Universitat Politècnica de València

Departamento de Sistemas Informáticos y Computación

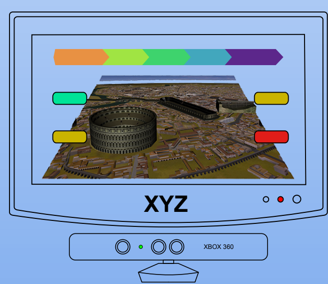
PhD Thesis

Using natural user interfaces to support learning environments

Author:
Juan Fernando Martín San José

Supervisor:
Prof. M. Carmen Juan Lizandra

Valencia, July 2015



Universitat Politècnica de València

Departamento de Sistemas Informáticos y Computación

PhD Thesis

Using natural user interfaces to support learning environments

Author:
Juan Fernando Martín San José

Supervisor:
Prof. M. Carmen Juan Lizandra

This thesis has been supported by the Spanish Ministry of Science and Innovation through the APRENDRA project (TIN2009-14319-C02-01).

To my family,
for being always there.

“Hizkuntza bat ez da galtzen ez dakitenek ikasten ez dutelako, dakitenek hitz egiten ez dutelako baizik.”

— Joxean Artze

Abstract

Education is a field of research in which Natural User Interfaces (NUI) have not been extensively exploited. NUI can help in the learning process, specially when used by children. Nowadays, children are growing up playing with computer games, using mobile devices, and other technological devices. New learning methods that use new technologies can help in the learning process.

The use of autostereoscopic vision is rising. At present, we can find many devices capables of rendering images that can give us a 3D perception without the use of glasses or any other wearable device. With this technology, new applications with more immersive environments can be developed and brought it to user's disposal.

Considering the importance of games and new technologies for learning, in this thesis, two different systems that use NUI for learning about a period of history were designed and developed. One of these systems uses autostereoscopic visualization, which lets the children see themselves as a background in the game, and that renders the elements with 3D sensation without the need for wearing special glasses or other devices. The other system uses frontal projection over a large-size tabletop display for visualization. The two systems have been developed from scratch. The Microsoft Kinect device is used in both systems for interaction.

A total of five studies were carried out to determine the efficacy of games with NUI interaction with regard to acquiring knowledge, ease of use, satisfaction, fun and engagement, and their influence on children.

In the first study, a comparison of the autostereoscopic system with the frontal projected system was carried out. This study analyzed different aspects such as engagement, increase of knowledge, or preferences. A total of 162 children from 8 to 11 years old participated in the study. From the results, we observed that the different characteristics of the systems did not influence the children's acquired knowledge, engagement, or satisfaction; we also observed that the systems are specially suitable for boys and older children (9-11 years old). The children had the depth perception with the autostereoscopic system. The children considered the two systems easy to

use. However, they found the frontal projection to be easier to use.

A second comparative study was performed to determine the mode in which the children learn more about the topic of the game. The two modes compared were the collaborative mode, where the children played with the game in pairs; and the individual mode, where the children played with the game solo. A total of 46 children from 7 to 10 years old participated in this study. From the results, we observed that there were statistically significant differences between playing with the game in the two modes. The children who played with the game in pairs in the collaborative mode got better knowledge scores than children who played with the game individually.

A third study that compares traditional learning with a collaborative learning method (in pairs and in large groups) using the game was carried out. A total of 100 children from 8 to 10 years old participated in this study. The results are in line with the second study. The children obtained higher score when collaborated in large groups or in pairs than attending to a traditional class. There were no statistically significant differences between playing in large groups and playing in pairs.

For personalized learning, a Free Learning Itinerary has been included, where the children can decide how to direct the flow of their own learning process. For comparison, a Linear Learning Itinerary has also been included, where the children follow a determined learning flow. A fourth study to compare the two different learning itineraries was carried out. A total of 29 children from 8 to 9 years old participated in this fourth study. The results showed that there were no statistically significant differences between the two learning itineraries.

Regarding the online formative assessment and multiple-choice questions, there is usually a question and several possible answers in questionnaires of this kind in which the student must select only one answer. It is very common for the answers to be just text. However, images could also be used. We have carried out a study to determine if an added image that represents/defines an object helps the children to choose the correct answer. A total of 94 children from 7 to 8 years old participated in the study. The children who filled out the questionnaires with imaged obtained higher score than the children who filled out the text-only questionnaire. No statistically significant differences were found between the two questionnaire types with images.

The results from the studies suggest that games of this kind could be appropriate educational games, and that autostereoscopy is a technology to exploit in their development. The following general conclusions have been extracted from the studies carried out:

- The new technologies are appropriated for developing educational games and autostereoscopy is a technology to exploit in their development.

- With applications like the ones presented in this thesis, children can learn using new technologies and, at the same time, they can have a good time playing the game, which promotes their learning.

- To play in a collaborative way facilitates the effectiveness of games with educational purposes. Children can learn a wide variety of educational topics by using new technologies and having fun with other children at the same time they are playing with them.

Resumen

La educación es un campo de investigación en el que las Interfaces de Usuario Naturales (NUI) no han sido extensamente explotadas. Las NUI pueden ser útiles en el proceso de aprendizaje, especialmente cuando se trata de niños. Hoy en día, los niños crecen jugando con juegos de ordenador, utilizando dispositivos móviles y otros dispositivos tecnológicos. Con nuevos métodos que utilicen alguna de estas nuevas tecnologías se podría mejorar el proceso de aprendizaje.

La utilización de la visión autoestereoscópica está aumentando. Hoy en día, podemos encontrar diversos dispositivos con la tecnología necesaria para mostrar imágenes en 3D sin necesidad de llevar gafas especiales u otros dispositivos. Con esta tecnología se pueden desarrollar entornos más inmersivos y ponerlos a disposición de los usuarios.

Teniendo en cuenta la importancia de los juegos y las nuevas tecnologías en el aprendizaje, en esta tesis se han diseñado y desarrollado dos sistemas diferentes que utilizan interfaces naturales de usuario para aprender los periodos de la historia. Uno de estos sistemas utiliza visión autoestereoscópica, la cual permite a los niños verse a ellos mismos dentro del juego y muestra los elementos del juego en 3D sin necesidad de llevar gafas especiales u otros dispositivos. El otro sistema utiliza proyección frontal como método de visualización. Los dos sistemas han sido desarrollados desde cero. Para los dos sistemas se ha utilizado el dispositivo Kinect de Microsoft para realizar la interacción.

Se han llevado a cabo un total de cinco estudios para determinar la eficacia de los juegos con interfaces de usuario naturales en cuanto al aprendizaje, facilidad de uso, satisfacción, diversión y su influencia en los niños.

En el primer estudio, se ha comparado el sistema autoestereoscópico con el sistema de proyección frontal. Este estudio ha tenido en cuenta diferentes aspectos como la satisfacción, cuánto han aprendido mientras jugaban o sus preferencias. Un total de 162 niños de 8 a 11 años han participado en este estudio. Por los resultados, observamos que las diferentes características de los sistemas no han influido en el aprendizaje, en la usabilidad o en la satisfacción; también observamos que los sistemas

son especialmente apropiados para chicos y niños mayores (de 9 a 11 años). Los niños tienen percepción de profundidad con el sistema autoestereoscópico. Los niños consideraron los dos sistemas fáciles de usar. Sin embargo, encontraron el sistema de proyección frontal más fácil de usar.

Se ha realizado un segundo estudio para determinar el modo con el que los niños pueden aprender en mayor medida el tema del juego. Los dos modos comparados han sido el modo colaborativo, en el que los niños jugaban por parejas; y el modo individual, en el que los niños jugaban solos. Un total de 46 niños de 7 a 10 años han participado en este estudio. Por los resultados, observamos que existen diferencias estadísticas significativas entre jugar al juego de un modo o de otro. Los niños que jugaron al juego en parejas en el modo colaborativo obtuvieron un mejor resultado que los niños que jugaron al juego en el modo individual.

Un tercer estudio compara el aprendizaje tradicional con el aprendizaje colaborativo (en parejas y en grupos grandes) utilizando el juego desarrollado. Un total de 100 niños de 8 a 10 años has participado en este estudio. Los resultados son similares al segundo estudio. Los niños obtuvieron una mejor puntuación jugando en colaboración que en el método tradicional en clase. No hubo diferencias estadísticas significativas entre jugar en grupos grandes y jugar en parejas.

Teniendo en cuenta el aprendizaje personalizado se ha incluido un itinerario libre de aprendizaje, en el cual los niños tienen la posibilidad de elegir de qué forma quieren dirigir su propio proceso de aprendizaje. A modo de comparación, se ha incluido también un itinerario lineal de aprendizaje, donde los niños siguen un flujo predeterminado. En este cuarto estudio han participado 29 niños de 8 a 9 años. Los resultados han mostrado que no hubo diferencias estadísticas significativas entre los dos itinerarios de aprendizaje.

En cuanto a la evaluación online con preguntas de test, normalmente, hay una pregunta y varias opciones de respuesta, donde se debe seleccionar solo una de ellas. Es muy común que la respuesta esté formada únicamente por texto. Sin embargo, también se pueden utilizar imágenes. En este quinto estudio se ha llevado a cabo una comparación para determinar si las imágenes que representan las respuestas son de ayuda para elegir la correcta. Un total de 94 niños de 7 a 8 años han participado en este estudio. Los niños que rellenaron los cuestionarios con imágenes obtuvieron una mejor puntuación que los niños que rellenaron los cuestionarios en los que solo había texto. No se encontraron diferencias estadísticas significativas entre los dos tipos de cuestionarios que incluían imágenes.

Los resultados de estos estudios sugieren que los juegos de este tipo podrían ser apropiados para utilizarlos como juegos educativos, y que la autoestereoscopía es una tecnología a explotar en el desarrollo de estos juegos.

De los estudios llevados a cabo podemos extraer las siguientes conclusiones generales:

- Las nuevas tecnologías son apropiadas para el desarrollo de juegos educativos, y la autoestereoscopía es una tecnología a explotar en su desarrollo.
- Con aplicaciones como las presentadas en esta tesis los niños pueden aprender utilizando nuevas tecnologías y, al mismo tiempo, pueden pasarlo bien jugando a juegos que mejoran su aprendizaje.
- Jugar de forma colaborativa mejora la efectividad de los juegos con propósitos educativos. Los niños pueden aprender una gran variedad de temas utilizando las nuevas tecnologías y divertirse con otros niños al mismo tiempo que juegan con ellos.

Resum

L'educació és un camp d'investigació en el qual les Interfícies d'Usuari Naturals (NUI) no han sigut extensament explotades. Les NUI poden ser útils en el procés d'aprenentatge, especialment quan es tracta de xiquets. Avui dia, els xiquets creixen jugant amb jocs d'ordinador, utilitzant dispositius mòbils i altres dispositius tecnològics. Amb nous mètodes que utilitzen alguna d'aquestes noves tecnologies es podria millorar el procés d'aprenentatge.

La utilització de la visió autoestereoscòpica està augmentant. Avui dia, podem trobar diversos dispositius amb la tecnologia necessària per a mostrar imatges en 3D sense necessitat de portar ulleres especials o altres dispositius. Amb aquesta tecnologia es poden desenvolupar entorns més immersius i posar-los a la disposició dels usuaris.

Tenint en compte la importància dels jocs i les noves tecnologies en l'aprenentatge, en aquesta tesi s'han dissenyat i desenvolupat dos sistemes diferents que utilitzen interfícies naturals d'usuari per aprendre els períodes de la història. Un d'aquests sistemes utilitza visió autoestereoscòpica, la qual permet als xiquets veure's a ells mateixos dins del joc i mostra els elements del joc en 3D sense necessitat de portar ulleres especials o altres dispositius. L'altre sistema utilitza projecció frontal com a mètode de visualització. Els dos sistemes han sigut desenvolupats des de zero. Per als dos sistemes s'ha utilitzat el dispositiu Kinect de Microsoft per a realitzar la interacció.

S'han dut a terme un total de cinc estudis per a determinar l'eficàcia dels jocs amb interfícies d'usuari naturals quant a l'aprenentatge, facilitat d'ús, satisfacció, diversió i la seua influència en els xiquets.

En el primer estudi, s'ha comparat el sistema autoestereoscòpic amb el sistema de projecció frontal. Aquest estudi ha tingut en compte diferents aspectes com la satisfacció, quant han après mentre jugaven o les seues preferències. Un total de 162 xiquets de 8 a 11 anys han participat en aquest estudi. Pels resultats, observem que les diferents característiques dels sistemes no han influït en l'aprenentatge, en la usabilitat o en la satisfacció; també observem que els sistemes són especialment

apropiats per a xics i xiquets majors (de 9 a 11 anys). Els xiquets tenen percepció de profunditat amb el sistema autoestereoscòpic. Els xiquets van considerar els dos sistemes fàcils d'usar. No obstant açò, van trobar el sistema de projecció frontal més fàcil d'usar.

S'ha realitzat un segon estudi per a determinar la manera amb el qual els xiquets poden aprendre en major mesura el tema del joc. Les dues maneres comparades han sigut la manera col·laborativa, en la qual els xiquets jugaven per parelles; i la manera individual, en la qual els xiquets jugaven sols. Un total de 46 xiquets de 7 a 10 anys han participat en aquest estudi. Pels resultats, observem que existeixen diferències estadístiques significatives entre jugar al joc d'una manera o d'una altra. Els xiquets que van jugar al joc en parelles en la manera col·laborativa van obtindre un millor resultat que els xiquets que van jugar al joc en la manera individual.

Un tercer estudi compara l'aprenentatge tradicional amb l'aprenentatge col·laboratiu (en parelles i en grups grans) utilitzant el joc desenvolupat. Un total de 100 xiquets de 8 a 10 anys ha participat en aquest estudi. Els resultats són similars al segon estudi. Els xiquets van obtindre una millor puntuació jugant en col·laboració que en el mètode tradicional en classe. No va haver-hi diferències estadístiques significatives entre jugar en grups grans i jugar en parelles.

Tenint en compte l'aprenentatge personalitzat s'ha inclòs un itinerari lliure d'aprenentatge, en el qual els xiquets tenen la possibilitat de triar de quina forma volen dirigir el seu propi procés d'aprenentatge. A manera de comparació, s'ha inclòs també un itinerari lineal d'aprenentatge, on els xiquets segueixen un flux predeterminat. En aquest quart estudi han participat 29 xiquets de 8 a 9 anys. Els resultats han mostrat que no va haver-hi diferències estadístiques significatives entre els dos itineraris d'aprenentatge.

Quant a l'avaluació online amb preguntes de test, normalment, hi ha una pregunta i diverses opcions de resposta, on s'ha de seleccionar solament una d'elles. És molt comú que la resposta estiga formada únicament per text. No obstant açò, també es poden utilitzar imatges. En aquest cinquè estudi s'ha dut a terme una comparació per a determinar si les imatges que representen les respostes són d'ajuda per a triar la correcta. Un total de 94 xiquets de 7 a 8 anys han participat en aquest estudi. Els xiquets que van emplenar els qüestionaris amb imatges van obtindre una millor puntuació que els xiquets que van emplenar els qüestionaris en els quals solament hi havia text. No es van trobar diferències estadístiques significatives entre els dos tipus de qüestionaris que incloïen imatges.

Els resultats d'aquests estudis suggereixen que els jocs d'aquest tipus podrien ser apropiats per a utilitzar-los com a jocs educatius, i que l'autoestereoscòpia és una tecnologia a explotar en el desenvolupament d'aquests jocs.

Dels estudis duts a terme podem extraure les següents conclusions generals:

- Les noves tecnologies són apropiades per al desenvolupament de jocs educatius, i la autoestereoscòpia és una tecnologia a explotar en el seu desenvolupament.
- Amb aplicacions com les presentades en aquesta tesi els xiquets poden aprendre utilitzant noves tecnologies i, al mateix temps, poden passar-ho bé jugant a jocs que milloren el seu aprenentatge.
- Jugar de forma col·laborativa millora l'efectivitat dels jocs amb propòsits educatius. Els xiquets poden aprendre una gran varietat de temes utilitzant les noves tecnologies i divertir-se amb altres xiquets al mateix temps que juguen amb ells.

Acknowledgements

First and foremost I want to thank my supervisor Prof. M. Carmen Juan. Her expert guidance, her patience and her stimulating ideas have been the basement onto which this thesis has been built.

Another important ingredient in the realization of this thesis has been the good atmosphere created by all my colleagues in the Universitat Politècnica de València: David Rodríguez, Vicente Broseta, Dr. Adrià Giménez, د. إيهاب خوري, Dr. Gustavo Rovelo, Emilio Granell, Sonia Cárdenas, Mauricio Loachamín, Dr. David Furió. Also the friends I made during my stay at Universidade Nova de Lisboa/Faculdade de Ciências e Tecnologia. Thanks to all of them. They have made me feel as at home.

There have been some other people who, at some way or other, have helped me in this journey but are left unmentioned. I hope I can name them on other occasions.

Finally, thanks to my parents, my sister, my family and friends for their positive attitude, understanding and support. This thesis is dedicated to all of them.

Definitions and abbreviations

Throughout this thesis several terms that have a specific meaning have been used. Following, there is a list of definitions and abbreviations ordered alphabetically.

Augmented Reality (AR) AR is a specific type of Mixed Reality where most of the information is real and virtual objects are coherently located onto the real scene.

Autostereoscopic vision Autostereoscopy is any method of displaying stereoscopic images (adding binocular perception of 3D depth) without the use of special headgear or glasses on the part of the viewer.

Cave Automatic Virtual Environment (CAVE) CAVE is an immersive virtual reality environment where projectors are directed to three, four, five or six of the walls of a room-sized cube.

Computer-Based Assessment (CBA) CBA is a kind of assessment that is performed by using computers and can be improved with the use of internet.

Fake Image A fake image is the one used in a questionnaire with images and did not appeared in the developed video game.

Free Learning Itinerary (FLI) FLI is a learning itinerary that offers users different alternatives in the path to follow through the contents of the itinerary and lets them decide.

Graphical Model Check (GMC) GMC is a checking model where the questions were grouped by raw scores and the ones which are higher than the mean are separated from the ones which are lower. The red lines represent the confidence bands.

Head-Mounted Display (HMD) HMD is a device that allows the rendering of computer generated imagery in a display close to the eyes. It can allow the visualization of AR from the user's perspective.

Human-Computer Interaction (HCI) HCI is a very active research area where the interaction with computers is studied, and usually involves highly multidisciplinary studies. It also refers to any kind of possible interaction and communication between a machine and a person.

Item Characteristic Curve (ICC) The curve indicates the probability that a child with each ability has to correctly answer a question.

Linear Learning Itinerary (LLI) LLI is a learning itinerary that integrates strong restrictions on the path that can be followed by the users. LLI are normally guided itineraries where the path is determined completely by the system. These types of itineraries can be seen as a collection of small pieces of information that are sequentially interconnected with each other.

Mixed Reality (MR) MR refers to the synthesis of virtual and real imagery that creates a combined scene of virtual and real information in any kind of proportion.

Natural User Interfaces (NUI) NUI is the common definition used by designers and developers of human-computer interfaces to refer to a user interface that is effectively invisible, or becomes invisible with successive learned interactions, to its users, and is based on nature or natural elements.

Open Graphics Library (OpenGL) OpenGL is a cross-language, multi-platform API for rendering 2D and 3D computer graphics. The API is typically used to interact with a GPU, to achieve hardware-accelerated rendering.

Open Natural Interaction (OpenNI) OpenNI is a framework that provides a set of open source APIs. These APIs are intended to become a standard for applications to access natural interaction devices.

Open Scene Graph (OSG) OSG is an open source 3D graphics application programming interface, used by application developers in fields such as visual simulation, computer games, virtual reality, scientific visualization and modeling. The toolkit is written in standard C++ using OpenGL.

Person-Item Map (PIM) PIM is a plot where the estimated ability of the user and question difficulty measures are placed side by side in one vertical dimension. The questions appear in order of difficulty.

Person-Parameter Distribution (PPD) PPD is a distribution of the users' abilities regarding the asked questions.

Radio Frequency Identification (RFI) RFI is the use of a wireless non-contact system that uses radio-frequency electromagnetic fields to transfer data from a tag attached to an object, for the purposes of automatic identification and tracking.

Rasch Model The Rasch Model measures a persons latent trait level from a probabilistic perspective. The probability of a user answering a question correctly relies on the user's underlying ability and the difficulty of question.

Real Image A real image is the one used in a questionnaire with images and did appeared in the developed video game.

Red Green Blue (RGB) RGB is an additive color model in which red, green, and blue light are added together in various ways to reproduce a broad array of colors.

Software Development Kit (SDK) SDK is typically a set of software development tools that allows for the creation of applications for a certain software package, software framework, hardware platform, computer system, video game console, operating system, or similar platform.

Stereoscopic vision Stereoscopy is a technique for creating or enhancing the illusion of depth in an image by means of stereopsis for binocular vision.

Traditional Learning Method (TLM) TLM is the learning method used for acquiring knowledge at the school, where a teacher uses elements of the classroom for teaching the students, like blackboards, chalks, or books.

Contents

List of Figures

List of Tables

List of Listings

I

Introduction

1 Introduction

"明人不用细说，响鼓不用重锤。"

— Chinese proverb

1.1 Motivation

The rapid development of technology has provided a lot of new and advanced systems that were unimaginable few years ago. Nowadays, the use of technological systems is common for daily tasks such as playing at home, watching the television, and several ways of entertainment. For example, the appearance of Microsoft Kinect revolutionized the gaming market, and it is also being a revolutionary device for Natural User Interfaces (NUI) which changes completely how the user interacts with the system. Microsoft Kinect is now widely used in video games by connecting it to an Xbox console, but this is not the only device it can be plugged in. The Microsoft Kinect driver, the OpenNI library, or Libfreenect (the software developed by the OpenKinect community) facilitate the programming of Microsoft Kinect applications and games for PCs. These facilities have led the natural user interaction to be incorporated in a large number of different types of applications (e.g., Chang et al. (2011b) developed a system for physical rehabilitation, which is an example of how devices of this kind are not only used for entertainment, but also for medical purposes). However, NUI have not been extensively exploited for creating learning environments for students. From our point of view, this technology is on the right track for being a good complement to the traditional educational approach, which have not changed too much over the years. Generally, traditional education consists basically in attending school and master lectures.

Nowadays, it is also common to have stereoscopic visualization. This means that a 3D sensation will be perceived by the user. To achieve stereoscopic visualization, three main technologies are used: passive, active, and autostereoscopic. The main difference between active/passive stereoscopy and autostereoscopy is that the autostereoscopic visualization generates the 3D sensation without the use of special glasses, headgear or other wearable devices. There are several previous works related to autostereoscopy, e.g., Arino et al. (2014) developed an autostereoscopic augmented reality system that compared with a virtual reality system. It was tested by children. In our opinion, this technology can also be used for developing educational systems.

With regard to the learning methods, the use of new technologies is causing a huge change in the way that people learn; the scientists studying human learning and thinking are focused on how the brain works while learning, and they research at places like workspaces or classrooms (e.g., Clark (1997); Lave and Wenger (1991)). The senses of the human body can also help to improve this learning process. The more senses the students use, the more they will learn, and this implies better scores. New systems using new technologies and tools that can stimulate more senses will help in the learning process (Sandor and Klinker, 2005). A good example of a tool that stimulate senses are video games. The use of video games implies that the user have fun, which is important for the learning process because when learning in a funny way it is less likely to forget the contents learned (Blecic et al., 2002). Kebritchi and Hirumi (2008) identified five reasons why the game-based learning is an effective learning tool. These reasons are:

1. The game uses action instead of explanation.

2. Playing creates engagement and satisfaction.

3. The game can be adjusted to several learning styles.

4. The game improves the abilities.

5. Playing provides an interactive context causing children make decisions.

Taking into account new tools for learning purposes, professionals of education are always searching for more ways of improving the way they teach (Tan et al., 2008). The contents of educational games are commonly developed using game technology and design principles whose primary purpose is to educate while entertaining the user. The use of game technology for learning purposes has been used, and its suitability for education has been supported by several previous works. For example, Virvou et al. (2005) evaluated an intelligent tutoring system as a virtual reality educational game, demonstrating that games of this kind can be very motivating while improving the learning effects on students; Mumtaz (2001) studied the nature and experiences of children's computer use at home and school, where she found out that 77% of them used computers every day for playing video games; Ebner and Holzinger (2007) designed an online game for higher education that is related to the theory of structures in civil engineering.

