Research article

Perceived satisfaction of university students with the use of chatbots as a tool for self-regulated learning

María Consuelo Sáiz-Manzanares [a,b,*], Raúl Marticorena-Sánchez [c,d], Luis Jorge Martín-Antón [e,f], Irene González Díez [a], Leandro Almeida [g]

[a] Universidad de Burgos. Facultad de Ciencias de la Salud. Departamento de Ciencias de la Salud, Burgos, Spain
[b] Grupo de Investigación: DATAHES, Unidad de Investigación Consolidada de la Junta de Castilla y León N° 348: EHPAIT, Spain
[c] Universidad de Burgos. Escuela Politécnica Superior. Departamento de Ingeniería Informática, Burgos, Spain
[d] Grupo de Investigación: ADMIRABLE, Spain
[e] Universidad de Valladolid. Facultad de Educación y Trabajo Social. Departamento de Psicología, Valladolid, Spain
[f] Grupo de Investigación: N° 179. Unidad de Investigación Consolidada de la Junta de Castilla y León N° 348: EHPAIT, Spain
[g] Escola de Psicologia, Campus de Gualtar, Universidade do Minho, Braga, Portugal

ARTICLE INFO

ABSTRACT

Chatbots are a promising resource for giving students feedback and helping them deploy meta-cognitive strategies in their learning processes. In this study we worked with a sample of 57 university students, 42 undergraduate and 15 Master's degree students in Health Sciences. A mixed research methodology was applied. The quantitative study analysed the influence of the variables educational level (undergraduate vs. master's degree) and level of prior knowledge on the frequency of chatbot use (low vs. average), learning outcomes, and satisfaction with the chatbot's usefulness. In addition, we examined whether the frequency of chatbot use depended on students' metacognitive strategies. The qualitative study analysed the students' suggestions for improvement to the chatbot and the type of questions it used. The results indicated that the level of degree being studied influenced the frequency of chatbot use and learning outcomes, with Master's students exhibiting higher levels of both, but levels of prior knowledge only influenced learning outcomes. Significant differences were also found in students' perceived satisfaction with the use of the chatbot, with Master's students scoring higher, but not with respect to the level of prior knowledge. No conclusive results were found regarding frequency of chatbot use and the levels of students' metacognitive strategies. Further studies are needed to guide this research based on the students' suggestions for improvement.

1. Introduction

Modern society is constantly evolving, technological advances and the COVID-19 crisis have highlighted the need for a change in the way the teaching-learning process is designed and implemented. This change, among others, focuses on the use of technological advances, specifically the use of the "Internet of Things" and application of the Blended Learning (b-Learning) methodology [1,2]. Possible resources include using interactive modular platforms such as Learning Management Systems (LMS). These systems allow teachers to insert various teaching support resources (conversational assistants, virtual and/or augmented reality, virtual laboratories,

etc.) [3]. Using these resources is part of the new concept of a university that has been defined as the Smart University [4,5]. However, in order to achieve effective learning outcomes for students, implementation of technological resources must be based on effective pedagogical designs [6–8]. Some authors even advocate the creation of specific learning scenarios [2,9].

For this reason, the Smart University faces many challenges, including incorporating technological resources to support teaching within the LMS [10]. The ultimate goal is to enhance Self-Regulated Learning (SRL—which is understood as students being encouraged by the teachers to reflect on their own knowledge and understanding) in real time and increase the number and type of metacognitive strategies throughout the learning process. Using intelligent personal assistants or chatbots [11] is one of the possible technological resources that shows promise in addressing these challenges. These aspects will be discussed below.

2. Literature review

2.1. Self-regulation and learning and Advanced Learning Technologies

The use of Advanced Learning Technologies (ALT) resources within virtual learning spaces, LMS, can improve the development of SRL processes [12]. This is because ALT encourages individualised learning [13], which increases motivation [14] and students' use of more complex metacognitive strategies [15]. According to Flavell [16], metacognition is defined as knowledge about one's own knowledge and the cognitive processes involved in it. He differentiated between declarative knowledge "knowing what" and procedural knowledge "knowing how". This definition, according to Veenman et al. [17] and Veenman & Beishuizen [18] falls within the framework of a mixed model explaining cognition and metacognition. This framework includes the work of Brown & DeLoache [19], who related metacognition to self-regulation strategies, i.e. to the procedural knowledge mentioned by Flavell [16]. In particular, according to Schellings et al. [20], there are four distinguishable types of metacognitive strategy [21].

a) Orientation: these are the strategies the subject uses to specify the demands of the task in cognitive terms. These strategies guide the resolution process and activate the prior knowledge needed to resolve a task.
b) Planning: these strategies allow the process of solving problems or tasks to be put in a sequence of steps.
c) Evaluation: these strategies facilitate monitoring throughout the problem-solving process, allowing evaluation of the steps in the resolution process, assessing whether they have been effective or not in order to redirect planning, if necessary.
d) Elaboration: these strategies are the highest level of metacognition and involve reflection on one's own practice. Depending on the conclusions, the learner can modify the process and procedure during a task or problem solving.

This classification refers to the instructional model proposed by Veenman [22] called WWWH (What, When, Why and How). Along the same lines, Román & Poggioli [23] proposed a classification of metacognitive strategies that includes metacognitive strategies *per se* (self-knowledge, self-planning and self-assessment) and what those authors called processing support strategies, which include
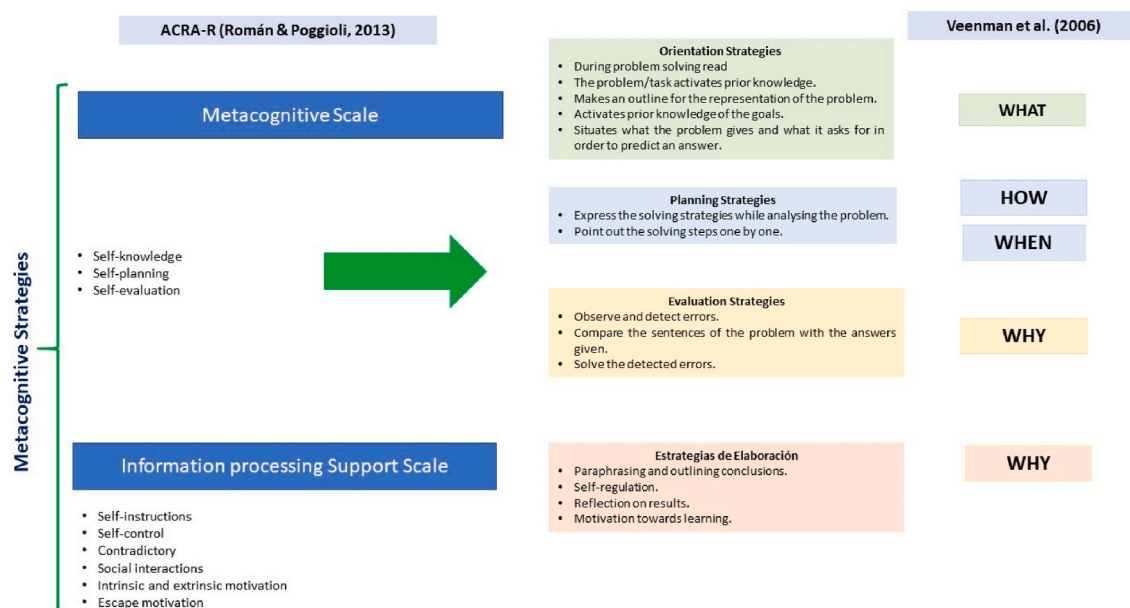

Fig. 1. Metacognitive and information support strategies: relating the theories of Roman and Poggioli (2013) and Veenman et al. (2006) (Note to editor: Fig. 1 in case of publication of the manuscript should be in colour).. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

self-regulation, self-control, motivation and social interactions. Fig. 1 presents a relationship between the studies by Veenman et al. [24] and the studies from Román & Poggioli [23] about classifying metacognitive strategies. This relationship will be followed in the present study.

It is therefore important to assess students' metacognitive strategies, as their use is directly related to effective learning outcomes [25] However, other research [22] has shown that there is a difference between the results of assessing metacognitive strategies using online methods (analysis of information during task resolution) and using offline methods (questionnaires on the perceived use of metacognitive strategies). According to studies by Van Der Stel & Veenman [21] the correlations between the measurement results from the two methods are weak, since the second method refers to the subject's perception of strategy use, and this may be influenced by aspects related to long-term memory. These may distort the actual perception of strategy use. However, the first method has more robust reliability indicators. Both methods should therefore be used to collect information and compare the results [7].

Another important aspect in the study of using metacognitive strategies is that using more complex strategies (Planning, Evaluation and Elaboration) seems to depend on the learners' levels of prior knowledge [15,26]. Prior knowledge about the field of study is associated with more effective learning. Although this relationship seems to change throughout the learning process. As it progresses, the use of metacognitive strategies increases, which refers to differential behaviour over time [27–30]. This reaffirms the need to analyse how SRL are used over time [31].

