




## **Designing Pedagogic Conversational Agents through Data Analysis**

### **Diseño de Agentes Conversacionales Pedagógicos usando análisis de datos**

Silvia Tamayo-Moreno <sup>1</sup> y  
Diana Pérez-Marín  <sup>2</sup>

Recibido: 02 de agosto de 2019  
Aceptado: 12 de diciembre de 2019

---

#### Cómo citar / How to cite

S. Tamayo-Moreno, y D. Pérez-Marín, "Designing Pedagogic Conversational Agents through Data Analysis". *TecnoLógicas*, vol. 23, no. 47, pp. 243-256, 2020. <https://doi.org/10.22430/22565337.1455>



<sup>1</sup> PhD. en Ingeniería en Computación, Big Data, Stratio, Madrid-España, silviatamayomoreno@gmail.com  
<sup>2</sup> PhD. en Ingeniería de computación, Centro de inteligencia artificial, Universidad Rey Juan Carlos, Madrid- España, diana.perez@urjc.es

## Abstract

Pedagogical Conversational Agents are systems or programs that represent a resource and a means of learning for students, making the teaching and learning process more enjoyable. The aim is to improve the teaching-learning process. Currently, there are many agents being implemented in multiple knowledge domains. In our previous work, a methodology for designing agents was published, the result of which was *Agent Dr. Roland*, the first conversational agent for Early Childhood Education. In this paper, we propose the use of Data Analytics techniques to improve the design of the agent. Two new techniques are applied: KDDIAE, application of (Knowledge Discovery in Databases) to the Data of the Interaction between Agents and Students – Estudiantes in Spanish, and BIDAÉ (use of Data Analytics to obtain information of agents and students). The use of KDDIAE and BIDAÉ proves the existence of a fruitful relationship between learning analytics and learning design. Some samples of rules related to learning analytics and design are the following: (Learning Analytics) Children who initially do not know how to solve the exercise, after receiving help, are able to understand and solve it → (Learning Design) An agent for small children should be able to provide help. In addition, help should be entertaining and tailored to their characteristics because it is a resource that children actually use; or (Learning Analytics) Younger children use more voice interaction → (Learning Design) An agent interface for young children must incorporate voice commands. A complete list of rules related to learning analytics and design is provided for any researcher interested in PCA design. 72 children were able to use the new Dr. Roland after applying the learning analytics-design rules. They reported a 100 % satisfaction as they all enjoyed interacting with the agent.

## Keywords

Pedagogic Conversational Agent, Learning Analytics, Knowledge Discovery in Databases, Learning Design.

## Resumen

Los Agentes Conversacionales Pedagógicos son sistemas informáticos que facilitan la enseñanza a los estudiantes y un recurso de apoyo para los profesores haciendo el proceso de enseñanza más agradable. El objetivo es mejorar el proceso de enseñanza-aprendizaje. Actualmente, hay muchos agentes que se implementan en múltiples dominios de conocimiento. En nuestro trabajo previo se publicó una metodología para diseñar agentes. Con esta metodología se diseñó el agente Dr. Roland, el primer agente conversacional para Educación Infantil. En este artículo, se propone el uso de técnicas de análisis de datos para mejorar el diseño de Dr. Roland. Se implementan dos técnicas nuevas: KDDIAE, aplicación de KDD (Descubrimiento de Conocimiento en Bases de Datos) a los Datos de Interacción entre Agentes y Estudiantes, y BIDAÉ (uso de Análítica de Datos para obtener información de la interacción entre Agentes y Estudiantes). El uso de KDDIAE y BIDAÉ prueba la existencia de la fructífera relación que puede darse entre analítica de aprendizaje y diseño de aprendizaje. Los niños que inicialmente no saben cómo resolver un ejercicio, después de recibir ayuda son capaces de comprender el ejercicio y solucionarlo → (Diseño de Aprendizaje) Un agente para niños pequeños debería poder proporcionar ayuda. Además, la ayuda debería ser entretenida y adaptada a las características de los niños; o (Análítica de Aprendizaje) Los niños pequeños usan más la interacción por voz → (Diseño de Aprendizaje) Una interfaz de agente para niños pequeños debe incorporar la posibilidad de interactuar por voz. En este artículo se proporciona una lista completa que relaciona analítica y diseño de aprendizaje para cualquier investigador que pueda estar interesado en APC. 72 niños pudieron usar el agente mejorado Dr. Roland al implementar las reglas. Reportaron un 100 % de satisfacción, ya que todos disfrutaron de la interacción con el agente.