Considering the importance of games and new technologies for learning, in this thesis, two different systems that use NUI for educational purpose were designed and developed. The Microsoft Kinect device was used in both of them to recognize the users' gestures, which provides natural interaction and let more human senses to get into the game; and, as said before, the human senses have an important role in the learning process. The first of these systems was based on an autostereoscopic display. This system merged the images from the real world with virtual elements, creating a 3D perception by using 8 different views. For merging the images, it used the camera provided by the Kinect device. Another system used a large-size projected surface as an interactive tabletop display. It is also provided with natural interaction and this interaction is performed by touching the tabletop. With these differences between the two systems, we can perform several comparisons by conducting different studies. Comparisons of this kind between two or more systems that are for the same purpose are quite common; different systems can have differences in the visualization (e.g., Juan and Calatrava (2011) compared video see-through and optical see-through HMD systems for the treatment of phobia to small animals).

Our two systems are distinguished by how the users interacts with them. In the autostereoscopic system, the user can interact by using gestures and selecting the elements that appear on the autostereoscopic display. Basically, the user has to move the hands in the air in front of the camera. On the contrary, the interaction for the frontal projection system is performed by interacting with a large-size tabletop display, which implies to touch the tabletop, as if the user was pressing buttons on top of the table. Another difference between the two systems is the visualization method.

In the autostereoscopic system, the visualization of the models in the screen had 3D sensation, and the visualization of the projected system was not stereoscopic. For these two systems (large-size tabletop display + Kinect and autostereoscopic display + Kinect) an educational game about historical ages was designed and developed. For the contents of the game, the selected historical ages, specifically, were: Prehistory, Ancient Times, the Middle Ages, the Early Modern Period and the Contemporary Period. With these two systems, five different studies were carried out. In the first study, the autostereoscopic display + Kinect and the projected surface + Kinect is compared. A second study compares the individual vs. the learning in pairs. A third study compares collaboration in large groups vs. collaboration in pairs or traditional methods. A fourth study compares different learning itineraries. Finally, a fifth study compares different types of questionnaires for assessment.

In one of our previous works (Furió et al., 2013b), the use of history as the subject of an educational computer game was one of the most preferred subjects. Students can learn about the past in an easier and ludic way using new technologies, such as tabletop displays, mobile devices, Virtual or AR.

1.2 Scientific goals and research hypotheses

The main objective of this thesis is to determine the effects of Natural User Interfaces and autostereoscopy in learning environments for children. This thesis focuses on studying how different factors affect learning and other aspects such as usability, fun, and overall satisfaction. These factors are:

- Natural User Interfaces (gestures vs. interaction over a table).
- Autostereoscopy (stereoscopy vs. non-stereoscopy).
- Number of participants (individual, in pairs or in large groups).

To achieve this objective, two different systems were designed and developed, and five different studies were carried out, as detailed next:

1. **Design and development of a new system that combines autostereoscopy and NUI**:

   A system that combines autostereoscopic technology and NUI for the interaction with the user has been developed from scratch. This development has been done by using the C++ programming language, OSG for the rendering, and other libraries for video and audio. Also, for the NUI part, the OpenNI has been used. The autostereoscopic visualization is achieved by using the Mirage SDK[1]. This SDK provides an OSG node that can be inserted in the graph scene and takes charge generating the eight views that the autostereoscopic visualization needs. The game developed consisted of five stages corresponding to five historical ages, and some mini-games inside each stage. When the children arrive at the final stage, they would be able to recreate the time line by ordering every historical age.

2. **Design and development of another system that combines frontal projection and NUI**:

   A system that combines frontal projection technology and NUI for the interaction with the user has been also developed from scratch. This system has been developed by using C# and the XNA framework with the official Kinect drivers from Microsoft. For the interaction Emgu.CV was used, which is a computer vision library based on the OpenCV library to manipulate complex graphics. Emgu.CV provided several functions that were useful for hand detection. GoblinXNA was used to display the 3D scene, which provides a scene graph that simplifies the implementation.

3. **Autostereoscopy vs. Frontal Projection (Study 1)**:

   One of the objectives of this study was to find out which system was most appreciated by the children. The first of our hypotheses is that the children will prefer the autostereoscopic system over the frontal projection system. The second of our hypotheses is that children will increase their knowledge about the subject treated in the game by using the two systems, and that the autostereoscopic system would lead to greater learning results.

   Some reasons that support these hypotheses are the following:

   (a) Although the two systems have the same NUI, the autostereoscopic display provides a 3D sensation that improves the immersion in the game.

   (b) While playing, the children can see themselves inside the game in the display, and this gives them a sensation of prominence that encourages them and, consequently, they are more motivated and involved in the game.

   (c) Since the size of the autostereoscopic display is 46 inches, the fact that the children can play video games using such a big TV makes a deep impression on them and they are eager to start playing.

[1] www.mirage-tech.com

The third hypothesis is that the frontal projection system will be easier to use. The support for this hypothesis is that nowadays children are accustomed to using actual gadgets and peripherals that are controlled in the same way, which is much different from using a gesture-oriented autostereoscopic system.

4. **Individual vs. Collaborative learning (in pairs) (Study 2)**:

   One of the objectives of this study was to find out which method the children acquired a higher increment of knowledge with.

   Our main hypothesis of this study (fourth hypothesis) is that the children will learn more by playing the game in the collaborative mode (in pairs) over the individual mode.

   Some reasons that support our hypothesis are the following:

   (a) Although the two modes have the same video game, when playing in the collaborative mode, the children can interact between themselves, and this can improve their learning outcomes.

   (b) When playing in the individual mode, the children are not able to interact with anybody and they have to answer every question.

5. **Collaboration in Large Groups vs. Collaboration in Pairs or Traditional Methods (Study 3)**:

   In this study, a comparison of the traditional learning method with a collaborative learning method (in pairs and in large groups) using the game was carried out.

   Our fifth hypothesis is that children who learn by playing a computer-based game collaboratively on a large-size tabletop display and involving up to 12 students will obtain significantly higher learning outcomes than those who learn by attending a traditional class.

   Our sixth hypothesis is that children who learn by playing collaboratively a computer-based game collaboratively on a large-size tabletop display and involving up to 12 students will obtain significantly higher learning outcomes than those who learn by playing with the same game, but playing in pairs.

   Our seventh hypothesis is that children who learn by playing a computer-based game collaboratively played on a large-size tabletop display and involving up to

12 students will show significantly higher satisfaction than those who learn by attending a traditional class.

Our eighth hypothesis is that children who learn by playing a computer-based game collaboratively played on a large-size tabletop display and involving up to 12 students table will show significantly higher satisfaction and usability than those who learn by playing with the same game, but playing in pairs.

6. **Free Learning Itinerary vs. Linear Learning Itinerary (Study 4)**:

   For personalized learning, a Free Learning Itinerary (FLI) has been considered, where the children can decide how to direct the flow of their own learning process. For comparison, a Linear Learning Itinerary (LLI) has also been included, where the children follow a determined learning flow. A study to compare the two different learning itineraries was carried out.

   The main objective of this study was to find out which of the two learning itineraries provides greater learning improvement. Our hypothesis (ninth hypothesis) is that the children will acquire more knowledge while playing with the Free Learning Itinerary. Some reasons that support our hypothesis are the following:

   (a) When playing with FLI, the children can choose the next historical age they want to learn.
   (b) Each child decides the itinerary that he/she wishes to follow.
   (c) The children can repeat the historical ages as many times as they want. If they do not want to learn about a historical age, they can opt out and ignore it.

7. **Text-only vs. Real Images vs. Fake Images (Study 5)**:

   In this study, we have focused on online formative assessment and multiple-choice questions. In this type of questionnaire, there is usually a question and several possible answers in which the student must select only one answer. It is very common for the answers to be just text. However, images could also be used. In this study, we have carried out a study to determine if an added image that represents/defines an object helps the children to choose the correct answer. The tenth hypothesis was that there would be significant differences between using only a text-only questionnaire and a questionnaire that, apart from the text also includes images. The eleventh hypothesis was that there would be significant differences between a questionnaire with images used during the learning process and a questionnaire with images that represent the item but that were not used during the learning process.

1.3 Thesis structure

The thesis document is structured as follows:

Chapter 1 introduces the thesis, including the motivation, the scientific goals, the research hypotheses, and this explanation of the thesis structure.

Chapter 2 shows the state of the art, reviewing the most relevant literature relative to learning environments, assessment, Natural User Interfaces, autostereoscopic vision, and frontal projection.

Chapter 3 provides an explanation of the game design and the game description.

Chapter 4 describes in detail the design and development of the autostereoscopic system.

Chapter 5 describes in detail the design and development of the frontal projection system.

Chapter 6 describes the first study of the thesis, where the autostereoscopic system was evaluated with children by comparing it with the frontal projection system.

Chapter 7 describes the second study of the thesis, where the autostereoscopic system was evaluated by the children in two different ways. Some of them played the game in pairs (collaborative version), and, the rest, played the game solo (individual version).

Chapter 8 describes the third study of the thesis, where the traditional learning method in class with a teacher in the classroom was compared with the frontal projection system played collaboratively in large groups or in pairs.

Chapter 9 describes the fourth study of the thesis, where two different learning itineraries are compared: the Linear Learning Itinerary and the Flexible Learning Itinerary.

Chapter 10 describes the fifth study of the thesis, where three different kinds of questionnaires are proposed and compared. One of them is made by only-text, another one contains images from the game (real images), and the last one contains images that are not included in the game (fake images).

Chapter 11 finalizes the work with the conclusions and future work, and enumerates the publications derived from this thesis.

In addition to this, there is an Appendix chapter A that shows the questionnaires that had been used in this thesis.

II

State of the art

State of the art

2

"To know the road ahead ask those coming back."

—

2.1 Introduction

Nowadays, children are growing up using computer games and other technologies that have changed their preferred leisure styles, their social interaction, and even their learning preferences (Bekebrede et al., 2011). This generation is commonly referred to as the "gamer generation" (Beck and Wade, 2004, 2006a), "digital natives" (Prensky, 2001), or the "net generation" (Tapscott, 1998). Teaching and learning are no longer restricted to traditional classrooms (Wang et al., 2007). Several authors have identified new technologies as learning tools that could help young children learn in relevant ways (Couse and Chen, 2010; Gimbert and Cristol, 2004). Game-based learning might be a more appropriate approach for teaching and engaging the children in a more successful way than traditional learning methods (Prensky, 2001). As Roman (2010) pointed out in the statement “The mouse’s days are numbered”, the current trend in new devices, games, and consoles is to get rid of all gamepads, joysticks, and other input methods. Largillier (2011) argued that technologies such as tablet interfaces should evolve with the user.

Considering this new generation of students, this thesis is focused on exploring the potential of new computer-based games to support learning. This chapter is divided into 5 sections: learning environments, assessment, natural user interfaces, autostereoscopic vision, and frontal projection. The learning environments section introduces game-based learning and its positive implications; in addition, the concept of e-Learning is presented, and some definitions are gathered. In this section, as subsections, traditional learning, collaborative learning, and learning itineraries are also discussed. In this thesis, we compare our developments with traditional learning. The influence of collaborative learning has been studied too. A comparison between a linear learning itinerary and a flexible learning itinerary has been carried out as well. Therefore, an introduction to all these concepts and a brief state of the art on each of them are included in this section. One of the studies is focused on assessment, hence, this aspect is also covered in this state of the art. The three technical components of the thesis to highlight are: natural user interfaces, autostereoscopic vision, and frontal projection. Accordingly, the last three sections present previous works related to these three components.

2.2 Learning environments

Students can learn in a wide diversity of settings, locations or cultures. The learning environments have included the culture of school or class since the beginning, and they have influence on the students’ education. Many computer games have been designed with an educational background and for learning purposes, but only few of them perform a deep analysis, as several studies determined (Connolly et al., 2011; Freitas and Campos, 2008; O’Neil et al., 2005). Most of the educational computer games have been developed with no underlying body of research or coherent learning theory (Shaffer et al., 2004).

According to Connolly et al. (2007), games-based learning can be defined as the use of computer games-based technology approach to deliver, support, and enhance teaching, learning, assessment and evaluation. Since playing games has been demonstrated to be highly related to learning, several models have been developed to identify the learning outcomes that can be achieved by playing digital games (Connolly et al., 2012). Games-based learning may also affect motivation, which is related to the children's willingness to participate in tasks and activities. According to Malone and Lepper (1987), there are seven factors that promote motivation: challenge, curiosity, control, fantasy, competition, cooperation, and recognition. Many of these are present in games (Prensky, 2001). As Garris et al. (2002) stated, the user gets hooked on cognitive processes that are triggered that have been proven to be beneficial for learning. Besides, games are extensively accepted by new generations of users. Another important item is that the effectiveness of game-based learning greatly depends on the acceptance by classroom teachers (Bourgonjon et al., 2013). This might create expectations about the learning environment that can be found in the classroom (Oblinger, 2004); however, the education is still using old methods (Beck and Wade, 2006b). Girard et al. (2013) reviewed the results of experimental studies that were designed to examine the effectiveness of serious games on the learning and engagement of players. They concluded that serious games might be powerful tools for learning; however, there are very few empirical studies that investigate the effectiveness of serious games in learning. One of the goals of this thesis is to reduce the dearth of studies in this area. Nevertheless, some previous works can be cited such as the work of Papastergiou (2009), where she developed and used two similar applications (a gaming one, and a non-gaming one) in order to compare them. The two applications were designed with the goal of introducing students to basic computer memory concepts. Her results showed that the students that had used the gaming application performed significantly better than those that had used the non-gaming one. Beale et al. (2007) used a game to learn about cancer, and they found better performance in the group that played the game when compared with a control group that did not play the game. Mayer et al. (2004) presented the design and evaluation of a gaming-scenario experiment for the exploration of development planning in an urban network. The results showed that the gaming-scenario approach generated new and critical insights in development planning. Guillén-Nieto and Aleson-Carbonell (2012) developed a serious game to teach intercultural business communication between Spaniards and Britons. This study showed that there were statistically significant differences between the pre- and post- knowledge tests. This means that, as a result of playing the communication game, the students improved their intercultural knowledge. Facer et al. (2004) reported a study to explore how using mobile technologies can create a powerful and engaging learning experience. They stated that gamers are expert when they control their own learning alongside more knowledgeable peers, and children as gamers are more likely to learn effectively by acting as mentors to novice learners.

One way in which students can control their own learning is choosing e-Learning. e-Learning presents the intersection between the world of information and communication technology and the world of education (Stankov et al., 2004), or even a virtual world (Monahan et al., 2008). There is no single definition for e-Learning. Selim (2007) defined e-Learning as the use of modern information and communications technology and computers to deliver instruction, information, and learning content. Arbaugh (2002) defined e-Learning as the use of the internet by users to learn specific content. For Stockley (2013), e-Learning involves the use of a computer or electronic device (e.g., a mobile phone) in some way to provide educational training, or learning material. e-Learning is not only about training and instruction, it is also about learning that is customized to individual needs and should be flexible and interactive. e-Learning systems cover a large variety of technology-based applications such as computer-based learning, virtual classrooms, digital collaboration, or web-based learning. A definition that is closely related to these ideas is personalized learning, which includes personalized material, personalized objectives, and personalized processes (Shi et al., 2002). Flexible learning itineraries are also related to this personalization, explained in subsection 2.2.3.

Different technologies such as AR have been used to develop educational systems. For example, Construct3D (Kaufmann, 2004) was developed to teach mathematics and geometry. Larsen et al. (Larsen et al., 2005) presented an AR system for learning how to play billiards. The most outstanding characteristic of this system was that the game was played on a real billiard table. Organic chemistry has also been taught using an AR system (Fjeld et al., 2007). In 2009, Wang et al. (2009) tested three user interface prototypes for learning about heritage temples. In the *Computer Graphics* research group several AR systems have been developed and tested: a storytelling system (Juan et al., 2008b); a system for learning the interior of the human body (Juan et al., 2008a); systems for learning about endangered animals (Juan et al., 2010b, 2011a), and words (Juan et al., 2010a); or a phone game for learning how to recycle (Juan et al., 2011b). As a result of the APRENDRA project, AR iPhone games for learning multiculturalism, solidarity, and tolerance (Furió et al., 2013b), and the water cycle (Furió et al., 2013a) have also been published. These two last games are shown in Figure 2.1.

2.2.1 Traditional learning

The traditional learning is the learning method used for acquiring knowledge at the school, where a teacher uses elements of the classroom for teaching the students, like

(a) Child finding an oxygen drop (b) Non-AR game with the Tablet PC

Figure 2.1: APRENDRA learning games

blackboards, chalks, or books. Before the introduction of the new technologies, it was the learning method most widely used at schools. With regard to studies in which traditional learning methods are compared with learning using new technologies, several works can be cited. Girard et al. (2013) analysed game-based learning tools versus traditional approaches. Their conclusion was that the users of game-based tools had the same learning improvement as with traditional approaches. In 2013, Al-Qahtani and Higgins investigated the effects of e-learning, blended learning, and classroom learning (Al-Qahtani and Higgins, 2013). Three groups were involved: the first group was taught by e-learning, the second group by blended learning (which combines e-learning and traditional teaching), and the third group was taught using the traditional teaching method. The sample consisted of 148 students: 43 students in the first group, 55 students in the second group, and 50 students in the third group. The ethics unit from an islamic culture course was selected for the study and adapted for the online course (the e-learning method). The results showed that there was a statistically significant difference between the blended learning method and the other two methods. However, no significant difference was found between the e-learning group and the traditional learning group. Furió et al. (2013b) carried out a study to determine whether an iPhone game had better learning outcomes than a traditional game for learning multiculturalism, solidarity and tolerance. In this study participated 84 children (from 8 to 10 years old). For learning outcomes, the results did not show significant differences between the two groups. For the satisfaction, 96% of the children indicated that they would like to play with the iPhone game again, and 90% indicated that they preferred the experience with the iPhone game over the traditional one. Figure 2.2a shows the real elements used in the traditional game. Figure 2.2b shows to children playing with the real labyrinth game. Furió et al. (2015) compared the learning effectiveness and satisfaction of children using an iPhone game vs. the traditional classroom lesson for learning the water cycle. Thirty-eight children from 8 to 10 years old participated in the study. The children made significant learning gains about the water cycle regardless of the method used. Even though the results showed that the iPhone method achieved higher knowledge results than the traditional classroom lesson, no statistically significant differences were found between the iPhone and the classroom lesson. Chuang et al. (2014) studied how happiness can improve learning performance by using interactive systems. The proposed system was a somatosensory system, which is a system that is made up of a number of different receptors like thermoreceptors, photoreceptors, mechanoreceptors, or chemoreceptors like the human body. With this system, users receive abundant feedback in the activity since the responding is direct. The main goal of the users was to solve a puzzle in two modes: scoring mode and timing mode. From the results, Chuang et al. determined that the learners maintained a positive attitude when using their approach for learning purposes. Of all the factors they studied, enjoyment was the most significant one.

(a) Elements used in a traditional game

(b) Two children playing with a real labyrinth game

Figure 2.2: Furió et al.'s study

2.2.2 Collaborative learning

For a single individual, it is hard to achieve complex tasks without the support of other people. Therefore, working together is nowadays highly valued in the workplace (Barron, 2000) and links group members together (Johnson et al., 1998). When compared with the individual learning working collaboratively in groups has been demonstrated to improve students' critical thinking skills, social skills, self-esteem, and problem-solving skills (Gokhale, 1995; Li, 2002) and even improve learning outcomes (Neo, 2003). This collaborative learning combined with the use of new technologies can benefit both students and teachers, when these technologies are used as communication, repository, or documentation tools (Kaptelinin, 1999).

With regard to the choice of the group size, according to Rau and Heyl (1990), smaller groups (i.e., three) contain less diversity and may lack divergent thinking styles and varied expertise that help to animate collective decision making. Conversely, in larger groups it is difficult to ensure that all members participate. A review of previous studies shows that the groups are normally made up of between 2 and 6 students. Elices et al. (2002) presented 5 studies of children between 10 and 13 years old. The children had to carry out a task, either individually or in pairs. In the work of Kirschner et al. (2009), the participants in the collaborative group worked in groups of three. The participants were high-school students and the learning problems were related to the field of biology. Gokhale (1995) and Li (2002) used a group of four. Twenty-four students participated in the collaborative learning group at the college level, and the subject content was series and parallel circuits. A traditional methodology that was comprised of lectures and worksheets was used. Neo (2003) used groups of 4-6 people. The students were undergraduates in their first year at the university. These students had to create a collaborative learning environment for a design project. The class was an 8-week course and the students were given lectures, tutorials, and labs.

2.2.3 Learning itineraries

A learning itinerary guides how students learn the content. Therefore, it implies a way to organize the learning sequence. It responds to the need of guiding students through the content, processes, and activities. Moreover, the level of interactivity of the materials is attached to the guidelines set out in the design phase. These guidelines will determine the learning itineraries. LLI are commonly used in educational environments or for educational purposes and represent the traditional approach for learning itineraries. The alternative to this traditional choice are FLI, which have both advantages and disadvantages when compared to LLI. An important advantage of FLI is that it allows students to adapt the itinerary to their own profile. Thus, several studies have focused on interpersonal differences (Kagan and Kogan, 1970) and field dependence (Witkin and Goodenough, 1977). Pask (1976) pointed out the individual capacity to switch from one strategy to another in order to adapt the prevailing conditions. Also, FLI have a greater capacity to adjust to users by being able to adapt the itinerary to the individual features and preferences of the student. The main disadvantage of FLI is that students may follow an incomplete or even incorrect itinerary. This can be on their own initiative, by mistake, or disorientation. Therefore, an appropriate design of the itinerary is particularly important in order to prevent these problems.

Previous experiences by researchers have focused on the benefits of integrating technology to give greater freedom to students in the learning process. Some initial experiments of applying technology to provide students with freedom in their learning itineraries include the use of hypermedia. Hypermedia is an extension of hypertext, which integrates multimedia elements to create a set of information items that are usually not offered linearly. As Grabinger (1996) affirmed, the major benefit of the hypermedia approach in learning environments may be that learners are invited to express their ideas and prior concepts and externalize them with hypermedia tools. Leclercq and colleagues carried out experiments that are related to hypermedia in the learning field (Leclercq and Pierret, 1989; Leclercq and Gilles, 1993). Leclercq and Pierret (1989) developed software based on hypermedia to respect both interpersonal and intrapersonal variation in learning strategies. Leclercq and Gilles (1993) also reviewed how hypermedia can enhance educational assessment. They concluded that these strategies of flexible learning increase the autonomy of the learners, which in turn increases their metacognitive activity. Adaptive hypermedia (Brusilovsky, 2001) adds new possibilities to the development of interactive learning systems. Adaptive hypermedia systems create a model with the features of each individual user (knowledge, preferences, ...) in order to adapt the system to the needs of each learner. Adaptive hypermedia and the introduction of the web have promoted the development of educational hypermedia systems such as ELM-ART (Brusilovsky et al., 1996a), 2L670 (De Bra, 1996) and InterBook (Brusilovsky et al., 1996b). These systems have been used in several experimental studies: for WWW-based tutoring systems (Weber and Specht, 1997); for designing hypermedia architectures consisting of Java servlets (De Bra and Calvi, 1998); or for creating tools for development adaptive courseware (Brusilovsky et al., 1998). These studies confirm the suitability of these learning strategies.

2.3 Assessment

The main role of a teacher is to guide students during their learning process. Another of the teachers' tasks is to determine if the students have acquired the defined learning goals. Students should demonstrate that they have acquired these defined learning goals. Rating students has been a research topic for more than 70 years. The design, development, use, and interpretation of student assessment is one of the most important topics in evaluation research (Arreola, 1995). Assessment can be defined as "the measurement of the learner's achievement and progress in a learning process" (Keeves, 1994; Reeves and Hedberg, 2009). The assessment of students is a core component for effective learning (Bransford et al., 2000). There are two main forms of assessment: summative and formative (Challis, 2005). Summative assessment measures what students have learned at the end of a course or after some defined period (Hargreaves, 2008). It can also refer to checking whether or not the students have met the desired learning goals or whether they have achieved the required levels of competence (Challis, 2005). Summative assessment usually includes scoring for validation or accreditation purposes. Formative assessment is applied as a source of continuous feedback to improve teaching and learning (Hargreaves, 2008). Formative assessment can also be seen as assessment for learning that takes place during instruction in order to support learning (Oosterhof et al., 2008; Vonderwell et al., 2007). Formative assessment activities are intrinsic parts of instruction that allow learning to be controlled and the instruction to be modified until the desired learning goals have been achieved (Gikandi et al., 2011). Hattie and Timperley (2007) and Nicol and Macfarlane-Dick (2006) stated that feedback is most effective when it is directly related to clearly defined learning goals, and that effective formative feedback is not only based on monitoring the progress towards those goals but that it must also encourage students to develop effective learning strategies.