From this framework, one of the ALT resources that shows promise—specifically in students in higher education—is the use of chatbots embedded in LMSs [32]. This is because these tools—conversational agents—encourage students to reflect on their own practice, since asking the chatbot questions already indicates thinking about the course content or the subject to be learned [33]. Likewise, analysing the chatbot's response helps the learner to use metacognitive strategies for evaluation and elaboration. All of which will enhance self-regulation in the learning process and autonomy in task resolution. Therefore, using chatbots is emerging as an opportunity to make it easier for learners to deploy metacognitive strategies during the learning process in b-Learning environments. While there are technical difficulties regarding the up-to-date technology that chatbots require, there is also a reluctance to use them [34], especially among university students, related to confidentiality of data, and technological skills that teachers and students need in order to apply these resources [35,36].

2.2. Use of intelligent personal assistants or chatbots in the teaching-learning process

Using conversational agents—including chatbots—to enhance learning is an emerging, but not yet widespread practice. Chatbots are tools that combine aspects of Artificial Intelligence (AI) and Natural Language Processing (NLP) and allow for some conversational interaction with a human being [37]. These interactions are mainly through text and voice, and may even be multimodal, combining image, video, etc. From the beginning, chatbots have been developed with a tutoring functionality to perform tasks [38]. The ultimate goal is to provide help to a human being via machine programming in order to provide a real-time interaction at any time [39]. At present, they are quite widely used in health contexts and their effectiveness has been demonstrated [40]. More specifically, chatbots using AI, NLP and Deep Learning (DL) seem to be the most effective [40,41] however, more studies are needed in this respect [42].

Using chatbots in educational contexts is just beginning [2], and they are being embedded in LMSs, such as Moodle (Modular Object-Oriented Dynamic Learning Environment). This is proving to be very effective for some students with special educational needs, e.g. people with visual impairments [43] or attention deficit disorders. This is due to the additional functionality chatbots offer as a guide for navigation on the platform or on the web, making it a resource that can facilitate inclusive learning [44]. It has also been found to be effective in learning to read and in increasing student motivation, specifically in primary education [33].

There are many types of chatbot, some of which include a voice response [11,39], but as indicated above, they present technological challenges that need to be addressed. Segedy et al. [45] emphasised the need to improve the feedback provided by the conversational agent. The way to improve this functionality is by studying the data collected in the interaction between the user and the conversational agent, in a question-answer binomial. This needs the application of data mining and Deep Learning techniques [40]. Another key element is the application of instructional designs that are conducive to incorporating technological resources that support the development of metacognitive strategies [46]. Instructional designs that facilitate self-regulation and motivation towards learning seem to be more effective, as they improve the application of metacognitive strategies [47]. Various studies along these lines [1] noted the need for pedagogical design [48] to consider the different levels of prior knowledge [15] and students' learning profiles or patterns [49]. Students' technology skills is another important variable for the effective use of chatbots. In addition, Tsivitanidou & Ioannou [2] indicated the importance of monitoring the learning process by offering personalised feedback to each student [1,49,50]. This means that the process of building a chatbot becomes a challenge of technological structure and instructional design.

2.3. Design and development of a chatbot for instructional support in higher education

There are many types of conversational assistant, which may be voice assistants, text assistants, multimodal assistants, etc. As noted above, the structure is based on the use of AI together with NLP [2]. The design of chatbots can be traced back to the work of Weizenbaum [51] which was developed from Turing's [52] emulation and which was applied to the context of patient-therapist interaction [53]. The structure Weizenbaum [51] proposed was based on natural language and a question-answer system. The question asked by the user enters the conversational agent's system and is understood as an *Intent*, then the Natural Language Understanding (NLU) system searches for similarity between the user's question and the available records. In the process of searching for similarity, the system analyses the literal sentence produced by the user (*Utterance*) with alternative sentences included in the system to establish correspondence with the *Intent*. The system includes different *Entities* that serve to extract common types of expressions. This would act

similarly to a category sorting process, e.g. @sys.colour would group different colours. These systems can also define the *Context*, the variables of a conversation in a given context can be specified, similar to a real conversation. This is the pragmatic part of the conversation, which is the most difficult aspect to implement. The system also contemplates *Fallback*, which refers to an *intent* that is given by default when the input has not been acknowledged. Finally, these systems can also include an *Event*, which is a process of automating actions that will trigger the execution of an *intent*.

Examples of this architecture can be found in conversational agents such as Apple's Siri [54] and Amazon's ALEXA [55]. However, they have yet to be incorporated successfully into the educational environment. Some studies have indicated the technical difficulties of linking to learning platforms [34] as well as with development and structuring of the database from the inclusion of Frequently Asked Questions (FAQ) [56]. Students have been found to be somewhat reluctant to use them related to concerns about privacy of conversations [35]. This is why the development of chatbots in educational environments is a complex process that requires further research to move forward, research that should focus on the study of user satisfaction and on users' suggestions for improving chatbots [41]. This all means that developing chatbots within LMSs and evaluating their effectiveness is a challenge in education, particularly in higher education.

In addition, Fidan and Gencel [57] noted the usefulness of chatbots as a means of regulation and feedback in learning, specifically in Blended learning (b-learning), to increase students' intrinsic motivation. They also found using chatbots to be beneficial for activating the establishment of teaching objectives and maintaining attention in e-learning or b-learning teaching.

Using technological resources—such as chatbots—that promote real-time interaction about the questions and issues students may have throughout their learning processes is important in effectively personalizing these processes [58]. This is still an emerging area, although the potential benefits of this technology are promising. Fig. 2 shows a diagram of the state of the art our study addresses.

Based on the above, the research questions for the present study were as follows.

2.4. Quantitative study

RQ1. Are there significant differences in frequency of chatbot use and in learning outcomes with regard to students' levels of prior knowledge (low vs. medium)?

RQ2. Are there significant differences in frequency of chatbot use and learning outcomes with regard to the level of course the students are doing (bachelor's vs. master's degrees)?

RQ3. Are there significant differences in the students' perceived levels of satisfaction with using the chatbot depending on their prior levels of knowledge (low vs. medium)?

RQ 4. Are there significant differences in students' perceived satisfaction with using the chatbot depending on the level of course they are doing (bachelor's vs. master's degree)?

RQ5. Is the frequency of chatbot use related to the percentile score in the Metacognitive Strategies Scale (MS)?

2.5. Qualitative study

RQ6. Are there differences between Bachelor's vs. Master's students' suggestions of positive aspects to add or negative aspects to remove from the chatbot technology?

RQ7. What kind of metacognitive strategies (orientation, planning, evaluation and/or elaboration) do the questions asked by Bachelor's and Master's students to the chatbot relate to?

3. Material and methods

3.1. Participants

We worked with a sample of 57 students from Health Sciences, 42 from the third year of the Bachelor's degree and 15 from the Master's degree. Convenience sampling was used to select the sample by taking students participating in a subject that used an Online Project Based Learning (OPBL) methodology. Participation was voluntary without any financial compensation. The inclusion criterion was that students signed the informed consent form; exclusion criterion was that they did not. Table 1 presents the participants according to age, degree type, and gender. The majority of the participants were women (this is common in Health Sciences degrees) and the mean ages were close between men and women within each degree type, [1] with a greater age range in the Master's students.[2]

3.2. Instruments

a. UBUVirtual Platform. This platform is an LMS developed in a Moodle environment, version 3.11.2 was used.
b. Learning strategies scale ACRA (r) by Román-Sánchez & Poggioli [23]. ACRA (r) is a well-tested instrument in research on learning strategies in Spanish-speaking populations (Carbonero et al., 2013). ACRA identifies 32 strategies at different stages of information processing. In this study, only the metacognitive scales were used (this included metacognitive and information processing scales).
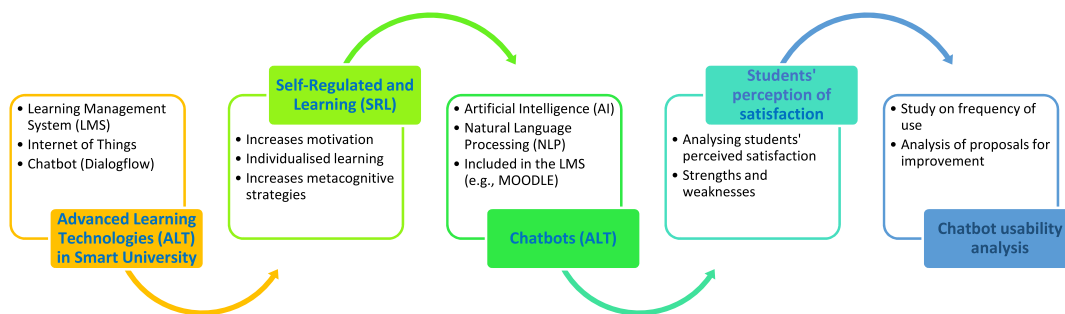
---

[1] https://github.com/.
[2] https://dialogflow.cloud.google.com.

Fig. 2. Diagram of the state of the art in ALT, SRL, Chatbots, Students 'satisfaction and Chatbot usability.

Table 1

Description of the participants with respect to the variables age and gender.

Degree type			Women			Men	
	N	N	M	SD	n	M	SD
Bachelor's	42	38	22.39	2.44	4	22.75	2.4
Master's	15	13	31.38	7.62	2	31.53	5.9

Note. M = Mean age; SD = Standard Deviation age.