## Palabras clave

Agente Conversacional Pedagógico, Análisis de datos, conocimiento en bases de datos, Diseño de aprendizaje.

## 1. INTRODUCTION

Technology is present in various areas of society, which increases its relevance and use in different areas, including education [1]. We are in an evolutionary cycle in which school systems, institutions, and teaching should adapt to society's demands. The role of teachers in the technological evolution of education, specifically in classrooms, is fundamental.

In addition, to adopt new educational models, areas where there are deficiencies can be identified [2]. This paper focuses on an educational technology called Pedagogical Conversational Agents (PCAs). PCAs are interactive systems that allow students to study in an entertaining and friendly manner [3].

Let us look at examples of agents classified by role:

1. Teacher: Autotutor, an intelligent tutoring system that makes use of an animated conversational agent with facial expressions, synthesized speech, and gestures [4]; Laura for learning Spanish [5], [6]; and Willow [7], [8].

2. Student: Betty [9], Challenging Teachable Agent (CTA) [10], Lucy [11-13], and The Teachable Agent Math Game [14].

3. Companion: MyPet [15], SBEL [16], and Troublemaker [17].

In a previous study [18], we proved the benefits of using PCAs for education. However, methodologies for the design, integration, and evaluation of agents in the classroom are not found in the state of the art. For that reason, we proposed a methodology called MEDIE (stands for Methodology to Design) to design PCAs [18]. MEDIE was used to create the agent named Dr. Roland [18], which was the first agent to be used in Pre-Primary Education (see Fig. 1).

This paper focuses on applying Data Analytics techniques to improve the design of Dr. Roland. KDD (KDDIAE) and BIDAIE are used to obtain information from agents and students.

The fruitful relationship between learning analytics and learning design is proven with an experiment, in which 72 children could use the new Dr. Roland.

They reported a 100 % satisfaction as they all enjoyed interacting with the agent.

This article is organized in 5 sections: Section 2 is a review of the related work.

Section 3 presents KDDIE and BIDAIE.

Section 4 focuses on the results.

Finally, Section 5 presents the conclusions and future work.



Fig. 1. Sample snapshot of Dr. Roland [18].

## 2. STATE OF THE ART

### 2.1 Pedagogical Conversational Agents

Pedagogical Conversational Agents (PCAs) are integrated into learning environments, and they can be defined as interactive systems that allow students to study in an entertaining and friendly manner [3].

Figure 1 shows a sample interface of an agent. Agents can be represented as people, animals, or things that can speak with sound or text. Agents are used in a wide range of applications, such as commercial enterprises, health, employee training, or education [19]. Due to the importance of technology in education, the integration of agents into learning environments is gaining increasing attention [20].

As the use of PCAs increases, understanding how to design these characters for teaching and learning becomes more relevant.

Designing Pedagogical Conversational Agents remains a challenge that has not been completely solved [21]. A bad design generates negative feelings in students, hinders communication and interaction, and, ultimately, the completion of learning tasks [20]. Therefore, design methodologies for PCAs should be improved.

### 2.2 Data Analytics

Data analytics can be defined as “high volume, high speed and/or wide variety of information assets that require new forms of processing to enable improved decision making, discovery of knowledge and optimization of processes” [22], [23].

The application of Data Analytics in education can be considered an excellent resource due to the possibilities it offers to analyze, visualize, understand, and even improve education. The traditional method of observation in the classroom is losing momentum as the most effective way to

understand and improve the educational process. As a result, other options such as Data Analytics are incorporated.

From Data Analytics, and its integration with different technologies, a series of methods have been derived. Such methods are being applied in the educational field. They include adaptive learning, which is based on the modification of the contents and forms of teaching according to the particular needs of each student; competency-based education [24], where the learning process is adapted to the pace and needs of each student; inverted classroom and Blended Learning [25], based on independent study and classroom practice; and gamification, which consists of using game mechanics in learning environments to stimulate the teaching-learning process among members of a student community [26].

## 3. KDDIAE AND BIDAÉ

Two analysis proposals are applied to PCAs: KDDIAE, application of KDD (Knowledge Discovery in Databases) to the Data of the Interaction between Agents and Students, and BIDAÉ (use of Data Analytics to obtain information of agents and students – Estudiantes in Spanish).