2.3.1 Computer-based assessment

As previously mentioned, teaching and learning are changing and are no longer restricted to traditional classrooms (Wang et al., 2007). Computer-based assessment (CBA) is not new. Computer-based assessment is performed by using computers and can be improved with the use of internet. Two of the first systems to support assessment were PLATO (Programmed Logic for Automatic Teaching Operations) and TICCIT (Time-shared Interactive Computer-Controlled Information Television) (Rota, 1981). From there, different tools for assessment such as the following have already been presented:

- MarkTool (Heinrich and Lawn, 2004), which allows teachers to annotate PDF documents sent by students with formative feedback (annotations can be textual and graphical.
- EAT (Electronic Assessment System) (Rashad et al., 2008), which allows teachers to modify the content taking into account the student's answers, answer-time, and student feedback.
- A flexible e-assessment system designed by Dube and Ma (2010), which can be adapted to different learning styles.
- GPAM-WATA, a Web-based dynamic assessment system, in which teachers can provide students with teaching assistance (Wang, 2010).
- FAML (Formative Assessment-based Mobile Learning) designed by Hwang and Chang (2011), which is a mobile system for local cultural learning that runs on mobile devices (PDAs).

Apart from tools, other initiatives have also been presented. In 2009, a consortium of Cisco, Intel, and Microsoft launched Transforming Education: Assessing and Teaching 21st Century Skills (Cisco et al., 2009) with the goal of mobilizing international educational, political, and business communities with regard to the needs and opportunities for transforming educational assessment and instructional practices. The JISC (Joint Information Systems Committee) published an overview of technologies, policies, and practices with e-assessment in further and higher education. The JISC is also undertaking efforts to standardize assessment. Along similar lines, the IMS Global Learning Consortium presented the IMS Question and Test Interoperability Specification (IMS Global Learning Consortium, 2008).

One of the most common CBA is performed online; it consists of a website where the students can reach the survey system and log in. Once they are in, they can select their answers from multiple items and they can write down open-ended questions in text boxes. When they have submitted their answers, they can also obtain a document with a statement of accomplishment about the evaluation made (Dommeyer et al., 2004). Some benefits of CBA are that evaluations of this kind eliminate paper costs, can be faster and easier to complete, allow efficient processing of data and are less vulnerable to influence by the faculty (Dommeyer et al., 2002b). Additionally, CBA allows adaptive testing based on the responses, which is not possible with paper-based assessments (Brown et al., 2008). Nevertheless, online assessment also has some disadvantages, such as requiring students to have technical access and to know their log-in information. Some students may also experience technical problems when accessing the evaluation (Anderson et al., 2005).

With regard to the preference of completing online evaluations over paper ones, there is no unanimity about this preference. Several works have indicated that students prefer completing online evaluations to paper ones (Layne et al., 1999; Dommeyer et al., 2004; Anderson et al., 2005). In the study carried out by Anderson et al. (2005) when asked about their preferred evaluation format (online or traditional), over 90% of the students selected agree or strongly agree in favour of the online format. Other studies contradict these data and mention that students prefer pen and paper exams to computer-based options (Llamas-Nistal et al., 2011, 2013). Other studies have argued that online evaluations tend to produce more written comments than traditional, in-class evaluations (Dommeyer et al., 2002a), and allow students to perform the evaluation collaboratively (Conejo et al., 2013), or even perform self-evaluation (Gathy et al., 1991) or personalized assessments based on their own knowledge and objectives (Lazarinis et al., 2010). Sorenson and Johnson (2003) determined that students give more and longer answers when they are performing an online assessment than when they are using a traditional paper-pencil system. Another study stated that the online tool was easy to use, students appreciated the anonymity of the online assessment, and that evaluations of this kind allowed students to offer more thoughtful remarks than performing the traditional evaluation (Ravelli, 2000). Blended approaches have also been taken into account in previous studies. Llamas-Nistal et al. (2013) combined the benefits of the digital world with the convenience of traditional evaluation and assessment sessions. This tool may also be seen as a cost-effective alternative to computer-supported e-assessment in those cases where the use of computers for performing assessment is not convenient or possible.

A few comparative studies have also been carried out. Wilson et al. (2011) studied the effectiveness of computer-assisted formative assessment in a large, first-year undergraduate geography course. Statistical analysis showed that the students who used the computer-assisted practice quizzes earned significantly higher grades than those students who did not. Wang (2014) performed a study in which four different e-Learning models were compared (with personalized dynamic assessment, without personalized dynamic assessment, with personalized e-Learning material adaptive annotation, and without personalized e-Learning material adaptive annotation). From their results, the e-Learning models compared without personalized dynamic assessment and the e-Learning models with personalized dynamic assessment were significantly more effective in facilitating student learning achievement and improvement of misconceptions.

2.3.2 Using images for assessment

It is generally accepted that images and graphics can communicate complex ideas with clarity, precision, and efficiency. For example, often the most effective way to describe, explore, and summarize a set of numbers is to look at pictures of those numbers (Tufte, 1989). Reports, executive summaries, and handouts or PowerPoint slides used in verbal presentations all benefit from accompanying graphics to capture attention, communicate key information at a glance, and increase understanding and memory retention. Think of graphics as giving the reader the greatest number of ideas, in the shortest time, with the least ink, in the smallest space (Kusek and Rist, 2004; Patton, 1997). It is important to present graphics with written or verbal explanations to ensure their correct interpretation (Torres et al., 2004). Several works have explored the role that images can play in the engagement of schoolchildren. For example, Busschots et al. (2006) explored this aspect for scientific discovery with an astronomy system. They described an online image analysis tool that was developed as part of an interactive, user-centered development of an online system. This system provided a suite of software tools used by schoolchildren and their teachers to study astronomy. In their case, the astronomical images were spectacular and had the ability to spark the imagination of the participants and, thus, provided a great medium for exploring the role that images can play in the engagement of schoolchildren in scientific discovery. Torres et al. (2004) stated that people learn more when they are engaged with the learning material, when they see, hear, and do something with the content, and when they integrate new knowledge with something they already know. There is also evidence that once an online system has been implemented, over time the response rate will gradually increase (Avery et al., 2006).

Although images are considered important for understanding and solving problems, very few previous works have studied their influence on item solving. One of the works to cite is the study of Dindar et al. (2013). They carried out a study with 112 students in which they compared animated questions vs. static graphic questions. No statistically significant difference was observed in terms of response accuracy between the static group and the animation group. The second work to cite is the study of Saß et al. (2012). They carried out a study with 158 students in which they included or did not include images in the stem and in the answer options. Their results indicated that images in the stem and in the answer options increased the number of correct answers.

2.4 Natural User Interfaces

NUI allow users to be the controller themselves by detecting the position of the different parts of their body. According to Fishkin (2004), NUI facilitate the acceptance of an application by users. However, adult people are more reluctant to accept the new advances, and more effort is required to introduce these technologies. For example, the study of Carvalho et al. (2012) presented a multi-touch game that was envisioned to encourage and teach digitally excluded people, as shown in Figure 2.3. From its results, they found that the use of NUI may be beneficial to help overcome some difficulties produced by the digital divide. As mentioned above, NUI are being

(a) First impact of the interface (b) Change in behaviour (c) Teamwork

Figure 2.3: Carvalho et al.'s experiment stages

incorporated in a large number of different types of applications: for physical rehabilitation (Chang et al., 2011b; Lange et al., 2011); for training individuals with cognitive impairments (Chang et al., 2011a); for navigating with Google Earth (Kamel Boulos et al., 2011); for videoconference in which depth perception was added to attendants (DeVincenzi et al., 2011); for transforming a paper based process to a NUI process in a chronic care hospital (Anacleto et al., 2012); 3D imaging for hand gesture recognition (Periverzov and Ilieş, 2012); or interaction with 3D objects from touchscreen inputs like the study of Cohé and Hachet (2012), in which non-technical users tended to interact with the objects by 3D cube manipulations. One of the objectives of their work was to find out the most widely used strategies for manipulating the 3D objects. The hardware used in that study was a TouchCo 13-inch multi-touch surface and an Optima video-projector, which was placed perpendicularly to the table, as shown in Figure 2.4. The users sat in front of the video-projector and were asked to interact with the objects. Many teachers have found that the systems that use interactive whiteboards are a great motivating teaching tool (Rudd, 2007; Warwick et al., 2010). Lien et al. (2012) developed an L-shape platform where the students could learn by moving their limbs in an easy way.

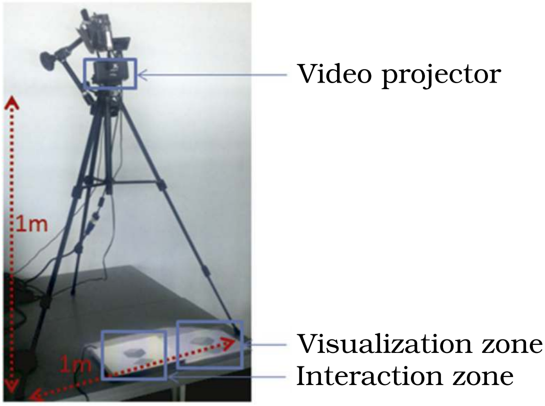
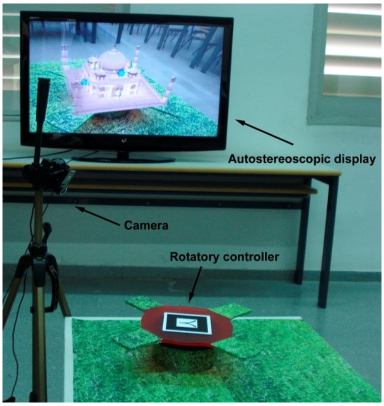


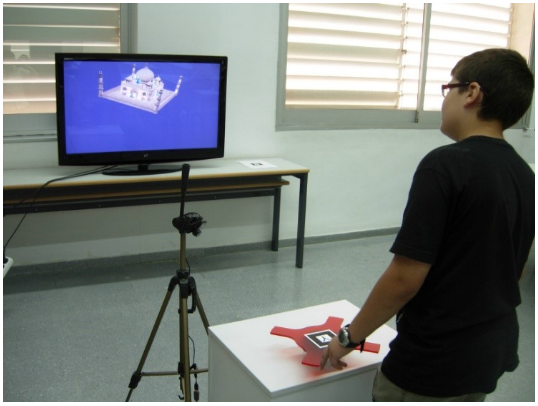
Figure 2.4: Cohé and Hachet's experimental setup

2.5 Autostereoscopic systems

Autostereoscopic systems are used for creating a stereoscopic image without the need of using external devices like glasses or other wearable elements. These systems are able to bring a different image to each eye, so the human brain can build the 3D scene. Several works can be cited regarding autostereoscopic systems. In 2012, Maimone et al. (2012) presented an enhanced personal autostereoscopic telepresence system using depth cameras. Kim et al. (2012) used this technology combined with NUI making an autostereoscopic display for multi-user collaboration. Kim et al.'s study proposed an autostereoscopic platform for sharing visual data with two or more users, which uses two displays. Our proposal shares this use of OpenNI and Kinect. Autostereoscopic systems have also been used for immersion platforms for the World Wide Web as in the work of Nocent et al. (2012) in which tracking devices were used. Taherkhani and Kia (2012) presented an eye-tracking display with autostereoscopy that used an LCD monitor by using the parallax barrier technique, where the two stereo images were rendered on the LCD simultaneously, but in different columns of pixels. Arino et al. compared augmented and virtual reality by using an autostereoscopic display, where 39 children from 8 to 10 years old tested both modes by playing a video game (Arino et al., 2014). From their study, no statistically significant differences were found between AR and VR. However, the children explicitly preferred the AR version (81%). Figure 2.5a shows the different elements used in the AR version. Figure 2.5b shows a child playing with the VR version.



(a) AR version. Table with the marker on the rotator support, camera, and autostereoscopic display

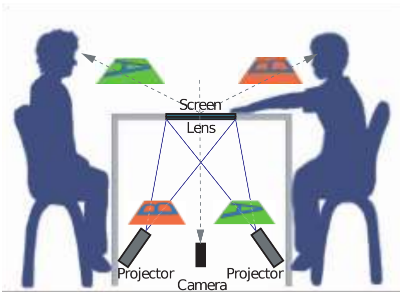


(b) A child plays with the VR version

Figure 2.5: Images from the Arino et al.'s study

2.6 Frontal Projection systems

Frontal projection systems are based on a commonly white surface where the images are visualized by using a projector. Several previous works can be cited for frontal projection systems. Chan and Lau (2012) presented a CAVE-like system with a projector and some infrared cameras on the ceiling of the CAVE structure. The user held the "Magicpad" in his hands (which mainly consists of one or more flat white surfaces). The user could use an infrared pen for tracking and 3D glasses to look at the Magicpad. Kubicki et al. (2013) presented another work involving interactive tables in which a traffic simulation was developed. A TangiSense interactive table was chosen for this. It is not a tactile table; the interaction is achieved via tangible objects that are placed on the table. It detects overlapping objects by using Radio Frequency Identification (RFID) technology. This technology lets the user handle objects equipped with RFID tags; in this case, for the simulation, these objects were traffic signals. Kakehi et al. (2005) developed an interactive view-dependent tabletop display. It was made up of some cameras and projectors installed under the table, which had a special screen system that allowed multiple images to be projected at the same time without the user having to wear specific devices. This experiment is shown in Figure 2.6. For entertainment purposes, in the work of Lam et al. (2006) a prototype



(a) Overview of the Lumisight Table architecture

(b) Four users' views at the Lumisight Table

Figure 2.6: Kakehi et al.'s interactive view-dependent tabletop display

of an AR table was designed for a card game. It consisted of an over-head camera to register card inputs and a plasma display to act as the game board, rendering 3D models and generating sounds. The system was able to recognize player inputs such as pressing command buttons and card inputs. This system needed a calibration process before playing in order to provide a stable environment. According to several works, interactive tables are enjoyable and engaging to use (Rick et al., 2009a). They also promote equity of participation (Harris et al., 2009) and encourage learning (Jamil et al., 2011; Falcão and Price, 2009).

This kind of technologies are used too nowadays for music purposes, for instance, as *Tangible music* with the *Reactable* system (Jordà et al., 2006; Jordà et al., 2007; Kaltenbrunner and Bencina, 2007; Kaltenbrunner, 2009) showed in Figure 2.7. This is an HCI AR-based interface that they used in their study for producing music; in addition, several musicians had used this system in concerts. It was developed to be used as an electronic multi-user musical instrument. It is based on a translucent round table that has a camera placed beneath it that analyzes the table's surface, tracking the position, orientation and type of objects on the surface.

In 2001, Dietz and Leigh (2001) presented DiamondTouch, which integrates a multi-user touch technology. This technology detects multiple and simultaneous touches and also identifies which user is touching each point. The system works

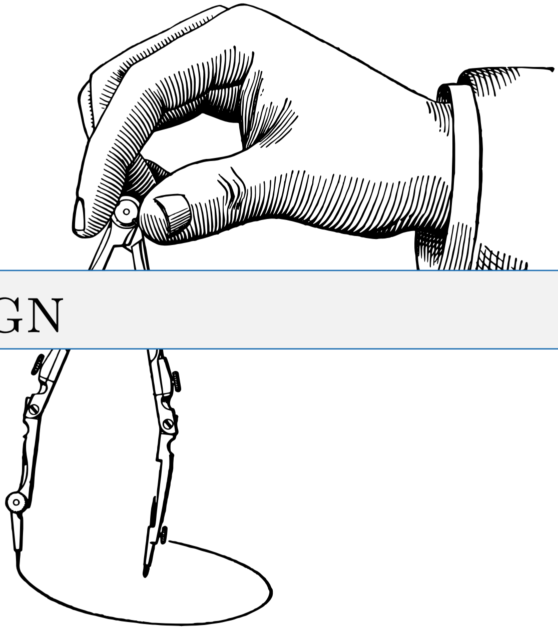
Figure 2.7: Reactable system

by transmitting signals that are capacitively coupled through the users and chairs, allowing user identification at each touch. DiamondTouch has been used in several studies (Smeaton et al., 2006; Esenther and Wittenburg, 2005; Kobourov et al., 2005). In 2010, Weiss et al. (2010) presented an interactive tabletop that allows interaction with complex physical controls. These authors use electromagnetic actuation techniques beneath the table to carry out the interaction. One of the drawbacks is the limited size of the display (24"), which limits the multi-user approach.

Several contributions have demonstrated the potential of interactive tables: they promote playfulness (Mansor et al., 2009; Marco et al., 2009), increase awareness (Hornecker et al., 2008), and are enjoyable to use (Do-Lenh et al., 2009). In the education field, these tables are especially suitable for learning purposes; for example, Numbernet, developed by Hatch et al.'s, an application in which a group of children use multi-touch technology to learn mathematics (Hatch et al., 2011). Digital Mysteries, created by Kharrufa et al., which is a collaborative learning application for school children (Kharrufa et al., 2010). Piper and Hollan compared the affordances of presenting educational material on a tabletop display and presenting the same material using traditional paper (Piper and Hollan, 2009). Falcão and Price studied collaborative activity on a tangible tabletop to learn about the physics of light (Falcão and Price, 2009). Suh et al. (2010) investigated the effectiveness of MMORPG-based instruction in elementary English education. Their findings suggest that games of this kind may be useful for improving the English abilities of students who study English as a second language. The students who used the game showed higher scores in listening, writing, and reading than those who learned with face-to-face instruction. Finally, we would like to highlight that some authors have already pointed out that one of the most promising uses of these interfaces is to assist learning through exploration (Rick et al., 2009b).

III

Design



Design

3

"خلف الصبر اشياء جميلة تنتظر"

— Arabic proverb

3.1 Introduction

In this chapter, a detailed explanation of the design process and the reasons why we chose this topic are exposed. After that, the description of the game is detailed.

3.2 Game design

The subject of the game that was chosen was a historical timeline, specifically, five historical ages (Prehistory, Ancient Times, the Middle Ages, the Early Modern Period and the Contemporary Period). The pedagogues of AIJU[1] were in charge of the selection of the theme. Based on their experience, they proposed four possibilities. Many other works have pointed out (De Freitas and Oliver, 2006; Law et al., 2008) the importance of considering the contents of the national curricula to develop educational computer games. According to this, the proposals of the game met the requirement that the topic should be one that is included in the primary education law of Spain[2]. The four proposals were:

1. Animals: Vertebrates and their reproduction
2. Animals: Nutrition
3. Plants: Parts, nutrition and their reproduction
4. The passage of time: Learning about different historical periods

We selected the last proposal, the historical period. The reasons for this selection were the following. First, in a meeting with the entire research group, the topics related to animals or plants were discarded because our group had already used these topics for different educational games (e.g., endangered animals (Juan et al., 2011a)). Second, in one of our previous works (Furió et al., 2013b), we carried out a study to determine the subject preferences for an educational computer game for children ranging in age from 8 to 10 years old. In this study, history was rated in fifth place among 12 possibilities. The 12 possibilities were the following: 1) nature and living organisms; 2) science and technology; 3) traffic rules; 4) sport and outdoor activities; 5) health and hygiene; 6) moral and ethical values; 7) multiculturalism, solidarity and tolerance; 8) calculation and reasoning; 9) music; 10) history; 11) language and understanding; 12) other. The five favorite subjects were the following, in order of preference from first to fifth: nature and living organisms; multiculturalism, solidarity and tolerance; science and technology; calculation and reasoning; and history.

The knowledge presented in the game is the same as what the children study at school. This knowledge was extracted from books used in the classroom. The primary education law of Spain divides these contents into three cycles. The first cycle includes basic notions of time (before – after, past – present – future) and time units (day, week, month, year); some events from the past and from the present and their relationship with historic topics appropriate for children; and the use of audio explanations to reconstruct the past. The second cycle contains the use of

[1] Centro Tecnológico del Juguete

[2] Boletín Oficial del Estado (BOE). Available at http://www.boe.es/boe/dias/2006/12/08/pdfs/A43053-43102.pdf

more time units (decade, century) and initiation to terms like succession, order or simultaneity; learning about ancient societies from daily aspects; recognition and meaning of traditions or historic buildings; and distinguishing the roles of men and women in history. Finally, the third cycle includes topics like data conventions such as (B.C., A.D., age); techniques to locate facts about the past on a timeline and notice the duration, simultaneity and relationships between historic events; characterization of some historical ages (Prehistory, Ancient Times, the Middle Ages, the Early Modern Period, the Contemporary Period) through their lifestyles; important people and events in the history of Spain; the use of some historical, geographical or artistic sources for making reports and other historical content; appreciation of men and women as subjects of history.

The design of the game is based on the experiential learning theory of Constructivism (Dewey, 1963). According to (Hernández, 2008), this method implies that the acquisition of knowledge can be focused on carrying out rich, context-based activities. Nowadays, this can be advanced with the use of the new technologies that have emerged in the last few years. These tools offer the students unlimited information immediately and the possibility to control their own learning directions themselves. Hernández proposed changing the classical paradigm of being in the classroom with the blackboard and chalk and using new ways to introduce new content. Following Hernández's proposal, other proposals, and our own personal experiences, we developed our game following the experiential learning theory of Constructivism. As a computer-supported group-based learning system, we also designed the game taking into account the approach proposed by Strijbos et al. (2004). This approach consists of five elements: three elements are shown as dimensions (learning objectives, task type, and level of pre-structuring), and two elements are shown in terms of discrete categories (group size and computer support). We followed the six steps suggested by Strijbos et al. (2004) to design our game. These steps are the following:

1. Determine the learning objectives: The learning objectives can range from "open skills" to "closed skills". Open skills are skills such as argumentation and negotiation. Closed skills are relatively fixed skills that can be learned separately. In our game, the skills are closed.

2. Determine the expected (changes in) interaction: focusing on feedback, exchanging ideas or discussion. Our game is focused on exchanging ideas.

3. Select the task type: The Task type can range from "well-structured tasks" to "ill-structured tasks". Well-structured tasks aim to convergence because there is only one correct solution. Ill-structured tasks have a considerable degree of

uncertainty regarding the rules and principles that can be applied and often have no clear-cut solution. In our game, the task type is well-structured.

4. Determine whether pre-structuring is needed and how much: The Level of pre-structuring can range from “high” to “low”. It addresses the level to which interaction is pre-structured in advance to ensure positive interdependence (Johnson, 1981) and individual accountability (Slavin, 1980). Our game has a high level of pre-structuring. If two participants are interacting with the system, half of the interactions with the system and half of the answers to the questions must be performed by each member in order to reach the next level.

5. Determine group size: Group size ranges from pairs (two members) to small groups (three to six members) or large groups (seven or more members). In the collaborative version of our game, we have tested the system with pairs and large groups (till 12 participants). We chose pairs for the following reasons: pairs are the smallest possible social unit; as the size of the group increases, it becomes progressively more difficult to identify the successful components of cooperative learning (O’Donnell and Dansereau, 1995). In order to determine the advantages of playing in pairs, we compare the game played in pairs with playing in large groups.

6. Determine how computer support can be applied (*with*, *at*, *through*): Interaction *with* computers refers to individual student interaction with a computer simulation. Interaction *at* computers represents a group of students interacting with a computer program; it can be either face-to-face or computer-mediated. Interaction *through* computers refers to interaction between group members via networked computers. In our game, the student’s interaction is at the same time and place (face-to-face) and *at* computer. Each child interacts with the system. To choose an answer or option, the children must talk to the other members in order to share information and to help them to achieve the solution.

The design guidelines for classroom collaborative games proposed by Villalta et al. (2011) were also taken into account to design our game. These design guidelines include the following features for collaborative games:

1. Interactivity and guidance. Our game offers guidance by the avatar guide. The user’s interaction with the game is simple and intuitive. It uses natural interaction (gestures without any special hardware).

2. Mechanics linked to learning objectives. In our game, the curricular content is embedded in our game. The success of the game is conditional to understanding its content.

3. Clear narrative. Our game has a base story that allows the immersion of the participants. The narrative is composed of challenges that define collaborative activities in a sequential and precise pattern.

4. Gradual increase in difficulty. This guideline was not taken into account for this version of our game.

5. Teacher mediation during the game. In our game, there is no participation by the teacher because it is a computer-supported learning game.

6. Organization of face-to-face interaction. Our game promotes communication between students. The students must agree before choosing an answer or an element.

7. Mechanics linked to collaboration. In our game collaboration is embedded. Success is conditional to having worked collaboratively.

8. Adequate spatial distribution. Our game distributes elements and activities on the screen in order to take advantage of the available space. The spatial distribution correctly relates aspects of the embedded knowledge to the connection with the real world. In the current version of our game, the player cannot control the camera. However, the elements that require a 360-degree visualization rotate continuously while onscreen.

9. Recognizable elements. In our game, the elements on the screen have distinctive traits that capture the players' attention. Moreover, the most important elements are displayed with stereoscopy in the autostereoscopic system.

