building chatbot is **marginally acceptable** and **ok/fair** to use with the SUS score **57.50**.

Considering the questionnaire items individually, the positive items #9, #7 and #3 had the higher mean values indicating that teachers felt confident using the system,

most people would learn to use it very quickly and the system perceived as easy to use. We can marginally infer that teachers would like to use the system frequently (item #1); however teachers have found that there was still some inconsistency in the system (item #6) despite the fact that various functions were well integrated (item #5).

The standard deviation of item #3 was relatively high indicating that not all teachers agreed with the level of easiness of the system, probably due to their own different digital skills level. The same holds for the standard deviation of item #10 indicating that teachers would need different levels of technical support before start using the system.

Finally, considering that SUS actually has two factors (Lewis & Sauro, 2009), i.e., Usability (items #1, #2, #3, #5, #6, #7, #8) and Learnability (items #4 and #9), interpretation of the participants' responses indicate that the system has high learnability since the median values for items 4 and 9 are quite low. Further studies can measure SUS in correlation to gender or digital skills.

In regard to the responses to the open questions related with teachers' perceptions on the potential use of chatbots on education, their benefits and challenges, teachers agreed that chatbots, is a promising educational technology and if used appropriately, can be a valuable educational tool because they can automate administrative tasks (e.g., inform about assignments, class timetables, syllabus, due dates, school policies, etc.) and teaching tasks (e.g., answer subject related questions, provide automated feedback on areas where students need most clarification) or assessment tasks (e.g., personalized assessment, automate marking and assessment feedback). Moreover, due to their interactivity they can engage students (Cunningham-Nelson et al., 2019). Our findings agree with previous research. Chatbots can offer many opportunities to the teaching and learning process (Hwang & Chang, 2021; Okonkwo & Ade-Ibijola, 2021; Wollny, et al., 2021). However, their deployment in education faces many challenges due to the technological limitations and appropriate pedagogical integration (Hwang & Chang, 2021; Huang, Hew & Fryer, 2020). Chatbot development is challenging, therefore teachers should have the support needed to integrate them in educational practice. The "Learning by Building" chatbot is a promising chatbot development environment for educational use.

References

1. Bangor, A., Kortum, P. & Miller, J. (2008). An Empirical Evaluation of the System Usability Scale. *International Journal of Human-Computer Interaction*, 24(6), 574–594. doi:10.1080/10447310802205776.
2. Bangor, A., Kortum, P. & Miller, J. (2009). Determining What Individual SUS Scores Mean: Adding an Adjective Rating Scale. *Journal of Usability Studies*, 4(3), 114–123.
3. Brooke, J. (1996). SUS: A 'quick and dirty' usability scale in Usability evaluation in industry (P. W. Jordan, B. Thomas, B. A. Weerdmeester, & A. L. McClelland eds.), pp. 189–194, London/UK: Taylor and Francis.
4. Cunningham-Nelson, S., Boles, W., Trouton, L., & Margerison, E. (2019). A review of chatbots in education: Practical steps forward. 30th annual conference for the australasian

association for engineering education (AAEE 2019): Educators becoming agents of change: Innovate, integrate. Motivate: Engineers Australia.