For KDDIAE (see Fig. 2), the following is an explanation of the steps applied to the interaction data between PCAs and students, following the practical approach of the process proposed by Brachman and Anand [25]:

1. Understand the application domain, relevant prior knowledge, and process goal identification. At this point, it is essential to have people with experience in the application area of PCAs (and all the areas involved) and, if possible, involve them in the process to better understand the context and the factors that may affect it and better interpret the results that should be obtained.



Fig. 2. KDDIAE. Source: Created by the authors.

2. Selection of the dataset, identification of target variables to be predicted, calculated or inferred and independent variables useful to calculate, process, or sample available records taking into account the conversational pedagogical agent and its characteristics, as well as those of the people with whom it interacts and the context in which occurs.
3. Data cleaning and pre-processing. The data should present significant values, and decisions should be made regarding noise (missing, atypical, or incorrect values). Otherwise, introducing the raw data into a data mining algorithm would lead to difficult learning processes or results that do not represent the real behavior.
4. Reduction and projection of data. It aims to identify the most significant characteristics to represent the data, depending on the purpose of the process. For this reason, transformation processes can be used to reduce the effective number of variables or find other representations of the data.
5. Establish an adjustment of the goals of the KDD process (Step 1) with a particular data-mining method.
6. Explanatory analysis and selection of hypotheses and model. Researchers should decide which models may be

suitable according to the general aim of the process.

7. Data mining. In this step, pattern search is performed over a given form of representation or set of representations.
8. Pattern interpretation, with a possible return to steps 1 through 7 for a complete iteration. The process can be feedback, repeating itself from the beginning or from any of the steps, until a valid model is obtained.
9. Acting on discovered knowledge, the model is ready for operation when it is considered to be acceptable with suitable outputs and/or permissible error margins.

BIDAE (see Fig. 3) aims to provide a practical vision of a process that can be followed to obtain information about the

interaction between PCAs and students based on the objective established at the beginning.

The type of information that BIDAE can serve to process could be: (i) to find out if variables (characteristics) that were initially thought to have a certain impact actually have it; (ii) to try to discover other variables that a priori were not considered significant in some respect; and, (iii) to try to reveal certain behaviors that maybe (based on the results) allow to generalize them, relationships between variables, or other aspects that are considered.

A first set of data related to PCAs (e.g., questionnaires) or results of the interaction of the agent with the students in its area of application should be available to start the process, and the following steps are followed:

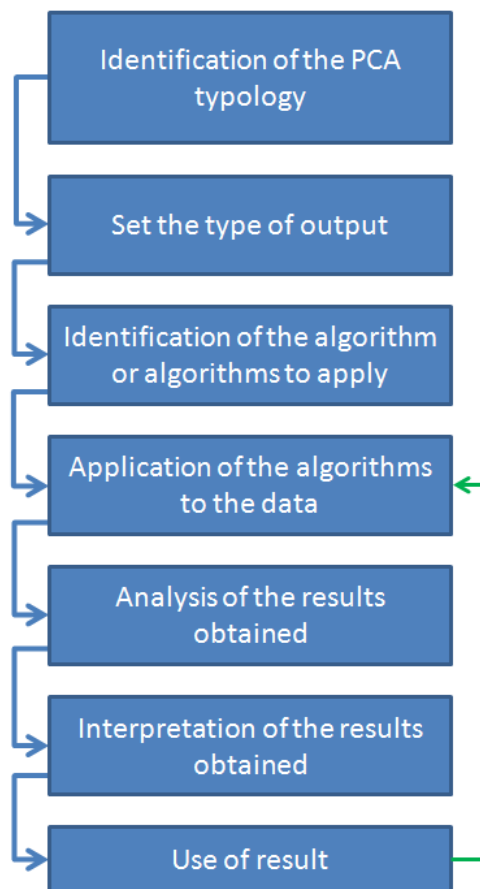


Fig. 3. BIDAE. Source: Created by the authors.

