

**Silvia Tamayo Moreno**  
**Diana Pérez Marín**

**Evaluación del Agente  
Conversacional Pedagógico  
Dr. Roland para el nivel de  
Educación Infantil mediante el  
método de medida de las  
prestaciones**

**Número 2017-10**

**Serie de Informes Técnicos DLSI1-URJC**  
**ISSN 1988-8074**  
**Grupo Docente de Lenguajes y Sistemas Informáticos I**  
**Universidad Rey Juan Carlos**



## Índice

|     |   |    |
|-----|---|----|
| 1   | Introducción.....   | 5  |
| 1.1 | Medida de las prestaciones, método de evaluación de la Interacción Persona Ordenador.....                       | 5  |
| 2   | Evaluación del agente Dr. Roland para Educación Infantil mediante el método de medida de las prestaciones ..... | 7  |
| 3   | Conclusiones.....   | 13 |
|     | Referencias .....   | 13 |



# Evaluación del Agente Conversacional Pedagógico Dr. Roland para el nivel de Educación Infantil mediante el método de medida de las prestaciones

Silvia Tamayo Moreno y Diana Pérez Marín

Dto. de CC Comp., Arq. Comp., LSI y EIO, Universidad Rey Juan Carlos,  
C/ Tulipán s/n, 28933, Móstoles, Madrid  
silviatamayomoreno@gmail.com, diana.perez@urjc.es

**Abstract.** In this paper the evaluation of the pedagogical conversational agent Dr. Roland is carried out for the level of Pre-Primary Education following the performance measurement method. The logging method is one of the test methods of the techniques of evaluation of the Human Computer Interaction.

**Keywords:** Medida de las prestaciones, Agente Dr. Roland, Educación Infantil, Interacción persona Ordenador.

## 1 Introducción

En primer lugar se va a hacer alusión a la interacción persona ordenador (IPO). Seguidamente a técnicas de evaluación de la IPO [1], centrando la atención en los métodos de test. Concretamente en el método de medida de las prestaciones.

A continuación se realizará la evaluación del Agente Conversacional Pedagógico Dr. Roland [11, 12] para el nivel de Educación Infantil usando el método de test de medida de las prestaciones.

### 1.1 Medida de las prestaciones, método de evaluación de la Interacción Persona Ordenador

Los sistemas interactivos son dispositivos hardware y software que actúan de interconexión entre las personas, permitiéndoles interactuar, y favoreciendo la realización de tareas y el alcance de los objetivos propuestos.

Con el desarrollo de los sistemas interactivos, y para tener una mejor comprensión de las relaciones persona-ordenador y mejorarla, surge la interacción persona ordenador (IPO). La IPO, conocida en la comunidad internacional como Human-Computer Interaction (HCI) o Computer-Human Interaction (CHI), según el grupo SIGCHI (Special Interest Group on Computer-Human Interaction, [2]) de ACM: “*Es la disciplina relacionada con el diseño, evaluación e implementación de sistemas informáticos interactivos para el uso de seres humanos, y con el estudio de los fenómenos más importantes con los que está relacionado*” [1].

Constantemente se presenta la situación en la que se diseñan productos que no se prueban con quienes van a ser los usuarios, sin prestar atención por tanto a la usabilidad de los productos y el software, como a las pruebas con usuarios. Los motivos suelen ser los costes y cambios que pueda suponer dicha evaluación. Se centran los esfuerzos y atención en la funcionalidad, y se tiene el pensamiento de que si quienes diseñan y programan, pueden usar el sistema y les gusta, es suficiente [3].

Evaluar algo implica hacer pruebas de funcionamiento y comprobar si cumple las expectativas, necesidades y requisitos. En el ámbito de los sistemas interactivos es clave y es importante que se realice durante todo el proceso de desarrollo. Los resultados de ello, obtenidos mediante la aplicación de una serie de técnicas, se pueden emplear para mejorar los sistemas. Lorés et al. [4] define la evaluación como “*La actividad que comprende un conjunto de metodologías y técnicas que analizan la usabilidad y/o la accesibilidad de un sistema interactivo en diferentes etapas del ciclo de vida del software*”.

Los prototipos se hacen para comprobar si el sistema reúne características que lo hagan usable para los usuarios finales. La evaluación del diseño conlleva comprobar si se adapta a lo esperado, permitiendo realizar las tareas, y comprobar el impacto en el usuario en aspectos como la facilidad de aprendizaje, manejabilidad, identificación de áreas que necesiten que sea recordada gran cantidad de información, etc. Con ello se identifican posibles problemas específicos del diseño, como todo aquello que cause resultados no esperados o confusos al usuario estando en su contexto de uso [5]. Según Dix [6] los objetivos de la evaluación son la comprobación de la funcionalidad del sistema, del efecto de la interfaz y la identificación de problemas específicos del sistema.