The Metacognitive Scale (MS) includes subscales of self-knowledge, self-planning, and self-evaluation. It has a Cronbach reliability coefficient of $\alpha = 0.82$ and Omega coefficient of $\Omega = 0.87$. The Information Processing Support Scale (IPSS) includes subscales of self-instruction, self-control, contradictory strategies, social interactions, intrinsic and extrinsic motivation, and escape motivation. It has a Cronbach reliability coefficient of $\alpha = 0.83$ and Omega coefficient of $\Omega = 0.82$. ACRA (r) provides a conversion of the direct scores (DS) from both scales into Percentiles (P) ranging from P_1 to P_{99}. In this study, the reliability indicators for the MS were $\alpha = 0.85$ and $\Omega = 0.85$, and for the IPSS, they were $\alpha = 0.86$ and $\Omega = 0.90$.

c. Applied teaching methodology. A methodology with OPBL [1] was used in b-Learning teaching environments. Students worked in groups of 2–5 members on a project based on a practical case. The work was done in the classroom and on the learning platform based on Moodle, UBUVirtual.

d. Monitor tool [59]. This is a free, open-source, Java desktop application (available at. It connects to the selected Moodle through web services and the REST API provided by the server. It has six modules: visualisation (which offers the representation of frequencies in different graphs: Heat Map, Boxplot, Dispersion, etc.), comparison, forums, dropout risk (detection of students who have not logged in for 7–15 days at certain times during the course), Calendar of events and Clustering (finding clusters by applying different algorithms such as *k-means* or Fuzzy *k-means*). In this study, Heat Map was used to monitor the frequency of participants' chatbot use. An example of monitoring student chatbot usage is shown in Fig. 3.

d Chatbot An *ad hoc* chatbot was developed for each subject in the Bachelor's and Master's degrees. DialogFlow technology was used to do this, which can be found at the link below This is a freely available technology. The chatbot was included as a resource in the UBUVirtual platform for each subject, an example of which is shown in Fig. 4. User identity was protected, as the only data sent was the question being asked in plain text. Technically it is not possible to identify who asked a question in the question itself, as the integration between Moodle and Dialogflow does not share web session data, nor does it attach unique user identifiers in Moodle. Dialogflow keeps records of questions asked in the Google Cloud Platform The logs can be downloaded in JSON format and it is possible to query these logs to determine which agent (subject) asked the question, the date/time of the question, and the automatic answer given by the chatbot. Similarly, through the Monitor application, the frequency of chatbot use can be recorded [60] in order to check how effectively the conversational assistant is being used. Supplementary Material 1 provides information about the process of integrating the chatbot—developed with DialogFlow technology—into a subject in the Moodle platform—in this case UBUVirtual.

e. Perceived Prior Knowledge Questionnaire (PPKQ) [11]. An *ad hoc* questionnaire was produced on concepts relevant to each of the subjects. Each questionnaire consisted of 12 questions related to relevant concepts in each subject with responses on a Likert-type scale from 1 to 5 (1 none to 5 all). The undergraduate questionnaire gave a reliability of $\alpha = 0.96$ and an Omega coefficient of $\Omega = 0.96$, whereas the master's questionnaire gave an $\alpha = 0.97$ and an Omega coefficient of $\Omega = 0.97$. To interpret the results, the students' scores were grouped into low prior knowledge (scores of 1–2.4 out of 5) and medium prior knowledge (scores of 2.5–3.9 out of 5), there were no scores higher than 3.9 out of 5. These questionnaires are provided in AppendixA.

f. Student's Perceived Satisfaction with the Use of the Chatbot Survey (SPSUCS) adapted by Huang & Chueh [41]. This is an *ad hoc* survey consisting of 10 closed questions measured on a Likert-type scale from 1 to 5, where 1 is strongly disagree and 5 is strongly agree. The following dimensions were analysed: Chatbot Performance Accuracy (CPA) 3 items; Perceived Benefit (PB) 4 items; Chatbot Usefulness (CHU) 1 item; Motivation (M) 1 item; and Overall Satisfaction (OS) 1 item. In addition, the questionnaire includes three open-ended questions relating to perceived chatbot usefulness, suggested additions and suggested items for removal.

Reliability indicators of α = 0.98 and Omega coefficient of Ω = 0.98 were found for the 10 Likert-format items. The survey is provided in AppendixB.

d. Learning outcomes. These were measured on the scale of 0–10 used in the Spanish education system. The following gradation was established: 0–4.9 Fail, 5 to 6.9 Pass, 7 to 8.9 Merit and 9 to 10 Outstanding.

3.3. Procedure

Prior to the start of the study, approval was obtained from the University of Burgos (Spain) Bioethics Committee (No. IO 03/2022).[3] A chatbot was then developed for each of the subjects in the study (see the Instruments section). Subsequently, students were informed of the aims of the study and written informed consent was obtained from those who agreed to participate. Then, during the first two weeks of the semester, students completed the ACRA (r) metacognitive strategies scale (comprising the ACRA (r) metacognitive and processing support scale) [23]. The chatbot was then explained to the students. Teaching was performed as usual during the semester. Finally, in week 9 (the last teaching week in the semester) the usability of the chatbot was assessed by applying the SPSUCS survey [11]. In addition, use of the chatbot on the UBUVirtual platform throughout the semester was monitored using the Monitor tool [60].

3.4. Designs

Recent studies [61,62] advise that empirical research using a b-Learning methodology should employ mixed research designs, as they provide important information for understanding the learning process. Accordingly, a mixed research methodology was applied in this study. In the quantitative study, following [63], to test RQ1, RQ2, RQ3 and RQ4 we used a 2 × 2 factorial design [type of degree (bachelor vs. master), degree of prior knowledge (low vs. average)] and a descriptive-correlational design to test RQ5. In the qualitative study following [64] a comparative design was applied to test RQ6 and RQ7.

3.5. Data analysis

Due to the characteristics of the sample (type of sampling and number of participants) in the quantitative study, non-parametric statistics were used to test the research questions. Specifically, a Mann-Whitney U test was used to test RQ1, RQ2 and RQ3, and a cross-tabulation and Cohen's Kappa coefficient were used to test RQ4. In addition, effect sizes were determined using the formula $\eta^2 =$ Z2/N-1, the interpretation following [65] were: η^2 between 0.2 and 0.3 small effect; η^2 from 0.50 to 0.8 η^2 moderate effect; η^2 higher than 0.8 large effect. The calculations were done using SPSS v. 28 [66]. For the qualitative study, before testing the RQs, the students responses in the SPSUCS were categorized, followed by an occurrence-document analysis using the qualitative analysis software Atlas. ti v. 22 [67].

Table 2 presents an overview of the research questions, the research method, the sample, the instruments used, the timing, and the data analysis applied.

4. Results

The results are presented in the order of the research questions.

4.1. Influence of prior knowledge on frequency of chatbot use and learning outcomes

RQ1. Are there significant differences in frequency of chatbot use and in learning outcomes with regard to students' levels of prior knowledge (low vs. medium)?

Significant differences were found between students with little prior knowledge and those with average prior knowledge, which better learning outcomes for the latter. There were no differences in the use of the chatbot. According to Cohen [65], the size of the effect was small (see Table 3).

4.2. Influence of the type of degree (bachelor vs. master) on frequency of chatbot use and learning outcomes

RQ2. Are there significant differences in frequency of chatbot use and learning outcomes with regard to the level of course the students are doing (bachelor's vs. master's degrees)?

Significant differences were found between bachelor's and master's degree students in both frequency of chatbot use and learning outcomes, with master's students scoring higher. In both cases, according to Cohen [65], the size of the effect was small (see Table 4).

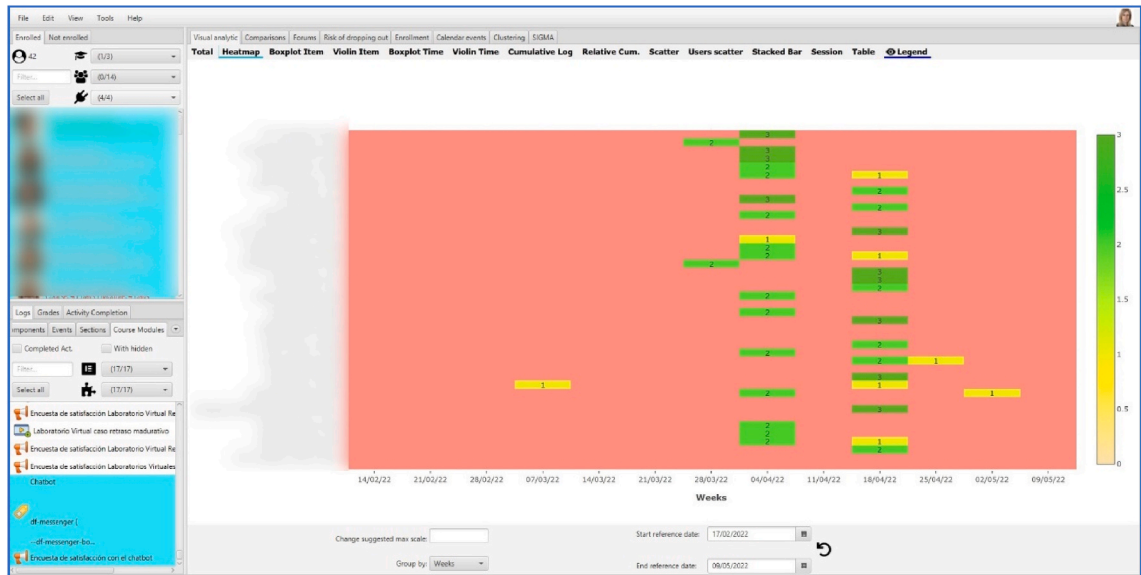[3] https://console.cloud.google.com.

