

Robotic - Cognitive Adaptive System for Teaching and Learning (R-CASTLE)*

Daniel C. Tozadore¹[0000-0003-0744-0132] and Supervisor: Roseli A. F. Romero¹[0000-0001-9366-2780]

Instituto de Ciências Matemáticas e de Computação - USP
Avenida Trabalhador Sancarlense, 400. São Carlos. Brasil.
tozadore@usp.br, rafrance@icmc.usp.br

Abstract. One of the biggest current challenges in education is to positively impact the teaching-learning process with technology. Two common reasons are the teachers' lack of preparation and the students' attention span. Several Human-Robot Interaction (HRI) studies are approaching these issues. However, very few of them are considering both teachers and students in a one and only application. Thus, the presented thesis had two objectives: to provide a unique and intuitive HRI tool for education and to evaluate its impact on the users. The resulting architecture was a Robotic - Cognitive Adaptive System for Teaching and Learning (R-CASTLE). R-CASTLE aims to provide customized interactions and personalized learning to the students through machine learning for autonomous vision and dialogue interactions. The methods are configured by the teachers in the windows of the system's graphical interface. Teachers can also have access to the system's evaluations in chart mode of the students' collective and individual performances. In end-to-end experiments, teachers and students claimed to experienced a sensitive potential of the system to support them. R-CASTLE was tested with other interactive devices in different applications and the results showed high performances in their activities design optimization. From the best of our knowledge, it is an innovative proposal implemented with collaborative assistance of institutes from Portugal, Italy and Japan. R-CASTLE is currently being adapted to support teachers in their remote classes during COVID-19 pandemic.

Keywords: Human-Robot Interaction · Artificial Intelligence on Education · Adaptive Systems · Personalized Learning.

1 Introduction

The R-CASTLE system, proposed in this thesis, stands for Robotic-Cognitive Adaptive System for Teaching and Learning. It has been developing in the last six years based on studies and methods of both the literature [10] and in the Robot Learning Laboratory, in ICMC, USP - São Carlos. As the name suggests,

* Supported by CNPq and CAPES.

it aims to offer an autonomous adaptation of system behavior and content difficulty, focused on the learning process, from designing the activities to their evaluation. Multiple open-source algorithms of image classification and verbal recognition are at the teachers's disposal to be used in the interactions, the student's database, for general knowledge and the previously performed activities and corresponding evaluations. In summary, the technical proposal of this thesis aims to deliver a system with the aforementioned features, while its application aims to support scientific studies regarding the users' experience with the system.

Although the R-CASTLE technical scheme is briefly presented later, its high-level functionalities of R-CASTLE are worthy to take note earlier. They can be divided for the teacher and for the students.

The teachers can interact with the system through the graphical interface to plan and execute their activities and make use of Artificial Intelligence without knowing how to program it. Unlike many intelligent tutoring systems whose content is specific, teachers can input the content they want to exercise with students. This brings great flexibility to the system, since it can be widely used in several areas of teaching. R-CASTLE automatically analyzes student behavior during student interaction. Thus, teachers can also receive individual reports of student performance in each activity. The data obtained is stored so that the teacher can have a overall view over a given period. The data stored by the R-CASTLE system is not only related to the student's performance, but also the video corresponding to the student's period of interaction with the system. This feature brings a wealth of information to the teacher. They can check the students' behavior during the interaction; for example if the task caused the student to feel any discomfort or if it went smoothly.

On the other hand, students can interact with the system through a social robot. It is believed that this fact can make the student exercise the content in a playful way, because, thanks to the vision and voice modules available in R-CASTLE, the student will be able to use voice to communicate with the system. It was observed that the students would be more willing to perform the exercises when the interaction is more authentic. Another interesting point is that the system adapts as the student reaches a certain level of difficulty. Therefore the student will be up for the challenge of improving themselves every time instead of feeling unmotivated. Students will be able to count on a system that supports in building their self confidence and learning skills through the cultivation of rapport with the robot and the robot's awareness of the student's preferences.

The collaborations from international institutes were made through: studies that resulted in published paper, as the case of the Waseda University and the study in [4]; exchange researching, as the case of Instituto Superior Técnico de Lisboa, where the PhD student (at that time) spent two months planning the R-CASTLE experimental setup with HRI specialists; and technical visit, as the case of Instituto Italiano de Tecnologia, where lectures and discussions were performed to improve R-CASTLE.