Los sistemas interactivos pueden ser evaluados en cualquier contexto que lo permita, donde se reúnan usuarios y evaluadores, desde espacios equipados como laboratorios, salas de reuniones o hasta en el entorno en el que los usuarios desarrollan sus tareas.

El abanico de métodos de para evaluar los diferentes aspectos es muy amplio, existiendo solapamiento entre algunos de ellos en cuanto a las actividades que desarrollan, no son independientes [7]. La elección de unos u otros depende de los costes y lo que se obtendrá con su uso.

**En cuanto al tipo de técnica de comprobación utilizada,** se distinguen:

- **Métodos de inspección:** Agrupa una serie de métodos que cuentan con expertos, llamados evaluadores, que explican el grado de usabilidad de un sistema. Se basan en la inspección de aspectos de la interfaz del sistema relacionados con la usabilidad ofrecida a los usuarios. Se considera opiniones, juicios o informes de los inspectores sobre elementos específicos de la interfaz como factor fundamental de la evaluación [8]. Permiten identificar, clasificar y contabilizar un gran número de errores potenciales de usabilidad a precio relativamente bajo, contribuyendo a ello el hecho de no usar usuarios [8, 9].

- **Métodos de indagación.** Consiste en alcanzar el conocimiento de una cosa mediante conjeturas y señales. Se llevan a cabo, hablando con los usuarios, observándolos, usando el sistema en tiempo real y observando las respuestas a preguntas, ya sea orales o escritas. La información sobre gustos del usuario, necesidades o identificación de requisitos es importante, especialmente en fases tempranas del proceso de desarrollo.
- **Métodos de test:** En ellos usuarios, normalmente representativos, usan el sistema o prototipo para tareas concretas. Los evaluadores usan los resultados para saber la forma en la que la interfaz de usuario da soporte a los usuarios con sus tareas.

Como parte de los métodos de test se encuentra el de medida de las prestaciones.

### **Medida de las prestaciones**

Este método se basa en tomar medidas sobre el rendimiento y otros tipos de aspectos subjetivos que afecten la usabilidad del sistema. Es importante la comprensión de lo que se pretende medir, se pueden seleccionar:

- Medidas de rendimiento, son medidas cualitativas, se puede contar las acciones, los comportamientos, número de personas, de errores, etc.
- Medidas subjetivas, es decir, percepciones de las personas, opiniones y juicios. Pueden ser cuantitativas o cualitativas.

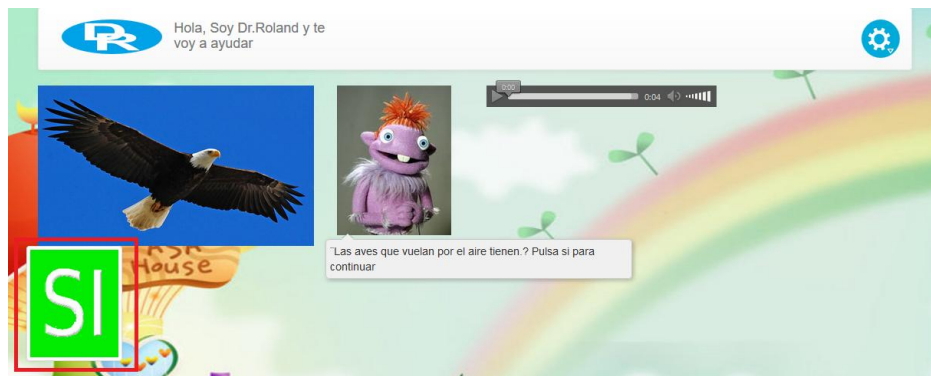
Por ello, se necesita disponer de un sistema implementado o un prototipo que permita la evaluación de dichos aspectos. Los participantes del test son usuarios realizando tareas, se analiza la forma en la que usan el producto y se mide el tiempo que emplean. Es útil para ser realizado en laboratorios de usabilidad, aunque también puede hacerse en un entorno real. Es importante la selección de tareas que los usuarios tendrán que llevar a cabo, estas deberán demostrar problemas de usabilidad, siendo el criterio más importante para la selección que prueben problemas potenciales de usabilidad del producto. Deberán ser tareas sugeridas por la propia experiencia, ya que los desarrolladores tienen ideas sobre dónde encontrar problemas, qué partes del producto fueron más difíciles de diseñar y cuáles son los problemas que se tienen que probar [10].