Fig. 3. Monitoring the use of the chatbot with the Monitor application throughout the semester (12 weeks)
(Note to editor: Fig. 2 in case of publication of the manuscript should be in colour)... (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Fig. 4. Chatbot functioning within a subject on the Moodle-based platform, UBUVirtual
(Note to editor: Fig. 3 in case of publication of the manuscript should be in colour)... (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)
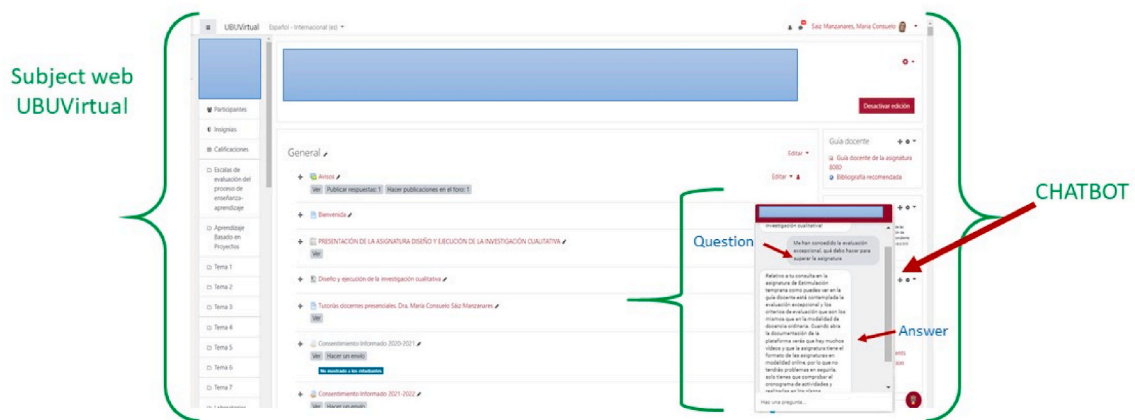
4.3. Influence of the type of prior knowledge on satisfaction with the use of the chatbot

RQ3. Are there significant differences in the students' perceived levels of satisfaction with using the chatbot depending on their prior levels of knowledge (low vs. medium)?

No significant differences were found in students' perceived satisfaction with using the chatbot with regard to their prior levels of knowledge (see Table 5).

4.4. Influence of the type of degree (bachelor vs. master) on satisfaction with the use of the chatbot

RQ 4. Are there significant differences in students' perceived satisfaction with using the chatbot depending on the level of degree course they are doing (bachelor's vs. master's degree)?

There were significant differences in all of the items in the SPSUCS scale except item 4 ("I can answer most of the important

Table 2
List of research designs applied to test the research questions.

Research Questions	Research method	Information Source	Sampling Method	Data Collection Tools	Data Collection Time	Data Analysis Method
RQ1	Quantitative	Students at the University of Burgos	Convenience	Monitor Learning outcomes PPKQ	The whole semester 9 weeks At the end of the semester 1–2 weeks	Mann-Whitney *U* test
RQ2	Quantitative	Students at the University of Burgos	Convenience	Monitor Learning outcomes Level of studies (bachelor vs. master)	The whole semesterAt the end of the semester While	Mann-Whitney *U* test
RQ3	Quantitative	Students at the University of Burgos	Convenience	SPSUCS Monitor Level of prior knowledge (low vs. medium)	9 weeks The whole semester	Mann-Whitney *U* test
RQ4	Quantitative	Students at the University of Burgos	Convenience	SPSUCS Level of studies (Bachelor vs. Master)	9 weeks 1–2 weeks	Mann-Whitney *U* test
RQ5	Quantitative	Students at the University of Burgos	Convenience	Monitor ACRA (r)	While 1–2 weeks	Cross-table Cohen's Kappa coefficient
RQ6	Qualitative	Students at the University of Burgos	Convenience	SPSUCS (answers to open questions) Level of studies (bachelor vs. master)	9 weeks The whole semester	Occurrence analysis-document Atlas.ti 22
RQ7	Qualitative	Students at the University of Burgos	Convenience	Student questions in DialogFlow. Level of studies (bachelor vs. master)	While	Analysis of WWWH questions (Veenman, 2011).

Table 3
Analysis of differences (Mann-Whitney U) with respect to prior knowledge in frequency of chatbot use and learning outcomes.

	Range average		*U*	*Z*	*p*	η^2
	1 n = 46	2 n = 11				
Frequency of Chatbot use	27.32	36.05	175.00	−1.63	0.13	0.05
Learning Outcomes	25.92	41.86	111.50	−2.87	0.004*	0.15

Note. 1 = Students with low prior knowledge; 2 = Students with average prior knowledge; effect size $\eta^2 = Z^2/N\text{-}1$.

Table 4
Analysis of differences (Mann-Whitney U) with respect to degree type in frequency of chatbot use and learning outcomes.

	Range average		*U*	*Z*	*P*	η^2
	1 n = 42	2 n = 15				
Frequency of Chatbot use	25.24	39.53	157.00	−2.98	0.003*	0.16
Learning Outcomes	25.25	39.50	157.50	−2.86	0.004*	0.15

Note. 1 = Bachelor's students; 2 = Master's students; effect size $\eta^2 = Z^2/N\text{-}1$.

questions about the subject through the chatbot") and item 5 ("Asking questions via the chatbot helps me to resolve my issues") which refer to the perceived benefits dimension. According to Cohen [65], the size of the effect was small in all cases (see Table 6).

4.5. Frequency of chatbot use and level of use of metacognitive strategies

RQ5. Is the frequency of chatbot use related to the percentile score in the Metacognitive Strategies (MS) Scale?

The vast majority (84.3%) of the students were in the lower bracket (P_1–P_{49}) in perception of use of MS. Of those, 66% had a low frequency of chatbot use (0–2 times), 31.4% had a moderate frequency (3–5 times), and 2% had a high frequency (7–12 times). 7% of the students had a moderate perception of use of MS (P_{50}–P_{75}). Of those, 25% had a low frequency of chatbot use (0–2 times), while 75% had a moderate frequency (3–5 times). Finally, 8.7% of the students had a high perception of MS use (P_{76}–P_{99}). Of those, 60% had a low frequency of chatbot use (0–2 times), 20% had a moderate frequency (3–5 times), and 20% had a high frequency (7–12 times). Cohen's Kappa coefficient was K = 0.17 p = 0.04 (see Table 7).

Table 5

Analysis of differences (Mann-Whitney U) with respect to students' levels of prior knowledge (low vs. average) in perceived satisfaction with the use of the chatbot.

Dimensions	SPSUCS	Range average		U	Z	P	η^2
		1 n = 46	2 n = 11				
CPA	Item 1	29.21	28.14	243.50	−1.98	0.84	0.070
CPA	Item 2	29.08	28.68	249.50	−0.07	0.94	0.000
PB	Item 3	29.13	28.45	247.00	−0.13	0.90	0.000
PB	Item 4	28.12	32.68	212.50	−0.84	0.40	0.010
PB	Item 5	28.73	30.14	240.50	−0.26	0.80	0.001
PB	Item 6	27.87	33.73	201.00	−1.10	0.28	0.020
CHU	Item 7	28.29	31.95	220.50	−0.68	0.50	0.008
CPA	Item 8	29.14	28.41	246.00	−0.14	0.89	0.000
Motivation	Item 9	27.74	34.27	195.00	−1.21	0.23	0.030
OS	Item 10	28.11	32.73	212.00	−0.86	0.39	0.013

Note. 1 = Students with low prior knowledge; 2 = Students with average prior knowledge; CPA = Chatbot performance accuracy; PB = Perceived Benefit; CHU = Chatbot Usefulness; OS = Overall Satisfaction; effect value $\eta^2 = Z^2/N\text{-}1$.

Table 6

Analysis of the differences (Mann-Whitney U) with respect to the type of degree (bachelor's vs. master's) in perceived satisfaction with the use of the chatbot.

Dimensions	SPSUCS	Range average		U	Z	p	η^2
		1 n = 42	2 n = 15				
CPA	Item 1	26.04	37.50	190.50	−2.32	0.02*	0.10
CPA	Item 2	26.23	36.77	198.50	−2.18	0.03*	0.08
PB	Item 3	26.17	36.93	196.00	−2.18	0.03*	0.08
PB	Item 4	26.50	36.00	210.00	−1.96	0.05	0.07
PB	Item 5	26.62	35.67	215.00	−1.90	0.06	0.06
PB	Item 6	26.14	37.00	195.00	−2.23	0.03*	0.09
CHU	Item 7	25.20	39.63	155.50	−2.97	0.003*	0.16
CPA	Item 8	24.80	40.77	138.50	−3.30	0.001*	0.19
Motivation	Item 9	25.18	39.70	154.50	−3.00	0.003*	0.16
OS	Item 10	26.43	36.20	207.00	−2.03	0.04*	0.07

Note. 1 = Bachelor's students; 2 = Master's students; CPA = Chatbot performance accuracy; PB = Perceived Benefit; CHU = Chatbot Usefulness; OS = Overall Satisfaction; effect value $\eta^2 = Z^2/N\text{-}1$.