1. Identification of the type of Pedagogical Conversational Agent in question in order to determine which techniques may be better for the results to be obtained. Any existing taxonomy can be used for this, such as the proposal in Pérez-Marín [7], which classifies agents using ten criteria.
2. Selection of the type of output in order to define the group of algorithms that best fits the data. For instance, one of the types of algorithms that are grouped in a taxonomy according to their output can be selected.
3. Identification of the algorithm or algorithms to apply in the group of algorithms identified in the previous step.
4. Application of the corresponding procedure of the algorithms to the data depending on the algorithm in question.
5. Analysis of the results obtained.
6. Interpretation of the results obtained and actions to be taken as a result.
  - a. If the results are as expected, continue to work using that information to provide feedback to the agent.
  - b. If the results are not as expected, analyze what may be failing by focusing on the following aspects: if the data obtained are sufficient and adequate to obtain the result that is intended, if the appropriate algorithm set is being applied, if the particular algorithm or algorithms in that previously selected group is suitable, or if the algorithm is being applied properly.
    - i. If the data with which the analysis is conducted are not correct or insufficient, the following steps are needed: (i) re-analyze what data are needed to obtain the desired result, (ii) identify the data to be captured, (iii) analyze what changes should be made in the conversational agent to capture that data, (iv) modify the agent to correct the capture of the erroneous data and/or to capture the new ones that are necessary.
    - ii. If the appropriate algorithm set is not being applied, the following steps are needed: (i) analyze which set of algorithms is best adapted by examining its characteristics and those of the conversational agent, (ii) select that set of algorithms to perform the analysis.
    - iii. If the particular algorithm or algorithms in that previously selected group is not suitable, the following steps are needed: (i) analyze which algorithm might be appropriate according to its characteristics, (ii) select an algorithm, (iii) apply this algorithm.
    - iv. If the algorithm is not being applied properly, the following steps are needed: (i) study how the algorithm is applied, (ii) repeat the process and apply the algorithm.
7. Use of result
8. If the results are as expected, the following steps are:
  - a. Analyze what additional or new knowledge should be included.

- b. Identify what data is needed for it.
  - c. Analyze the agent and its characteristics to identify what modifications could be made to adapt it to the aim.
  - d. Make the modifications in the agent.
  - e. Repeat the experimental phase for data collection.
  - f. When sufficient data are available, return to step one of the process.
9. If they were not as expected although the process in Step 6 was followed, return to Step 4.

#### 4. RESULTS

MEDIE has been implemented to design Dr. Roland for several education levels. In this study, Dr. Roland was used for Pre-Primary Education. Figure 1 shows a sample snapshot of Dr. Roland with big buttons, a colorful interface, and many multimedia elements such as images, videos, and sounds. Given that young children cannot read, all the written text is also spoken, and the keyboard is shown as requested by teachers, who want to introduce children aged 5–6 to reading and writing.

Dr. Roland was used by 72 children to learn about animals in three schools in Spain. In the first school, there were three sessions in which Dr. Roland was used by 24 children (ages 4–5 years). In the second school, there were two sessions in which Dr. Roland was used by 25 children (ages 2–3 years). In the third school, there were two sessions in which Dr. Roland was used by 23 children (ages 2–3 years). KDDIAE and BIDAIE were used to design the agent.

The procedure to apply KDDIAE followed the steps described in Section 3.

Since the beginning, Pre-Primary Education teachers were involved in the design of Dr. Roland, providing their

expertise and advice so that programmers could understand the learning goal (in this case in the field of natural sciences; in particular, animals).

The dataset in this study is composed of students' answers, student-agent interaction, and student-agent conversations. The target variables are the influence of the agent's help to improve student's ability to solve and/or understand the exercises, the dialogue that should be used to get the best results, and average times. The independent variables are exercises, questions, type of questions, methods of interaction and solution, help, key dialogue structures to understand exercises, and interface features.

The data was processed to identify the most significant characteristics in this case. Thus, we found the most relevant ones: exercise duration, number of actions between the user and the agent, average time between actions, type of interaction (spoken or written), help used, exercise understanding, completed exercises, attempted exercises, solved exercises, number of attempts to solve an exercise, if once users have been given help they are able to solve an exercise, if once an exercise is completed students keep working with the agent, if the students receive some explanation at the end of an exercise, and if students see the correct answers to exercises. All the times were measured in seconds.

The method adopted in this study is an Expectation–Maximization algorithm [27].

The goal is to establish if the agent captures the information as intended by the teachers, if this information is meaningful, if there is other information that should be included, or which information that has not been captured by the agent should be captured in the following design cycle of the agent. That information can be used to improve the design of the agents. In this way, the information provided by teachers and Data Analytics techniques would confirm the



successful relationship between data analytics and the design of educational computer systems.

Table 1 lists the variables defined to obtain information for the design of the agent as explained in [27]. (In the previous paper, the focus was on gathering data.

The focus of this paper is on the fruitful relationship between learning analytics and learning design.).