## **2 Evaluación del agente Dr. Roland para Educación Infantil mediante el método de medida de las prestaciones**

Se ha llevado a cabo la técnica de test de medida de las prestaciones [10] para el agente Dr. Roland para Educación Infantil. El agente para dicho nivel puede verse, junto con un ejemplo de interacción entre el agente y el usuario en las figuras 1, 2 y 3.

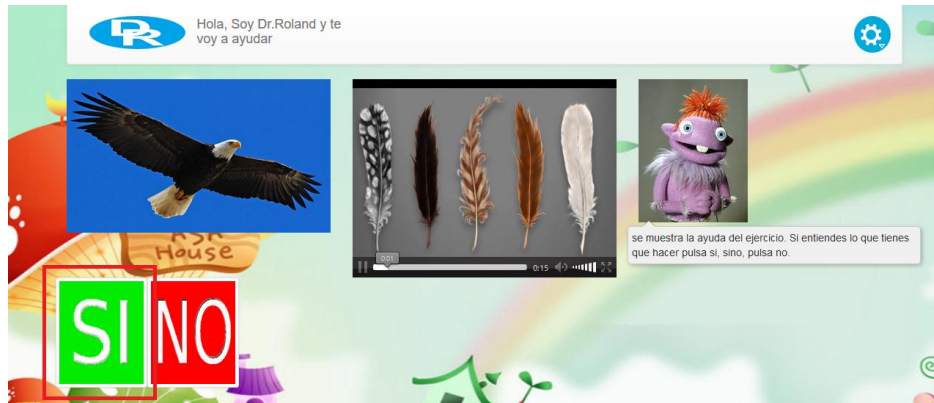


**Fig. 1.** Agente Dr. Roland, ejemplo de interacción entre el estudiante y el agente (1)



**Fig. 2.** Agente Dr. Roland, ejemplo de interacción entre el estudiante y el agente (2)





**Fig. 3.** Agente Dr. Roland, ejemplo de interacción entre el estudiante y el agente (3)

En este apartado se va a hacer alusión a las medidas de las prestaciones del uso del agente Dr. Roland.

La Tabla 1 recoge una serie de medidas por experiencia (Inf-1, Inf-2 e Inf-3), correspondiendo cada experiencia a cada uno de los colegios en los que ha sido puesto en práctica. Tres colegios en este caso (Inf-1, Inf-2 e Inf-3). Recoge números de ejercicios realizados (17, 16 y 16 respectivamente), respondidos correctamente (15, 14 y 10), tiempos de uso total en cada colegio (40, 65 y 75 minutos), tiempos medios de uso por sesión en cada colegio (20; 32,5; 37,5 minutos), tiempos medios de ejercicios y una serie de porcentajes como el de ejercicios respondidos correctamente sobre el total hechos por experiencia (88,23%; 81,25%; 62,5%). Indicar que en las dos primeras se usó pizarra digital y en la última una tableta.

Se recoge una serie de medidas por sesiones de cada una de las experiencias en la Tabla 2, entendiendo como experiencia aquello que se mencionó previamente, y como sesión cada uno de los días en los que se ha puesto en práctica el agente. Tres experiencias (Inf-1, Inf-2 e Inf-3), con tres sesiones, la primera (Inf-1) y dos en cada una de las restantes (Inf-2 e Inf-3). Recoge los ejercicios realizados, tiempos de uso, dificultades en la interacción, porcentajes de ejercicios resueltos de voz o por teclado, uso de reintentos en los ejercicios, uso de la ayuda, ejercicios respondidos correctamente o entendimiento de los ejercicios.