As the contributions of R-CASTLE are several, the objectives and hypothesis addressed in this thesis were divided into three aspects: its technical approaches (e.g. the analysis of the algorithms); its impact on each user (e.g. the teachers and students perceptions regarding their experience with the system); and its impact on the educational process (e.g. how the system may affect the teaching/learning factors). The R-CASTLE validation in these aspects based in the experiments results is summarized in Section 4.

The remainder of this paper briefly describes: the R-CASTLE technical scheme in Section 2, the performed experiments in Section 3, the discussions of the based on the results in Section 4, and the conclusion in Section 5.

2 R-CASTLE technical overview

Although is hard dealing with AI at its technical level, managing data and interacting with devices through software windows are more intuitive and an alternative path to overcome the teachers' lack of technical knowledge and training. Similarly, interacting with robots in human-like manners makes more natural to the students the process of adapting to new methodologies. In R-CASTLE, both of these advantages are placed together in a robotic architecture, which is capable to be used with other interactive devices as well. The Figure 1 shows the scheme of R-CASTLE and its proposed communication flow to teachers and students through social robots and AI techniques.

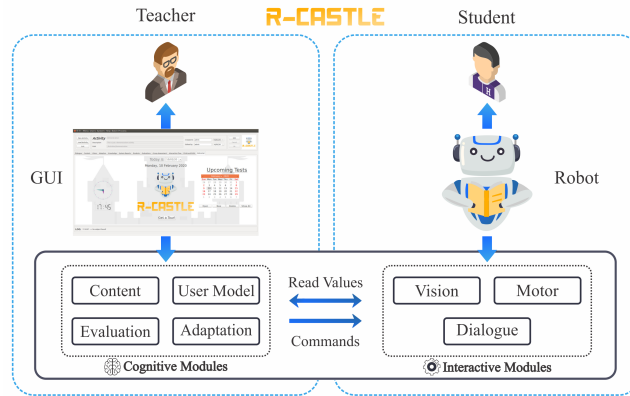


Fig. 1. R-CASTLE scheme.

The communication of the teacher and the R-CASTLE is made through the Graphical User Interface whereas the R-CASTLE communicates to the students through an interactive device, usually a social robot (that is why the "R" of "Robot" in the name).

There are two groups of modules in R-CASTLE. One is the Interactive group and the other is the Cognitive. This division is only made for scope matters. The Cognitive group encloses the modules that process the outputs provided by the modules of the Interactive group, with the result of changing them into useful information to store and make decisions based on them. They are also used to store information manually inserted by the designers or autonomously acquired during the activities. They are the *Content*, the *Adaptation*, the *Student Model* and the *Evaluation* modules. The Interactive group are the modules responsible for controlling the resources of the robot and the methods to exchange information with the environment. These modules are the *Dialogue*, the *Vision* and the *Motor*. The Dialogue and the Vision modules have a fundamental role in this proposal as described in the next sections. The Motor module only controls the robot's moves, which in some cases can be dismissed. Python 2.7 was used in all the algorithm due to the compatibility with NAOqi ¹, the operational system for Softbank Aldebaran robots. A video about R-CASTLE's main features can be seen at <https://www.youtube.com/watch?v=G1Nj98L1Mrc>.

2.1 Graphical User Interface

Teachers can interact with R-CASTLE through a Graphical User Interface (GUI), showed in Figure 2. It was implemented to operate over the technical implementations of the system. R-CASTLE allows the teachers to create and manage activities as work-spaces through the windows. This is due to the fact that every activity has different setups for some specific modules, such as the Dialogue, the Content and the Vision modules. However, they can also share common data, such as the stored information of the users, definitions, evaluations and most of the things that are the same for the majority of the activities. The modules are available for configuration in their corresponding tabs in the middle section of the window. It is believed that the features of this interface constitute a potential solution to the question with regard to the teachers' access to AI techniques intuitively.

2.2 Content

The GUI makes the content insertion easy for teachers. The content is stored in the system database and can be approached in any late activity. The content is approached in Topics. Each Topic is an entry in the Content module which has a concept about it (an explanation of the topic from which the questions will be derived) and a lot of questions regarding it.

Once the system has the capabilities of both speech and image recognition, the expected answers can be a sentence or an image. In the case of a sentence, the system analyzes the answer by the dialogue system, as shown in [7]. On the other hand, the answers which are expected as images are classified by the Vision module, as shown in [3].