The number of clusters was 2, cluster (or group) 0 and cluster (or group) 1, using the variables in Table 1. From the results, it can be deduced that group 1 is the largest, concentrating 78 % of the information. Regarding numerical

attributes, the information provided focuses on the mean and standard deviation. Table 2 presents the most significant results produced with both groups to find design rules.

As can be seen in Table 2, there is a significant variability in the duration of the exercises of both groups. Regarding the number of attempts to resolve the exercise, there is a fairly homogeneous situation in both groups: it is close to 1 and there is a relatively small dispersion in each group.

In relation to the number of actions of the agent with the user, there are differences between the groups that represent the largest deviation.

Table 1. Variables for the analysis. Source: Created by the authors.

Variable	Description
Exercise-duration	Duration of each exercise in seconds
Number-actions-student-agent	Number of actions between the agent and the student during an exercise
Average-time-between-actions	Average time between actions in seconds
Type-interaction	Using a keyboard or the voice
Help	Flag to indicate whether the student has received help
Initial-understanding	Flag to indicate whether the student has understood the exercise as it was initially shown to him/her
Ending-exercise	Flag to indicate whether the student has been able to finish the exercise
Resolution	Variable to indicate whether the student has tried to solve the exercise (without taking into account if it has correctly been solved). The possible values for this variable are yes, no, or N/A (not applicable).
Resolution-correct	Variable to indicate whether the exercise has been correctly solved. The possible values for this variable are yes, no, or N/A (not applicable).
Resolution-attempts	Number of times the exercise has been attempted
Understanding-after-help	Flag to indicate whether, once the student has been given help, s/he is able to understand the exercise
Number-repetition-exercise	Variable to count the number of times that the exercise has been repeated
Continuation-work-with-agent	Flag to indicate whether, once the student has finished the exercise, s/he keeps working with the agent
Explanation-resolution	Flag to indicate whether the user has received the explanation of how to solve the exercise
Show-right-response	Flag to indicate whether the student has seen the correct answer of the exercise

Table 2. Most relevant results of the EM analysis. Source: Created by the authors

Variable	Group 0	Group 1
Exercise-duration	173.23	56.66
Number-actions-student-agent	11.72	6.66
Type-interaction	68.44 % voice, rest keyboard	81.71% voice, rest keyboard
Help	84.2 % received help	23.1 % did not receive help
Initial-understanding	83.81 % did not understand	87.66 % understood
Resolution	63.45 % solved the exercise	89.41 % solved the exercise
Resolution-attempts	1.56	1.04
Understanding-after-help	70.74 % understanding after help	69.43 % understanding after help
Explanation-resolution	31.56 % used the explanation	14.63 % used the explanation
Show-right-response	14.61 % showed the correct answer	3.53 % showed the correct answer

With regard to interaction type, the voice predominates in both groups. Hence, this population should be characterized using their voices (which makes sense since young children cannot write well).

There is a difference in the initial understanding: group 1 is able to understand at the beginning, while group 0 is not. This is also related to reception of help since group 0 exhibits a higher value of reception of help, while group 1 did not receive as much help. As for the understanding after the help, group 0 has a relatively higher value, which makes sense since they initially did not understand, and the fact that they received the corresponding help implies an improvement in understanding. The children in both groups solved the exercises correctly in most cases; therefore, the options of explaining the solution and showing the correct response were used less.

The patterns were interpreted and the algorithms were applied in the previous step. Consequently, what was said in terms of interpretation in Step 7 is extrapolated to Step 8. In addition, in the light of the results, the provided information is considered sufficient for the initial aim; therefore, it is not necessary to go back.

With the information that can be obtained from the model, a detailed analysis of the variables is carried out to

reveal which relationships should be maintained and/or strengthened, which changes can be made in those that are unrelated and what is wanted, or what new variables, information or relationships should be obtained. Thus, taking into account the characteristics of the agent, the algorithm analyzes and identifies the necessary changes and how they would have to be applied to the agent in the next phase of the iterative and incremental development process.

BIDAE is applied to data from agent Dr. Roland and children in Pre-Primary Education. In the taxonomy of roles of conversational pedagogical agents, Dr. Roland has a teacher role, since it tries to have students learn, it teaches them.

The selection of the group of algorithms is supported because the conversational pedagogical agent uses the responses and reactions of students at each interaction to produce the following response; to adapt to students' actions, reactions, needs, and responses; and to offer appropriate responses in the interaction. In this study, classification algorithms will be applied as a refinement strategy in the analysis.