4.6. Qualitative study

4.6.1. Students' opinions (Bachelor's vs. Master's) about the chatbot and suggestions for additions or removals

RQ6. Are there differences between Bachelor's vs. Master students' suggestions of positive aspects to add or negative aspects to remove from the chatbot technology?

4.7. Positive aspects

There were differences between Bachelor's and Master's degree students about the positive aspects of using the chatbot for their learning. Undergraduate students reported making little use of the chatbot (30.83%, compared to 0% reported by Master's students). The reason given by undergraduate students was that they did not know how to use it and that they needed more information (8.26%, compared to 0% for Master's students). Likewise, Bachelor's students indicated that they preferred to use e-mail (17.62%) or class interaction (22.36%) to ask questions, compared to 0% of Master's students. Similarly, Master's students perceived the chatbot as easy

Table 7

Crosstabulation of the variables, frequency of chatbot use and percentile in perceived use of MSEs.

	Frequency of Chatbot use							
Percentile in MS	F1	%	F2	%	F3	%	Total	%
P1	32	66.6	15	31.4	1	2.0	48	84.3
P2	1	25.0	3	75.0	0	0	4	7.0
P3	3	60.0	1	20.0	1	20	5	8.7
Total	36	63.2	19	33.3	2	3.5	57	100

Note. MS percentile: low P1 = P_1–P_{49}; medium P2 = P_{50}–P_{75}; high P3 = P_{76}–P_{99}; Frequency of Chatbot use F1 = 0–2 times, F2 = 3–5 times, F3 = 7–12 times; MS = Metacognitive Strategies Scale.

to use (35.24%) compared to 6.06% for Bachelor's students. Finally, Master's students indicated that the best functionality of the chatbot was the immediacy of the response (36.78%) compared to 3.74% of undergraduates giving this response. Fig. 5 presents a Sankey diagram analysis of the positive aspects of chatbot use for undergraduate and master's students.

4.8. Aspects to add

Undergraduate students indicated that they needed more information on how the chatbot worked (18.4%) compared to 0% of Master's students. Undergraduate students also indicated that it would be useful to include an obligation to use the chatbot at least once a week (8.8%) compared to 0% of Master's students. In contrast, Master's students indicated that more questions should be included in the chatbot (48.8%) compared to 0% of Bachelor's students, and that is should include more responses (48.8%) compared to 1.6% of Bachelor's students giving this response. The undergraduate students also thought that the chatbot should be more visible in the LMS (20.4%) compared to 0% of Master's students. Fig. 6 shows a Sankey diagram of the analysis of the aspects to be included in the chatbot for Bachelor's and Master's degree students.

4.9. Aspects to be removed

The Master's students indicated that nothing needed to be removed (52.15%), compared to 14.52% of Bachelor's students giving this response. Similarly, undergraduate students indicated that they would not know what to delete (16.67%) compared to 0% of Master's students. Undergraduate students again emphasised (41.40%) that they did not know how to use the chatbot as opposed to 0% of Master's students giving this response. The undergraduate students also noted that the chatbot should be more specific for each of the elements in the LMS (7.53%) compared to 0% of Master's students giving this response. Fig. 7 shows a Sankey diagram of the analysis of the aspects to be removed from the chatbot for Bachelor's and Master's students.

4.10. Students' questions to the chatbot: analysis of the type of metacognitive strategies they apply

RQ7. What kind of metacognitive strategies (orientation, planning, evaluation and/or elaboration) do the questions asked by Bachelor's and Master's students to the chatbot relate to?

Supplementary Material 2 lists the questions asked by Bachelor's and Master's degree students in the chatbot for each subject. Undergraduate students provided 219 questions that were collected in the chatbot, almost all of which (99.09%) referred to meta-cognitive orientation strategies, and 0.91% to metacognitive planning strategies. Master's students provided 77 questions that were recorded in the chatbot. The vast majority (97.40%) referred to metacognitive orientation strategies, and 2.59% were about meta-cognitive planning strategies.
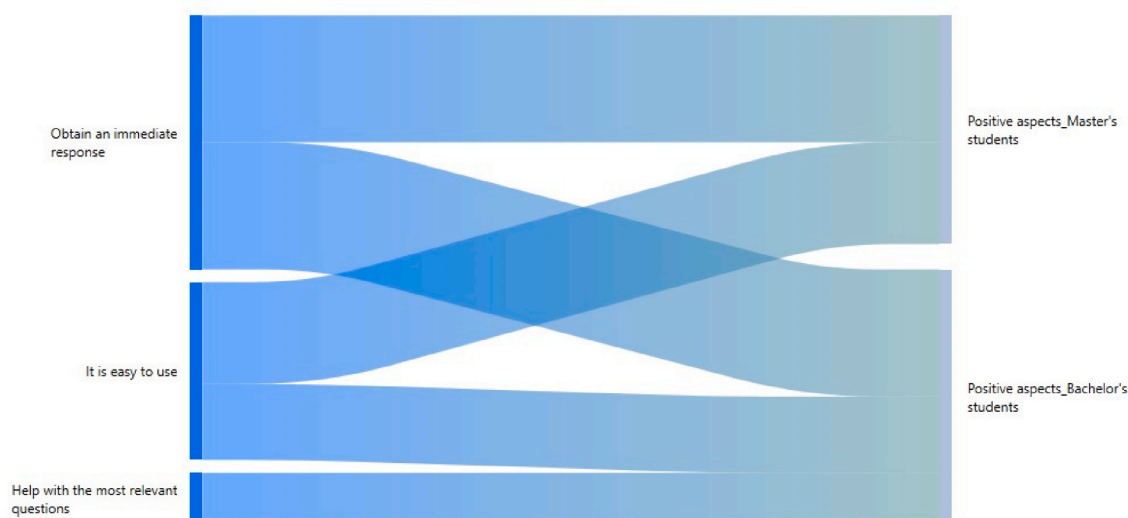

Fig. 5. Sankey plot of the categorisation of the answers given to the open question "positive aspects of chatbot use"
(Note to editor: Fig. 4 in case of publication of the manuscript should be in colour)... (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)
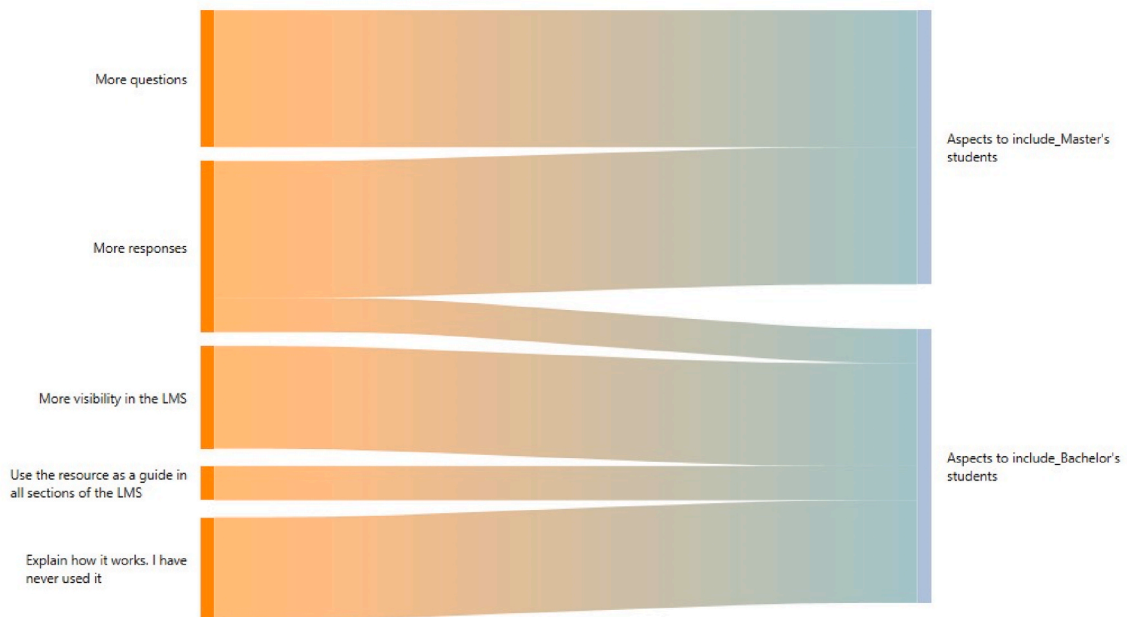
Fig. 6. Sankey diagram of the categorisation of the answers given to the open question "aspects to be included in the chatbot"
(Note to editor: Fig. 5 in case of publication of the manuscript should be in colour)... (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)
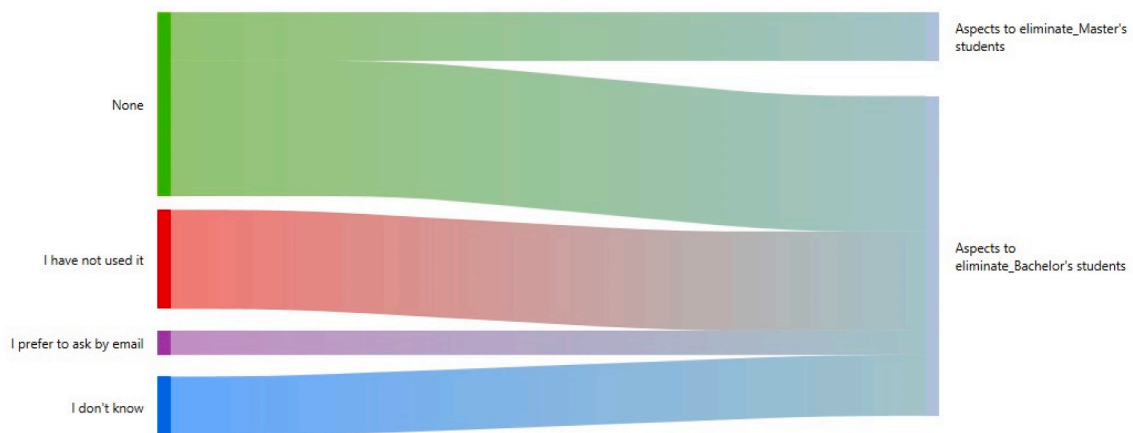