¹ http://doc.aldebaran.com/2-5/dev/programming_index.html

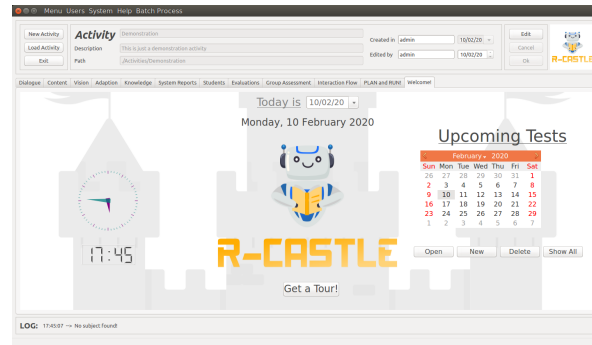


Fig. 2. R-CASTLE window.

2.3 Adaptation

The Adaptation module (or adaptive system) is responsible for adapting the robot behavior and the difficulty level of the approached content based on the information it receives from the vision and dialogue module. The indicative values autonomously took from the students are mapped into three groups with regard to the skills of the users considered by the system for adaptation optimization. The measured skill are attention, Communication and Learning.

There are two ways to achieve adaption based on the read values of the users in R-CASTLE. The first one is to use a Ruled-Based algorithm for an adaptive function to set the final value of adaptation. The second is to make the system to project the threshold values into Fuzzy Rules and use it to compute the adaptive function. The emotion recognition through the users' facial expressions and the face gaze counting are provided by the vision system. The read values of number of spoken words, answers' correctness and time to answer are provided by the dialogue system.

2.4 User Model

The User Model, or Student Model module, stores information about every student that interacted with the robot. The information is regarding their first name, family name, age, school year, birthday, user's pictures and eight users interest sport, dance, team, music, toy, hobby, game and food. Transactions in the database can be manually operated by the designer in the corresponding window. The students' interests are supposed to be used in small talks at the time that it was previously set or autonomously, when a high frequency of bad readings (a high disattention level or a high number of wrong answers) is detected by the system. More details about this module can be seen in [6].

2.5 Evaluation

The Evaluation Module reports to the teacher the students' performance during the interaction with R-CASTLE. For each activity, the teacher can access

different graphs about the system assessment of each student or the whole class report. Another functionality is that the system can report the performance of Machine Learning algorithms' accuracy based on the teachers' validation.

Recorded videos made by the frontal camera of the robot are available to be watched later on. With the videos and the verbal answers stored, the sessions are available for as many reassessments as wanted in the tab "Off-Line Evaluation". Therefore, beyond the functionality of keeping the teachers aware of the students' performance, the Evaluation module can work as a laboratory to retest and maintain the Machine Learning methods in their high performance.

Once an efficient automation of the mechanical evaluation can save teachers' time, this module has a sensitive potential for helping IA in Education from a very popular perspective of pressing buttons and checking graphs, instead of analyzing raw data.

2.6 Dialogue

One of the most important communication forms of humans is the verbal one. Thus, the dialogue system has the biggest responsibility in keeping the communication flowing with efficiency throughout its interactions with the students. Speech recognition and Text-to-Speech are allowed through, respectively, python *SpeechRecognition* ²⁾ library and *Softbank Text-to-Speech API* ³⁾. If any NAO robot is available, the text to speech from Aldebaran is switch to the *Google Text-to-Speech* (gTTS) ⁴⁾ library.

The sentence matching algorithm [7] is an important part of the system because it evaluates the correctness level of the users' answers, based on the expected answers registered by the teachers.

2.7 Vision

As a primary responsibility, the vision system manages the device's camera and recognition algorithms. Secondly, it also sends information to the modules which will process this information. Images of users' faces, for example, are sent to the User Model system whereas the information of the users' facial expressions such as emotion and face gaze are used by the Adaptation module.

For the image recognition in the tasks, some Machine Learning methods are available. They are the Multilayer Perceptron (MLP) Networks, K-Nearest Neighbor (KNN), Support Vector Machines (SVM) and Convolutional Neural Networks (CNN). The teacher can choose which one of them will be used in the next session and change their parameters before training in the algorithm settings section. For user recognition, the system uses the Python Face Recognition ⁵⁾ The Haar Cascade algorithm, implemented in *OpenCV* library, is used for

²⁾ <https://pypi.org/project/SpeechRecognition/> Accessed in 30/01/2020

³⁾ <http://doc.aldebaran.com/2-1/naoqi/audio/alttexttospeech-tuto.html> Accessed in 30/01/2020

⁴⁾ <https://pypi.org/project/gTTS/> Accessed in 30/01/2020

⁵⁾ https://github.com/ageitgey/face_recognition Accessed in 20/01/2020.

face gaze detection, as described in [8]. Finally, the emotion recognition through facial expression is the result of a study where seven emotional states of Ekman's model were trained to be detected by a CNN [5].