The evaluation can be conducted using different options: a training set (on the same set on which the predictive model is built to determine the error), a supplied test set (on a separate set), cross-validation, or

percentage split (dividing the data into two groups, according to the indicated percentage).

In this case, we used the cross-validation mode, dividing the instances into as many folders as the folds parameter indicates (10); and, in each evaluation, the data of each folder is taken as test data, and the rest of the data is training. The calculated errors are the average of all the executions. Among the possible algorithms, a decision tree type C4.5-J48 will be applied to predict attributes.

For the following variables, no results were obtained: duration of the exercise (exercise-duration), number of agent actions with the user during interaction with an exercise (number-actions-student-agent), average time between actions (time-average-between-actions), number of times that the exercise is performed (number-repetition-exercise), and attempts to solve the exercise (resolution-attempts).

In relation to the analysis of the data obtained and the confusion matrices, the diagonal values are the successes and the others are errors. In this way, the percentage of the total number is known for each value (how many were well classified and how many with errors).

As for relations, their result and interpretation can be seen in their corresponding column. For example, all those who did not initially understand, except in a case of error, understood once they received the help (understanding-after-help = yes: no (11.0 / 1.0)). Moreover, no participant, after receiving the help, did not understand (understanding-after-help = no: yes (0.0)). The comprehension question after help did not apply to those who initially understood (understanding-after-help = na: yes (25.0 / 2.0)). All those who requested help understood after the help (understanding-after-help = yes: yes (11.0)) (no student claimed s/he did not understand the exercise once s/he requested help (understanding-after-help = no: no (0.0))).

Those who completed the exercise (ending-exercise), except one case of error, did so in two or fewer attempts to solve it (resolution-attempts <= 2: yes (34.0 / 1.0)).

The exercises of those who kept working with the agent (continuation-work-with-agent), except for three cases of error, lasted less than 196 seconds (exercise-duration <= 196: yes (34.0 / 3.0)) (for the others greater (exercise-duration > 196: no (2.0))). Those who understood after the help (understanding-after-help), having requested it (help = yes), except for an error case, performed more than 7 actions with the agent (number-actions-user-agent > 7: yes (12.0 / 1.0)). Those who did not request to show the correct response (show-right-response) made 2 or fewer attempts to solve the exercise (resolution-attempts <= 2: no (34.0)). Therefore, in Steps 6 and 7, all of this will be incorporated into the agent.

Regarding the interpretation of the results and the actions to be performed accordingly, the algorithm will use those expected results to further consolidate and use that information in order to provide the agent with feedback.

As for the incorrect and insufficient data, the steps that should be taken are the following: analyze again what data is needed to obtain the desired result, identify the data to be captured, analyze what changes need to be made in the conversational agent to capture that data, modify the agent to correct the capture of the erroneous data, and/or capture new ones as necessary.

Regarding the use of the results, the following steps should be taken: analyze what additional or new knowledge should be included, identify what data is needed for it, analyze the agent and its characteristics to identify what modifications could be made to adapt it to the intended aim, make modifications to the agent, and repeat the experimental phase for data collection.

## 5. CONCLUSIONS

The use of Data Analytics techniques has allowed us to design a Pedagogic Conversational Agent for Pre-Primary Education following the MEDIE methodology. The use of KDDIE and BIDAÉ has highlighted the fruitful relationship between Learning Analytics and Learning Design, as shown below:

-(Learning Analytics) Children who initially do not know how to solve the exercise, after being given help, are able to understand it and solve it → (Learning Design) An agent for small children should be able to provide help. Furthermore, help should be entertaining and adapted to their characteristics because it is a resource that children actually use.

-(Learning Analytics) Younger children use more voice interaction → (Learning Design) An agent interface for small age children must incorporate voice commands.

-(Learning Analytics) Children who completed the exercise needed two or fewer attempts to solve it → (Learning Design) Students should be given the possibility of solving the exercises and trying again if they want to.

-(Learning Analytics) Most children who solved correctly the exercises used the option that explained how to solve the exercise, and less of them showed the correct answer → (Learning Design) Agents should allow a flexible use of the explanation and visualization of the correct response depending on the case.

-(Learning Analytics) Children who initially did not understand and, therefore, used more help and a greater number of interactions with the agent spent more time in each exercise than those who understood from the beginning → (Learning Design) Exercises should not be too long.

