Conversational Agents in Distance Education: Comparing Mood States with Students' Perception

Aliane Loureiro Krassmann, Fábio Josende Paz, Clóvis Silveira, Liane Margarida Rockenbach Tarouco, Magda Bercht

Programa de Pós-Graduação em Informática na Educação (PPGIE), Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, Brazil
Email: alkrassmann@gmail.com

Abstract

One of the advantages of a Conversational Agent is that it stimulates social presence among Distance Education students, providing continuous psychological and pedagogical support. However, the interaction between students and virtual agents is often functionally ineffective. In this regard, this study aims to investigate students' mood states, inferred by chat log analysis of interactions between students and a Conversational Agent, relating such mood states with the students' perceptions about the tool. The study was conducted with a sample of students attending a Distance Education post-graduation course. The results showed the relationship between interest, utility and satisfaction perceived by the students with mood states detected by the log analysis. Also, the underutilization of the Conversational Agents was diagnosed, and improvements to be implemented so that negative mood states which can be overcome were identified. The research brings contributions on building better Conversational Agents for educational purposes.

Keywords

Conversational Agents, Distance Education, Students' Perception, Mood States, Affective Computing

1. Introduction

One of the biggest technological trends nowadays is the intelligent virtual assistants (e.g. Apple's Siri, Microsoft's Cortana, Amazon's Alexa and Google's Assistant), also called chatbots, Conversational Agents (CA), conversational enti-

ties, among other names, which are softwares used for various practical purposes, for instance, assistance and information acquisition by conversing with a machine in a dialogic fashion, using natural language (Dale, 2016). It can complement or even replace traditional information, communication and sales channels like newsletters, websites, sales desks or hotlines (Zumstein & Hundertmark, 2017).

The use of CA can support education in a variety of ways, for example, enabling greater interactivity, facilitating sociability and increasing the level of use of Virtual Learning Environments (VLE) (Kang, Nah, & Tan, 2012; Abushawar & Atwell, 2015), making it more interactive and dynamic (Griol & Callejas, 2013), and allowing instant retrieval of information without the student having to search or browse multiple web pages in order to look for answers to Frequently Asked Questions (FAQ) (Ghose & Barua, 2013). Xie & Luo (2017) argue that CA can improve the individual's skills, promote task completion and users' satisfaction by providing immediate assistance. Fryer et al. (2017) stress that they are a potential source of motivation for sustained communication in learning.

Distance Education (DE) is a modality of education in which the mediation occurs mostly online, with the support of digital technologies. It is the fastest growing educational modality in the world (Online Learning Consortium, 2017), mainly due to the flexibility of time and space that it provides to the student. However, authors such as Leonhardt et al. (2007), Heuvelman-Hutchinson (2012) and Rush (2015) emphasize that one of the main difficulties of DE students is the feeling of isolation that they experience, caused not only by the lack of face-to-face contact with the teacher, but also because the times they access the VLE may not be synchronous with other users' times. CA appear as an alternative to this impasse, since, according to Griol & Callejas (2013), they give the student the sensation of interacting with another user, which would be a service equivalent to having full-time tutoring. However, although CA are not new to most of people, just a minority is using them regularly and intensively (Zumstein & Hundertmark, 2017). We here ask: why?

According to Jaques & Vicari (2007), in order for a VLE to interact effectively with the users, it must recognize their emotion to respond to them appropriately. Picard et al. (2004) emphasize that emotional awareness, that is, being aware of one's affective state, can be instrumental in helping to deal with that state productively. In this sense, Danilava et al. (2012) argue that long-term interaction with a CA depends on the user's continuous motivation to interact with it, and such interaction is influenced by trust, sympathy, positive emotional bond and/or utility. However, as Mou & Xu (2017) well highlight, humans may not be able to find appropriate motivation to develop social relationships with machines.

The aforementioned studies are inside the Affective Computing area, which, according to Picard (2003), emphasizes the need for a balance, that is, machines are not meant to seem "emotional", but effective, in the sense of knowing the

appropriated time to analyze what the user is feeling. Thus, we consider that the investigation of emotional aspects among users of CA can be a remarkable factor to improve such interactions.

This study was conducted around the hypothesis that the students' emotional states when interacting with a CA can impact on the quality of the conversation and, mainly, on the user's perception about the tool in terms of interest, utility and satisfaction. To investigate it, DE students' chat logs are analyzed and, by means of a questionnaire, compared with their personal opinions about the experience of using a CA.

2. Background and Related Work

Natural language processing dates back to the 1960s, with the emergence of chatbot Eliza, which simulated a psychoanalyst in conversation with patients (Weizenbaum, 1966). It became more popular in the 1990s with the launching of the inference machine A.L.I.C.E. (Artificial Linguistic Internet Computer Entity), an open source project that until nowadays promotes the dissemination of the Artificial Intelligence Markup Language (AIML) (Wallace, 1995). It won the Loebner Prize (Turing test) in 2000, 2001 and 2004.

Among the Conversational Agents with educational purposes developed with AIML technology, to name a few, are: Doroty, which trains users on computer networks administration (Leonhardt et al., 2005); Blaze, created to improve user's cognitive skills in resolving mathematical problems (Aguiar, Tarouco, & Reategui, 2014); Geranium, used as a tool for learning about urban ecosystem (Griol & Callejas, 2013); and Mentor Chat, developed for collaborative language learning (Tegos et al., 2014).