10. Accessible language (the text on the screen must have a clear message and be concise and easy to read). In our game, the text only appears in the buttons and in very specific information on the screen. Villalta et al. (2011) stated that spoken information should be preferred over written text because it induces less cognitive load. In our game, we include a guide avatar that offers the children audio explanations about the steps to follow.

11. Avoid information overload. Our game avoids information overload by limiting the information on the screen.

12. Action guide. Our game includes educational and playful aspects. It includes a script that specifies action sequences and events that can take place in both the virtual and real world.

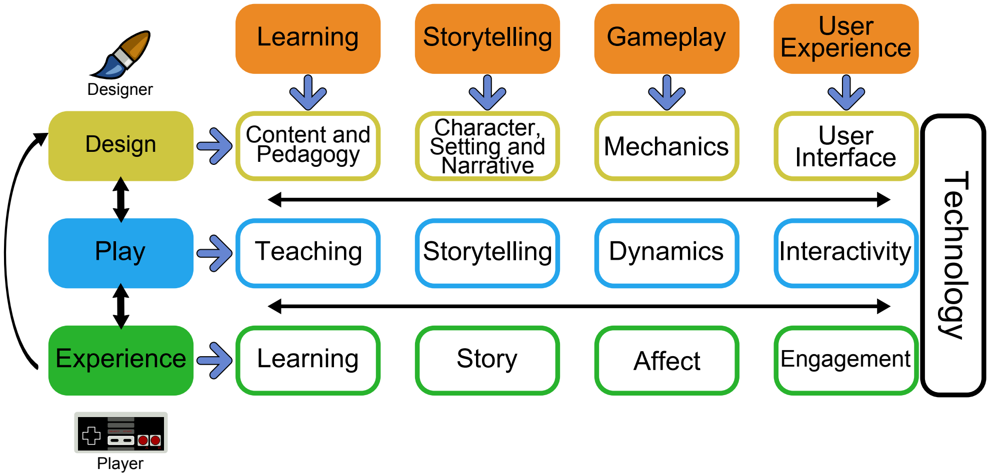


Figure 3.1: Expanded DPE Framework

The *design, play and experience* (DPE) framework proposed by Winn (2008) had a great influence on our design. This framework, shown in Figure 3.1, was created as an extension of the *mechanics-dynamics-aesthetics* (MDA) framework (Hunicke et al., 2004) to address the needs of serious games design for learning. The MDA framework focuses on the design of games for entertainment. Both frameworks take into account the relationship between the designer and the player. The DPE framework layers are the following:

- The learning layer. Our design defines the content, the pedagogy and a set of learning outcomes, i.e., the knowledge that is transmitted by playing the game. The main purpose of the game was to help and motivate children to learn about the main historical periods of the time line. The contents of the game were based on the textbooks that children use in the classroom, which are determined by the primary education law of Spain. The periods of history used in this game were: Prehistory, Ancient Times, the Middle Ages, the Early Modern Period, and the Contemporary Period.

- The storytelling layer. It creates a story to transmit content. In our history game, there is one character, the avatar guide, which guides the children during the game. With his narrative, the avatar transports the children to other historical ages so that they can think and act as if they were there (e.g., painting in a Prehistoric cave).

- The gameplay layer. It defines what the player does in the game. The gameplay layer is composed of mechanics, dynamics, and affects. The mechanics are the rules that define the functionality of the game, what the player can do, and the player's challenges and goals. The dynamics are the resulting behavior when the rules are instantiated over time with the player's interactions. The affects are the resulting experiences or emotions that are induced in the player. As Hunicke et al. (2004) indicated, the gameplay layer most closely resembles the original MDA framework, but with a change in terminology (specially, the change of aesthetics by affect). At first (from a player's perspective), the game can be considered as fun or not fun. However, several researchers have tried to identify the specific aspects that create a fun experience. For example, Hunicke et al. (2004) identified eight kinds of fun as aesthetic goals. Heeter et al. (2005) proposed sixteen forms of fun that include the following: beauty, immersion, intellectual problem-solving, competition, social interaction, comedy, thrill of danger, physical activity, love, creation, power, discovery, advancement and completion, application of an ability, altruism, and learning. Of the goals on Heeter et al.'s list, we consider that our game contemplates the following nine aesthetic goals: beauty, immersion, competition, social interaction, physical activity, creation, discovery, advancement and completion, and learning. With regard to the rules, there are two types of interaction depending on the mini-game that the children were playing: Press button and Drag & drop. In the first type of interaction (Press button), the possible answers to a question are displayed in the screen (in the case of the autostereoscopic system) and on the large-tabletop display (in the case of the frontal projection system) as buttons and the children have to choose the right one. The children have to consecutively select the correct answers to the questions that the avatar asks. When the children have answered all the questions correctly, the game continues to the next mini-game. When the children select an incorrect choice, the button turns from yellow to red; if they select the correct answer, the button turns from yellow to green. In the second type of interaction (Drag & drop), there are several drop areas on the screen and their corresponding images. Using their hands, the children have to drag the images and drop them into their drop area.

Each element that the children can interact with is placed near the edges of the screen so that the children can easily reach the buttons or the draggable images. Once all the mini-games of a historical age are done, the game presents the next historical age and its first mini-game. As the game advances through all the historical ages, a time line that is placed in the upper part of the screen indicates the current historical age that the children are in. According to the theory of flow of Csíkszentmihályi (1990), in order to achieve a state of flow, the level of challenge must match the players' abilities as their skills increase. If the challenge is too ambitious, the player could become frustrated and might give up. If the challenge is too small, the player could quickly become bored and might quit playing. In our game we achieve this balance by establishing the most appropriate age range. Another form of gameplay balancing from the theory of flow is related to the frequency of rewards given to the player. In our game, we have incorporated this gameplay balancing by showing the historical ages that the children have successfully passed in the time line.

- The user experience layer. The user interface represents the mode of communication between the player and the game. As Hunicke et al. (2004) indicated, the ultimate goal of the designer is to develop a game that immerses the players in the game world and engages them in the play experience. Good user interfaces should be transparent, in other words, the players should not have to focus their attention on how to play the game but rather focus on the gameplay, the storytelling, and the learning experience. This was one of our main goals in designing the game. In our game, the player does not have to wear extraneous devices such as headgear or special glasses and the interaction is achieved using natural gestures. Therefore, in our game, the players' attention is on the gameplay and the learning experience. In order to ensure that the children listen to the entire audio and video explanations, the interaction is disabled during the explanation so that the children cannot select elements or drag objects.

Considering all the aforementioned theories and guidelines, the design of the game was an iterative process, in which several iterations were held. Not all the periods required the same number of iterations. Figure 3.2 shows the three iterations performed during the design process for the cave painting mini-game. Iteration A shows a sketch where the main elements of the interface are placed. This sketch gives an idea of the purpose of the mini game and the actions that will be carried out in that mini game. Iterations B and C distinguish the two different layouts of the autostereoscopic and frontal projection systems, and show where the children will be located while playing the video game.

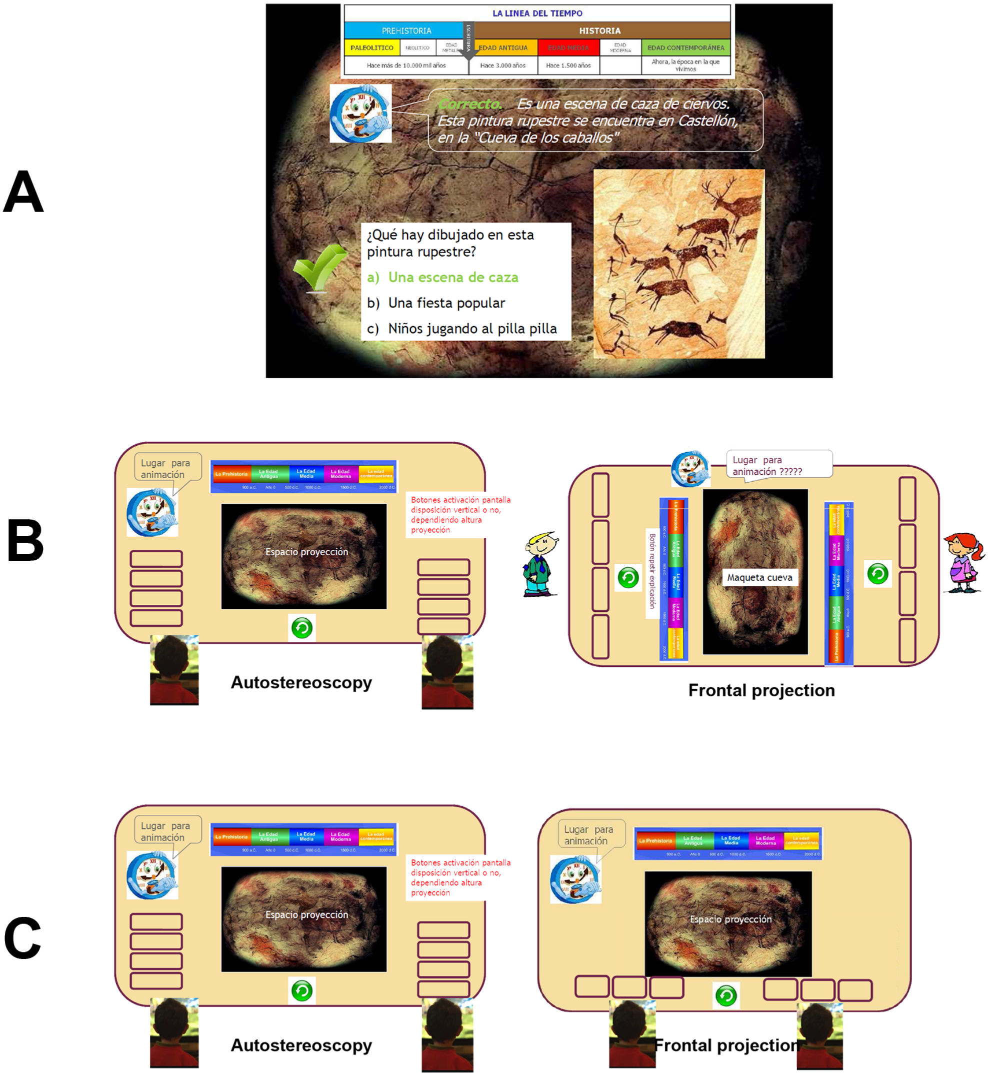


Figure 3.2: Design evolution of the cave painting mini-game

3.3 Game description

The aim of the game is to reinforce the learning of the concept of time line, including its order, and the characteristics of each historical age. The children assumed the mission of completing the time line by travelling through the different historical ages. The game is divided into a series of mini-games, several of which pertain to each time period on the time line. The order of these ages in the time line and the events that start and end each historical age are emphasized. There are also video and audio explanations at the beginning of the mini-games to introduce the historical ages and to give more detailed information in each mini-game.

The game had the same stages and order in both the frontal projection and autostereoscopic configurations. The children played the game from Prehistory to the present day. All the mini-games shared some common characteristics. The children had to use their own hands to interact with the games, searching for shapes or pressing buttons by moving their hands to the active area. In the case of frontal projection, the buttons were placed at the bottom of the screen, and, in the autostereoscopic case, they were placed on the sides of the display. The position of the buttons was different in the two configurations since the position of the children and the place where they put their arms in a stand pose were different. The game consisted of seven mini-games distributed into five historical ages: Prehistory, Ancient Times, the Middle Ages, the Early Modern Period, and the Contemporary Period. In the frontal projection configuration, the children could interact with whichever hand they preferred. However, in the case of the autostereoscopic configuration, the child playing on the right had to use his/her right hand, and the child playing on the left had to use his/her left one. This choice helped the children to interact with the buttons that were close to them.

The flowchart of the mini-games that shaped the game is the following (see Figure 3.3):

1. At the beginning, the children heard the voice of an avatar introducing them to the game. The avatar guided them through the process of user calibration by adopting the *Psi* pose (Ψ), only for the autostereoscopic version.

2. Once they were ready to start, they had to select the first historical age from the time line, Prehistory, by pressing the correct buttons. After a video explanation of Prehistory, they played two mini-games from this time period; the first consisted of finding some cave paintings and using the colours that the cavemen used for that purpose. In the second mini-game, the children had to select a colour and leave an imprint of the shape of their hand in the cave.

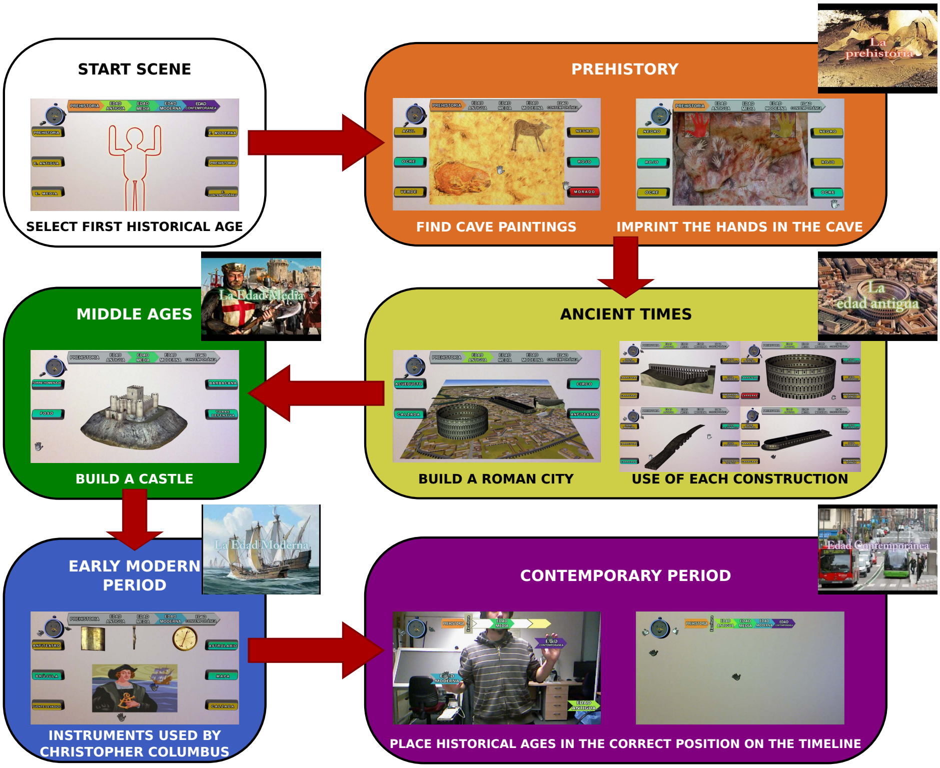


Figure 3.3: Flowchart of the mini-games integrated into the game of history

3. When all this was done, they had to select the next historical age the same way as previously; this time, it was the turn for Ancient Times. In this mini-game, the children had to reconstruct a Roman city by placing an amphitheatre, an aqueduct, a Roman circus, and a Roman road in it. Afterwards, the game asked them some questions about the use of the buildings they had just used to construct the Roman city. The Roman city stage for the two systems is shown in Figure 3.4, where the button disposition between the two systems is distinguished.

4. The next historical age was the Middle Ages. Here, the children had to build a medieval castle by correctly answering the questions the game asked. By

choosing the correct answer, one more piece of the castle was added to the structure. At the end, the whole castle was visible, as shown in Figure 3.5, and the children could go on to the next historical age.

5. After completing the Middle Ages, the children began the Early Modern Period, where they had to find the objects that Christopher Columbus used in his journeys to discover the American continent. This task was completed by finding a compass, a map and an astrolabe.

6. When all these objects were found, the children reached the final historical age (and last stage) of the game, the Contemporary Period. In this part of the game, the children had to complete a puzzle that recreated the timeline by moving their hands. Once this puzzle was complete, a final audio speech was played telling them that they had reached the end of the game.

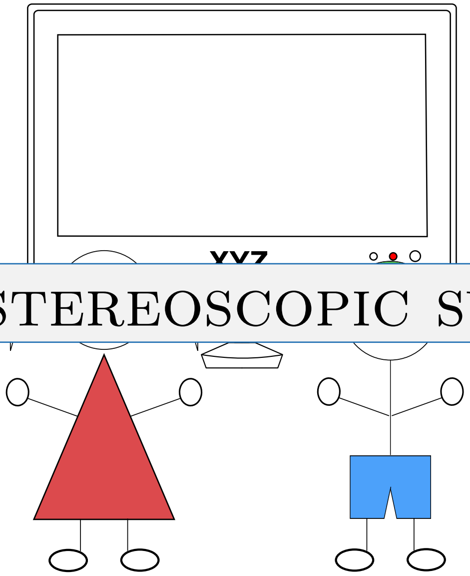
(a) Frontal Projection (b) Autostereoscopic system

Figure 3.4: Roman city stage and button disposition

Figure 3.5: Medieval castle completely built

IV

Developments



4 Autostereoscopic system

> *"I haven't failed, I've found 10000 ways that don't work."*
>
> — Thomas Alva Edison

4.1 Introduction

In this chapter, the autostereoscopic system is presented. Technical details about the hardware and the software used, and about interaction are found in this chapter.

4.2 Technical details

This system combines autostereoscopic visualization and natural interaction. The real world and users are captured by the camera. This image appears in the background without stereoscopy. The virtual elements are the objects with stereoscopic perception. Therefore, the children could see themselves in the screen. They were able to interact with the game by moving their hands. This interaction was achieved using the Kinect camera, which detects their movements. Figure 4.1 shows a couple of children playing with the autostereoscopic configuration. As Figure 4.4 shows, the autostereoscopic display was placed on a mid-height table which allowed the children to see their entire body on the screen. The Microsoft Kinect was placed in front of the 3D display, which was centered relative to the 3D display. Also, there were two numbered markers at a distance of about 2 m. from the display to let the children know the surface area where they should stand.

Figure 4.1: Children playing with the autostereoscopic system

4.2.1 Hardware

To capture the image of the real world and to track the children's bodies, a Microsoft Kinect device was used, represented in Figure 4.2. Microsoft Kinect consists of two frontal cameras (one is a RGB camera similar to a webcam, and the other is an infrared camera) and one infrared emitter. The RGB camera had a 640×480 pixel image resolution. The autostereoscopic rendering was made possible by using an XYZ display. The specific model was XYZ3D8V46, which had a screen size of 46" and full HD resolution (1920×1080 pixels). This display renders the eight views of the 3D vision. To do this, it uses a technology known as LCD/lenticular (Omura et al., 1998).

4.2.2 Software

The OpenSceneGraph (OSG) toolkit 3.0.1 was used to render the 3D models and the virtual world. It is an open source graphics toolkit that is written in Standard C++ and OpenGL. In OSG the run method calls the main loop. The run method internally calls the frame method which updates the next frame of the graph scene. The problem with this approach is that the behaviour of this loop cannot be modified. To solve this problem, the run method has been replaced by the code shown in Listing 4.1. For our system, with this loop, the scene to render was updated with the children's pose, the sound and the video states.

```
while !done
  update children's pose
  update sound and video
  perform interaction
  frame
end while
```

Listing 4.1: Update main loop

Figure 4.2: Microsoft Xbox Kinect

The autostereoscopic rendering was performed by using the Mirage SDK (`http://www.mirage-tech.com`). This SDK provides an OSG `Node` that is able to calculate the eight different views needed by the 3D perception. With this node, an OSG scene can be defined by adding cameras, 3D models (in format `.osg` and `.osgt`), transformation matrices, etc... Finally, this node must be established as the root node of the scene graph. Once this scene graph is complete, a 3D perception can be perceived in a display without using any glasses or external devices. Figure 4.3 shows how this integration was achieved.

To render the scene, in our system, the image captured by the camera was shown as a background image in each of the 8 views. To make these 8 views work fluently, the application required the image captured by the Kinect camera to be developed using a separated thread to the update main loop. This process is shown in Listing 4.2. Additionally, each of the 8 views renders the calculated image for each point of view of the 3D models.

```
while !done
  disable z-buffer
  get pixel array from Kinect
  texturize quad at bottom
end while
```

Listing 4.2: Separated thread for retrieving the image of the Kinect camera

OpenNI and the Kinect drivers for Windows were used for registration and video capture. OpenNI allowed different users to be detected and it also returned the position of the possible *SkeletonJoints* of the user (hands, elbows, neck, head centre of mass, etc.). With the *SkeletonJoint* of the hands, it was possible to know whether or not the user was pressing the buttons.

In order to differentiate the users that were playing the game and any other people that could be nearby, the children had to be calibrated by the application. By doing this, the game was able to track and capture only the players' movements and to ignore any other people that were moving inside the game area. The children were calibrated at the beginning of the game after an audio explanation that told the children to adopt the *Psi* pose for a second. Figure 4.4 shows a silhouette of the pose. Its name comes from the resemblance that it has with the greek letter Ψ (Psi). For our development, we integrated this scene graph, the OpenNI library (which provided NUI support and video capture from Kinect), and the Mirage SDK (which provided the autostereoscopic views).

The captured videos were rendered in the background of the game and had no 3D effect. Similarly, the explanation videos were rendered in the first plane in full screen and had no 3D effect, either. These videos were decoded using the ffmpeg library (http://ffmpeg.org) and the Simple DirectMedia Layer (SDL) library (http://www.libsdl.org) for synchronizing the video files with their audios. All the video files were in `.mpg` format. Furthermore, the FMOD audio library (http://www.fmod.org/) was used to play the audio files. To make this library easier to use, the `Sound` singleton class was developed. In this way, with the help of a few methods from this class, the system was able to preload, load, and play any sound necessary for the correct guidance of the game. The singleton pattern made the code of the `Sound` class easy to be accessed from any part of the code with a really good performance. All the audio files were in `.wav` format. The system was coded in C++. Figure 4.5a shows a hierarchical diagram of the architecture of the autostereoscopic configuration. Aside, Figure 4.5b shows the architecture for the frontal projection system that the autostereoscopic system will be compared with (see subsection 5.3).

4.2.3 Interaction

Figure 4.6 shows how the buttons were located throughout the game; hand-shaped pointers for hand guidance can also be seen in Figure 4.6a. The avatar that guided the children during the whole game is represented by an alarm clock figure, shown in the upper-left corner. He guided the children telling them what they must do in each part of the mini-games.

For pressing the buttons the children had to locate the hands on the buttons, as shown in Figure 4.6b. In this way, children had to select the options that appeared throughout the whole game.