**Tabla 1.** Evaluación - Medidas de las prestaciones en Infantil: medidas por experiencia

| <b>Medidas</b>   | <b>Experiencia Inf-1 (3 sesiones)</b> | <b>Experiencia Inf-2 (2 sesiones)</b> | <b>Experiencia Inf-3 (2 sesiones)</b> |
|--|---------------------------------------|---------------------------------------|---------------------------------------|
| <b>Número Ejercicios totales hechos en cada experiencia</b>  | 17                                    | 16                                    | 16                                    |
| <b>Número Ejercicios medios hechos en cada experiencia por sesión</b>                                      | 5,66                                  | 8                                     | 8                                     |
| <b>Tiempo total de uso por experiencia (minutos)</b>   | 40                                    | 65                                    | 75                                    |
| <b>Tiempo medio de uso por experiencia por sesión (minutos)</b>  | 20                                    | 32,5                                  | 37,5                                  |
| <b>Dispositivo usado</b>   | Pizarra digital                       | Pizarra digital                       | tableta                               |
| <b>Número Ejercicios respondidos correctamente por experiencia</b>   | 15                                    | 14                                    | 10                                    |
| <b>Número de Ejercicios respondidos correctamente por experiencia por sesión</b>                           | 5                                     | 7                                     | 5                                     |
| <b>Porcentaje de ejercicios respondidos correctamente sobre el total hechos por experiencia</b>            | 88,23%                                | 81,25%                                | 62,5%                                 |
| <b>Porcentaje de ejercicios respondidos correctamente sobre el total hechos por experiencia por sesión</b> | 88,34%                                | 87,5%                                 | 62,5%                                 |
| <b>Tiempo medio de uso por ejercicio por experiencia (minutos)</b>   | 2,35                                  | 4,06                                  | 4,68                                  |
| <b>Tiempo medio de uso por ejercicio por experiencia por sesión (minutos)</b>                              | 3,53                                  | 3,25                                  | 3,69                                  |

Cabe destacar que se ha hecho uso de la ayuda en todos los ejercicios, ya que aunque entendiesen lo que les pedía el ejercicio y supiesen resolverlo, siempre veían los videos por que les gustaba.

En cuanto al uso del reintento, se ha hecho uso de ello en todos los casos en los que se falló al resolver bien la primera vez. Es decir, se ha considerado que no han sido respondidos correctamente aquellos ejercicios que en una misma realización de los mismos, no se han resuelto bien a la primera, y es en estos casos en los que se ha hecho uso de reintento. En este sentido añadir, que se ha hecho así, porque dadas las características de niños tan pequeños, es importante para mantener su atención elementos o situaciones que la capten, y el refuerzo positivo es uno de ellos para que no se frustren. Por lo que cuando resolvían incorrectamente la primera vez, el resto de niños o la profesora ayudaban para que finalmente resolvieran bien el ejercicio, por lo

que finalmente todos se resolvían correctamente. Y es por ello, que se ha considerado ejercicios que no han sido resueltos correctamente, aquellos que no se resolvieron a la primera en una misma realización de los mismos.

En relación al porcentaje de ejercicios respondidos correctamente sobre el total hechos, fueron para la experiencia Inf-1 en cada una de sus tres sesiones del 75%, 80% y 100%, para la experiencia Inf-2 en cada una de sus dos sesiones del 75% y 87,5, para la experiencia Inf-3 en cada una de sus dos sesiones del 50% y 75%, respectivamente. Se puede apreciar una mejoría en el porcentaje en toda ellas. El tiempo medio de uso por ejercicio por sesión (minutos) contrastando la situación inicial y final para cada experiencia, también se redujo en los dos primeros colegios, por el otro se mantiene constante, lo que puede indicar una mejora en la destreza en el uso o incluso en el aprendizaje, teniendo que hacer menos reintentos por ejemplo, siendo para Inf-1 (2,5; 3 y 1,875), para Inf-2 de (5,83 y 4,286).

Se hizo uso de la ayuda en los ejercicios, lo que tiene sentido también ante la demanda de los niños de querer verlo porque les gustaban. Los vídeos e imágenes fue de las cosas que manifestaron que más les gustaron. En cuanto al porcentaje de ejercicios resueltos de voz en Inf-1 (100%, 60% y 37,5%), en Inf-2 (62,5% y 50%) y en Inf-3 (75% y 50%) respectivamente, mientras que para el porcentaje restante fueron resueltos por teclado. A la vista de estos resultados, puede apreciarse que se ha hecho un uso mayoritario de resolución por voz, lo que tiene sentido, dadas las características de ser niños tan pequeños, que saben ni leer ni escribir.