Fig. 7. Sankey diagram of the categorisation of the answers given to the open question "aspects to eliminate in the chatbot"
(Note to editor: Fig. 6 in case of publication of the manuscript should be in colour)... (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

5. Discussion

We found an influence of the students' prior knowledge variable on learning outcomes, with students with greater prior knowledge having better learning outcomes, but no influence on the frequency of chatbot use. There were also significant differences between Bachelor's and Master's students in the learning outcomes and in the frequency of chatbot use, with Master's students scoring higher in both. These results confirm the findings from other studies regarding the importance of prior knowledge on learning outcomes [15]. However, the lack of influence of that knowledge on the use of the chatbot can be explained by the need to improve the feedback provided by the conversational agent [45] by increasing the types of questions and answers. These question-answer pairs could be tailored to different levels of prior knowledge [1,50,56].

There were also significant differences in students' perceived satisfaction with the use of the chatbot in all dimensions, with Master's students scoring higher, except for two items (4 and 5) in the "Perceived Benefit" dimension where there were no differences,

which referred to resolving the most important issues and the chatbot's help in clarifying students' questions. Furthermore, there were no significant differences in students' perceived satisfaction with the use of the chatbot with respect to levels of prior knowledge. These results indicate the need for the design of the chatbot to differentiate between types of users. It seems that students who have already graduated, such as Master's students, are more mature in approaching subject knowledge and probably also have more technical skills, which means they are likely to use the tool differently and also value its function and potential, which is in line with what [2] reported.

Another important aspect we noted was that 84.3% of the students perceived low use of metacognitive strategies (P_1–P_{49}). Within this group, 66.6% exhibited a low frequency of chatbot use (0–2 times). Likewise, 7% of the students reported moderate use of MS (P_{50}–P_{75}), and within these, 75% exhibited medium use of the chatbot. This opens the debate on students' actual use of metacognitive strategies and the relationship with using technological feedback tools such as chatbots. Perhaps there needs to be training in the use of the chatbot from the WWWH instructional approach proposed by Veenman [22] in order to increase its use and effectiveness with respect to the use of more complex metacognitive strategies. Along these lines, it is possible that using an off-line MS measurement instrument may be distorting the results. Future research will also include online assessment methods for metacognitive strategy use.

The results of the qualitative study reaffirmed the results of the quantitative study. The undergraduate students reported less use of the chatbot. The reasons for this were that many of these students did not know how to use the chatbot, something that no Master's students reported. Bachelor's students said that they preferred to use email or face-to-face interaction with the teacher. Similarly, Master's students indicated the benefits of the chatbot and that its greatest functionality was the possibility of making real-time queries, aspects that were hardly valued by Bachelor's students. In terms of suggested additions, the undergraduate students indicated a need for more information on how to use the chatbot and to include an obligation to make at least one weekly query in the chatbot, aspects that were not reported by the Master's students. The Master's students noted a need to include more questions and answers in the chatbot; this aspect was given low consideration by Bachelor's degree students. On similar lines, the undergraduate students indicated that the chatbot should be more specific and that it should appear in the virtual learning platform not in a general way, but specifically every time the student consulted a specific aspect. These results support [2] regarding the weight of students' technological skills for effective use of the chatbot. All these contributions underscore the importance of considering users' suggestions and improving chatbot functionality [41].

While there seem to have been differences between the profiles of Bachelor's and Master's students, these differences were not completely conclusive or generalisable. Therefore, other variables may have an influence, such as learning history, learning profiles, and students' task resolution patterns for different tasks, in line with findings from Binali et al. [47]. This highlights the need to consider the pedagogical design of the chatbot [48], the different levels of prior knowledge [15] and student learning profiles [49]. Future studies will explore these aspects.

It is also important to bear in mind that the questions that both types of students asked in the chatbot were mainly metacognitive orientation questions (focused on declarative knowledge) and to a lesser extent planning questions. This opens up the question of why students did not use metacognitive evaluation or elaboration questions (procedural knowledge) in the chatbot. The answer probably lies in the design of the chatbot itself and/or the novelty of this tool. Future studies will explore this question further.

Finally, the results confirm the importance of using a mixed research methodology, as it provides a more complete test spectrum that broadens the researcher's knowledge and illuminates future research [61,62].

6. Conclusions

Although this is a complex topic that requires further research, the results can be understood from study approaches that advocate analysing metacognitive strategy use over time [30,31]. Similarly, learners have been found to need adaptation to use ALT tools and that specific scaffolding is likely to be required for each phase of the learning process [27–29]. The use of novel technological resources seems to be affected by the technological competences of the students, who identify these challenges themselves [2,35,36], which means future studies need to include more thorough training plans. On the other hand, the hypothesis that prior knowledge would have more weight on the frequency of chatbot use was not confirmed. This may be due to the type of statistics applied, since due to the sample characteristics, we used non-parametric statistics and it was not possible to apply multivariate parametric tests. Therefore, in future studies, the sample will be expanded and these multivariate analysis techniques will be applied.

Another important aspect is students' perceptions of the use of metacognitive strategies, with a high proportion of students demonstrating low percentiles. This result opens up a new research focus related to the use of metacognitive strategies in university students. On the one hand, there is the dilemma raised by Veenman and colleagues [21,22] about the differences in measuring metacognitive strategies between online and offline methods. In this study, both methods were applied—an off-line instrument, the ACRA (r) scale [23], and an on-line method, analysing the questions asked by the students in the chatbot—and no differences were found, since in both cases low use of complex metacognitive strategies was recorded in most cases. This may be related to other variables, such as students' habituation to the use of chatbot technology. Therefore, in future research, in addition to recording the queries made through the chatbot, we will also analyse the queries made by students through other channels (e.g. emails, face-to-face questions) in order to confirm the types of metacognitive strategies through online analysis of methods.

In terms of satisfaction using the chatbot, students judged that the chatbot helped them to focus their questions about the conceptual content of the subject. This perception in itself is promising, as it confirms that the chatbot helps students to reflect on their own practice. This reflection enabled the students to focus on the elements of the chatbot that, in their opinion, did not work or that needed to be improved [32].

7. Limitations and future lines of research

The results of this should be considered with caution because the sample was obtained through convenience sampling and covered students from a specific knowledge branch—health sciences—at a specific university. Therefore, in subsequent studies, more knowledge branches and more universities will be examined.

In addition, the need to improve the feedback provided by the chatbot was apparent. The key lies in improving the question-answer database [49,56]. In this regard, the instructional design of the chatbot should be improved to include more questions that encourage the deployment of more complex metacognitive strategies [46]. One possibility would be to create specific chatbots for each of the resources included in the LMS along the lines of the studies by Binali et al. [47]; Gupta et al. [48]; and Kinnebrew et al. [46] with different levels of difficulty depending on students' prior knowledge [15].

However, this functionality means more technological and computational complexity (adding more sophisticated AI resources) that needs to be addressed through interdisciplinary work (instructional psychologists and IT and computer professionals) [34–36]. Future chatbots will include improvements in this regard [40,41].

Another notable aspect is the need to improve how students' use of the chatbot is monitored throughout the semester. In this study there was weekly monitoring, but perhaps there would need to be daily monitoring and analysis of changes over smaller timescales [50].

Finally, it is worth emphasizing positive aspects of this work, such as the analysis of multiple variables: prior knowledge, meta-cognitive strategies measured with online and offline methods, micro-analytical analysis of students' satisfaction with the use of the chatbot, and educational level. This analysis made it possible to identify important aspects affecting future research along the lines indicated by Huang & Chueh [41]. However, the inclusion of chatbots in LMSs is still in its infancy and there is a long way to go for research in this field, which looks promising [2].

Future studies in this area face two significant challenges. One is the technological improvements that will provide better usability in common education contexts. This should include conversational assistants with human-interaction structures, something that will no doubt help improve student interactions. Along similar lines, more use should also be made of artificial intelligence techniques applied to improving the conversational thread of the student-machine interaction. Another important aspect in future research will be training teachers in the use of this technology.

Credit author statement

María Consuelo Sáiz-Manzanares: Conceptualization, Investigation, Data curation, Methodology, Formal analysis, Resources, Visualization, Validation, Writing original draft, Writing - review & editing; **Raúl Marticorena-Sánchez:** Conceptualization, Visualization, Software, Writing - review & editing; **Luis Jorge Martín-Antón**: Methodology, Formal analysis, Validation, Writing - review & editing; **Irene González Díez:** Conceptualization, Writing - review & editing; **Leandro Almeida:** Methodology, Formal analysis, Validation, Writing - review & editing.

Funding statement

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix B. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.heliyon.2023.e12843.