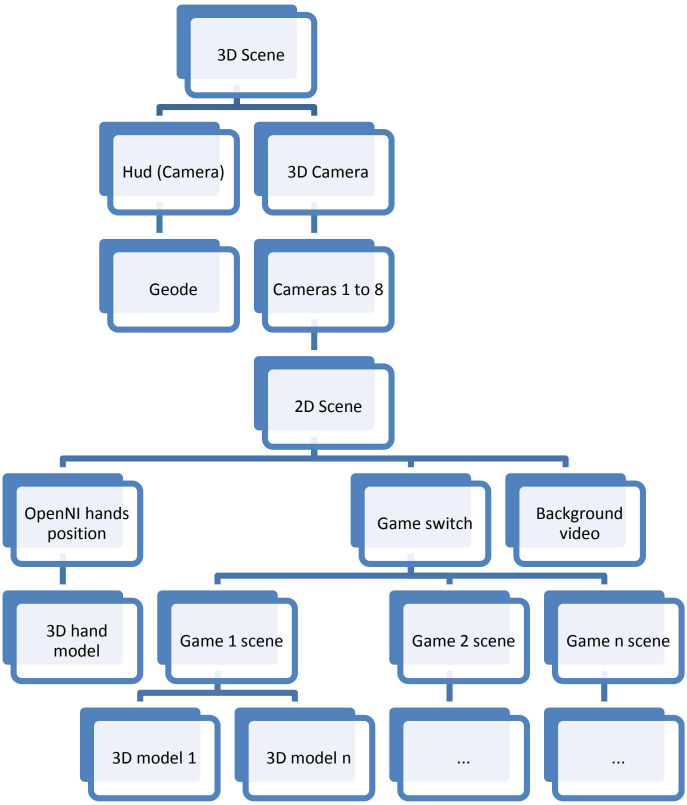


Figure 4.3: OSG graph scene used to create the 3D scene integrating OpenNI with a 3D layer specific for the autostereoscopic display

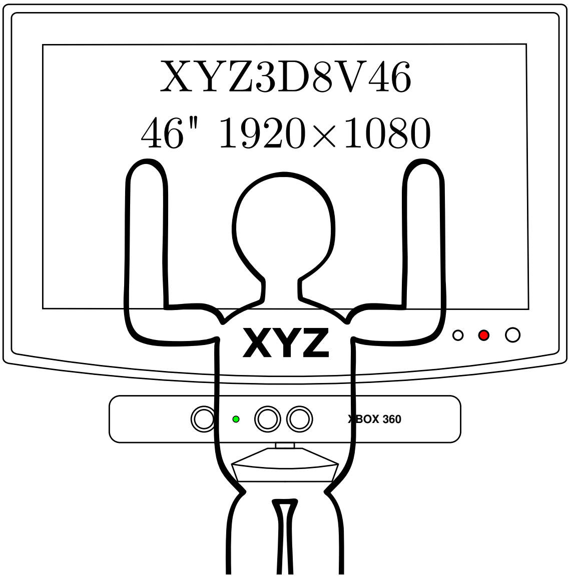
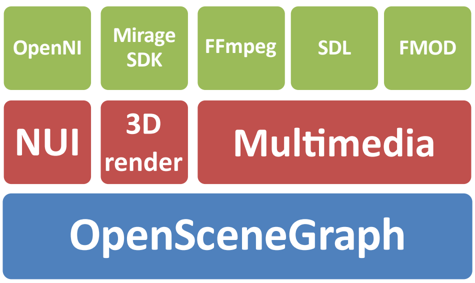
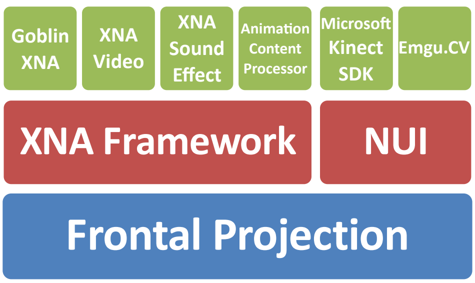


Figure 4.4: Autostereoscopic configuration. 3D display with Kinect and child position



(a) Autostereoscopic architecture



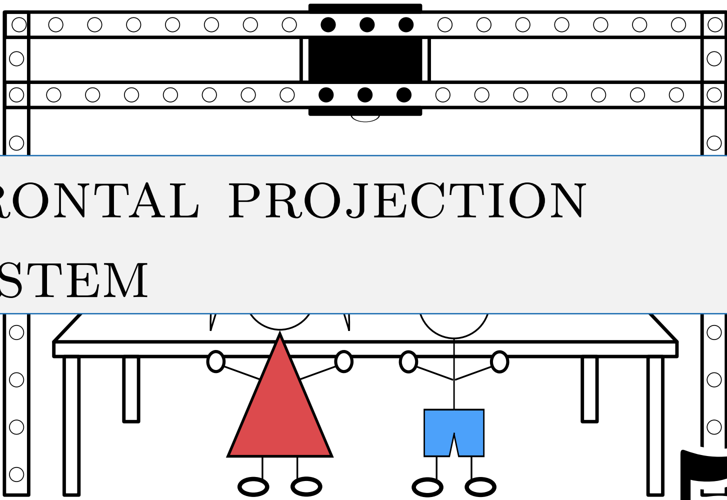
(b) Frontal Projection architecture

Figure 4.5: Systems architectures comparison

(a) Autostereoscopic configuration

(b) Frontal projection

Figure 4.6: Button interaction



Frontal projection system

5

"The mind is everything. What you think you become."

— Buddha

5.1 Introduction

In this chapter, the frontal projection system is presented. The game design and contents are the same as in the autostereoscopic game. Therefore, the design guidelines and the description of the game were explained in sections 3.2 and 3.3. Detailed information about the hardware and the software used and technical details of the system built are found in this chapter. This system works as a large-size tabletop display. To simulate a tabletop display, a Kinect device and a projector were combined with a metal structure. This system provides a table as a large tabletop display with natural interaction.

5.2 Hardware

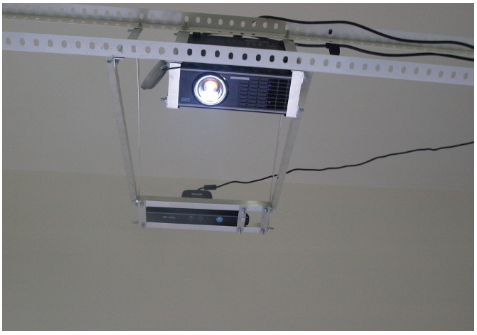
For the user interaction, a Microsoft Kinect device and an InFocus IN1503 short throw projector were used for the projection. This projector could generate an image of 177×111 cm. at a throw distance of 140 cm. It produced a brightness of 3000 ANSI lumens and had a resolution of 1280×800 pixels. A table covered with a white cardboard was used for the projection area. A steel support was used to place the Kinect device and the projector vertically as shown in Figure 5.1a. The table surface was used for capturing and display at the same time. However, since Kinect had to capture at least the entire width of the screen area, a problem arose because the projector and the Kinect were placed together on the metallic support. They cast shadows on the screen or occlude part of the capture area. The solution was to determine the distance required between Kinect and the Projector. Figure 5.1b shows this distance. Notice that the Microsoft Kinect and the projector were not at the same height and were separated from each other. When mounting this configuration, it is necessary to keep this distance in mind in order to obtain concordance between the projected images and the points captured by the Microsoft Kinect. Figure 5.2 shows a configuration schema of this system.

5.3 Software

The programming language that was used to develop the game was C#. We also used the XNA Framework with the official Kinect drivers from Microsoft. This makes the XNA Framework a quick and easy way to develop applications using the Microsoft Kinect device. Emgu.CV which is a computer vision library based on the OpenCV library was used to manipulate complex graphics. This provided several functions that were useful for hand detection. GoblinXNA was used to display the 3D scene, which provides a scene graph that simplifies the implementation. Figure 4.5b shows a hierarchical diagram of the architecture of the frontal projection configuration.

To calculate the equivalence between camera pixels and screen pixels, the system had to be calibrated. This process had to be performed each time the cardboard position or the size of the table was changed. A second application was developed to perform this calibration. This application facilitated calibration, but it had to be

(a) Two children playing with the frontal projection configuration. The steel structure can also be seen



(b) Detail of the aluminum support for the projector and Kinect

Figure 5.1: Image of the frontal projection structure and detail of the projector and Kinect

done manually. In order to define the area to be used as screen we had to use this application.

Commonly, the games that use depth information from Kinect use the distance from the camera to the user's body. Our system required a different distance, which was the distance from the children's hand to the table. To achieve this, a reference image (captured before starting the game) was subtracted from each frame. As a result, the depth image obtained took the floor and the table as a reference instead of the Kinect position.

Our system also required knowing the position of the hands. This was easy using the Microsoft Kinect SDK with the camera in the horizontal position. However, since the Kinect device was pointing to the floor and skeleton detection did not work, we developed manual recognition from scratch. To do this, each frame captured by the depth camera was subtracted from the reference image and a threshold was applied. Each white blob generated by this process was a hand candidate. The last step consisted of checking whether the white blob was an object introduced to the scene (e.g., a box or a child's baseball cap) or if in fact it was a hand. A heuristic was used, if the blob was quite big and had at least one part extended out from the table (an arm), it was considered to be a hand. In this case, the hand position was calculated as the point that was further from the table border and was contained on the blob.

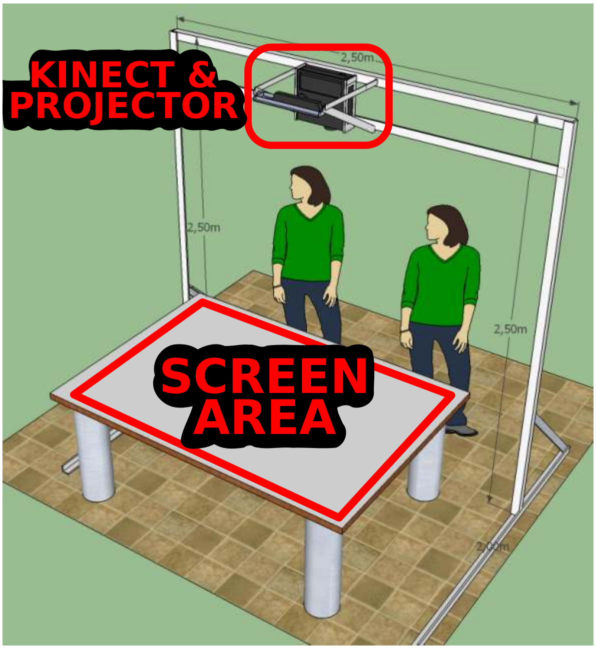


Figure 5.2: Configuration schema

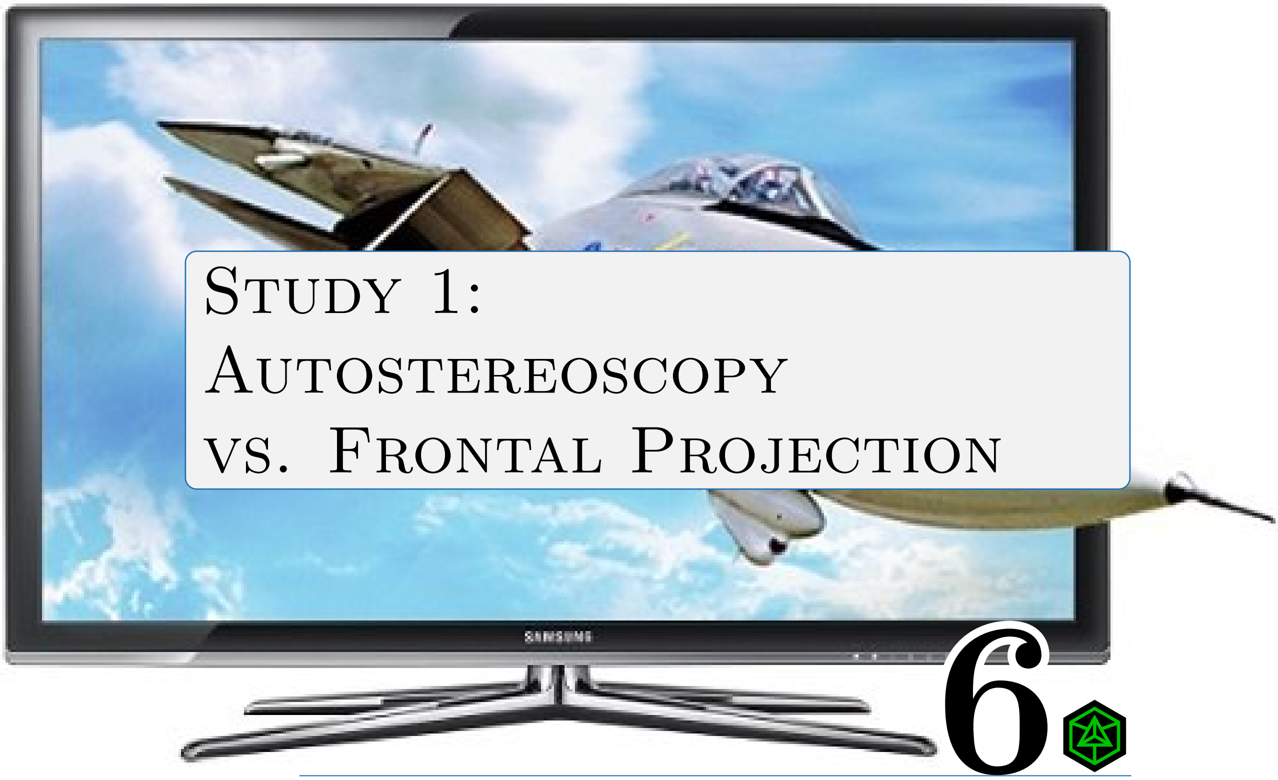
For button interaction, if a hand was close to the table and on the button, the button was considered to be pressed. The button area was partially covered by the hand at a predefined height. Figure 5.3 shows an example of a button being pressed by a child. Figure 5.4 shows the different areas the children could interact with.

Figure 5.3: Two children playing the Prehistory mini-game

Figure 5.4: Ancient Times mini-game

V

Validation



Study 1: Autostereoscopy vs. Frontal Projection

6

"Choose a job you love, and you will never have to work a day in your life."

— Confucius

6.1 Introduction

In this chapter a novel study that emphasizes the use of an autostereoscopic display as a complement for learning environments is presented. The autostereoscopic system is compared with the frontal projection system with the very same game developed inside. Both autostereoscopic and frontal projection systems use Natural User Interfaces as input method.

6.2 Design of the evaluations

The developed games described in the previous chapters (see Chapter 4 and Chapter 5) were extensively played by 162 children. This section explains in detail the participants, the measurements and the procedure carried out during the evaluations.

6.2.1 Participants

A total of 162 children participated in this study. There were 84 boys (51.85%) and 78 girls (48.15%). They were between seven and eleven years old, and they had already finished their academic course between the second and fifth grades of primary school. The mean age was 8.81 ± 1.03 years old. The children were students from three different summer schools in Valencia.

6.2.2 Measurements

To retrieve data for the analysis, three different questionnaires were used. There was a pre-test questionnaire with only thirteen questions of plain text. The pre-test questionnaire is shown in Table A.1. This test was used to evaluate the children's knowledge before they started playing the games. There was a second post-test questionnaire. This questionnaire had the same thirteen questions from the pre-test, and thirteen additional questions related to different aspects including usability (Table A.2). By comparing the pre-test and this post-test, it was possible to determine if there had been an increase in knowledge. There was a last questionnaire that the children filled out (Table A.2) once they had played with the two system configurations (frontal projection and autostereoscopy). This questionnaire was used to determine which of the two configurations they preferred. This questionnaire had nineteen questions; ten questions obtained information about the last configuration played and the last nine questions compared the two configurations.

6.2.3 Procedure

The participants were assigned to one of the following two groups:

- Group A: Participants that played with the autostereoscopic configuration first and afterwards played with the frontal projection configuration.
- Group B: Participants that played the frontal projection configuration first and afterwards played with the autostereoscopic configuration.

The A and B groups were balanced by grouping the children into pairs (1 boy + 1 girl, 2 boys, 2 girls), with the same number of pairs for each combination. Figure 6.1 shows graphically the procedure for both groups. The following protocol was used:

1. A pair of children filled out the pre-test questionnaire in a web-based form (PreAuto for Group A and PreFrontal for Group B).
2. These children played one configuration (frontal projection or autostereoscopy).
3. Then, they filled out the post-test questionnaire on-line (Pos1Auto for Group A, and Pos1Frontal for Group B).
4. Then, they played a short version of the other configuration. This short version corresponded to the preferred period.
5. Finally, they filled out the final questionnaire (Pos2Auto for Group A, and Pos2Frontal for Group B).

Figure 6.2 shows a pair of children playing the two different systems. Since the questionnaires were filled out on-line, the answers were automatically stored in a remote database.

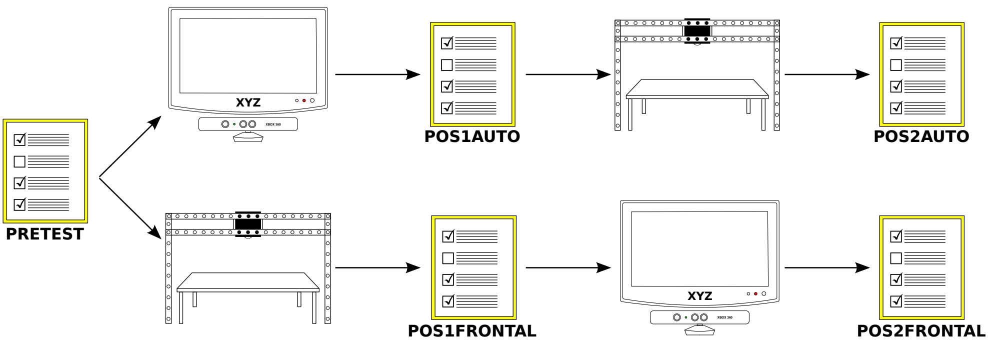


Figure 6.1: Study procedure

(a) Autostereoscopic system (b) Frontal projection system

Figure 6.2: Example of children playing the two different systems

For the creation of the questionnaires, we used the Gandia Qüest tool, which was developed by the Tesigandia company[1]. This tool allows the data of the results to be stored in several formats and also allows data processing. This tool presents a user-friendly interface for creating the forms, and it also makes it easier to add multimedia content to the questionnaires such as images, music, video, and flash applications. The tool can also manage different languages to facilitate the creation of the same form in different languages. In our case, the data was stored in XLSX format (Excel 2010) for the data processing. For two of the questions that required drag and drop interaction, an embedded flash program was also used. In order to maintain data integrity, the data retrieved from each user was only stored when the whole questionnaire had been completed, otherwise no data was stored. The questionnaires created using Gandia Qüest tool and filled out via web were used to carry out all the studies of this thesis.

We used the Gandia Qüest tool, but many other tools can also be used for the same purpose; for example, Website Analysis and MeasureMent Inventory[2], Survey Monkey[3], Formstack[4]. Even Google Drive[5] can be used to create a form survey.

[1] http://www.tesigandia.com/
[2] http://www.wammi.com
[3] http://www.surveymonkey.com
[4] http://www.formstack.com
[5] http://www.drive.google.com

6.3 Results

The data from the study were analyzed using the statistical open source toolkit *R*[6] with the *RStudio*[7] IDE. The *R* toolkit has been used for the analysis of all the studies of this thesis. The following facts should be kept in mind regarding the number of participants and the number of different questions in the questionnaires.

- When completing the on-line questionnaires, the children moved from one question to the next, and the system did not monitor if they had answered all the questions. Therefore, there were children who had not answered all the questions. This means that the number of children who answered each question was not the same.

- We considered the number of participants that used the two systems (one after the other) for the comparison questions. However, there were also participants that only used one of the systems for questions related to usability and knowledge. Therefore, the number of participants was different for several questions.

6.3.1 Learning outcomes

Several t-tests were performed to determine if there were statistically significant differences in the knowledge acquired. In these tests, the knowledge variable was analyzed, which took into account all the knowledge questions (Q1 to Q13 of Pre and Pos1 in Table A.1) and summed up the correct answers. They were multiple-choice questions with 4 to 6 possible answers that were scored as success or fail. As an example, one of these questions was: *Ancient Times started with the: a) The invention of the wheel; b) The invention of writing; c) The discovery of America; d) The fall of the Roman Empire; e) The invention of the compass.* The knowledge variable was compared in the Pre and Pos1 questionnaires. Figure 6.3 shows the box plot for the scores before and after playing with the two systems. A high dominance of correct questions after playing the first game (Pos1Auto and Pos1Frontal) over the pre-test (PreAuto and PreFrontal) can be observed. All t-tests are shown in the format: *(statistic [degrees of freedom], p-value, Cohen's d)*; and ** indicates the statistical significance at level $\alpha = 0.05$. From a paired t-test, the ratings of the knowledge variable between PreAuto (mean 3.41 ± 1.88) and Pos1Auto (mean 7.81 ± 2.71) showed that there was a statistically significant difference ($t[80] = -17.62$, $p < 0.001$**, Cohen's

[6] Available at http://www.r-project.org

[7] Available at http://www.rstudio.com

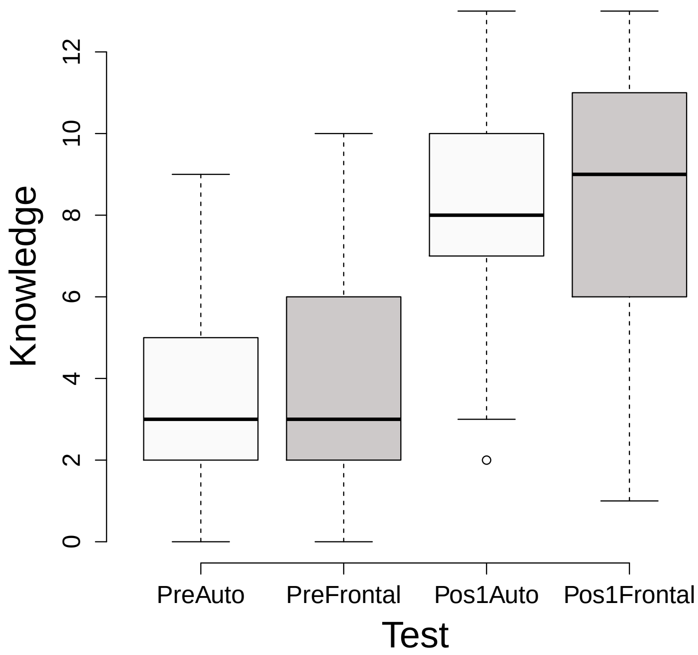


Figure 6.3: Scores of the knowledge variable in the Pre and Pos1 questionnaires for the autostereoscopic system and for frontal projection

$d = -1.96$). Another paired t-test revealed that there was a statistically significant difference between the ratings of the knowledge variable in PreFrontal (mean 3.54 $\pm$ 2.48) and Pos1Frontal (mean 7.91 $\pm$ 3.52) ($t[80] = -14.96$, $p < 0.001$**, Cohen's $d = -1.66$). To determine whether or not there was difference between the initial knowledge of the two groups, two t-test were performed, between the knowledge variable in PreAuto (mean 3.41 $\pm$ 1.88), and the knowledge variable in PreFrontal (mean 3.54 $\pm$ 2.48) ($t[160] = -0.39$, $p = 0.697$, Cohen's $d = -0.06$). These results revealed that there was no statistically significant difference between the knowledge in the two pre-tests. To determine whether or not there was difference between the acquired knowledge in the two groups, a t-test was performed between the knowledge variable in Pos1Auto (mean 7.81 $\pm$ 2.71) and the knowledge variable in Pos1Frontal (7.91 $\pm$ 3.52) ($t[160] = -0.20$, $p = 0.843$, Cohen's $d = -0.03$), which also revealed that there was no statistically significant difference between the acquired knowledge using the two systems. To complete the analysis and to determine in which questions there were statistically significant differences, we performed a paired t-test for each question between Pre and Pos1 for the autostereoscopic system (shown in Table 6.2). A similar analysis was performed for the frontal projection system, in which statistically significant differences were obtained for each question, showed at Table 6.3. A t-test for each knowledge question comparing the acquired knowledge using the two systems revealed that there was only statistically significant difference for Q1 ($t[160] = -2.69$, $p = 0.008^{**}$, Cohen's $d = -0.42$) where the autostereoscopic system got a better score, and for Q13 ($t[160] = 3.24$, $p = 0.001^{**}$, Cohen's $d = 0.51$) where the frontal projection system got a better score.