3 Experiments

Experiments were conducted to evaluate the algorithms' performances and the users' perception regarding the system. Tests in laboratory consisted of testing several parameter variations in the methods and compare their results with the ones from the literature, whereas experiments with users consisted of several activities with teachers and students in an elementary school, as show in Figure 3. Results from both evaluations are presented in Section 4.

Tests performed in ideal conditions (laboratory) showed high and satisfactory results achieved by the used algorithms, as shown in [8]. Conversely, results of experiments in real conditions (school) showed a decreasing tin the system's performance. However, the lower achieved performance was considered yet acceptable, evaluated by the scores the students gave to the questionnaire after the sessions with the robot.

Experiments performed with teachers pointed out an optimistic time to get used to the R-CASTLE's. This conclusion was stated by teachers themselves in [12]. It was analyzed the teachers' perceptions regarding the system's usefulness, intuitiveness, potential, efficient, time to get used, learning enhancement and a comparison between R-CASTLE and the traditional methods.

Also, a high acceptance of the students to the system's performance was also reached in experiments with them [9]. The study analyzed the students perceptions of the system regarding their: enjoyment, learning, empathy, intelligence and content difficulty.



Fig. 3. Teachers (left picture) and students (middle and right pictures) participating in the experiments.

Furthermore, studies have been conducted testing the architecture with other devices as output. In one of them, two groups of children were analyzed comparing a NAO robot with a tablet in storytelling activities for foreign language

teaching [2]. In another experiment, two groups of graduate students were compared, with the difference of the system output as a NAO robot in a group and a low-cost domestic robot LARA [1].

Results from both studies showed the same performance of the architecture in both conditions, as well as in the classifications performance and in the students' perceptions of the system's efficiency using interactive devices that are less expensive than the currently available social robots. It means, R-CASTLE has potential to provide acceptable communication with users not only with high-tech robots but also with cheaper interactive devices.

4 Results and Discussion

As previously said, the objectives and hypothesis addressed in this thesis were divided into three aspects: its technical approaches, its impact on each user, and its impact on the educational process. Discussions of each objective and hypothesis - based on the experiment results - are presented in the following subsections.

4.1 Technical approaches

The objective (Obj) and hypothesis (H) regard to the technical issues in question are:

Obj I *To evaluate algorithms for automatic users' measure extraction and multimodal adaptation that take into account a teachers-friendly setup and Social Robotics issues.*

H1 *Algorithms without previous data dependency for multimodal adaptation present a good trade-off between intuitive setup and performance compared to supervised algorithms.*

Algorithms showed high performances in ideal conditions, with accuracy above 90% most of the time [8]. In real conditions - school environments - they showed acceptable performances, taking into account the challenges in facing noisy environments, but obviously their accuracy was reduced in such conditions [9, 11].

In relation to the adaptation algorithms, a Rule-Based and a Fuzzy Modeling were proposed and compared [11]. Both of them had their own advantages but they presented a very similar overall performance in dataset that was tested. Their performances were also compared to supervised ML methods. Supervised methods showed better performances, with the worst case being a greater F-1 of 20%. Although they had insignificant training time, they depended on acquired data to set their prediction models. Conversely, no data dependency and modeling as close as possible to human interpretation through linguistic variables and sets were the strongest features of the proposed methods.

Thereby, **Objective I** was achieved since the evaluations of the algorithms were presented and their advantages were explored by the R-CASTLE implementation. On the other hand, even though the **Hypothesis H1** points out encouraging initial outcomes, it lacks studies and evidence to affirm a real worthy trade-off between the analyzed methods, mainly from the teachers' point of view.

It's important to note that results obtained here were from a specific database. More studies and data are required to accomplish relevant generalization and analysis on this point.