The aforementioned CA run on VLEs hosted on the web, with 2D interfaces and/or visually represented by the bust or just the head. Nowadays, it is also possible to find them embodied in 3D immersive VLE, such as Virtual Worlds (VW). An example of a VW CA is Atena (translated from Portuguese, acronym for: Tutor Agent for Teaching and Navigating the Environment) (Figure 1): an NPC (Non-player Character), i.e. an automated avatar that accompanies the student in a journey through the 3D scenario. Atena's knowledge base is on the teaching of Physics (Krassmann et al., 2017).

Despite the mentioned benefits and the flexibility of environments and systems in which CA can be integrated, there are still communication difficulties

Figure 1. Student avatar interacting with chatbot Atena in a 3D VW (Source: the authors).

between agents and humans, due to several factors, such as limitations in natural language recognition capacity, difficulties in stimulating dialogue continuation (Leonhardt et al., 2007), and lack of control of repeated sentences and treatment of unknown sentences (Neves et al., 2006). Fryer et al. (2017) also point out that there is a difficulty in maintaining interest in the tool after the "novelty effect" dissipates.

Besides that, Savin-Baden et al. (2015) suggest that the greater the emotional engagement between the user and the CA, the more positive will be the experience. However, the study of Mou & Xu (2017) showed that people use different communication strategies in human-machine communication; when interacting with a machine, some may feel more confident while others may feel confused and even intimidated. Burden (2009) emphasizes that CA have typically limited ways to express emotion, which might result in less acuity in the overall emotion analysis. Hill et al. (2015) observed that people use more positive emotion words when communicating with another person as opposed to a CA. The authors found that messages to chatbots contained fewer words per message, more negative emotion and sexual words. Those assumptions suggest that for a CA it is even more difficult to understand students' intents.

From the earliest times, human relationships are permeated by affective states and feelings. Scherer (2005) argues that emotional states can include a set of phenomena with different origins, intensity, duration, and bodily reactions. Discoveries in neuroscience have revealed that affect and cognition are appropriately integrated with one another (Picard et al., 2004). Affective phenomena contribute to regulating and guiding attention, helping humans select next moves away from negative or harmful choices (Picard, 2003).

Studies in the area of Affective Computing have shown that it is possible to recognize the mood states of a student in a VLE by means of a model to correlate variables that can influence it (e.g. personality traits, motivational factors, and affective subjectivity identified in texts), which can be used to assist teachers and promote better teaching practices (Longhi et al., 2012).

In view of the wide range of human emotional states and feelings, it is necessary to narrow the scope for a more accurate vision. Tran (2004) considers that the analysis of "mood" rather than "emotion" is more convenient, since it is more representative of daily commonplace feelings, and therefore easier to measure. Davidson also states that emotion influences behavior, while mood influences cognition. Scherer (2005) defines emotion as an occasional phenomenon, with high intensity and brief duration, being characterized as a dynamic process, while the mood is diffuse, with low intensity and long lasting. Considering these characteristics, in this research it was decided to analyze the mood of the user interacting with the CA instead of the emotion.

Regarding the analysis of mood, Tran (2004) introduced the Geneva Emotion Wheel (GEW), with the dimensions Satisfied/Dissatisfied and Enthusiastic/Unenthusiastic, organized along two appraisal criteria: Pleasantness/Unpleasantness and Low Control/High Control (Figure 2) in a circular
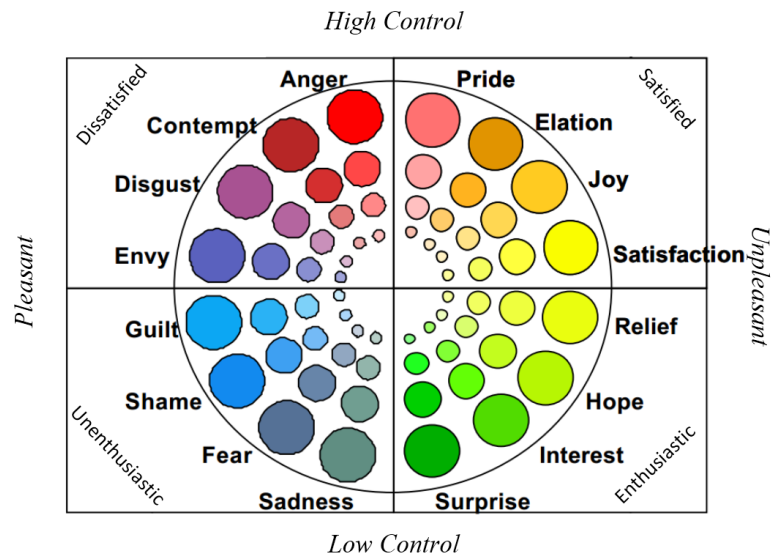
Figure 2. Geneva emotion wheel (Source: Tran, 2004).

form, each mood state with its four levels of intensity, forming a radiant. This is the model that was used as the basis for the representation of students' mood states in the present study.

Considering the expression of affect manifested by means of texts (natural language), Scherer (2005) proposed 36 categories, indexing a series of adjectives and nouns that denote an affective phenomenon. Neviarouskaya et al. (2010) proposed the @AM (Attitude Analysis Model) system, which classifies sentences according to fine-grained attitude labels (nine affect categories (Izard, 1971)): anger, disgust, fear, guilt, interest, joy, sadness, shame, surprise, using the original version of SentiFul database (created by the authors), which contains (in English) sentiment conveying adjectives, adverbs, nouns, and verbs.