References

[1] M.C. Sáiz-Manzanares, R. Marticorena-Sánchez, J.J. Rodríguez-Díez, S. Rodríguez-Arribas, J.F. Díez-Pastor, Y.P. Ji, Improve teaching with modalities and collaborative groups in an LMS: an analysis of monitoring using visualisation techniques, J. Comput. High Educ. 33 (2021) 747–778, https://doi.org/10.1007/s12528-021-09289-9.

[2] O. Tsivitanidou, A. Ioannou, Envisioned pedagogical uses of chatbots in higher education and perceived benefits and challenges, in: P. Zaphiris, A. Ioannou (Eds.), Learning and Collaboration Technologies: Games and Virtual Environments for Learning. HCII 2021, Lecture Notes in Computer Science, 12785, Springer, Cham, 2021, pp. 230–250, https://doi.org/10.1007/978-3-030-77943-6_15.

[3] M. Hernandez-de-Menendez, C. Escobar Díaz, R. Morales-Menendez, Technologies for the future of learning: state of the art, Int. J. Interact. Des. Manuf. 14 (2) (2019) 683–695, https://doi.org/10.1007/s12008-019-00640-0.

[4] J.P. Bakken, V.L. Uskov, S.V. Kuppili, A.V. Uskov, N. Golla, N. Rayala, Smart Universities 70 (2018), https://doi.org/10.1007/978-3-319-59454-5.

[5] D. Rutkauskiene, D. Gudoniene, R. Maskeliunas, T. Blazauskas, The gamification model for E-learning participants engagement, in: V.L. Uskov, et al. (Eds.), Smart Education and E-Learning 2016, Smart Innovation, Systems and Technologies, vol. 59, 2016, pp. 291–301, https://doi.org/10.1007/978-3-319-39690-3_26.

[6] D.S. Brogan, D. Basu, V.K. Lohani, A virtual learning system in environmental monitoring, in: M. Auer, K.S. Kim (Eds.), Engineering Education for a Smart Society. GEDC 2016, WEEF 2016. Advances in Intelligent Systems and Computing vol. 627, Springer, Cham, 2018, pp. 352–367.

[7] M.C. Sáiz-Manzanares, C.I. García Osorio, J.F. Díez-Pastor, L.J. Martín-Antón, Will personalized e-Learning increase deep learning in higher education? Information Discovery and Delivery 47 (1) (2019) 53–63, https://doi.org/10.1108/IDD-08-2018-0039.

[8] M. Schophuizen, K. Kreijns, S. Stoyanov, M. Kalz, Eliciting the challenges and opportunities organizations face when delivering open online education: a group-concept mapping study, Internet High Educ. 36 (2018) 1–12, https://doi.org/10.1016/j.iheduc.2017.08.002.

[9] H.B. Essel, D. Vlachopoulos, A. Tachie-Menson, et al., The impact of a virtual teaching assistant (chatbot) on students' learning in Ghanaian higher education, Int J Educ Technol High Educ 19 (2022) 57, https://doi.org/10.1186/s41239-022-00362-6.

[10] M.E. Auer, K.-S. Kim, Advances in Intelligent Systems and Computing Engineering Education for a Smart Society, Springer, Cheonan, Korea, 2016, https://doi.org/10.1007/978-3-319-60937-9.

[11] M.C. Sáiz-Manzanares, J.R. Casanova, J.A. Lencastre, L. Almeida, L.J. Martín-Antón, Student satisfaction with online teaching in times of COVID-19, Comunicar 30 (70) (2022) 31–40, https://doi.org/10.3916/C70-2022-03.

[12] M. Taub, R. Sawyer, J. Lester, R. Azevedo, The impact of contextualized emotions on self-regulated learning and scientific reasoning during learning with a game-based learning environment, Int. J. Artif. Intell. Educ. 30 (1) (2020) 97–120, https://doi.org/10.1007/s40593-019-00191-1.

[13] M. Taub, R. Azevedo, R. Rajendran, E.B. Cloude, G. Biswas, M.J. Price, How are students' emotions related to the accuracy of cognitive and metacognitive processes during learning with an intelligent tutoring system? Learn. InStruct. 72 (2021), 101200 https://doi.org/10.1016/j.learninstruc.2019.04.001.

[14] B.J. Zimmerman, D.H. Schunk, Self-regulated learning and performance: an introduction and an overview, in: B.J. Zimmerman, D. H Schunk (Eds.), Handbook of Self-Regulation of Learning and Performance, Routledge/Taylor & Francis Group, 2011, pp. 1–12.

[15] M. Taub, R. Azevedo, How does prior knowledge influence eye fixations and sequences of cognitive and metacognitive SRL processes during learning with an intelligent tutoring system? Int. J. Artif. Intell. Educ. 29 (1) (2019) 1–28, https://doi.org/10.1007/s40593-018-0165-4.

[16] J.H. Flavell, Cognitive Development, second ed., Prentice-Hall, Inc, 1985.

[17] M.V.J. Veenman, J.J. Beishuizen, Intellectual and metacognitive skills of novices while studying texts under conditions of text difficulty and time constraint, Learn. InStruct. 14 (6) (2004) 621–640, https://doi.org/10.1016/j.learninstruc.2004.09.004.

[18] M.V.J. Veenman, P. Wilhelm, J.J. Beishuizen, The relation between intellectual and metacognitive skills from a developmental perspective, Learn. InStruct. 14 (1) (2004) 89–109, https://doi.org/10.1016/j.learninstruc.2003.10.004.

[19] A.L. Brown, J.S. DeLoache, Skills, plans, and self-regulation, in: R.S. Siegler (Ed.), Children's Thinking: what Develops?, Lawrence Erlbaum Associates, Inc, 1978, pp. 3–35.

[20] G.L.M. Schellings, B.H.A.M. Van Hout-Wolters, M.V.J. Veenman, J. Meijer, Assessing metacognitive activities: the in-depth comparison of a task-specific questionnaire with think-aloud protocols, Eur. J. Psychol. Educ. 28 (3) (2013) 963–990, https://doi.org/10.1007/s10212-012-0149-y.

[21] M. Van Der Stel, M.V.J. Veenman, Metacognitive skills and intellectual ability of young adolescents: a longitudinal study from a developmental perspective, Eur. J. Psychol. Educ. 29 (1) (2014) 117–137, https://doi.org/10.1007/s10212-013-0190-5.

[22] M.V.J. Veenman, Learning to self-monitor and self-regulate, in: R. Mayer, P. Alexander (Eds.), Handbook of Research on Learning and Instruction, Routledge, New York, 2011, pp. 197–218, https://doi.org/10.4324/9780203839089.ch10.

[23] J.M. Román, L. Poggioli, ACRA (r): Escalas de Estrategias de Aprendizaje [Learning Strategies Scales], UCAB Publications (Postgraduate Doctorate in Education), 2013.

[24] M.V.J. Veenman, B.H.A.M. Van Hout-Wolters, P. Afflerbach, Metacognition and learning: conceptual and methodological considerations, Metacognition and Learning 1 (1) (2006) 3–14, https://doi.org/10.1007/s11409-006-6893-0.

[25] N. Reoyo, M.Á. Carbonero, L.J. Martín-Antón, Characteristics of teaching effectiveness from the perspectives of teachers and future secondary school teachers, Rev. Educ. 376 (2017) 62–84, https://doi.org/10.4438/1988-592X-RE-2017-376-344.

[26] R. Azevedo, A Johnson, A. Chauncey, A. Graesser, Use of hypermedia to assess and convey self-regulated learning, in: D.H. Schunk, B. Zimmerman (Eds.), Handbook of Self-Regulation of Learning and Performance, Routledge, New York, 2011, pp. 102–121, https://doi.org/10.4324/9780203839010.

[27] R. Azevedo, Issues in dealing with sequential and temporal characteristics of self- and socially-regulated learning, Metacognition and Learning 9 (2) (2014) 217–228, https://doi.org/10.1007/s11409-014-9123-1.

[28] R. Azevedo, D.C. Moos, A.M. Johnson, A.D. Chauncey, Measuring cognitive and metacognitive regulatory processes during hypermedia learning: issues and challenges, Educ. Psychol. 45 (4) (2010) 210–223, https://doi.org/10.1080/00461520.2010.515934.

[29] A.F. Hadwin, Commentary and future directions: what can multi-modal data reveal about temporal and adaptive processes in self-regulated learning? Learn. InStruct. 72 (2021), 101287 https://doi.org/10.1016/j.learninstruc.2019.101287.

[30] Y. Zhang, L. Paquette, N. Bosch, J. Ocumpaugh, G. Biswas, S. Hutt, R.S. Baker, The evolution of metacognitive strategy use in an open-ended learning environment: do prior domain knowledge and motivation play a role? Contemp. Educ. Psychol. 69 (2022), 102064 https://doi.org/10.1016/j.cedpsych.2022.102064.

[31] I. Molenaar, S. Järvelä, Sequential and temporal characteristics of self and socially regulated learning, Metacognition and Learning 9 (2) (2014) 75–85, https://doi.org/10.1007/s11409-014-9114-2.

[32] V. Cabales, Muse: scaffolding metacognitive reflection in design-based research, in: Proceedings of CHI '19: CHI Conference on Human Factors in Computing Systems (CHI '19), 2019 (Glasgow, Scotland UK. ACM, New York, NY, USA). (Accessed 4 May 2019).