5. Removed for blind review
6. Daud, S. H. M., Teo, N. H. I., & Zain, N. H. M. (2020). E-java Chatbot for Learning Programming Language: A post-pandemic Alternative Virtual Tutor. *Int. J. Emerging Trends Eng. Res.* 8(7). 3290–3298. doi:10.30534/ijeter/2020/ 67872020
7. Galko, L., Porubän, J., & Senko, J. (2018). Improving the User Experience of Electronic University Enrollment," in 16th IEEE International Conference on Emerging eLearning Technologies and Applications, ICETA 2018, Stary Smokovec, Slovakia, Nov 15–16, 2018. Editors F. Jakab, 179–184. doi:10.1109/ICETA.2018.8572054
8. Hwang, W. and Salvendy, G. (2010). Number of people required for usability evaluation: The 10±2 rule. *Communications of the ACM,* 53(5), 130-133. Retrieved from https://dl.acm.org/doi/pdf/10.1145/1735223.1735255
9. Hassenzahl, M. (2005). The thing and I: Understanding the relationship between user and product. In *M. A. Blythe, K. Overbeeke, A. F. Monk, & P. C. Wright (Eds.), Funology: from usability to enjoyment* (pp. 31–42). Springer Science + Business Media Inc
10. Hassenzahl, M. (2013). User experience and experience design. *The Encyclopedia of Human-Computer Interaction. Interaction Design Foundation*. https://www.interaction-design.org/literature/book/the-encyclopedia-of-humancomputer-interaction-2nd-ed/user-experience-and-experience-design
11. Huang, W., Hew, K. F., & Fryer, L. K. (2022). Chatbots for language learning—Are they really useful? A systematic review of chatbot-supported language learning. *Journal of Computer Assisted Learning*, 38(1), 237– 257. https://doi.org/10.1111/jcal.12610
12. Hwang, G.-J. & Ching-Yi Chang, C.-Y. (2021). A review of opportunities and challenges of chatbots in education, *Interactive Learning Environments*, DOI: 10.1080/10494820.2021.1952615
13. ISO.2018. 9241-11:2018. Ergonomics of Human-System Interaction–Part11: Usability: Definitions and concepts. *International Standardization Organization(ISO)*
14. Kim, N.-Y. (2019). A Study on the Use of Artificial Intelligence Chatbots for Improving English Grammar Skills. J. Digital Convergence 17, 37–46. doi:10.14400/JDC.2019.17.8.037
15. Kortum, P., & Bangor. A. (2013). Usability Ratings for Everyday Products Measured with the System Usability Scale. *International Journal of Human-Computer Interaction,* 29 (2), 67–76. doi:10.1080/10447318.2012.681221.
16. Lewis, J. R., & Sauro, J. (2009). The Factor Structure of the System Usability Scale. *In Proceedings of the 1st International Conference on Human Centered Design*: Held as Part of HCI International, 94–103. doi:10.1007/ 978-3-642-02806-9_12
17. Luo, X., Tong, S., Fang, Z., & Qu. Z. (2019). Frontiers: Machines Vs. Humans: The Impact of Artificial Intelligence Chatbot Disclosure on Customer Purchases. *Marketing Science*, doi:10.1287/mksc.2019.1192
18. Nielsen, J. (2012). How many test users in a usability study? Retrieved from https://www.nngroup.com/articles/how-many-test-users/
19. Okonkwo, C.W. & Ade-Ibijola, A. (2021). Chatbots applications in education: A systematic review, Computers and Education: Artificial Intelligence, 2, 100033.
20. Sandoval, Z. V. (2018). Design and Implementation of a Chatbot in Online Higher Education Settings. Issues Inf. Syst. 19, 44–52. doi:10.48009/ 4.iis.2018.44-52
21. Sauro, J. (2018). 5 Ways to Interpret a SUS Score. MeasuringU. Access: https://measuringu.com/interpret-sus-score/

22. Vijayakumar, B., Höhn, S., and Schommer, C. (2019). “Quizbot: Exploring Formative Feedback with Conversational Interfaces,” in 21st International Conference on Technology Enhanced Assessment, TEA 2018, Amsterdam, Netherlands, Dec 10-11, 2018. Editors S. Draaijer, B. D. Joosten-ten, and E. Ras, (Springer), 102–120. doi:10.1007/978-3-030-25264-9
23. Winkler, R., & Soellner, M. (2018). Unleashing the Potential of Chatbots in Education: A State-Of-The-Art Analysis. in Academy of Management Annual Meeting Proceedings 2018 (1), 15903.
24. 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. 4:654924. doi: 10.3389/frai.2021.654924.