Although we cannot accurately recognize an individual's mood state by means of words, as they can be considered discrete whereas emotions can be thought of as both discrete and continuous (Picard, 2003), this article sought to examine this aspect in human-machine interaction (Conversational Agent), using the model proposed by Tran (2004) as a reference.

To infer mood states on textual inputs, as there is no Portuguese database to automatically classify terms, similarly to Scherer's (2005) study, subjective textual evidence, such as punctuation, interjections and chat context, were empirically considered. To do so, we used conversation analysis techniques, which "examines how participants manage interaction as it proceeds: how they make sense of the moment-by-moment unfolding of interaction" (Wooffitt, 2005: p. 90).

Studies as Derrick et al. (2013) considered subjective aspects to identify the mood "deception" in text, as typing cues (response time and the number of edits), and messaging cues (lexical diversity and word count). The authors searched to find the relationship between spontaneous deception and the num-

ber of edits (e.g., backspaces, deletes), response time, word count, and lexical diversity in chat-based communication.

In this research we focus on more objective data, analyzing just what is clearly expressed in the text. It was then compared with students' opinion about the tool in terms of interest, satisfaction and utility. Such supplementary data is gathered to investigate, for example, if signs of negative moods may be related to student dissatisfaction with the CA, and consequently identify possible improvements to resolve such drawbacks.

3. Method

We have conducted an exploratory mixed methods research, involving the participation of students of a Distance Education post-graduation course from a Brazilian public university. The web-based Conversational Agent METIS (an acronym for Mediator of Education in Technology of Information and Socializer), previously built on AIML technology, was used.

The data were collected within a period of 12 weeks (between April and June of 2017), and the participants received access for the CA since six months before it, being incentivized to use it since then. To compose the sample, chat logs of 30 different users were analyzed, and 17 students volunteered to answer the questionnaire. For an isonomic analysis of the content, the demographic data of the respondents were not disclosed.

Two instruments were used for data collection, as follows.

- Instrument 1—Questionnaire of student perception

Composed of nine questions (displayed in Table 1): seven objective questions grouped into axes related to interest, utility and perceived satisfaction, and two open-ended questions for comments and subjective perceptions of the participants about the CA.

Longhi et al. (2012) define interest as the mood state that drives (or not) someone towards pursued objectives. Therefore, interest was considered an

Table 1. Summary of objective responses given to the Instrument 1 (Source: the authors).

Axis	Objective questions	% Ratings				
		1	2	3	4	5
Interest	Q1—How often do you access the Conversational Agent?	29%	59%	12%	0%	
	Q2—In your opinion, is the Conversational Agent interesting?		18%	41%	41%	
Utility	Q3—Do you consider the Conversational Agent useful for your learning?		30%	23%	47%	
Satisfaction	Q4—In your opinion, is the Conversational Agent intelligent?		41%	41%	18%	
	Q5—Are the Conversational Agent's answers coherent?		42%	29%	29%	
	Q6—Are the Conversational Agent's answers relevant?		42%	29%	29%	
	Q7—How satisfied are you with the Conversational Agent?		42%	24%	34%	
General	Q8—What do you expect from the Conversational Agent?	*Open response*				
	Q9—How could the Conversational Agent be improved?	*Open response*				

important aspect to evaluate (Q1, Q2). According to Danilava et al. (2012), utility may represent the frequency that a Conversational Agent is accessed by the users. So we include one question about utility (Q3). Burden (2009) clarifies that the most immediate test of a CA's salience is the satisfaction of the customers using it. In this way, four questions concerned about satisfaction (Q4 to Q7). Q8 and Q9 were more general user perception related questions.

The objective questions, with the exception of Q1, were given five-point Likert scale response options, with extremes representing strongly disagree (1) and strongly agree (5). Q1 offered the following alternatives: 1) I've never accessed it, 2) I access it less than once a month, 3) I access it once in a while (more than twice a month), 4) I regularly access it (at least once a week), and 5) I access it very frequently (more than 3 times a week).

Q1, Q2, Q3 and Q7 also provided a blank space next to it for participants' comments, in an attempt to collect some more information that might justify their answers.

- Instrument 2—Analysis of the logs recorded by the CA

Picard et al. (2004) point out that despite the convenience and widespread acceptance of questionnaires, the use of self-report information is considered unreliable when it comes to emotions. In order to minimize this difficulty, a sample of chat logs of interactions with the CA was considered, assigning a mood state to each user input.

The 30 longest conversation logs (in number of text lines) were selected for the sample, from 30 different IP addresses to ensure that the logs came from different users, resulting in a total of 250 lines (average of 8 lines per conversation log).

The log analysis was carried out incrementally (one by one) by the authors of the study, following two steps:

1) *Appraisal Extraction*: method used by Longhi et al. (2012), classifying words that have affective connotations in the groups of emotions that determine the student mood state. For example, the emotion family of "happy" comprises terms like "fond", "elated", "caring", "cheerful" and "delighted", among many others. However, they used text-based chats among humans from VLE forums as their textual source. In the present study, the analysis uses chat logs of interactions of humans with a computational tool, so other forms of verification were also required.

2) *Subjective Textual Evidence*: such as punctuation, interjections and chat context.

Following these two steps, a mood state was assigned to each student's textual input in the CA interface, according to Tran's (2004) Geneva Emotion Wheel, including the option neutral when a given mood state was not found in the model or it was not possible to identify it.

In a similar manner, each CA response was also evaluated regarding utility, with ratings ranging from:

- 0 (zero) when considered totally useless, inadequate or incoherent;
- 50 (fifty) when partially useful, adequate or coherent;
- 100 (one hundred) when fully useful, adequate or coherent.