[33] C. Liu, M. Liao, C. Chang, H. Lin, An analysis of children ' interaction with an AI chatbot and its impact on their interest in reading, Comput. Educ. 189 (300) (2022), 104576, https://doi.org/10.1016/j.compedu.2022.104576.

[34] J. Ochoa-Orihuel, R. Marticorena-Sanchez, M.C. Saiz-Manzanares, Moodle LMS integration with Amazon alexa: a practical experience, Appl. Sci. 10 (6859) (2020) 1–21, https://doi.org/10.3390/app10196859.

[35] M.C. Sáiz-Manzanares, R. Marticorena-Sánchez, J. Ochoa-Orihuel, Effectiveness of using voice assistants in learning: a study at the time of covid-19, Int. J. Environ. Res. Publ. Health 17 (15) (2020) 1–20, https://doi.org/10.3390/ijerph17155618.

[36] N. Stathakarou, S. Nifakos, K. Karlgren, S.T. Konstantinidis, P.D. Bamidis, C.S. Pattichis, N. Davoody, Students' perceptions on chatbots' potential and design characteristics in healthcare education, Stud. Health Technol. Inf. 272 (2020) 209–212, https://doi.org/10.3233/SHTI200531.

[37] D. Griol, Z. Callejas, A neural network approach to intention modeling for user-adapted conversational agents, Comput. Intell. Neurosci. (2016) 8402127, https://doi.org/10.1155/2016/8402127.

[38] A. Söderström, A. Shatte, M. Fuller-Tyszkiewicz, Can intelligent agents improve data quality in online questiosnnaires? A pilot study, Behav. Res. Methods 53 (2021) 2238–2251, https://doi.org/10.3758/s13428-021-01574-w.

[39] X.L. Pham, T. Pham, Q.M. Nguyen, T.H. Nguyen, T.T.H. Cao, Chatbot as an intelligent personal assistant for mobile language learning, in: Proceedings of the 2018 2nd International Conference on Education and E-Learning, 2018, pp. 16–21, https://doi.org/10.1145/3291078.3291115, 10.1145/3291078.3291115.

[40] U. Hemavathi, A.C.V. Medona, AI-based interactive agent for health care using NLP and deep learning, in: A. Joshi, M. Mahmud, R.G. Ragel (Eds.), Information and Communication Technology for Competitive Strategies (ICTCS 2021), Springer Nature Singapore, 2023, pp. 11–18.

[41] D.H. Huang, H.E. Chueh, Chatbot usage intention analysis: veterinary consultation, J. Innovat. Knowl. 6 (3) (2021) 135–144, https://doi.org/10.1016/j.jik.2020.09.002.

[42] P. Parmar, J. Ryu, S. Pandya, J. Sedoc, S. Agarwal, Health-focused conversational agents in person-centered care: a review of apps, Npj Dig. Med. 5 (21) (2022), https://doi.org/10.1038/s41746-022-00560-6.

[43] T. Kita, C. Nagaoka, N. Hiraoka, M. Dougiamas, Implementation of voice user interfaces to enhance users' activities on Moodle, in: Proceedings of 2019 4th International Conference on Information Technology: Encompassing Intelligent Technology and Innovation Towards the New Era of Human Life, InCIT, 2019, 2019, pp. 104–107, https://doi.org/10.1109/INCIT.2019.8912086.

[44] S. Gupta, Y. Chen, Supporting inclusive learning using chatbots? A chatbot-led interview study, J. Inf. Syst. Educ. 33 (1) (2022) 98–108. https://aisel.aisnet.org/jise/vol33/iss1/11.

[45] J.R. Segedy, J.S. Kinnebrew, G. Biswas, The effect of contextualized conversational feedback in a complex open-ended learning environment, Educ. Technol. Res. Dev. 61 (1) (2013) 71–89, https://doi.org/10.1007/s11423-012-9275-0.

[46] J.S. Kinnebrew, J.R. Segedy, G. Biswas, Integrating model-driven and data-driven techniques for analyzing learning behaviors in open-ended learning environments, IEEE Trans. Learn. Technol. 10 (2) (2017) 140–153, https://doi.org/10.1109/TLT.2015.2513387.

[47] T. Binali, C.C. Tsai, H.Y. Chang, University students' profiles of online learning and their relation to online metacognitive regulation and internet-specific epistemic justification, Comput. Educ. 175 (2021), 104315, https://doi.org/10.1016/j.compedu.2021.104315.

[48] S. Gupta, K. Jagannath, N. Aggarwal, R. Sridar, S. Wilde, Y. Chen, Artificially Intelligent (AI) Tutors in the Classroom: A Need Assessment Study of Designing Chatbots to Support Student Learning, vol. 213, 2019. PACIS 2019 Proceedings, https://aisel.aisnet.org/pacis2019/213.

[49] G. Biswas, J.R. Segedy, K. Bunchongchit, From design to implementation to practice a learning by teaching system: betty's brain, Int. J. Artif. Intell. Educ. 26 (1) (2016) 350–364, https://doi.org/10.1007/s40593-015-0057-9.

[50] T.V. Dobudko, A.V. Ochepovsky, S.V. Gorbatov, W. Hashim, A. Maseleno, Functional monitoring and control in electronic information and educational environment, Int. J. Recent Technol. Eng. 8 (2) (2019) 1383–1386, https://doi.org/10.35940/ijrte.B2030.078219.

[51] J. Weizenbaum, ELIZA - a computer program for the study of natural language communication between man and machine, Commun. ACM 9 (1966) 36–45.

[52] A.M. Turing, Computing machinery and intelligence, Mind ume LIX (236) (1950) 433–460, https://doi.org/10.1093/mind/LIX.236.433.

[53] P. Smutny, P. Schreiberova, Chatbots for learning: a review of educational chatbots for the Facebook Messenger, Comput. Educ. 151 (2020), 103862, https://doi.org/10.1016/j.compedu.2020.103862.

[54] A.S. Tulshan, S.N. Dhage, Survey on virtual assistant: Google assistant, Siri, cortana, alexa, in: S. Thampi, O. Marques, S. Krishnan, K.C. Li, D. Ciuonzo, M. Kolekar (Eds.), Advances in Signal Processing and Intelligent Recognition Systems. SIRS 2018, Communications in Computer and Information Science, vol. 968, Springer, Singapore, 2019, pp. 190–201, https://doi.org/10.1007/978-981-13-5758-9_17.

[55] M. Melton, J. Fenwick, Alexa skill voice interface for the Moodle learning management system, J. Comput. Sci. Coll. 35 (4) (2019) 26–35.

[56] S. Han, M.K. Lee, FAQ chatbot and inclusive learning in massive open online courses, Comput. Educ. 179 (2022), 104395, https://doi.org/10.1016/j.compedu.2021.104395.

[57] M. Fidan, N. Gencel, Supporting the instructional videos with chatbot and peer feedback mechanisms in online learning: the effects on learning performance and intrinsic motivation, J. Educ. Comput. Res. 60 (7) (2022) 1716–1741, https://doi.org/10.1177/07356331221077901.

[58] K.F. Hew, W. Huang, J. Du, D. Jia, Using chatbots to support student goal setting and social presence in fully online activities: learner engagement and perceptions, J. Comput. High Educ. (2022), https://doi.org/10.1007/s12528-022-09338-x.

[59] R.Y.P. Ji, R. Marticorena-Sánchez, C. Pardo-Aguilar, C. López-Nozal, M. Juez-Gil, Activity and dropout tracking in Moodle using UBUMonitor application. IEEE, Revista Iberoamericana de Tecnologias del Aprendizaje 17 (3) (2022) 307–317, https://doi.org/10.1109/RITA.2022.3191279.

[60] R. Marticorena-Sánchez, C. López-Nozal, Y.P. Ji, C. Pardo-Aguilar, Á. Arnaiz-González, UBUMonitor: an open-source desktop application for visual E-learning analysis with Moodle, Electronics 11 (6) (2022) 954, https://doi.org/10.3390/ELECTRONICS1106095.

[61] B. Birgili, Ö. Demir, An explanatory sequential mixed-method research on the full-scale implementation of flipped learning in the first years of the world's first fully flipped university: departmental differences, Comput. Educ. 176 (2022), 104352, https://doi.org/10.1016/j.compedu.2021.104352.

[62] M. Bond, Facilitating student engagement through the flipped learning approach in K-12: a systematic review, Comput. Educ. 151 (2020), 103819, https://doi.org/10.1016/j.compedu.2020.103819.

[63] D.F. Campbell, J. Stanley, *Experimental And Quasi-Experimental Designs In Social Research* (9a reimp.) [original Work Published Experimental and Quasi-Experimental Design for Research in 1966, Amorrortu, Buenos Aires, 2005.

[64] U. Flick, Designing Qualitative Research, SAGE, 2011, https://doi.org/10.4135/9781849208826.

[65] J. Cohen, Statistical power analysis, Curr. Dir. Psychol. Sci. 1 (1992) 98–101, https://doi.org/10.1111/1467-8721.ep10768783.

[66] I.B.M. Corporation, Statistical Package for the Social Sciences, IBM, 2022 (SPSS) (Version 28) [Software], https://www.ibm.com/es-es/products/spss-statistics.

[67] Atlas ti, Software Package Qualitative Data Analysis (Atlas.Ti), 2021 [Software]. Atlas.ti, version 22. https://atlasti.com/es.