#	PreAuto	PreFrontal	t	p	Cohen's d
Q1	0.12 ± 0.33	0.21 ± 0.41	-1.48	0.142	-0.23
Q2	0.19 ± 0.39	0.16 ± 0.37	0.41	0.680	0.06
Q3	0.10 ± 0.30	0.15 ± 0.36	-0.95	0.342	-0.15
Q4	0.33 ± 0.47	0.27 ± 0.44	0.85	0.396	0.13
Q5	0.32 ± 0.47	0.26 ± 0.44	0.86	0.390	0.14
Q6	0.26 ± 0.44	0.30 ± 0.46	-0.52	0.601	-0.08
Q7	0.01 ± 0.11	0.07 ± 0.26	-1.94	0.054	-0.31
Q8	0.73 ± 0.44	0.67 ± 0.47	0.85	0.396	0.13
Q9	0.67 ± 0.47	0.56 ± 0.50	1.45	0.149	0.23
Q10	0.36 ± 0.48	0.44 ± 0.50	-1.12	0.265	-0.18
Q11	0.07 ± 0.26	0.09 ± 0.28	-0.29	0.774	-0.05
Q12	0.21 ± 0.41	0.22 ± 0.42	-0.19	0.850	-0.03
Q13	0.04 ± 0.19	0.15 ± 0.36	-2.47	0.015**	-0.39

Table 6.1: Means and standard deviations of questions for the PreAuto and PreFrontal, t-test analysis, and Cohen's d, d.f. = 160

A multifactorial ANOVA test was also performed to take into consideration several factors simultaneously (age, game and gender). The effect size used was the eta-squared (η^2). The results of the analysis in Table 6.5 show that there were statistically significant differences in the gender and the age factors. The p-values in these cases showed statistically significant differences, and the effect sizes revealed that the most influential factor was age, followed by gender. There was also some interaction between the gender and age factors, but there were no statistically significant differences. A Tukey post-hoc test showed that the acquired knowledge was significantly

#	PreAuto	Pos1Auto	t	p	Cohen's d
Q1	0.12 ± 0.33	0.75 ± 0.43	-11.09	<0.001**	-1.23
Q2	0.19 ± 0.39	0.35 ± 0.48	-2.97	0.004**	-0.33
Q3	0.10 ± 0.30	0.73 ± 0.44	-10.59	<0.001**	-1.18
Q4	0.33 ± 0.47	0.40 ± 0.49	-0.96	0.339	-0.11
Q5	0.32 ± 0.47	0.56 ± 0.50	-3.98	<0.001**	-0.44
Q6	0.26 ± 0.44	0.52 ± 0.50	-4.50	<0.001**	-0.50
Q7	0.01 ± 0.11	0.47 ± 0.50	-8.20	<0.001**	-0.91
Q8	0.73 ± 0.44	0.79 ± 0.41	-1.09	0.278	-0.12
Q9	0.67 ± 0.47	0.81 ± 0.39	-2.53	0.013**	-0.28
Q10	0.36 ± 0.48	0.67 ± 0.47	-4.44	<0.001**	-0.49
Q11	0.07 ± 0.26	0.65 ± 0.48	-10.02	<0.001**	-1.11
Q12	0.21 ± 0.41	0.80 ± 0.40	-10.27	<0.001**	-1.14
Q13	0.04 ± 0.19	0.32 ± 0.47	-5.63	<0.001**	-0.63

Table 6.2: Means and standard deviations of questions for the PreAuto and Pos1Auto, t-test analysis, and Cohen's d, d.f. $= 80$

#	PreFrontal	Pos1Frontal	t	p	Cohen's d
Q1	0.21 ± 0.41	0.56 ± 0.50	-6.50	<0.001**	-0.72
Q2	0.16 ± 0.37	0.26 ± 0.44	-1.92	0.059	-0.21
Q3	0.15 ± 0.36	0.77 ± 0.42	-11.36	<0.001**	-1.26
Q4	0.27 ± 0.44	0.47 ± 0.50	-2.96	0.004**	-0.33
Q5	0.26 ± 0.44	0.67 ± 0.47	-6.76	<0.001**	-0.75
Q6	0.30 ± 0.46	0.42 ± 0.49	-2.59	0.012**	-0.29
Q7	0.07 ± 0.26	0.53 ± 0.50	-6.70	<0.001**	-0.74
Q8	0.67 ± 0.47	0.74 ± 0.44	-1.18	0.242	-0.13
Q9	0.56 ± 0.50	0.73 ± 0.44	-3.32	0.001**	-0.37
Q10	0.44 ± 0.50	0.77 ± 0.42	-4.91	<0.001**	-0.55
Q11	0.09 ± 0.28	0.74 ± 0.44	-12.31	<0.001**	-1.37
Q12	0.22 ± 0.42	0.70 ± 0.46	-8.22	<0.001**	-0.91
Q13	0.15 ± 0.36	0.57 ± 0.50	-7.61	<0.001**	-0.85

Table 6.3: Means and standard deviations of questions for the PreFrontal and Pos1Frontal, t-test analysis, and Cohen's d, d.f. $= 80$

#	Pos1Auto	Pos1Frontal	t	p	Cohen's d
Q1	0.75 ± 0.43	0.56 ± 0.50	2.69	0.008**	0.42
Q2	0.35 ± 0.48	0.26 ± 0.44	1.20	0.234	0.19
Q3	0.73 ± 0.44	0.77 ± 0.42	−0.54	0.590	−0.08
Q4	0.40 ± 0.49	0.47 ± 0.50	−0.95	0.344	−0.15
Q5	0.56 ± 0.50	0.67 ± 0.47	−1.45	0.149	−0.23
Q6	0.52 ± 0.50	0.42 ± 0.49	1.26	0.210	0.20
Q7	0.47 ± 0.50	0.53 ± 0.50	−0.78	0.435	−0.12
Q8	0.79 ± 0.41	0.74 ± 0.44	0.74	0.461	0.12
Q9	0.81 ± 0.39	0.73 ± 0.44	1.31	0.192	0.21
Q10	0.67 ± 0.47	0.77 ± 0.42	−1.39	0.165	−0.22
Q11	0.65 ± 0.48	0.74 ± 0.44	−1.20	0.234	−0.19
Q12	0.80 ± 0.40	0.70 ± 0.46	1.46	0.147	0.23
Q13	0.32 ± 0.47	0.57 ± 0.50	−3.24	0.001**	−0.51

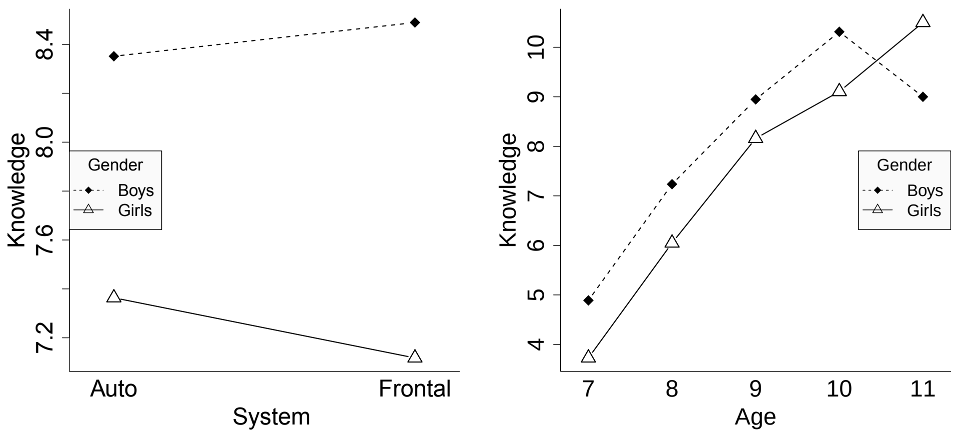
Table 6.4: Means and standard deviations of questions for the Pos1Auto and Pos1Frontal, t-test analysis, and Cohen's d, d.f.=160

different between children of ages 7 and 9, 7 and 10, 7 and 11, 8 and 9, and 8 and 10.

Factor	d.f.	F	p	η^2
Gender	1	4.35	0.039**	0.029
Age	4	16.08	<0.001**	0.309
Game	1	0.0005	0.982	<0.001
Gender:Age	4	0.23	0.922	0.006
Other interactions	3	< 0.85	>0.464	<0.017

Table 6.5: Multifactorial ANOVA for the knowledge variable, $N = 162$

For the acquired knowledge variable, Figure 6.4a shows the interaction plot between gender and the two systems. Boys acquired more knowledge than girls using the two systems. Figure 6.4b shows the interaction plot between gender and age. From these figures, it can be observed that the score means at older ages were higher than at younger ages with significant differences among the age groups. Also, the knowledge score with the frontal projection system was a bit higher than with the autostereoscopic configuration in the case of boys; in the case of girls the knowledge score was a bit higher with the autostereoscopic configuration, having statistically significant differences between genders as indicated by the above ANOVA analysis.



(a) By system (b) By age

Figure 6.4: Interaction plots for the acquired knowledge

6.3.2 System comparison outcomes

Several *t*-tests were performed to determine if there were statistically significant differences in the opinions of the children depending on which game configuration was played first. Table A.2 shows the questions that were used to perform the test (Q14-Q25). The results of these tests are shown in Table 6.6, Table 6.7 and Table 6.8. First, the data of the children that played the autostereoscopic system first versus the children that played the frontal projection system first were analyzed (Table 6.6). The scores of each child playing with one system first and later with the other (Pos1Auto versus Pos2Frontal (Table 6.7), and Pos1Frontal versus Pos2Auto (Table 6.8)) there were also compared. From the analysis of Q14 (how much fun the children had playing the game for the first time), no statistically significant differences were found. Nevertheless, when the same child played first with one of the two games, he/she scored the first time statistically significantly higher than the second time. This result was the same independently of whether the child played with the autostereoscopic configuration first or with frontal projection configuration first. We would like to highlight that the second time they played the game was shorter. The analysis of Q16 (ease of use) for between subjects showed that the children that played with the autostereoscopic system (3.94 ± 0.80) gave a statistically significant higher score to the ease of use than the ones who played with frontal projection (3.74 ± 0.84). However, when a within subjects analysis was performed, the results showed that there was a statistically significant difference in favour of the system played the second time for both the autostereoscopic system and the frontal projection system. Our explanation for these results is that the second time they played, they found the game easier because they had already played before and they already knew what they had to do in the game even though the interaction was not exactly the same. Something similar happened with Q18 (ease of selecting the answers). The first time they played, no statistically significant differences were found; however, the second time they played they gave a statistically significant higher score to the second system used. Q19 asked children if they liked the images shown in the game. The analysis showed that no statistically significant differences were found either between subjects or within subjects. Statistically significant differences were found in the autostereoscopic vision-oriented questions (Q23 and Q24). The analysis of Q23 showed that statistically significant differences were found between subjects for the first time they played in favour of the autostereoscopic system. If the children played the frontal projection first and the autostereoscopic system later, statistically significant differences were also found in favour of the autostereoscopic system. In contrast, playing with the autostereoscopic system first and with frontal projection second, the children scored the autostereoscopic system higher, but this difference was not statistically significant. The analysis of Q24 showed that there were statistically significant differences between and within subjects in favour of the autostereoscopic system. These results reveal that with autostereoscopy the children had the feeling of being able to touch the 3D elements like the medieval castle or the Roman road. Finally, another t-test was performed for Q25 which asked the children for a global score for the game from 1 to 10. The results of this question showed that there were no statistically significant differences between subjects. However, when the analysis was within subjects, there was a statistically significant difference in favour of the autostereoscopic system when it was played first. No statistically significant difference was found when the children played with the frontal projection system first.

#	Pos1Auto	Pos1Frontal	t	d.f.	p	Cohen's d
Q14	4.85 ± 0.35	4.79 ± 0.45	1.26	241	0.211	0.16
Q15	4.45 ± 0.68	4.45 ± 0.83	0.01	234	0.994	0.00
Q16	3.94 ± 0.79	3.74 ± 0.84	1.98	240	0.049**	0.25
Q17	4.61 ± 0.56	4.48 ± 0.65	1.66	240	0.098	0.21
Q18	3.98 ± 0.89	4.07 ± 0.91	-0.72	242	0.473	-0.09
Q19	4.63 ± 0.51	4.58 ± 0.72	0.65	242	0.514	0.08
Q22	4.38 ± 0.80	4.42 ± 0.80	-0.30	241	0.761	-0.04
Q23	5.45 ± 1.56	4.38 ± 2.22	4.27	233	<0.001**	0.56
Q24	4.95 ± 1.96	4.27 ± 2.18	2.50	229	0.013**	0.33
Q25	9.62 ± 0.77	9.45 ± 1.26	1.28	242	0.202	0.16

Table 6.6: Means and standard deviations of questions for the Pos1Auto and Pos1Frontal tests, t-test analysis, and Cohen's d

#	Pos1Auto	Pos2Frontal	t	d.f.	p	Cohen's d
Q14	4.84 ± 0.37	4.67 ± 0.56	2.38	62	0.021**	0.30
Q15	4.51 ± 0.66	4.42 ± 0.78	1.35	64	0.182	0.17
Q16	4.12 ± 0.62	4.45 ± 0.86	-2.76	65	0.008**	-0.34
Q17	4.69 ± 0.53	4.80 ± 0.44	-1.47	63	0.146	-0.18
Q18	4.15 ± 0.76	4.48 ± 0.66	-3.07	65	0.003**	-0.38
Q19	4.62 ± 0.55	4.62 ± 0.60	0.00	64	1.000	0.00
Q22	4.29 ± 0.81	4.15 ± 1.00	1.22	65	0.228	0.15
Q23	5.58 ± 1.52	5.14 ± 2.02	1.83	65	0.071	0.23
Q24	4.90 ± 1.96	4.35 ± 2.24	2.16	61	0.035**	0.27
Q25	9.59 ± 0.76	9.20 ± 1.54	2.69	65	0.009**	0.33

Table 6.7: Means and standard deviations of questions for the Pos1Auto and Pos2Frontal tests, t-test analysis, and Cohen's d

6.3.3 Satisfaction outcomes

In order to measure the satisfaction that the children had while playing the game, several Chi-squared tests were performed for the satisfaction questions. Table A.2 shows the questions that were used to perform the test (Q27-Q34). The results of these tests are shown in Table 6.9. The Chi-squared tests performed revealed

#	Pos1Frontal	Pos2Auto	t	d.f.	p	Cohen's d
Q14	4.88±0.32	4.75±0.51	2.05	58	0.045**	0.27
Q15	4.48±0.80	4.45±0.79	0.36	57	0.718	0.05
Q16	3.84±0.65	4.41±0.80	−4.80	55	<0.001**	−0.64
Q17	4.54±0.65	4.78±0.41	−2.69	58	0.009**	−0.35
Q18	4.07±0.84	4.38±0.70	−2.99	55	0.004**	−0.40
Q19	4.64±0.60	4.69±0.56	−0.72	58	0.472	−0.09
Q22	4.44±0.79	4.36±0.82	0.82	58	0.416	0.11
Q23	4.68±1.97	5.74±1.63	−3.56	56	<0.001**	−0.47
Q24	4.41±1.89	5.52±1.85	−4.21	55	<0.001**	−0.56
Q25	9.56±0.85	9.51±1.14	0.33	58	0.745	0.04

Table 6.8: Means and standard deviations of questions for the Pos1Frontal and Pos2Auto tests, t-test analysis, and Cohen's d

that there was a statistically significant difference for Q27 only, which refers to the preference for the system they liked the most. After analyzing the results, we could see that the children tended to choose the system they had used the first time. In this question, 45% of the children preferred both systems, 40% of them preferred the autostereoscopic configuration, 14% chose frontal projection, and 1% none of them. Several other Chi-squared tests were performed to analyze the interaction with the systems. For ease of use (Q28), the percentages were 41% for frontal projection, 31% for both systems, 27% for autostereoscopy, and 1% for none of them. For comfort (Q29), the results revealed that 36% of the children chose both systems, 35% frontal projection, 28% autostereoscopy, and 1% none of them. For the controlling variable (Q30), the frontal projection offered the best result with 44% of the scores, followed by autostereoscopy with 28%, and both of them with 26% of the scores. Only 2% of them chose none of systems. For Q31 which asked which system had the best-looking images, the results obtained in this question were 42% for autostereoscopy, 31% for both systems, 26% for frontal projection, and 1% for none of them. The Chi-squared test for Q32 (if children would recommend some of the systems to their friends), obtained the results of 67% for both systems, 21% for autostereoscopy, 11% for frontal projection, and 1% for none of them. The question about satisfaction (Q33), which asked about which of the two systems was the most fun, showed 54% for both systems, 31% for autostereoscopy, and 15% for frontal projection. Note that, in this question, none of the children chose the "None of them" choice. Q34 asked the children if they would like to use any of these systems in their classrooms. The results obtained by this question were 45% for both systems, 41% for autostereoscopy, 13% for frontal projection, and 1% for none of them. In order to determine which of all the mini-games was liked the most, Q26 was asked after playing with the game for the first time. In that question, the children could select the mini-games they preferred, and they could select more than one option. The mini-game with the highest score was Prehistory (find cave painting and place an imprint of the shape of your hand in the cave) with 66.39% of votes. The second highest was Ancient Times (construct a Roman city and answer questions about Roman construction) with 55.73%. The next preferred mini-game was the Middle Ages (build a medieval castle) with 53.27%. Following, 47.13% of children selected the Contemporary Period (solve the timeline puzzle) as one of their favourites. Finally, the mini-game with the least votes was the Early Modern Period (find objects used by Christopher Columbus) with 39.75% of votes.

#	Auto	Frontal	χ^2	d.f.	N	p	Cramer's V
Q27	1 Auto	3 Both	8.79	3	128	0.03**	0.26
Q28	2 Frontal	3 Both	2.47	3	127	0.48	0.14
Q29	2 Frontal	3 Both	1.39	3	126	0.71	0.11
Q30	2 Frontal	2 Frontal	4.40	3	126	0.22	0.19
Q31	1 Auto	1 Auto	2.69	3	127	0.44	0.15
Q32	3 Both	3 Both	5.52	3	126	0.14	0.21
Q33	3 Both	3 Both	3.98	2	125	0.14	0.18
Q34	1 Auto	3 Both	2.58	3	127	0.46	0.14

Table 6.9: Modes of questions for the Pos1Auto and Pos1Frontal tests, Chi-squared analysis, and Cramer's V

With regard to Q35 (changes in the game), some comments included the following: "I wouldn't change anything in the game", "It would be better to have more difficult activities", "The game could be last longer", "To have more topics to learn".

6.3.4 Avatar outcomes

As the guide/avatar is a principal part of the game, two questions were asked to determine the children's opinion about him. These questions followed a Likert scale ranging from 1 to 5 and Chi-squared tests were performed to analyze them. Table 6.10 shows the results of these tests. The first of these two questions was Q20, which asked how much they liked the avatar. The results of the test showed that there were no statistically significant differences for the avatar between subjects. The mean answer of the children for the autostereoscopic system was a little bit higher than the mean for the frontal projection, but there were no statistically significant differences. From the scores, it can be deduced that most of the children liked the avatar character. The second question about the avatar was Q21. It asked how much the avatar helped the children during the game. This test obtained results that showed there were no statistically significant differences between the two systems. This reveals that the avatar character was helpful for the children while playing the game.

#	Auto	Frontal	χ^2	d.f.	N	p	Cramer's V
Q20	4.46 ± 0.76	4.19 ± 1.18	7.88	4	244	0.096	0.18
Q21	4.10 ± 1.03	4.22 ± 0.97	1.74	4	242	0.783	0.08

Table 6.10: Chi-squared analysis for avatar questions

6.3.5 Correlation analysis

The correlation analysis for the Pos1 questionnaire is shown in Figure 6.5. A correlation between Q14 and Q25 was found. This correlation means that the more fun the children had, the higher the score they gave to the game. Another correlation was found between Q23 and Q24. Q23 was related to depth perception and Q24 was related to the sense of presence. This result indicates that viewing the 3D elements as coming out of the screen is closely related to the feeling of being able to touch these elements.

6.4 Conclusions

In this study, Natural User Interfaces were used as a complement for interaction with children in learning environments. Two different systems were compared. The autostereoscopic system allows the users to have a complete experience without having to carry devices or wires on their bodies. To our knowledge, this is the first time this combination has been presented, especially for education. In the autostereoscopic system, the children could see themselves in the autostereoscopic 3D display, and the game was controlled by gestures. In contrast, the frontal projection system simulated a large-size tabletop display.

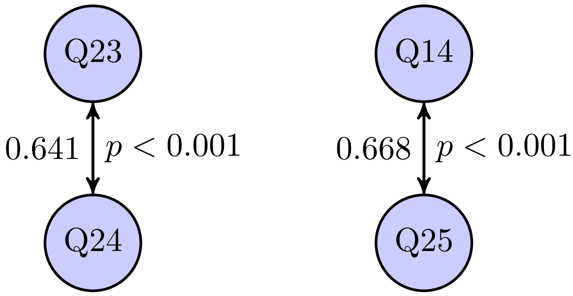


Figure 6.5: Significant correlations for questions

The second of our hypotheses was that children would increase their knowledge about the subject of the game by using the two systems, and that the autostereoscopic system would lead to greater learning results. Comparing their initial knowledge and their knowledge after playing, statistically significant differences were obtained, which corroborates the first part of the second hypothesis. Different arguments have been offered to justify this trend. One of them is that video games promote a positive attitude towards learning (Durkin and Barber, 2002), especially due to their motivational nature (Papastergiou, 2009; Anneta et al., 2009; Rosas et al., 2003). Other authors have also attributed positive learning effects to video games (Egenfeldt-Nielsen, 2007).

Differences in age, gender, and which system was played first were also considered. These analyses revealed that boys improved their knowledge a small amount more than girls, and indicate that older children improved their knowledge quite a bit more than younger children. These results indicate that systems of this type are more suitable for older children (9-11 years old) than for younger children 7-8 years old. However, there was no statistically significant difference between the acquired knowledge using the two systems. Therefore, the second part of our second hypothesis (the autostereoscopic system will obtain greater learning results) was not corroborated. Although unexpected, it is an excellent result because it means that the game is well suited for learning outcomes and that the two systems can be used for this purpose. For the system the children liked the most, the children preferred both configurations (45%), followed by the autostereoscopic system (40%), and then the frontal projection (14%). In our first hypothesis, we predicted that children would prefer the autostereoscopic system. From the percentages, we can affirm that this hypothesis has been corroborated (both + autostereoscopy > both + frontal projection). For depth perception, the results showed that the 3D sensation (Q23) was mainly perceived and appreciated, being more evident when the children played with the autostereoscopic system after playing with the frontal projection system. The results revealed that autostereoscopy gave the children the feeling of being able to touch the 3D elements (Q24). Q24 was related to presence. Q23 and Q24 were correlated. From our point of view, these results are important and can be exploited for the development of educational games. For ease of use, when Q16 was analyzed, a statistically significant difference was found in which the autostereoscopic system was scored higher. However, when the children were asked explicitly about the easiest system to use, they preferred the frontal projection system (41%), followed by both systems (31%), and the autostereoscopic system (27%). From these results, we consider that the frontal projection system is easier to use which corroborates our third hypothesis (the frontal projection system will be easier to use). Our opinion is that, in the frontal projection system, the children interacted easily and fast simply by placing their hands over the buttons. However, they had some problems using the autostereoscopic configuration when trying to get the correct position of the hand on the 3D world by moving their arms in the air, forward and backwards. More studies should be carried out to assure that the frontal projection system is the easiest to use. For the system they control better, again, the children chose the frontal projection system (44%), followed by the autostereoscopic system (28%), and both systems (26%). In this case, the children maintained their preference for the frontal projection system. Our opinion is that the easier the game is to use, the better they control the game. For the most comfortable system, the children chose the frontal projection system (44%), followed by the autostereoscopic system (28%), and both systems (26%). For the most fun experienced (Q14), the results showed that the children gave high scores when asked about fun. For an analysis between subjects, no statistically significant differences were found, but there were statistically significant differences for analysis within subjects in favour of the system played the second time (either autostereoscopy or frontal projection). We would like to highlight that the second time the children played, they chose the time period of the history and they did not play long enough to the complete game. This fact could have influenced this result. A correlation between Q14 and Q25 (global score) was found. This correlation means that the more fun the children had, the higher the score they gave to the game. For the role of the avatar, the children liked the avatar and they also thought he was helpful. These results reveal the importance of having some kind of character guiding the children throughout the game (someone to tell the children what they must do in each part of the game / lesson). With regard to the topic of the game, some knowledge questions revealed that data like dates or the names of historic events are the most difficult for children to remember.

Based on our study, we considered that, playing games using the entire body as controller and an autostereoscopic vision is metaphorically similar to the real-world experience. In this situation, the selection of elements is done by using your hands and interacting by yourself with no external devices such as glasses or controllers. In addition, the user has depth perception. With regard to the interaction with the Kinect in the autostereoscopic system, some 7-year-old and 8-year-old children had trouble with being calibrated by the Kinect sensor due to their short height (Kinect needs a minimum user height in order to work properly). This problem did not appear in frontal projection because the skeleton of the user was not used for hand detection. The user calibration in this case did not depend on the height of the user because the only element the Kinect camera could detect were the arms of the children and their distance from the camera.

Study 2: Individual vs. Collaborative learning

7

"Be yourself; everyone else is already taken."

— Oscar Wilde

7.1 Introduction

In this chapter a study that emphasizes the use of collaborative interactions (in pairs) between the children as a complement for learning environments is presented. The collaborative method, where the children have to play with the game in pairs, is compared with an individual method, where the children have to play with the game alone. These two methods are played with the autostereoscopic system of the game detailed in Chapter 4.

7.2 Design of the evaluations

The developed game described in the previous chapter (see section 3.2) was played by 46 children. This section explains in detail the participants, the measurements and the procedure carried out during the evaluations.

7.2.1 Participants

A total of 46 children participated in this study. There were 22 boys (47.83%) and 24 girls (52.17%). They were between seven and ten years old, and they had already finished their third grade academic course of primary education. The mean age was 8.52 ± 0.58 years old. The children were attending three different summer schools in Valencia. Since the children attended summer school, they knew each other before the study; however, we did not take into account whether or not they were friends.

7.2.2 Measurements

To retrieve data for the analysis, two questionnaires in a web-based form were designed to obtain the knowledge that the children acquired while playing with the game. These questionnaires consisted of thirteen questions, showed in Table A.1, about the contents that the children learned while playing with the game developed in this thesis. Comparing the answers given by the children in the pretest with the answers given in the post-test (which had the same thirteen questions), it was possible to find out whether had been an increase of knowledge.

7.2.3 Procedure

The study lasted over one month. Each child played only once and only in one day. The entire activity (pre and post questionnaires and game play) lasted 30 minutes. The participants were assigned to one of the following two groups:

- Group A: Participants that played in pairs (collaborative mode). These pairs were made by one boy and one girl, two boys, or two girls.
- Group B: Participants that played solo (individual mode) with no more company than the monitor. The children of this group had to select the options that appeared in the two sides of the screen by using their two hands.

The person that monitors the activity only guides the children in the steps to follow. In other words, the person tells the children to sit down and to fill out the questionnaires and accompanies the children to the playing area. This person does not interfere during the activity unless the children have interaction problems (do not know how to select the buttons) or technical problems (game failure). Figure 7.1 shows graphically the procedure of both groups. The following protocol was used:

1. A pair of children from Group A, or a child from Group B, filled out the pre-test questionnaire (PrePair for Group A and PreIndiv for Group B).