Figure 3 displays a summary of the proposed evaluation of CA.

In addition to the general analysis of the logs, a specific study of three randomly selected students was performed, comparing Instruments 1 and 2 directly. This association was possible because sometimes the CA asked the user's name in the first interaction, registering it in the logs.

4. Data Analysis

In order to facilitate the analysis of results, subsections were created for each data collection instrument.

4.1. Analysis of Instrument 1

To estimate the reliability of the objective part of the questionnaire, the Cronbach's Alpha coefficient (Cronbach, 2004) was used, which allows measuring the correlation among the answers given by the respondents. Reliability $\alpha = 0.88$ was obtained, which is considered a high reliability.

Table 1 shows a summary of the ratings given by the students to each response option. In order to make the analysis more concise, the negative ratings (1 and 2, strongly or partially disagree) and the positive ratings (4 and 5, strongly or partially agree) were grouped.

The first question (Q1) showed that 12% of students access the Conversational Agent more than twice a month, 59% access it less than once a month, and 29% have never accessed it. According to the students, among the reasons why they do not frequently access it are: they cannot establish a dialogue, they do not obtain the necessary answers, and they do not have free time due to professional activities. Four students claimed to be unaware of the existence of the METIS and one participant said that it was strange and impersonal to ask questions to a CA.
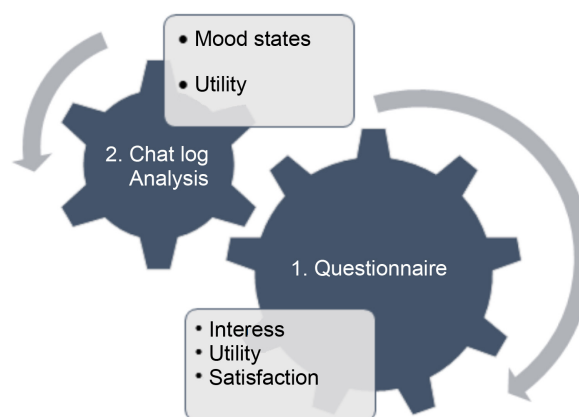

Figure 3. Summary of the proposed conversational agent evaluation (Source: the authors).

In Q2, about whether the students consider the CA interesting, 41% of the participants agreed partially or strongly, attributing this to the different and quick way of solving questions and providing information, besides being an interactive way of learning. However, this was the same proportion of students who gave a neutral rating to this question, associating this fact with the limitations on the CA's responses. The other 18% that gave a negative rating affirmed that the CA is either uninteresting to them, or they do not understand what it is for, or they still don't know about this tool.

Q3 asked the students about CA's utility for learning. Of the 17 respondents, 47% partially or strongly agreed that the METIS is useful because, according to them, it helps in a DE courses by making learning more meaningful when they are challenged, and because "all knowledge is worth having". However, 30% partially or strongly disagreed, saying again that they did not know how to use it or did not know what it is for. Still, 23% remained neutral in their ratings, stressing the need for a larger "answers database" (knowledge base) or unfamiliarity with the tool.

Q4 to Q6 received no comments, only the objective answers. The students were asked whether the CA is intelligent (Q4), and 41% of the answers were neutral, to the same extent of those who disagreed partially or strongly, leaving 18% of students who considered it intelligent. Only one participant gave the maximum rating, strongly agreeing. In Q5, about whether the students considered CA's responses coherent, 42% expressed disagreement, 29% were neutral and 29% agreed. However, no participant strongly agreed to this item. When they were asked whether CA's responses were relevant (Q6), the same ratings as those of Q5 were obtained.

Regarding students' satisfaction with the CA (Q7), a small increase was observed in the positive ratings. Although the same 42% rated it negatively, attributing this to the CA's incoherent responses, or for not knowing the tool enough to give an opinion about it, 24% remained neutral and 34% said they were satisfied, agreeing partially or strongly. Some of them commented that the CA is interesting and a good idea as a mediator in the DE course they were taking. One participant gave it a maximum rating, affirming to be very satisfied despite having performed few tests.

The last two items (Q8 and Q9) were open-ended questions. In Q8, the students were asked about what they expect from the CA. They said that expected it to be effective, challenging, and able to bring coherent and helpful information, thus being a learning aid. Students also said they expect it to help them do the course assignments and to have a broader "database", referring to its knowledge base. Some participants said that they expect to become more familiar with the tool and, therefore, they have not created expectations about it yet.

In Q9 students were asked on how the CA could be improved, where they suggested that it should give more coherent answers, learn from the user's feedback and have a bigger "database", emphasizing the same topics raised in the

previous question. They also highlighted that a "brief instruction" or "introductory approach" is needed until they become familiar with it. Some participants affirmed, again, that they could not make comments because they had little experience using the tool.

4.2. Analysis of Instrument 2

Figure 4 shows a summary of the mood states identified in the logs of conversations between the Conversational Agent and the students. On the left is a table containing the total inputs (lines) counted for each mood state dimension (quadrants) and emotion families (sub quadrants), and on the right is a graph with the percentage ratings for each quadrant (including the Neutral dimension).

In general terms, it is possible to observe in Figure 4 that 17% of the inputs did not fit into any mood state, so they were categorized as Neutral. The predominant mood state was Enthusiastic, with 160 lines (64%), and the students expressed mainly the emotion family of Interest (134). The chat log analysis demonstrated that Interest was subdivided in two categories: 1—Interest in the subjects or activities inherent or related to the course (99); 2—Interest in the CA's skills (35). Some examples of Interest expressed by the interlocutors are presented below.