4.2 Users analysis

The objectives and hypotheses regarding the analyses of the user's perceptions are:

Obj II *To present an easy way for teachers to manage their educational activities through technology.*

Obj III *To deliver a natural user-experience with social robots that motivates the students.*

H2 *Students would present higher scores in the personal assessment items regarding their enjoyment and rapport building in questionnaires after performing activities with systems using personalization skills compared to their last meeting with the robot.*

Participant teachers gave scores above the average (above 3 out of 5) to every item with regards to their impression about the system. A controversial point was noted in open questions, where 5 out of the 14 teachers (35%) said that the high amount of R-CASTLE variables may be a setback. They also said this could be resolved by using the system and getting used to deal with it [12].

However, the most important discovery was that the two teachers which used the R-CASTLE to perform their activities raised their scores by one point on items regarding the system's advantages of intuitiveness, the amount of time it took to get used to it and learning enhancement. This result leads to believe that other teachers would also increase their scores if they had also used the full system's potential in some of their activities. Unfortunately, the time demanded to achieve a complete study in this sense played against this thesis.

In the experiments with students [9], their motivation and natural interactions were observed. Students' motivation was intended to be autonomously measured by their eye gaze, however, lightning problems made it hard. By human observation, an excitement pointed out by the students' happy emotions and reactions was noted after personal interactions. One example of this is when the robot said it "tried the student's favorite food and it went pretty bad". This fact was also indicated by the high classification of happy emotion.

Furthermore, from one meeting to another, only 4 out of 36 students (11%) did not perform the activity again with the robot due to a loss of interest.

The rest of the class was labeled by the teachers as "joyful and engaged while interacting with the system".

Although a slight decrease in the students' enjoyment score in the second meet was noted compared to the first one, they still gave high scores to this item in the second interaction (a mean of 4.6 out of 5 in the worst case). On the other hand, the rapport building had a slight increase (even though it wasn't statistically significant). These outcomes suggests that the potential of the adopted techniques impacted the students' perceptions.

Hence, no evidence to support **Hypothesis H2** was found in this study, since the students' perception of rapport feeling decreased. Specific studies comparing these points with a larger number of participants would benefit a better analysis regarding H2. Yet, they still gave high scores for these items in the next section and they were able to communicate with the system and presented high expectations in interacting with the robot again, supporting the goals of **Objective III**. Teachers also showed high scores and the observed items regarding familiarization in dealing with R-CASTLE, suggesting the accomplishment of **Objective II**.

4.3 Impact on the educational process:

Obj IV *To increase the teaching experience by providing a scalable resource that can help teachers with personalized tutoring for their students.*

Obj V *To increase the learning experience through Personalized Learning and Social Robotic techniques.*

H3 *R-CASTLE would give more support and motivation to the teachers than traditional methods by assisting with planning, performing and evaluations.*

H4 *Students would present increasing scores in the items regarding learning statements in questionnaires after performing activities with the systems using adaptive behaviors compared to their last meeting with the robot.*

Significant findings were observed in the educational process. The system provided an interesting alternative to the teaching process since the teachers did note its support and potential in helping them in their activities. These are exciting outcomes of **Objective IV**.

Conversely, the results studying the **Hypothesis H3** were not supportive. Mainly because it was hard to measure how much of the teachers' workload was reduced and how much of their motivation was increased when using the system with such few interactions with it. However, two points are worthy to highlight. First, and again, teachers who used the system claimed to have more ease in their work when using R-CASTLE. In the open questions, 5 out of 14 teachers (35%) claimed that an increase in motivation was an important strength of working with the R-CASTLE.

A significant difference was found on the content difficulty: the students rated the second session more difficult than the first. However, they also increased their scores on perception and the robot's intelligence in their personal statements. Although these factors do not necessarily mean a learning gain, they are often

pointed out as factors strongly related to the improvement of the learning gain. The fact that teachers used the first activity’s feedback to program the second one seemed to influence the students’ experience in the second time as well. This is an inspiring observation since the R-CASTLE’s goal was to help teachers achieve a greater influence on their students through its features. Therefore, by analyzing the presented points it is possible to conclude that **Objective V** was achieved and that **Hypothesis H4** was supported.

Finally, a summary of this thesis’ objectives and hypotheses as well as their outcomes can be seen in Table 1.

Table 1. Summary of Objectives and Hypotheses.