2. These children, or this child, played with the developed autostereoscopic game.

3. Then, they / he / she filled out the post-test questionnaire (PosPair for Group A and PosIndiv for Group B).

The children were not helped with the answers. The children could not talk with each other while they were filling out the questionnaires. Since the questionnaires were filled out on-line, the answers were automatically stored in a remote database. Figure 7.2 shows a child playing our system individually.

7.3 Results

As in the previous study, the data from this study were analyzed using the statistical open source toolkit *R* with the *RStudio* IDE.

7.3.1 Learning outcomes

In order to measure how much the children learned, the *knowledge* variable was analyzed before playing (Pre-test) and after playing (Post-test). The *knowledge* variable was created to condense the thirteen knowledge questions (Table A.1) by counting the number of correct answers. Several t-tests were performed to determine if there were statistical differences in the knowledge acquired. When the answers given by the children before playing the game in the two modes were analyzed, no statistically significant differences were found. This means that the two groups of children had similar knowledge about the topic of the game before playing, which therefore assures that the two groups come from a normal distribution. Figure 7.3 shows the box plot for the scores before and after playing. A high dominance of correct answers after playing the game (PosPair and PosIndiv) over the pre-test (PrePair and PreIndiv) can be observed.

All t-tests are shown in the format: *(statistic [degrees of freedom], p-value, Cohen' s d)*; and ** indicates the statistical significance at level $\alpha = 0.05$. Firstly, to determine

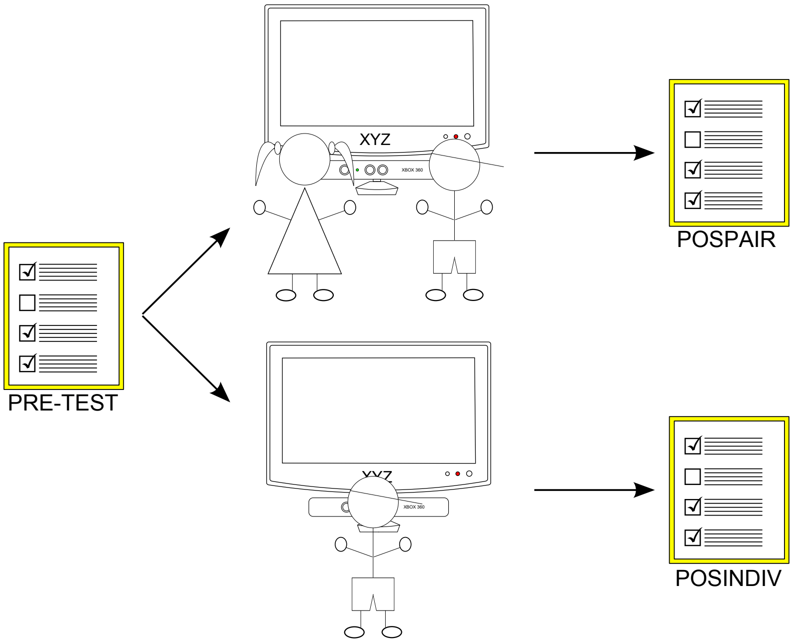


Figure 7.1: Study procedure

whether or not there was difference between the initial knowledge in both pre-tests, an unpaired t-test was performed between PrePair (2.97 $\pm$ 1.56) and PreIndiv (2.81 $\pm$ 1.47) ($t[44] = 0.32$, $p = 0.751$, Cohen's $d = 0.10$) where no statistically significant differences were found. From a paired t-test, the scores of the knowledge variable between PrePair (2.97 $\pm$ 1.56) and PosPair (7.70 $\pm$ 2.44) showed that there were statistically significant differences ($t[29] = -10.65$, $p < 0.001^{**}$, Cohen's $d = -1.94$). Another paired t-test between the PreIndiv (2.81 $\pm$ 1.47) and the PosIndiv (6.00 $\pm$ 2.62) questionnaires revealed statistically significant differences ($t[15] = -4.58$, $p < 0.001^{**}$, Cohen's $d = -1.14$). Finally, to determine whether or not there were statistically significant differences between the acquired knowledge in the two groups, another unpaired t-test was performed between the knowledge in PosPair (7.70 $\pm$ 2.44) and the knowledge in PosIndiv (6.00 $\pm$ 2.62) ($t[44] = 2.15$, $p = 0.037^{**}$, Cohen's $d = 0.66$) showing that the knowledge gained while playing in pairs was significantly higher than the knowledge acquired while playing alone. To complete the analysis and

Figure 7.2: Child playing our system individually

determine which questions had statistically significant differences, several t-tests were performed for each question between PrePair - PreIndiv, PrePair – PosPair, PreIndiv – PosIndiv, and PosPair – PosIndiv questionnaires.

Table 7.2 shows that children who played with the game in pairs acquired more knowledge in questions Q1, Q2, Q3, Q5, Q7, Q10, Q11, Q12 and Q13. This can be compared with the results in Table 7.3 provided by children who played the game individually. In this case, questions Q1, Q3, Q6, Q7, Q11 and Q12 were the questions with statistically significant differences.

Finally, to compare the knowledge acquired in each question after playing with the game, the results between the two post-tests were compared. Results are shown in Table 7.4, where Q8 stands out, being the only question with statistically significant differences.

A multifactorial ANOVA test was also performed to take into consideration several factors simultaneously. The factors of gender, age, and game mode were between subjects. The effect size used was the eta-squared (η^2). The results of the analysis in Table 7.5 show that there were statistically significant differences in the gender factor. The p-values in these cases showed statistically significant differences, and the effect sizes revealed that the most influential factor was age, followed by gender. There was also some interaction between the gender and age factors, but there were no statistically significant differences. A Tukey post-hoc test showed that the acquired

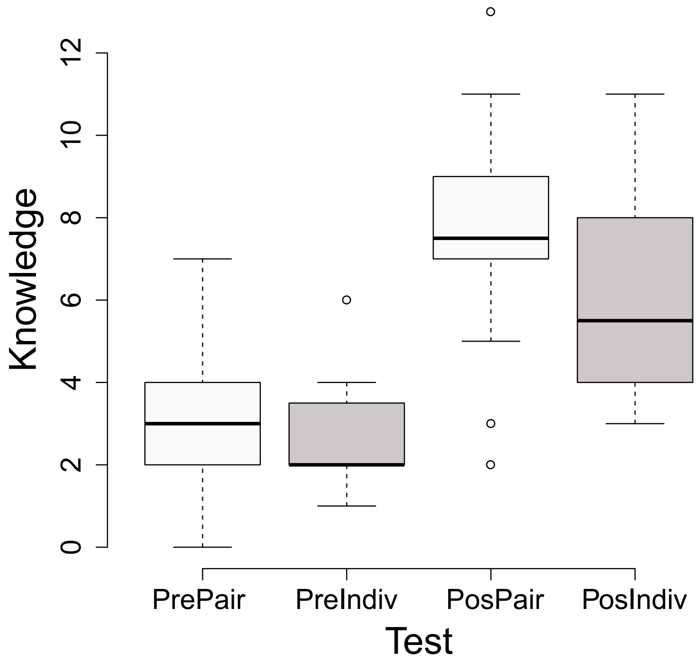


Figure 7.3: Scores of the knowledge variable in questionnaires, before and after play with the game, by pairs and individually

knowledge was significantly different between children of ages 8 and 9, and between the interactions Boy:9-years-old - Girl:8-years-old, Pair:9-years-old - Individual:8-years-old, and Boy:9-years-old:Pair - Girl:8-years-old:Individual.

For the acquired knowledge variable, Figure 7.4a shows the interaction plot between gender and the two game modes. Boys acquired more knowledge than girls playing the two modes. Figure 7.4b shows the interaction plot between gender and age. From these figures, it can be observed that the score means at older ages were higher than at younger ages with statistically significant differences among the age groups. Also, the knowledge score playing the game by pairs was a bit higher than playing the game individually, having statistically significant differences between genders as indicated by the above ANOVA analysis.

#	PrePair	PreIndiv	t	p	Cohen's d
Q1	0.13 ± 0.34	0.19 ± 0.39	-0.48	0.635	-0.15
Q2	0.20 ± 0.40	0.12 ± 0.33	0.63	0.533	0.19
Q3	0.10 ± 0.30	0.31 ± 0.46	-1.84	0.073	-0.57
Q4	0.27 ± 0.44	0.25 ± 0.43	0.12	0.905	0.04
Q5	0.13 ± 0.34	0.12 ± 0.33	0.08	0.938	0.02
Q6	0.30 ± 0.46	0.00 ± 0.00	2.56	0.014**	0.79
Q7	0.03 ± 0.18	0.06 ± 0.24	-0.45	0.653	-0.14
Q8	0.77 ± 0.42	0.50 ± 0.50	1.87	0.069	0.58
Q9	0.57 ± 0.50	0.69 ± 0.46	-0.79	0.435	-0.24
Q10	0.20 ± 0.40	0.25 ± 0.43	-0.38	0.703	-0.12
Q11	0.03 ± 0.18	0.19 ± 0.39	-1.79	0.080	-0.55
Q12	0.20 ± 0.40	0.00 ± 0.00	1.96	0.057	0.61
Q13	0.03 ± 0.18	0.12 ± 0.33	-1.19	0.240	-0.37

Table 7.1: Means and standard deviations of questions for PrePair and PreIndiv questionnaires, t-test analysis, and Cohen's d, d.f. = 44

#	PrePair	PosPair	t	p	Cohen's d
Q1	0.13 ± 0.34	0.73 ± 0.44	-6.60	$<$0.001**	-1.20
Q2	0.20 ± 0.40	0.47 ± 0.50	-2.80	0.009**	-0.51
Q3	0.10 ± 0.30	0.80 ± 0.40	-7.17	$<$0.001**	-1.31
Q4	0.27 ± 0.44	0.33 ± 0.47	-0.63	0.536	-0.11
Q5	0.13 ± 0.34	0.53 ± 0.50	-3.89	$<$0.001**	-0.71
Q6	0.30 ± 0.46	0.47 ± 0.50	-1.72	0.096	-0.31
Q7	0.03 ± 0.18	0.40 ± 0.49	-4.10	$<$0.001**	-0.75
Q8	0.77 ± 0.42	0.90 ± 0.30	-1.28	0.211	-0.23
Q9	0.57 ± 0.50	0.73 ± 0.44	-1.54	0.134	-0.28
Q10	0.20 ± 0.40	0.60 ± 0.49	-3.03	0.005**	-0.55
Q11	0.03 ± 0.18	0.63 ± 0.48	-5.83	$<$0.001**	-1.07
Q12	0.20 ± 0.40	0.83 ± 0.37	-6.24	$<$0.001**	-1.14
Q13	0.03 ± 0.18	0.27 ± 0.44	-2.97	0.006**	-0.54

Table 7.2: Means and standard deviations of questions for PrePair and PosPair questionnaires, t-test analysis, and Cohen's d, d.f. = 29

#	PreIndiv	PosIndiv	t	p	Cohen's d
Q1	0.19 ± 0.39	0.50 ± 0.50	-2.61	0.020**	-0.65
Q2	0.12 ± 0.33	0.19 ± 0.39	-0.44	0.669	-0.11
Q3	0.31 ± 0.46	0.75 ± 0.43	-3.42	0.004**	-0.85
Q4	0.25 ± 0.43	0.31 ± 0.46	-0.37	0.718	-0.09
Q5	0.12 ± 0.33	0.38 ± 0.48	-1.73	0.104	-0.43
Q6	0.00 ± 0.00	0.25 ± 0.43	-2.24	0.041**	-0.56
Q7	0.06 ± 0.24	0.44 ± 0.50	-2.42	0.029**	-0.61
Q8	0.50 ± 0.50	0.62 ± 0.48	-0.81	0.432	-0.20
Q9	0.69 ± 0.46	0.75 ± 0.43	-0.56	0.580	-0.14
Q10	0.25 ± 0.43	0.50 ± 0.50	-1.73	0.104	-0.43
Q11	0.19 ± 0.39	0.56 ± 0.50	-3.00	0.009**	-0.75
Q12	0.00 ± 0.00	0.62 ± 0.48	-5.00	<0.001**	-1.25
Q13	0.12 ± 0.33	0.12 ± 0.33	0.00	1.000	0.00

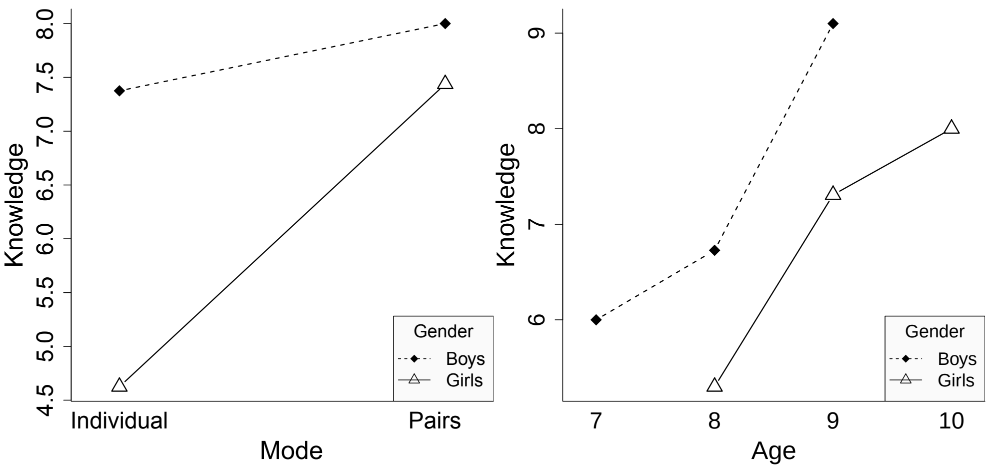
Table 7.3: Means and standard deviations of questions for PreIndiv and PosIndiv questionnaires, t-test analysis, and Cohen's d, d.f. = 15

#	PosPair	PosIndiv	t	p	Cohen's d
Q1	0.73 ± 0.44	0.50 ± 0.50	1.59	0.119	0.49
Q2	0.47 ± 0.50	0.19 ± 0.39	1.90	0.064	0.59
Q3	0.80 ± 0.40	0.75 ± 0.43	0.38	0.703	0.12
Q4	0.33 ± 0.47	0.31 ± 0.46	0.14	0.889	0.04
Q5	0.53 ± 0.50	0.38 ± 0.48	1.01	0.317	0.31
Q6	0.47 ± 0.50	0.25 ± 0.43	1.44	0.158	0.44
Q7	0.40 ± 0.49	0.44 ± 0.50	-0.24	0.811	-0.07
Q8	0.90 ± 0.30	0.62 ± 0.48	2.32	0.025**	0.72
Q9	0.73 ± 0.44	0.75 ± 0.43	-0.12	0.905	-0.04
Q10	0.60 ± 0.49	0.50 ± 0.50	0.64	0.525	0.20
Q11	0.63 ± 0.48	0.56 ± 0.50	0.46	0.648	0.14
Q12	0.83 ± 0.37	0.62 ± 0.48	1.59	0.120	0.49
Q13	0.27 ± 0.44	0.12 ± 0.33	1.10	0.277	0.34

Table 7.4: Means and standard deviations of questions for PosPair and PosIndiv questionnaires, t-test analysis, and Cohen's d, d.f. = 44

Factor	d.f.	F	p	η^2
Gender	1	4.54	0.040**	0.087
Age	3	1.75	0.174	0.101
Pos (Pair/Ind.)	1	1.63	0.210	0.043
Gender:Age	1	0.80	0.376	0.015
Age:Pos	1	1.30	0.261	0.025
Gender:Pos	1	2.20	0.261	0.042
Gender:Age:Pos	1	0.15	0.700	0.002

Table 7.5: Multifactorial ANOVA for the *knowledge* variable, $N = 46$



(a) By game mode (b) By age

Figure 7.4: Interaction plots for the acquired knowledge

7.3.2 Satisfaction outcomes

The aspects not related to knowledge were determined by the questions Q14 to Q24 shown in Table A.3. Their analysis is shown in Table 7.6. As Q14 shows, most of the children stated that they had a lot of fun while playing the game. Moreover, the children scored the game with 9.57 over 10 (Q23). We also asked the children to rate how difficult the game was (Q16); the results indicated that they found the game easy to play. The questions about the avatar (Q20 and Q21) were also scored highly; most of the children liked the avatar and they also thought that the avatar helped them a lot during the game.

#	Bounds	$\mu \pm \sigma$	Answer
Q14	[1-5]	4.85 ± 0.36	Very much
Q15	[1-5]	3.47 ± 0.69	To several friends
Q16	[1-5]	3.96 ± 0.78	Easy
Q17	[1-5]	4.63 ± 0.53	Good
Q18	[1-5]	4.02 ± 0.82	Easy
Q19	[1-5]	4.67 ± 0.51	A lot
Q20	[1-5]	4.48 ± 0.83	A lot
Q21	[1-5]	4.22 ± 0.83	A lot
Q22	[1-5]	4.46 ± 0.71	A lot
Q23	[1-10]	9.57 ± 0.77	Very good

Table 7.6: Satisfaction results

To determine if there were statistically significant differences between the two game modes with respect to satisfaction questions, a t-test analysis was performed. Table 7.7 shows the results of this test. Statistically significant differences were only found in Q18, where the children were asked about the difficulty of selecting the answers. As the results indicate, the children who played the collaborative mode found it easier to select the answers than the children who played the individual mode.

In order to have a global score for the questions included in the satisfaction questionnaire, we used a new variable, called *satisfaction*. This variable consists of the sum of all the values of questions Q14 - Q23 given by each child. To analyze the *satisfaction* variable, a multifactorial ANOVA test was performed. Table 7.8 shows the results. In this case, no statistically significant differences were found between the different factors.

For Q24, which asked the children about the mini-games they liked the most, a vote count was performed. The children could vote for more than one mini-game. When the results between the children who played the individual mode and the children who played in pairs are compared, some differences can be observed as shown in the bar plots in Figure 7.5. The preferred historical ages were Prehistory and Ancient Times.

A Chi-squared test was performed to analyze the children's preferred mini-game.

#	Collaborative	Individual	t	p	Cohen's d
Q14	4.83 ± 0.37	4.88 ± 0.33	-0.37	0.715	0.11
Q15	4.40 ± 0.76	4.60 ± 0.49	-0.91	0.368	0.29
Q16	4.07 ± 0.57	3.75 ± 1.03	1.31	0.197	0.41
Q17	4.70 ± 0.46	4.50 ± 0.61	1.22	0.228	0.38
Q18	4.21 ± 0.71	3.60 ± 1.02	2.24	0.030**	0.71
Q19	4.67 ± 0.54	4.69 ± 0.46	-0.13	0.898	0.04
Q20	4.53 ± 0.67	4.38 ± 1.05	0.61	0.547	0.19
Q21	4.17 ± 0.97	4.31 ± 0.46	-0.56	0.581	0.17
Q22	4.40 ± 0.66	4.56 ± 0.79	-0.72	0.473	0.22
Q23	9.57 ± 0.72	9.56 ± 0.86	0.02	0.986	0.01

Table 7.7: Means and standard deviations for satisfaction, t-test analysis, and Cohen's d. d.f. = 44

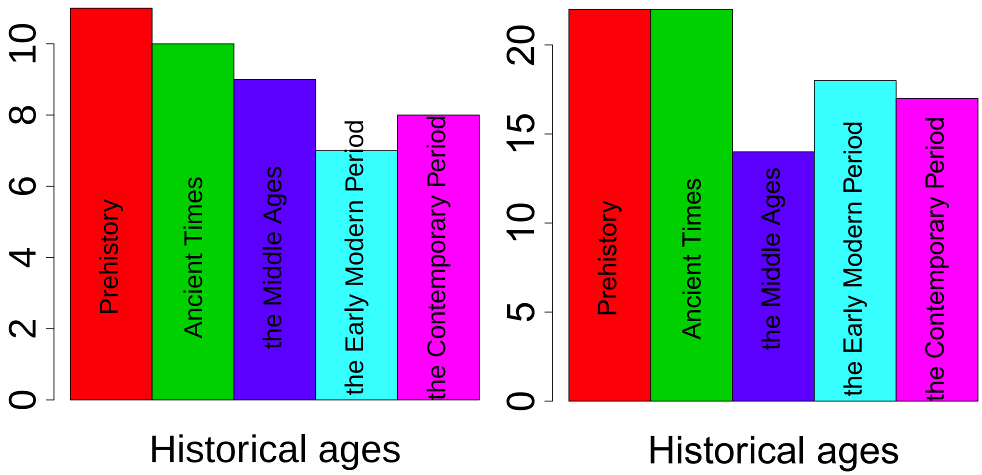
Factor	d.f.	F	p	partial η^2
Gender	1	0.003	0.950	0.0001
Age	3	0.882	0.459	0.068
Pos (Pairs/Ind.)	1	1.821	0.185	0.048
Gender:Age	1	0.667	0.419	0.018
Other interactions	1	<1.510	>0.227	<0.040

Table 7.8: Multifactorial ANOVA for the *satisfaction* variable, $N = 46$

Table 7.9 shows the results of this test. These results show that there were no statistically significant differences between the two game modes. The Prehistory and the Ancient Times mini-games were the most voted as the preferred mini-games in the two groups (collaborative and individual).

7.3.3 Correlation analysis

An analysis to determine if there was a correlation among any of the questions was performed. When analyzing the two groups together, a correlation between Q20 (avatar) and Q23 (score the game) was found ($0.667, p < 0.001$). This means that the avatar character was an important factor in the children liking the game. When analyzing the groups separately, the same correlation was found in the individual



(a) Individual mode (b) Pair mode

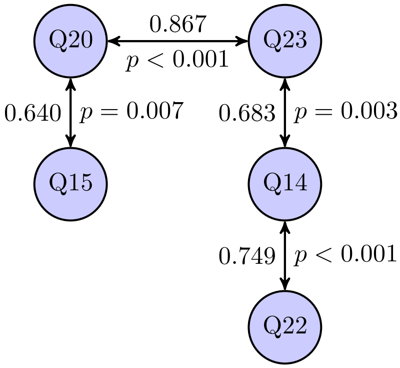
Figure 7.5: Comparative frequencies of the mini-games that the children liked the most

#	Collaborative	Individual	χ^2	p	Cramer's V
Prehistory	1	1	0.00	1.00	0.05
Ancient Times	1	1	0.18	0.48	0.14
Middle Ages	0	1	0.10	0.76	0.09
Early Modern	1	0	0.55	0.46	0.16
Contemporary	1	0	0.01	0.90	0.06

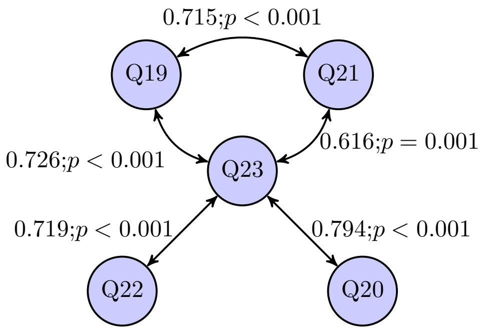
Table 7.9: Modes of preferred mini-game, Chi-squared analysis, and Cramer's V, d.f. = 1, $N = 46$

group. All the correlations found for the individual group are shown in Figure 7.6a. From these results, we can state that the level of fun the children had while playing the game is related to the score they gave to the game. We can also affirm that the more they learned, the more fun they had. It can also be observed that the avatar plays an important role, since there is a high correlation with the score that the children gave the game.

When considering groups formed only by boys and groups formed only by girls separately, there were different correlation results. In the group of boys, there was a correlation between Q15 and Q23 ($0.648, p = 0.001$). For the group of girls, the correlations found are shown in Figure 7.6b. It can be observed that the score (Q23) is directly related to the children's perception of having learned (Q22), having liked the images (Q19) and the avatar (Q20), and children's perception that the avatar had helped (Q21).