Interest 1—Conversation 27:

"Can you tell me which of my activities are behind schedule?"

Interest 2—Conversation 26:

"Tell me what you can do…"

The second mood state most identified was Satisfied, with 35 lines (13%), predominating the emotion family of Joy (15). An example of this mood state in the conversations is transcribed below.

Joy—Conversation 15:

"Ok Thanks Metis we're both in the same boat… Good Night"

The mood state Dissatisfied was identified in 16 chat lines, corresponding to 6% of the total logs, being 12 inputs associated with the emotion family of Anger. An example is given as follows.
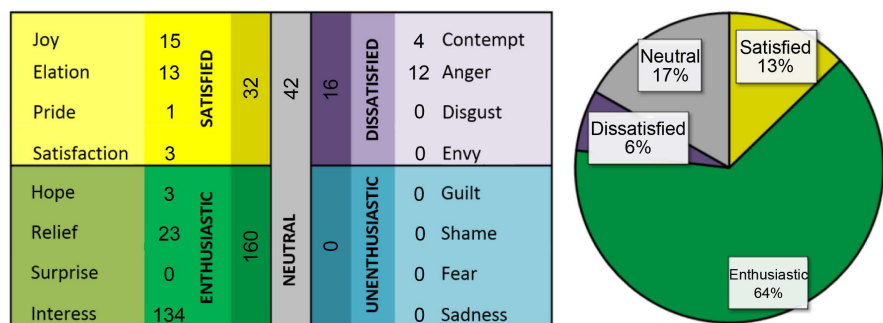
Anger—Conversation 23:

"What's your problem?"

Figure 4. Mood states identified in the chat logs (Source: the authors).

It was evinced that many users abruptly quitted the conversation, without saying goodbye to the CA or devoid of the answers they were searching for, which may be another indication of dissatisfaction. However, this was not computed because, besides not being an input line, it may have been caused by other factors, as internet connection dropout. Analyzing the log files, one of the problems observed was that CA responses were very direct and objective, so it did not stimulate the continuation of the dialogue. However, there was no evidence of Unenthusiastic mood states in the chat logs.

Regarding the utility of the answers in the chat logs in general, a total of 4.150 points was scored, representing approximately 17% of useful answers provided to the user by the CA. This means that, on average, at least one line (out of 8) in the chat was useful for the user or coherent with the subject being dialogued. In addition to being a low rating, most of those useful responses were related to greetings when the student started or left the chat. The study included a target population that was unfamiliar with the use of CA and perhaps that is the reason why complex sentences were frequently used by the participants, typical of human interaction but difficult for the CA to analyze appropriately.

In some extent, this also denotes an overly high expectation of the user regarding the CA's ability to understand the asked questions. The recurring times when the input was not understood may have provoked some frustration. According to Burden (2009), users always have high expectations and expect the bot to be able to do many "common sense" things, even if the bot is within a constrained role.

The chat log analysis of the three specific users identified in the sample facilitated a direct association between the instruments, as follows.

- Student 1

It was observed that the first student (S1), with a log of 6 lines, received 33% of useful answers, and the mood states were 50% Enthusiastic, 33% Neutral and 17% Dissatisfied. The median of S1's answers in the questionnaire was 3 (Neutral), and the frequency of access was less than once a month, the same as that of the other two users individually analyzed. This student made a comment saying "I'm still exploring it" (the CA), but did not make comments in the other objective questions. In Q8, the student said that expected METIS could "help with the course assignments so they could be more easily done", and in Q9 the student said that it should be improved to provide more useful answers.

- Student 2

The chat logs (9 lines) of the second student (S2) contained 22% of useful answers. In 45% of the conversation, the Enthusiastic mood state was observed, 33% was Neutral and 22% Dissatisfied, and S2 also had a median of 3 in the questionnaire. This student showed low frequency of access and said it was because the CA was "limited"; the same argument was given in questions Q2 and Q7. In spite of that, S2 considered METIS useful for learning because of its rapid responses (Q3). When asked about expectations (Q8), S2 said to be expecting

something like "Google" (web search tool), but took some responsibility for not getting all the expected answers, saying in Q9: "maybe I need to do better when I talk to her".

- Student 3

The third student (S3) had a chat log of 16 lines, with 25% of useful answers, and interactions showing the mood states Enthusiastic (68%), Neutral (19%) and Dissatisfied (13%). Besides having the highest inference rating of the Enthusiastic mood, S3 was also the one with the highest questionnaire median: 4 (partially agree). The low frequency of access was attributed to lack of curiosity about the tool and for being busy with professional activities in the period of data collection. In Q2, S3 commented that METIS is interesting because "it makes them speak straight", and in Q3 expressed a neutral opinion about the CA's utility, reporting lack of familiarity with the tool. As for satisfaction (Q7), S3 found the CA interesting but does not know what to expect from it (Q8). This student also commented that the CA can be improved by learning from users' responses (Q9).

5. Discussion

This section discusses the main findings of the research, comparing them with other studies in the area, and highlighting some of the contributions. Initially, it is important to clarify that not all the 17 questionnaire respondents may have been considered in the chat log analysis, because it was noted that some of them said they did not know or had never accessed the tool. However, in general, it is possible to effectively relate Instruments 1 and 2, as the questionnaire reliability was considered high, and the inferences carried out from the chat logs did not show great discrepancies.