**Tabla 2.** Evaluación - Medidas de las prestaciones en Infantil: medidas por sesión

| <b>Medidas</b>   | <b>Exp<br/>Inf-1:<br/>sesión 1</b> | <b>Exp<br/>Inf-1:<br/>sesión 2</b> | <b>Exp<br/>Inf-1:<br/>sesión 3</b> | <b>Exp<br/>Inf-2:<br/>sesión 1</b>   | <b>Exp<br/>Inf-2:<br/>sesión 2</b> | <b>Exp<br/>Inf-3:<br/>sesión 1</b>   | <b>Exp<br/>Inf-3:<br/>sesión 2</b> |
|--|------------------------------------|------------------------------------|------------------------------------|--|------------------------------------|--|------------------------------------|
| <b>Número Ejercicios hechos en cada sesión</b>                                       | 4                                  | 5                                  | 8                                  | 8  | 8                                  | 8  | 8                                  |
| <b>Tiempo total de uso por sesión</b>  | 10                                 | 15                                 | 15                                 | 35   | 30                                 | 30   | 45                                 |
| <b>Número Ejercicios respondidos correctamente</b>                                   | 3                                  | 4                                  | 8                                  | 6  | 7                                  | 4  | 6                                  |
| <b>Porcentaje de ejercicios respondidos correctamente sobre el total hechos</b>      | 75%                                | 80%                                | 100%                               | 75%  | 87,5%                              | 50%  | 75%                                |
| <b>Tiempo medio de uso por ejercicio por sesión (minutos)</b>                        | 2,5                                | 3                                  | 1,875                              | 5,83   | 4,286                              | 7,5  | 7,5                                |
| <b>Porcentaje de ejercicios en los que se ha hecho uso de la ayuda del ejercicio</b> | 100%                               | 100%                               | 100%                               | 100%   | 100%                               | 100%   | 100%                               |
| <b>Uso de reintento en la respuesta</b>  | Sí                                 | Sí                                 | No                                 | Sí   | Sí                                 | Sí   | Sí                                 |
| <b>Dificultad en la interacción (si/no)</b>  | No                                 | No                                 | No                                 | Al inicio  | No                                 | Al inicio  | No                                 |
| <b>Porcentaje de ejercicios en los que no entendían lo que se preguntaba</b>         | 0%                                 | 0%                                 | 0%                                 | No entendieron en 2 ejercicios (suponen el 25%) pero con la ayuda y la profesora entendieron | 0%                                 | No entendieron en 3 ejercicios (suponen el 37,5%) pero con la ayuda y la profesora entendieron | 0%                                 |
| <b>Porcentaje de ejercicios resueltos de voz</b>                                     | 100%                               | 60%                                | 37,5%                              | 62,5%  | 50%                                | 75%  | 50%                                |
| <b>Porcentaje de ejercicios resueltos por teclado</b>                                | 0%                                 | 40%                                | 62,5%                              | 37,5%  | 50%                                | 25%  | 50%                                |

### 3 Conclusiones

El método de test de medidas de las prestaciones de la Interacción Persona Ordenador fue usado para la evaluación de Agente Dr. Roland para el nivel de Educación Infantil. Tras la evaluación puede concluirse, que el agente fue integrado en el aula y usado por los alumnos. De ello puede extraerse, que los estudiantes fueron capaces de interactuar con el agente, siendo el algoritmo RUANLP el que guía la interacción y el diálogo entre estudiantes y el agente.

### Referencias

1. Hewett, T., Baecker, R., Card, S., Carey, T., Gasen, J., Mantei, M., Perlman, G., Strong, G. & Verplank, W. (1992). ACM SIGCHI Curricula for Human-Computer Interaction. Technical Report. ACM, New York, NY, USA.
2. SIGCHI Special Interest Group in Computer Human Interaction. Última fecha de consulta: 20-04-2017.  
<http://www.sigchi.org/>
3. Granollers i Saltiveri, T., Lorés Vidal, J., & Cañas Delgado, J. J. (2012). Diseño de sistemas interactivos centrados en el usuario.
4. Lorés, J., Granollers, T. y Lana, S. (2002). Introducción a la interacción persona-ordenador. En J. Lorés (ed.) La interacción persona-ordenador. Lérida, 2001-2002, 20-40.
5. Saltiveri, G. (2007). MPIu+ a. Una metodología que integra la ingeniería del software, la interacción persona-ordenador y la accesibilidad en el contexto de equipos de desarrollo multidisciplinares. Universitat de Lleida.
6. Dix, A. (1993). Human computer interaction. Prentice Hall, Englewood Cliffs, NJ.
7. Rodeiro, J. (2001). Representación y Análisis de la componente visual de la interfaz de usuario. Tesis doctoral. Universidad de Vigo.
8. Nielsen (1994), "Heuristic evaluation", Usability Inspection Methods, John Wiley & Sons, New York, NY.
9. Bias, R. & Mayhew D. (1991). Cost-justifying usability. IEEE Software.
10. Hassan Montero, Y. & Ortega Santamaría, S. (2009). "Informe APEI sobre usabilidad".
11. Tamayo-Moreno, S & Pérez-Marín, D. (2016). Analizando la interacción de estudiantes de educación Infantil y Primaria con un agente conversacional. Proceedings of IkaSNABAR-GUIDE 2016, 9th International Conference. Gorra J. Palazio (ed.), servicio editorial de la Universidad del País Vasco.
12. Tamayo-Moreno, S & Pérez-Marín, D. (2016). Adapting the design and the use methodology of a Pedagogical Conversational Agent of Secondary Education to Childhood Education