-(Learning Analytics) Children who initially understood used less help → (Learning Design) Agents should allow a flexible use of the help depending on the case.

-(Learning Analytics) Children who understood the exercise after the help, having requested it, performed more than 7 actions with the agent → (Learning Design) Providing help where and when it is necessary is an important part of the process.

-(Learning Analytics) Children who did not know how to solve the exercise used the option to show the correct answer → (Learning Design) It is important to include, in the exercises, the correct answer, and, if possible, a visible explanation of it for children.

(Learning Analytics) The duration of the exercise was less than 196 seconds → (Learning Design) The duration of each exercise should not be very long.

One hundred percent of the 72 Spanish children (aged 2–5 years) who used Dr. Roland claimed that they enjoyed interacting with the agent. Since the MEDIE proposes an iterative and incremental process, the comments of teachers and students will be applied to future versions of the agent.

Nevertheless, these procedures suffer from a general disadvantage: their dependency on the data provided by teachers and students. This is because, in some cases, it may be difficult to obtain enough information to apply the procedures. This is another research line for future work.

## 7. REFERENCIAS

- [1] J. Salinas, “Innovación docente y uso de las TIC en la enseñanza universitaria,” *RUSC. Univ. Knowl. Soc. J.*, vol. 1, no. 1, pp. 1–16, Nov. 2004.  
<https://doi.org/10.7238/rusc.v1i1.228>
- [2]. B. Garza González and A. G. Solís Hernández, “Uso Pedagógico de las TIC en el Aula / Pedagogical use of ICT in the classroom,” *RECI Rev. Iberoam. las Ciencias Comput. e Informática*, vol. 1, no. 2, pp. 19-37, Jul. 2014.  
<https://doi.org/10.23913/reci.v1i2.13>
- [3] W. L. Johnson, J. W. Rickel, and James C. Lester, “Animated Pedagogical Agents: Face-