Ghose & Barua (2013) discussed the difficulty in maintaining a dialogue with the CA for a sustained period of time, where the participants interacted for an average of 10 lines. In the present study, this characteristic of limited interactions was also observed, with an average of 8 lines per conversation. Considering that this value was calculated by a sample comprising the 30 longest chat logs, the overall average number of interactions is probably even smaller. Thus, it can be concluded that the tool is underutilized, that is, the students are not very often accessing the CA METIS.

It was found that users' perceptions about the CA reasonably correspond to the criteria of interest and satisfaction, as diagnosed by Instrument 1. But the low utility of the CA was evident both in the questionnaire (47% of negative opinions in this aspect) and in the chat logs (17% of useful answers per conversation). Therefore, it is possible to observe that, despite the perceived and actual low utility, the levels of interest and satisfaction were higher than those of utility, with negative ratings of only 41% (Q2), 18% (Q4), 29% (Q5), 29% (Q6) and 34% (Q7), allowing to infer that this aspect (utility) did not totally affect the student's view of the tool. In other words, users have noted that the CA had limitations

but accept it fairly positively.

Dale (2016) highlights that the next milestone in the CA area is on making truly conversational interactions, by which is meant the ability to take account of discourse context, rather than just treating a dialog as a sequence of independent conversational pairs. In this sense, to overcome the impasse of low utility diagnosed, the use of the AIML tag <topic> in the CA's knowledge base is suggested to keep the user in the expected direction. This tag limits the responses to a set of categories linked to a key concept, reducing the chances of a totally random and out-of-context response. Since a great deal of interest in the subjects and activities of the course was noted in the chat logs, the use of topics may help to avoid the distraction of the user, always responding to something related to the subject in question until another topic is detected, requested or mentioned. Also, as recommended by Fryer et al. (2017), it would be useful if learners had logged in the system before using it, so it would afford the system to remember users' past interactions and use these data to become familiar to users.

Among the participants' comments in the questionnaire, suggestions for CA improvements were identified, essentially including a bigger repertoire of useful answers, solutions to unproductive interactions and the recurrent request for knowledge base expansion. Abdul-Kader & Woods (2015) advise that developing a perfect CA is very difficult because it needs a very large database and it must give reasonable answers to all interactions.

Moreover, the need to implement strategies that encourage students to know and interact with the CA was diagnosed, since many of the participants remained neutral in some answers, affirming they could not give an opinion or did not have sufficient knowledge about the tool. Mou & Xu (2017) discuss the effects of novelty experience with sophisticated technological tasks, emphasizing the need for actions that facilitate user familiarization to avoid difficulties in these terms.

Regarding the chat logs, it was identified that despite the low utility of the CA answers, the predominant mood state was Enthusiastic, with 64%, showing that students were very interested mainly in obtaining information about the activities or subjects of the course. Thus, it can be inferred that users have a good perception and believe in the potential of the tool for learning purposes, also evidenced by the absence of the Unenthusiastic mood state in the logs. Emotional engagement in this experience is an important factor to stimulate the students' social presence. This result corroborates with Zumstein & Hundertmark (2017), when they say that CA generally get great acceptance from most users.

Some participants attributed some responsibility to themselves in what concerns improvements in the CA, in a collaborative or solidary view about the system. Supporting this assertion, Longhi et al. (2012) stated that the Enthusiastic mood expresses positivity to face the challenges of learning, which lead to collaboration and cooperation.

The Neutral dimension was observed in 17% of the student's chat logs, considered an acceptable value when it comes to human-machine interactions. Still,

the dissatisfied mood was present in 6% of the logs. According to Longhi et al. (2012), this mood state is evident when there is expression of anger, contempt, disgust and/or envy. In this context, it may indicate moments where the CA responses end up leading the students to externalize a negative emotion, which in turn may lead them to quit the CA environment and, when recurrent, reflected on their general behavior.

The analysis of three specific students allowed to directly relate their perception expressed in the questionnaire (Instrument 1) with the mood states and utility inferred through the chat logs (Instrument 2). It was observed that S3, who maintained a more positive posture (median 4 in the questionnaire), was more Enthusiastic (68%) and interacted twice the average (16 lines) with the CA, retrieving 25% of useful answers. The other two students (S1 and S2) obtained a neutral median (3) in the questionnaire, and expressed 50% and 45% in the Enthusiastic dimension, with logs of 6 and 9 lines, and answers with 33% and 22% of utility, respectively. These data demonstrate that the Enthusiastic mood state, in this case, was not directly related to the utility of the answers, allowing to consider the possibility that it may be related to the student's personal traits.

It was observed that S1, who obtained the highest rating of useful answers (33%), gave the lowest rating (2) in this item, partially disagreeing with it, and seemed to be more Dissatisfied (22%) than S2 and S3 (17% and 13%). However, S1 was also the one who least interacted with the tool (6 lines). Therefore, it is possible to infer that perhaps positive mood states like the Enthusiastic state can be triggered to the same extent that the frequency of user interaction increases, and that the perceived utility may be inversely related to the occurrence of the dissatisfied state in the conversations.

6. Conclusion

Dale (2016) presumes that very soon we'll be in a world where some of our conversational partners we'll know to be humans, some we'll know to be Conversational Agents, and probably some we won't know either way, and we may not even care.

In recent years, the research on CA has been growing, expanding its potential to provide interactivity to students and bringing great benefits especially to those who are remote, as Distance Education students. In this case, they are seen as supporting tools for teachers, reducing students' sense of social isolation by being available to amicably interact with the them (in natural language), in a continuous way.