Name	Analysis	Evaluation Method	Result	Reason
Obj I	Multimodal adaptation algorithms evaluation	Accuracy, training time, parameter configuration, execution time and application suitability	Achieved	Analysis and discussion of several algorithms were presented
Obj II	Teacher friendly system	Users’ answers to the survey: Time to get used, intuitiveness, open answers.	Achieved	Scores above neutral state
Obj III	Student-Robot natural communication	Users’ answers to the survey: Enjoyment, Rapport Building	Achieved	High scores achieved in the related items
Obj IV	Scalable resource	Users’ answers to the survey: System’s usefulness, System’s potential	Achieved	High scores achieved in the related items
Obj V	Personalized Learning and Social Robotics	Users’ answers to the survey: Learning perception, Robot’s Intelligence, Content difficulty	Achieved	High scores achieved in the related items
H1	Better advantages of algorithms set by the user	F-Measure (or F-1), modeling intuitiveness and training time.	Not Supported	Measures 20% lower in worst case
H2	Robot personalization in students experience	Users’ answers to the survey: Enjoyment, Rapport Building.	Not Supported	Measures have decreased
H3	Enhancement in the teaching process	Users’ answers to the survey: System’s usefulness, System’s potential	Supported with no statistical significance	Measures have increased
H4	Enhancement in the learning process	Users’ answers to the survey: Learning perception, Robot’s Intelligence, Content difficulty	Supported	Measures have increased

5 Conclusion

This paper presented the R-CASTLE proposal, a HRI framework for supporting educational activities in a autonomous and intuitive way. Experiments in laboratory presented high performance of the used algorithms whereas experiments with teachers and students showed a Overall, the project achieved the majority of their objectives and hypotheses.

As R-CASTLE provide features to HRI experiments in other areas, it has a good potential to also support researches in futures studies. Currently, it is being adapting to support teachers in remote classes during COVID-19 pandemic. It also provides a solutions to the classroom lessons when classes return, once the human-human interaction would be avoid until full safety.

References

1. Pinto, A.H.M., *et al.*: Users' perception variance in emotional embodied robots for domestic tasks. In: 2018 Latin American Robotic Symposium, 2018 Brazilian Symposium on Robotics (SBR) and 2018 Workshop on Robotics in Education (WRE). pp. 476–482 (Nov 2018). <https://doi.org/10.1109/LARS/SBR/WRE.2018.00090>
2. Tozadore, D.C., *et al.*: Tablets and humanoid robots as engaging platforms for teaching languages. In: 2017 Latin American Robotics Symposium (LARS) and 2017 Brazilian Symposium on Robotics (SBR). pp. 1–6 (Nov 2017). <https://doi.org/10.1109/SBR-LARS-R.2017.8215290>
3. Tozadore, D., Romero, R.: Comparison of image recognition techniques for application in humanoid robots in interactive educational activities. from portuguese: Comparação de técnicas de reconhecimento de imagens para aplicação em robô humanoides em atividades interativas educacionais. In: XXII Conferência Internacional sobre Informática na Educação, Fortaleza - CE (2017)
4. Tozadore, D., *et al.*: Wizard of oz vs autonomous: children's perception changes according to robot's operation condition. In: 2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN). pp. 664–669. IEEE (2017)
5. Tozadore, D., *et al.*: Effects of emotion grouping for recognition in human-robot interactions. In: 2018 7th Brazilian Conference on Intelligent Systems (BRACIS). pp. 438–443. IEEE (2018)
6. Tozadore, D., *et al.*: Towards adaptation and personalization in task based on human-robot interaction. In: 2018 Latin American Robotic Symposium, 2018 Brazilian Symposium on Robotics (SBR) and 2018 Workshop on Robotics in Education (WRE). pp. 383–389. IEEE (2018)
7. Tozadore, D., *et al.*: Matching sentences in semantic and syntax level for human-robot dialogues. To be published. (2019)
8. Tozadore, D., *et al.*: Project r-castle: Robotic-cognitive adaptive system for teaching and learning. IEEE Transactions on Cognitive and Developmental Systems **11**(4), 581–589 (2019)
9. Tozadore, D., *et al.*: When social adaptive robots meet school environments (2019). https://doi.org/https://aisel.aisnet.org/amcis2019/cognitive_in_is/cognitive_in_is/4/
10. Tozadore, D.C., Romero, R.A.F.: Facing the challenges of artificial intelligence in education through adaptive systems and social robotics. Computer and Education (2020)
11. Tozadore, D.C., Romero, R.A.F.: A fuzzy decision-making system using multimodal objective measures for educational adaptive systems. Expert Systems with Applications (2020)
12. Tozadore, D.C., Romero, R.A.F.: Graphical user interface for educational content programming with social robots activities and how teachers may perceive it. Revista Brasileira de Informática na Educação **28**, 191 (2020)