- to-Face Interaction in Interactive Learning Environments,” *Int. J. Artif. Intell. Educ.*, vol. 11, pp. 47–78, 2000. Available: [URL](#)
- [4] N. K. Person and A. C. Graesser, “Designing AutoTutor to be an Effective Conversational Partner,” in *Fourth International Conference of the Learning Sciences*, Michigan, 2000, pp. 246–253. Available: [URL](#)
- [5] K. Theodoridou, T. Yerasimou, “Learning Spanish with ‘Laura’: The Role of an Intelligent Agent in a Spanish Language Course,” in *Proceedings of ED-MEDIA 2008--World Conference on Educational Multimedia, Hypermedia & Telecommunications*, Waynesville, 2008, pp. 4907–4912. Available: [URL](#)
- [6] K. D. Theodoridou, “Learning with Laura: Investigating the Effects of a Pedagogical Agent on Spanish Lexical Acquisition,” (Tesis Doctoral), The University of Texas at Austin, 2009. Available: [URL](#)
- [7] D. Pérez Marín, “Uso de agentes conversacionales pedagógicos en sistemas de aprendizaje híbrido (b-learning),” *Actas del IV Semin. Investig. en Tecnol. la Inf.*, vol. 79–94, 2011. Available: [URL](#)
- [8] I. Pascual Nieto, “Una metodología para gestión de la interacción entre los estudiantes, los profesores y el contenido en aplicaciones en línea de Aprendizaje Híbrido usando modelos conceptuales,” (Tesis Doctoral) Universidad Autónoma de Madrid Escuela Politécnica Superior, Madrid, 2009. Available: [URL](#)
- [9] K. Leelawong and G. Biswas, “Designing Learning by Teaching Agents: The Betty’s Brain System,” *Int. J. Artif. Intell. Educ.*, vol. 18, no. 3, pp. 181–208, 2008. Available: [URL](#)
- [10] C. Kirkegaard, A. Gulz, and A. Silvervarg, “Introducing a Challenging Teachable Agent,” in *Learning and Collaboration Technologies. Designing and Developing Novel Learning Experiences. LCT 2014*, Cham: Springer, 2014, pp. 53–62. [https://doi.org/10.1007/978-3-319-07482-5\\_6](https://doi.org/10.1007/978-3-319-07482-5_6)
- [11] N. Matsuda, W. W. Cohen, K. R. Koedinger, G. Stylianides, V. Keiser, and R. Raizada, “Tuning Cognitive Tutors into a Platform for Learning-by-Teaching with SimStudent Technolog,” in *1st APLEC Workshop Proceedings*, 2010. Available: [URL](#)
- [12] N. Matsuda *et al.*, “Cognitive anatomy of tutor learning: Lessons learned with SimStudent,” *J. Educ. Psychol.*, vol. 105, no. 4, pp. 1152–1163, Jan. 2013. <https://doi.org/10.1037/a0031955>
- [13] N. Matsuda, W. W. Cohen, and K. R. Koedinger, “Teaching the Teacher: Tutoring SimStudent Leads to More Effective Cognitive Tutor Authoring,” *Int. J. Artif. Intell. Educ.*, vol. 25, no. 1, pp. 1–34, Mar. 2015. <https://doi.org/10.1007/s40593-014-0020-1>
- [14] L. Pareto, “A Teachable Agent Game Engaging Primary School Children to Learn Arithmetic Concepts and Reasoning,” *Int. J. Artif. Intell. Educ.*, vol. 24, no. 3, pp. 251–283, Sep. 2014. <https://doi.org/10.1007/s40593-014-0018-8>
- [15] Z.-H. Chen, C. C. Y. Liao, T.-C. Chien, and T.-W. Chan, “Nurturing My-Pet: Promoting Effort-Making Learning Behavior by Animal Companions,” in *In 16 th International Conference on Computers in Education, Nurturing*, 2008, pp. 27–34. Available: [URL](#)
- [16] E. Reategui, E. Polonia, and L. Roland, “The role of animated pedagogical agents in scenario-based language e-learning: a case-study,” in *Conference ICL2007*, Villach, 2007, pp. 1–8. Available: [URL](#)
- [17] E. Aimeur and C. Frasson, “Analyzing a new learning strategy according to different knowledge levels,” *Comput. Educ.*, vol. 27, no. 2, pp. 115–127, Sep. 1996. [https://doi.org/10.1016/0360-1315\(96\)00018-8](https://doi.org/10.1016/0360-1315(96)00018-8)
- [18] S. Tamayo- Moreno, “Propuesta de Metodología para el Diseño e Integración en el Aula de un Agente Conversacional Pedagógico desde Educación Secundaria hasta Educación Infantil,” (Tesis Doctoral), Universidad Rey Juan Carlos de Madrid, Madrid, 2017. Available: [URL](#)
- [19] A. Kuz, M. Falco, L. Nahuel, and R. Giandini, “Agent SocialMetric: herramienta de asistencia al docente para determinar el clima social y la estructura del aula,” *IE Comun. Rev. Iberoam. Informática Educ.*, no. 22, pp. 16–29, 2015. Available: [URL](#)
- [20] G. Veletsianos, C. Miller, and A. Doering, “Enali: A Research and Design Framework for Virtual Characters and Pedagogical Agents,” *J. Educ. Comput. Res.*, vol. 41, no. 2, pp. 171–194, Oct. 2009. <https://doi.org/10.2190/EC.41.2.c>
- [21] S. van Vuuren, “Technologies that power pedagogical agents and visions for the future,” *Educ. Technol.*, vol. 47, no. 1, pp. 4–10, 2007. Available: [URL](#)
- [22] M. Beyer and D. Laney, “Gartner Research,” *The Importance of “Big Data”: A Definition*. 2012, ID: G00235055, Available: [URL](#)
- [23] O. Maimon and L. Rokach, *Data Mining and Knowledge Discovery Handbook*. Springer., 2005. Available: [URL](#)
- [24] Y. Vasquez, “Educación basada en competencias,” *Educ. Rev. Educ. nueva epoca*, no. 16, p. 1, Mar. 2001. Available: [URL](#)
- [25] J. F. Strayer, “How learning in an inverted classroom influences cooperation, innovation and task orientation,” *Learn. Environ. Res.*,

- vol. 15, no. 2, pp. 171–193, Jul. 2012.  
<https://doi.org/10.1007/s10984-012-9108-4>
- [26] M. Area Moreira and C. S. González González, “De la enseñanza con libros de texto al aprendizaje en espacios online gamificados,” *Educ. Siglo XXI*, vol. 33, no. 3, pp. 15-37, Nov. 2015.
- <https://doi.org/10.6018/j/240791>
- [27] A. P. Dempster, N. M. Laird, and B. Rubin, “Maximum Likelihood from Incomplete Data via the EM Algorithm,” *J. R. Stat. Soc. Ser. B*, vol. 39, no. 1, pp. 1–38, 1977. Available: [URL](#)