On the other hand, as Mou & Xu (2017) point out, humans may not be able to find appropriate motivation to develop social relationships with machines, so it is necessary to use strategies to encourage them to develop such interactions, triggering positive moods that may predispose them to positively receive the information, hence favoring learning. In addition, Hill et al. (2015) argue that the obstacle for computers is not just in understanding the meanings of words, but

in the endless variability of expression in how those words are collocated in language use to communicate meaning, which makes this interaction more difficult.

This study investigated mood states inferred by chat log analysis of interactions among students and a CA (METIS), and related it with the students' perceptions about the tool. The analysis allowed to accept the hypothesis that students' emotional states when interacting with a CA can impact on the quality of the conversation and on the user's perception about the tool in terms of interest, utility and satisfaction. Also we verified some causes underlying the underutilization of the CA, such as user unfamiliarity or limitations of the knowledge base, which made possible to identify improvements to be implemented so that negative mood states, such as Dissatisfied, can be overcome.

One of the main contributions of this study is the way by which the CA evaluation was conducted, making use of two instruments for data collection (questionnaire and chat logs) in a complementary way.

As future work, we intend to outline and test strategies to improve students' mood states and perceptions, in addition to using text mining techniques for the chat logs analysis, in order to automate the emotion lexical inference, making it more dynamic and fast to allow proactive actions to reduce the user dispersion and/or distraction, and consequently the underutilization of the tool.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

Abdul-Kader, S. A., & Woods, J. (2015). Survey on Chatbot Design Techniques in Speech Conversation Systems. *International Journal of Advanced Computer Science and Applications, 6*, 72-80.

AbuShawar, B., & Atwell, E. (2015). ALICE Chatbot: Trials and Outputs. *Computación y Sistemas, 19*, 625-632. https://doi.org/10.13053/cys-19-4-2326

Aguiar, E. V. B., Tarouco, L. M. R., & Reategui, E. (2014). Supporting Problem-Solving in Mathematics with a Conversational Agent Capable of Representing Gifted Students' Knowledge. 2014 47*th Hawaii International Conference on System Sciences* (pp. 130-137), Waikoloa, 6-9 January 2014.

Burden, D. J. (2009). Deploying Embodied AI into Virtual Worlds. *Knowledge-Based Systems, 22*, 540-544. https://doi.org/10.1016/j.knosys.2008.10.001

Cronbach, L J. (2004). My Current Thoughts on Coefficient Alpha and Sucessors Procedures. *Educational and Psychological Measurement, 64*, 391-418. https://doi.org/10.1177/0013164404266386

Dale, R. (2016). The Return of the Chatbots. *Natural Language Engineering, 22*, 811-817. https://doi.org/10.1017/S1351324916000243

Danilava, S., Busemann, S., & Schommer, C. (2012). Artificial Conversational Companions: A Requirements Analysis. *Proceedings of 4th International Conference on Agents and Artificial Intelligence* (pp. 282-289). http://www.dfki.de/web/forschung/iwi/publikationen/renameFileForDownload?filena

me=CompanionsICAART_1122.pdf&file_id=uploads_1366

Derrick, D. C., Meservy, T. O., Jenkins, J. L., Burgoon, J. K., & Nunamaker Jr., J. F. (2013). Detecting Deceptive Chat-Based Communication Using Typing Behavior and Message Cues. *ACM Transactions on Management Information Systems, 4*, 9. https://doi.org/10.1145/2499962.2499967

Fryer, L. K., Ainley, M., Thompson, A., Gibson, A., & Sherlock, Z. (2017). Stimulating and Sustaining Interest in a Language Course: An Experimental Comparison of Chatbot and Human Task Partners. *Computers in Human Behavior, 75*, 461-468. https://doi.org/10.1016/j.chb.2017.05.045

Ghose, S., & Barua, J. J. (2013). Toward the Implementation of a Topic Specific Dialogue Based Natural Language Chatbot as an Undergraduate Advisor. 2013 *International Conference on Informatics, Electronics & Vision* (*ICIEV*), Dhaka, 17-18 May 2013. https://doi.org/10.1109/ICIEV.2013.6572650

Griol, D., & Callejas, Z. (2013). An Architecture to Develop Multimodal Educative Applications with Chatbots. *International Journal of Advanced Robotic Systems, 10*, 1-15. https://doi.org/10.5772/55791

Heuvelman-Hutchinson, L. (2012). *The Effect Different Synchronous Computer Mediums Have on Distance Education Graduate Students' Sense of Community and Feelings of Loneliness*. Doctoral Thesis, Lynchburg: Liberty University. http://digitalcommons.liberty.edu/doctoral/547/

Hill, J., Ford, W. R., & Farreras, I. G. (2015). Real Conversations with Artificial Intelligence: A Comparison between Human-Human Online Conversations and Human-Chatbot Conversations. *Computers in Human Behavior, 49*, 245-250. https://doi.org/10.1016/j.chb.2015.02.026

Izard, C. E. (1971). *The Face of Emotion*. New York: Appleton-Century-Crofts. https://www.amazon.com/face-emotion-Century-psychology/dp/B0006CFW2S

Jaques, P. A., & Vicari, R. (2007). A BDI Approach to Infer Student's Emotions in an Intelligent Learning Environment. *Computers and Education, 49*, 360-384. https://doi.org/10.1016/j.compedu.2005.09.002

Kang, Y., Nah, F. F., & Tan, A. (2012). Investigating Intelligent Agents in a 3D Virtual World. *Thirty Third International Conference on Information Systems*. Orlando. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.656.8513&rep=rep1&type=pdf

Krassmann, A. L., Rossi Filho, T. A., Tarouco, L. M. R., & Bercht, M. (2017). *Initial Perception of Virtual World Users: A Study about Impacts of Learning Styles and Digital Experience* (pp. 95-112). International Educative Research Foundation and Publisher. http://www.ijier.net/index.php/ijier/article/view/684/574

Leonhardt, M., Dutra, R. L. S. D., Granville, L. Z., & Tarouco, L. M. R. (2005). DOROTY: An Extension in the Architecture of a ChatterBot for Academic and Professional Training in the Field of Network Management. *IFIP World Conference on Computers in Education*. Cape Town, 4-7 July 2005 http://www.sbs.co.za/wcce2005/

Leonhardt, M., Tarouco, L. M. R., Vicari, R., Santos, E. R., & Da Silva, M. D. S. (2007). Using Chatbots for Network Management Training through Problem-Based Oriented Education. In 7*th IEEE International Conference on Advanced Learning Technologies* (Vol. 5, pp. 845-847). Piscataway, NJ: Institute of Electrical and Electronics Engineers. https://doi.org/10.1109/ICALT.2007.275

Longhi, M. T., Behar, P. A., & Bercht, M. (2012). Mood Inference Machine: Framework to Infer Affective Phenomena in ROODA Virtual Learning Environment. *International Journal of Advanced Corporate Learning, 5,* 8-16. https://doi.org/10.3991/ijac.v5i1.1740

Mou, Y., & Xu, K. (2017). The Media Inequality: Comparing the Initial Human-Human and Human-AI Social Interactions. *Computers in Human Behavior, 72,* 432-440. https://doi.org/10.1016/j.chb.2017.02.067

Neves, A. M. M., Barros, F. A., & Hodges, C. (2006). Iaiml: A Mechanism to Treat Intentionality in Aiml Chatterbots. In *18th IEEE International Conference on Tools with Artificial Intelligence* (pp. 225-231). Piscataway, NJ: Institute of Electrical and Electronics Engineers. https://doi.org/10.1109/ICTAI.2006.64

Neviarouskaya, A., Helmut, P., & Mitsuru, I. (2010). Recognition of Affect, Judgment, and Appreciation in Text. In *Proceedings of the 23rd International Conference on Computational Linguistics* (pp. 806-814). Association for Computational Linguistics. http://www.aclweb.org/anthology/C10-1091

Online Learning Consortium (2017). *The Distance Education Enrollment Report 2017.* https://onlinelearningconsortium.org/read/digital-learning-compass-distance-education-enrollment-report-2017/

Picard, R. W. (2003). Affective Computing: Challenges. *International Journal of Human-Computer Studies, 59,* 55-64. https://doi.org/10.1016/S1071-5819(03)00052-1

Picard, R. W., Papert, S., Bender, W., Blumberg, B., Breazeal, C., Cavallo, D., & Strohecker, C. (2004). Affective Learning—A Manifesto. *BT Technology Journal, 22,* 253-269. https://www.media.mit.edu/publications/bttj/Paper26Pages253-269.pdf https://doi.org/10.1023/B:BTTJ.0000047603.37042.33

Rush, P. (2015). Isolation and Connection: The Experience of Distance Education. *International Journal of E-Learning & Distance Education, 30.* Http://Www.Ijede.Ca/Index.Php/Jde/Article/View/936

Savin-Baden, M., Tombs, G., & Bhakta, R. (2015). Beyond Robotic Wastelands of Time: Abandoned Pedagogical Agents and New Pedalled Pedagogies. *E-Learning and Digital Media, 12,* 295-314. https://doi.org/10.1177/2042753015571835

Scherer, K. R. (2005). What Are Emotions? And How Can They Be Measured? *Social Science Information, 44,* 695-729. https://doi.org/10.1177/0539018405058216

Tegos, S., Demetriadis, S., & Tsiatsos, T. (2014). A Configurable Conversational Agent to Trigger Students' Productive Dialogue: A Pilot Study in the CALL Domain. *International Journal of Artificial Intelligence in Education, 24,* 62-91. https://link.springer.com/article/10.1007/s40593-013-0007-3?no-access=true

Tran, V. (2004). *The Influence of Emotions on Decision-Making Processes in Management Teams.* Thèse No. 323, Genève: Faculté De Psychologie et des Sciences de l'Education, Université de Genève.

Wallace, R. (1995). *A.L.I.C.E.: Artificial Linguistic Internet Computer Entity.* https://www.chatbots.org/chatbot/a.l.i.c.e/

Weizenbaum, J. (1966). Eliza—A Computer Program for the Study of Natural Language Communication between Man and Machine. *Communications of the ACM, 9,* 36-45. https://doi.org/10.1145/365153.365168

Wooffitt, R. (2005). *Conversation Analysis and Discourse Analysis: A Comparative and Critical Introduction.* New York, NY: Sage. https://doi.org/10.4135/9781849208765

Xie, T., & Luo, L. (2017). Impact of Prompting Agents on Task Completion in the Virtual World. *International Journal of Online Engineering, 13,* 35-48. http://www.online-journals.org/index.php/i-joe/article/view/6850

Zumstein, D., & Hundertmark, S. (2017). Chatbots—An Interactive Technology for Personalized Communication, Transactions and Services. *IADIS International Journal on WWW/Internet, 15,* 96-109.