(a) Correlations found in the individual group



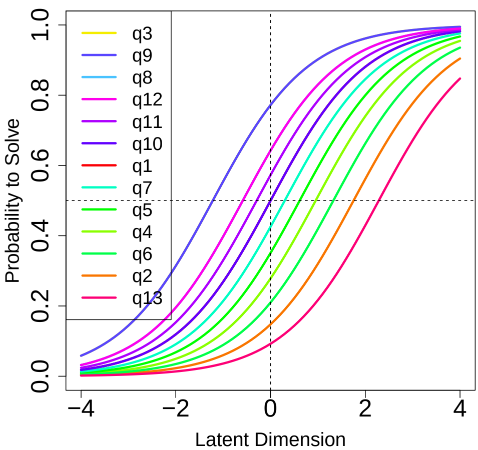
(b) Correlations found in the group of girls

Figure 7.6: Correlations

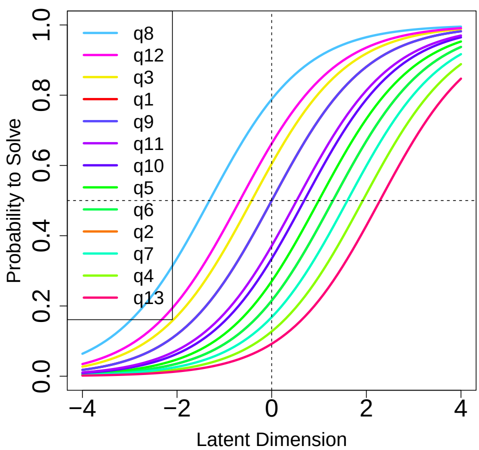
When considering groups organized by age separately, in the group of 8-year-old children, the strongest correlations were between Q20 and Q23 ($0.574, p = 0.006$), and between Q20 and Q21 ($0.552, p = 0.009$). For the group of 9-year-olds, the most correlated question was about the fun experienced (Q14). The more fun they had, the more they would recommend the game to friends ($0.665, p < 0.001$), and the more they liked the avatar ($0.729, p < 0.001$).

7.3.4 Rasch model

In order to complete the analysis, the dichotomous Rasch model proposed by Georg Rasch was used. This model measures a person's latent trait level from a probabilistic perspective (Rasch, 1960). The probability of a user answering a question correctly relies on the user's underlying ability and the difficulty of the question (Fischer, 2006). Figure 7.7 shows the Item Characteristic Curve (ICC) for every question. The latent dimension shows the ability of the children measured in the interval [-4, 4], with 0 being a child with medium ability. The curve indicates the probability that a child of each ability has to correctly answer a question. The dotted lines represent the medium values of each axis (0 for ability and 0.5 for probability). Figure 7.7a shows the ICC for the group of children who played individually. It can be observed that for the individual group, the hardest question was Q13, where it was necessary for a child to have an ability value of 2 in order to have a probability of 0.5 to answer the question correctly. The easiest questions were Q3 and Q9 (which had the same value), where a child with an ability value of -1 was enough to have a probability of 0.5 to answer correctly. The most balanced question in this group was Q10, which needed an ability of 0 (the medium value) to have a probability of 0.5. Figure 7.7b shows the group of children who played collaboratively. The order of the questions changed a little. The hardest question was also Q13, but the easiest one was Q8. The most balanced questions were Q1 and Q9.



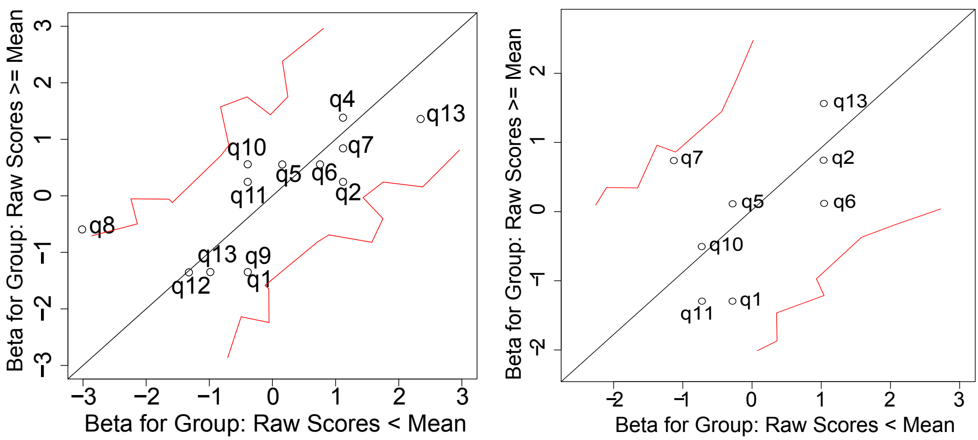
(a) Individual



(b) Pair

Figure 7.7: Item Characteristic Curve for all questions

A graphical model check was also performed. In this model the questions were grouped by raw scores and the ones which are higher than the mean are separated from the ones which are lower. The red lines represent the confidence bands. The results of the questions for the collaborative group are shown in the graph in Figure 7.8a. In this case, every question was inside the confidence bands, except Q8. This indicates that Q8 is an easy question. In fact, this question was answered correctly by almost every child in the collaborative group. In the individual group, all the questions were inside the confidence bands, as shown in Figure 7.8b. This result is in line with the result obtained when each knowledge question was analyzed, in which the t-test between PosPair – PosIndiv indicated that the children acquired more knowledge in only one question (Q8).



(a) Collaborative (b) Individual

Figure 7.8: Graphical Model Check

Based on the results of the ICC and the graphical model check, it can be concluded that the questions are appropriate for the assessment of the acquired knowledge for both game modes.

In order to visually check the children and the questions, a Person-Item Map was plotted, where the estimated ability of the child and question difficulty measures are placed side by side in one vertical dimension. The questions appear in order of difficulty. The Person-Parameter Distribution, which is at the top of the graph, is a distribution of the children's abilities. For the collaborative group, it can be observed in Figure 7.9 that the hardest question (Q13) was easier than the ability of 10.34% of the children. On the other hand, the easiest question (Q8) was more difficult than the ability of 3.44% of the children. The question in the middle (Q11) was easier than the ability of 75.86% of the children and more difficult than the ability of 24.14%. For the individual group, the hardest question (Q13) was easier than the ability of 6.25%

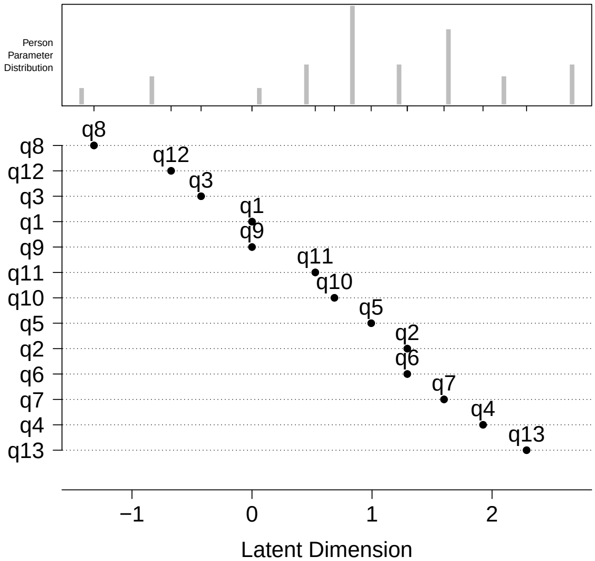


Figure 7.9: Person-Item Map for the collaborative group

of the children, and the easiest questions (Q3 and Q9) were more difficult than the ability of 18.75% of the children. The question in the middle (Q5) is easier than the ability of 68.75% of the children, and more difficult than the ability of 31.25% of the children.

When the Person Parameter Distribution of the children who played individually and in pairs were compared separately, we obtained the distributions shown in Figure 7.10. In the distribution of children who played collaboratively, most of the values are situated in higher values of ability, between 0 and 2. In contrast, the opposite happens when looking at the distribution of children who played individually, where most of the values are situated in values of less ability. This indicates that the collaborative mode provides more ability to correctly answer the questions.

The Rasch model is also based on the idea that the conditional maximum likelihood estimation of the item parameters is independent of the actual values of the individual parameters. This means that it can be expected to get the same item parameter estimations from individuals with high raw scores as from individuals with low raw scores; however, this only happens when the model is true. To determine if the model is true, the test proposed by Andersen (1973) can be used. This test is based on a

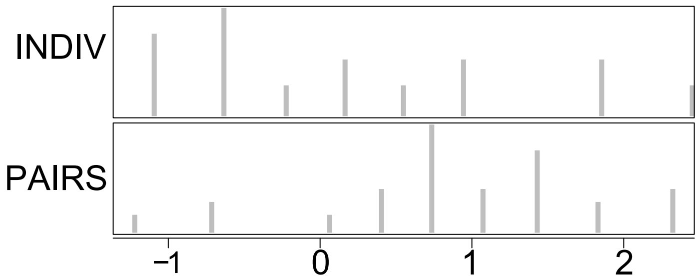


Figure 7.10: Comparison of individual and pairs Person Parameter Distribution

comparison between the difficulties estimated from different score groups and overall estimates, resulting in a conditional likelihood ratio. Andersen stated that 2 times the logarithm of this ratio is χ^2-distributed when the Rasch model is true. In our study, this test offered the values, LR-value = 8.828, df = 12, $p = 0.718$, which fit the χ^2 distribution. Therefore, in this study the Rasch model is true.

7.4 Conclusions

In this study, the educational video game presented in section 3.2 was used for test the level of the knowledge acquired by the children when playing in two different modes. The first mode was the collaborative mode, where the children had to play with the game in pairs with other children. The second mode was the individual mode, where only one child played at a time.

Statistically significant differences were found when the knowledge variable of the children before and after playing with the two modes (collaborative and individual) was analyzed. Therefore, it can be concluded that play our historical content game positively affected the learning outcomes of the participants. This indicates that the children remembered quite a lot of the knowledge transmitted in the game. For this reason, we can say that our game has been effective when it comes to transmitting knowledge in the short-term.

When the results obtained after playing the two modes were compared, statistically significant differences were found in favor of the collaborative mode. If the Person Parameter Distribution is taken into account, it also indicates that the collaborative mode provides more ability to correctly answer the questions. In other words, the ability to correctly answer the learning questions was higher in the group of children who played the game collaboratively. Therefore, we can conclude that the collaborative mode facilitates learning to a greater extent than the individual mode.

The results show that the game transmits knowledge effectively both collaboratively and individually. At this point, we would like to highlight that the percentage of correct answers for most of the questions (11 of 13) was higher in the collaborative mode. Our fourth hypothesis was that the children would increase more their knowledge about the topic of the game when they play the game in the collaborative mode. Comparing their initial knowledge before playing with the game, we observed a very little difference in favour of the collaborative mode; however, when comparing their final knowledge after playing with the game, a more notorious difference arose. The results showed that the children who played in collaboration with another children, acquired more knowledge that the children who played alone. This result indicates that our hypothesis has been corroborated. Finally, we believe that this result is very encouraging because it implies that games of these characteristics are suitable for both collaborative and individual learning, and that collaboration positively affects learning outcomes.

In the analyses of other aspects such as experienced fun, usability, or preferences, there were no statistically significant differences for the 10 analyzed questions, except for Q18. The results indicate that the game was easy to use and the children had fun while playing it. Most of the children would recommend the game to their friends and, generally, they scored the game very high. It is also important to have some kind of guide such as the figure of the avatar throughout the game. From the results, we observed that the children valued this character highly. They thought the avatar helped them a lot during the game. The high scores assigned to the questions related to the avatar and the correlations found corroborate the importance of the two design guidelines indicated by Villalta et al. (2011) (Interactivity and guidance, and accessible language).

We would also like to discuss the ease of use and its implications. Several authors argued that perceived ease of use is an important technical factor that affects educational effectiveness (Jones et al., 1999; Mayes and Fowler, 1999; Squires, 1999). Based on their findings, Sun et al. (2008) stated that learning systems that are easy to use help students to focus their attention on the learning content and they are more motivated to learn. Since our game was easy to use, we consider that the game does help students focus their attention on the learning content.

Differences in interactions like gender, age and game mode (collaborative or individual) were also considered. These analyses revealed that boys improved their knowledge more than girls, and indicate that older children improved their knowledge more than younger children, specially between 9 and 10-year-old children.

Study 3: Collaboration in Large Groups vs. Collaboration in Pairs or Traditional Methods

8

"Education is the movement from darkness to light."

— Allan Bloom

8.1 Introduction

In this study, a comparison of the traditional learning method with a collaborative learning method using the game was carried out. The collaborative part can be performed in two different ways: in pairs (two children) or in a large group (up to 12 children). Our fifth hypothesis is that children who learn by playing a computer-based game collaboratively on a large-size tabletop display and involving up to 12 students will obtain significantly higher learning outcomes than those who learn by attending a traditional class.

Our sixth hypothesis is that children who learn by playing a computer-based game collaboratively on a large-size tabletop display and involving up to 12 students will obtain significantly higher learning outcomes than those who learn by playing with the same game, but playing in pairs.

Our seventh hypothesis is that children who learn by playing a computer-based game collaboratively played on a large-size tabletop display and involving up to 12 students will show significantly higher satisfaction than those who learn by attending a traditional class.

Our eighth hypothesis is that children who learn by playing a computer-based game collaboratively played on a large-size tabletop display and involving up to 12 students table will show significantly higher satisfaction and usability than those who learn by playing with the same game, but playing in pairs.

8.2 Design of the evaluations

The frontal projection game described in the Chapter 5 was played by a total of 100 children. This section presents the characteristics of the children that played the game, the measurements used during the experiment, and the steps followed.

8.2.1 Participants

A total of 100 children participated in this study. Two different schools were involved. At the first school, a total of 82 children participated in the study. There were 43 boys (52.44%) and 39 girls (47.56%). They were between eight and eleven years old. The mean age was 8.83 ± 0.72 years old. The children were in 3rd grade (40) and 4th grade (42). There were four classes. Two for 3rd grade and two for 4th grade. These children were attending the Engeba school of Valencia. At this first school, the children participated by playing the game in large groups (shown in Figures 8.1 and 8.2)

Figure 8.1: Room and material used for the study

and by attending to a traditional class.

At the second school, a total of 18 children participated. These children are a subset of the participants from the study presented in Chapter 7. There were 13 boys (72.22%) and 5 girls (27.78%). They were between eight and nine years old, and they had already finished the third grade of primary education. The mean age was 8.72 ± 0.46 years old. At this second school, the children participated by playing the game played in pairs.

8.2.2 Measurements

To retrieve the data for the analysis, 5 different questionnaires were used. There was a pre-test (A) with only thirteen knowledge questions. This pre-test was designed to measure the knowledge the children had before playing the game or receiving the traditional learning method in the classroom. Afterwards, the knowledge questions were also asked in the B and C questionnaires. Satisfaction and usability questions were also included in these questionnaires. Finally, the children filled out two more different post-tests (D, E) once they received the two kinds of learning. These questionnaires include satisfaction and usability questions and also questions to determine which of

Figure 8.2: Children playing the game

the two learning methods they preferred. The questions asked in the questionnaires are detailed in Table A.4.

8.2.3 Procedure

For the comparison between learning in large groups and learning in a traditional class, there were 82 participants divided into four classes. These four classes belonged to two groups for 3rd grade (40 children from 3A and 3B) and two groups for 4th grade (42 children from 4A and 4B). The children in 3A and 4B were assigned to Group α. The children in 3B and 4A were assigned to Group β. The students that had played with the game were also divided into two groups with a maximum of 12 students.

The school does not have a policy of dividing the children into different classes based on their academic achievement. Therefore, a priori, we consider the two classes for 3th grade and the two classes for 4th grade (i.e., third grade 3A and 3B, and fourth grade 4A and 4B), to be homogeneous in knowledge. To corroborate this homogeneity, the previous knowledge of the two classes in each grade was analyzed in Section 8.3.1.

The procedure to compare between learning in large groups and learning in a traditional class can be summarized as follows:

- Group α: Participants that played the learning game in large groups first and afterwards received traditional learning in the classroom.
- Group β: Participants that received traditional learning in the classroom and afterwards played the learning game in large groups.

Figure 8.3 shows graphically the procedure for both groups. The following protocol was used:

1. The children filled out the pre-test questionnaire (A).
2. The children from Group α played the learning game in groups collaboratively and children from Group β received traditional learning in the classroom.
3. Then, all the children filled out the first post-test questionnaire (B for Group α and C for Group β).
4. Afterwards, they received the other learning method: the children from Group α received traditional learning in the classroom and children from Group β played the learning game in groups collaboratively.
5. Finally, all the children filled out the final questionnaire (D for Group α and E for Group β).

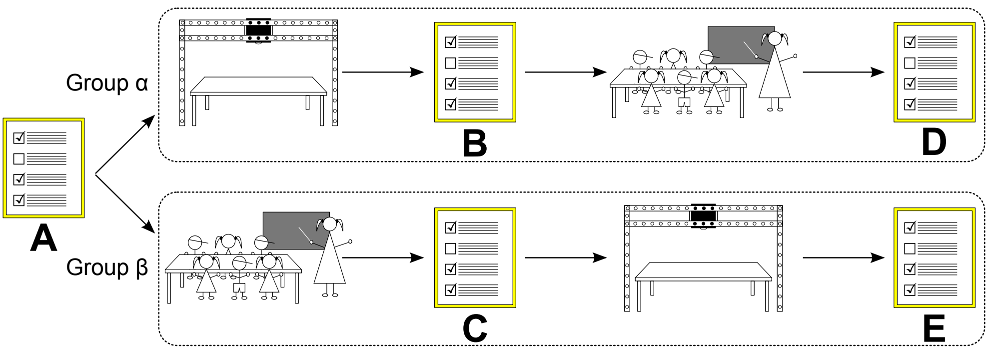


Figure 8.3: Study procedure

For the pair mode (18 participants), the protocol was the following:

1. The children filled out the pre-test questionnaire (A).
2. The children played the learning game in pairs.
3. The children filled out the post-test questionnaire (B).
4. Afterwards, the children filled out the final questionnaire (D, but only the questions not included in B).

The homogeneity of the pair mode and the 3A group (the large group used for the comparison) was also analyzed and is included in Section 8.3.1.

The game can be played by one or two players at a time. When there is only one player, the player is placed in front of the table in the middle and uses both hands to select the buttons on the right/left sides. When there are two players, one player is placed in front of the table on the left side and the other player is placed on the right side. The child on the left side must use his/her left hand, and the child on the right side must use his/her right one. This placement helps the children to interact with the buttons close to them. In an attempt to balance the participation of the two children in the pair mode, half of the interactions with the game and half of the answers to the questions should be performed by each child in order to reach the next level. If one of the children is much more active than the other, the most active child could have more interactions than the other child. To limit this situation to the maximum, the buttons for each child are on his/her side and one child would have to invade the physical space of the other child in order to perform that selection. When there are more than two players, they must play collaboratively in pairs by turn, the children have to be placed around the table where the game is played (Figure 8.1). Since there are 6 different activities possible (5 historical ages and a final puzzle), one pair of children interacts physically with the game while the rest of the other children are around the table. The children that are interacting are responsible for that historical age, but they have to speak with all of their classmates in order complete the activity. The children collaborate to answer correctly in order to pass to the next historical age. Since the game is collaborative and not competitive, the children do not compete to be the first to answer. If the group is composed of ten children, one of the pairs plays the contemporary period game and also completes the puzzle. If the number of children is not even, then, one of the children plays twice. The traditional classroom lesson was given in a traditional classroom. The traditional classroom lesson had the same learning content as the game. The children were not informed about whether or not their answers on the pre-test were correct. Thus, the children did not acquire any knowledge by answering the pre-test; they only learned during the game.

8.3 Results

This section presents the analysis of the data collected from this study to corroborate the four hypotheses of this study. In order to explore means, standard deviations, and other measurements, an initial descriptive analysis was carried out. Then, data normality was checked and, the pertinent statistical tests were carried out based on those results. Before using inferential tests, Shapiro-Wilk and Anderson-Darling tests were performed to check data normality. Both tests reported that our data did not fit the normal distribution. For this reason, the tests used were non-parametric (the Mann-Whitney U test for unpaired data, the Wilcoxon Signed-rank sum test for paired data, and the Kruskal-Wallis test instead of ANOVA). As in the previous studies, the data were analyzed using the statistical open source toolkit R with the *RStudio* IDE.

8.3.1 Learning outcomes

In order to measure how much the children had learned, the knowledge variable was analyzed before playing (Pre-test) and after playing (Post-test). The knowledge variable was created to condense the thirteen knowledge questions by counting the number of correct answers.

Large Group vs. Traditional Class

The descriptor of each group is presented in the format (*median; interquartile range*), and all tests are presented in the format (*statistic U/W, normal approximation Z, p-value, r effect size*); ** indicates the statistical significance at level $\alpha = 0.05$.

To determine whether or not there were differences between the initial knowledge of those two groups, an unpaired test was performed between the knowledge variable in PreLGroup (3; 2) and the knowledge variable in PreTClass (2; 1) ($U = 1031$, $Z = 1.833$, $p = 0.067$, $r = 0.202$). These results revealed that there were no statistically significant differences between the knowledge of the two groups in the pre-test. We also checked if there were differences between the initial knowledge of the children from each class (3A, 3B, 4A, and 4B). The results of these tests were that there were no statistically significant differences between the classes 3A (3; 1.25) and 3B (2; 2) ($U = 276.5$, $Z = 1.422$, $p = 0.159$, $r = 0.219$), or between the classes 4A (3; 2) and 4B (3; 1) ($U = 224.5$, $Z = 0.105$, $p = 0.927$, $r = 0.016$).

To determine whether or not there were differences between the initial knowledge and the knowledge after playing the game, a paired test was performed between

PreLGroup (3; 2) and PostLGroup (8; 5) that showed that there were statistically significant differences between the scores that the children obtained before and after playing the game in large groups ($W = 2.5$, $Z = -5.481$, $p < 0.001^{**}$, $r = 0.605$). Another paired test revealed that there were also statistically significant differences between the scores of PreTClass (2; 1) and PostTClass (4; 4) ($W = 41.5$, $Z = -4.639$, $p < 0.001^{**}$, $r = 0.512$), which refers to before and after receiving the traditional learning lesson in the classroom. Finally, to determine whether or not there were differences between the two learning methods, an unpaired test was performed between PostLGroup (8; 5) and PostTClass (4; 4) ($U = 1226.5$, $Z = 3.603$, $p < 0.001^{**}$, $r = 0.398$). These results showed that there were statistically significant differences in favour of the game learning method playing in large groups. Figure 8.4 shows the box plot for the scores before and after playing for the group of children that played the game in a large group and the children that received the traditional class. Therefore, our fourth hypothesis has been corroborated.

The Kruskal-Wallis test was also performed to take into consideration several factors in the study: gender, age and learning method. The results, which shown in Table 8.1, revealed that the knowledge acquired was independent from the gender and the age factors, and a significant effect of the learning method on knowledge ($\chi^2[1] = 12.978$, $p < 0.001^{**}$, $r = 0.398$) was found. The age factor obtained a very small p-value, although we cannot assume statistically significant differences at level $\alpha = 0.05$.

Factor	Kruskal-Wallis χ^2	d.f.	p
Gender	0.403	1	0.525
Age	5.938	2	0.051
Learning method	12.978	1	<0.001**

Table 8.1: The Kruskal-Wallis tests for the *knowledge* variable

For the knowledge variable, Figure 8.5a shows the interaction plot between the gender and learning method factors. It can be seen graphically that the acquired knowledge was very similar for boys and girls regardless of the learning method they used. Figure 8.5b shows the interaction plot between the gender and age factors, where children from the 9-year-old and 10-year-old groups scored better that the children from the 8-year-old group, but not significantly better.

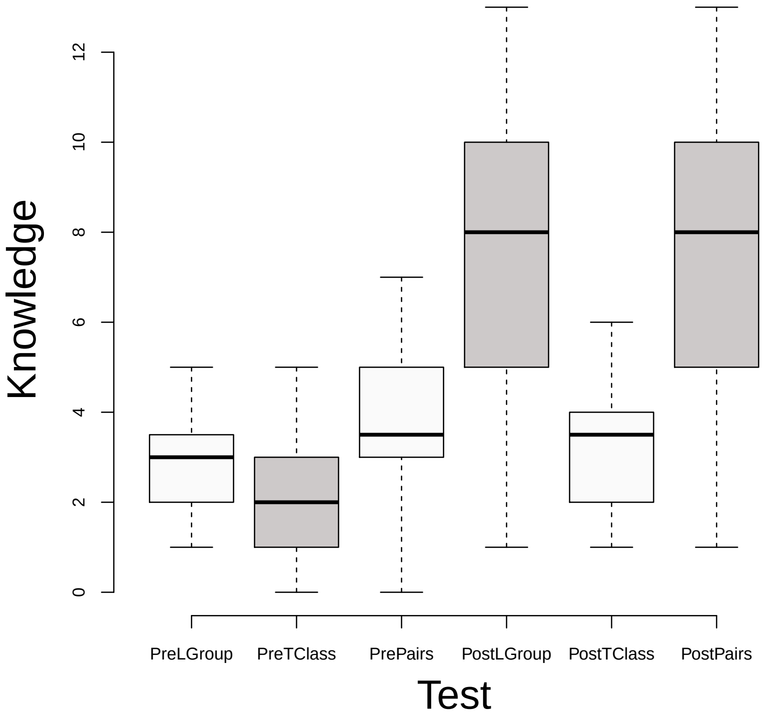
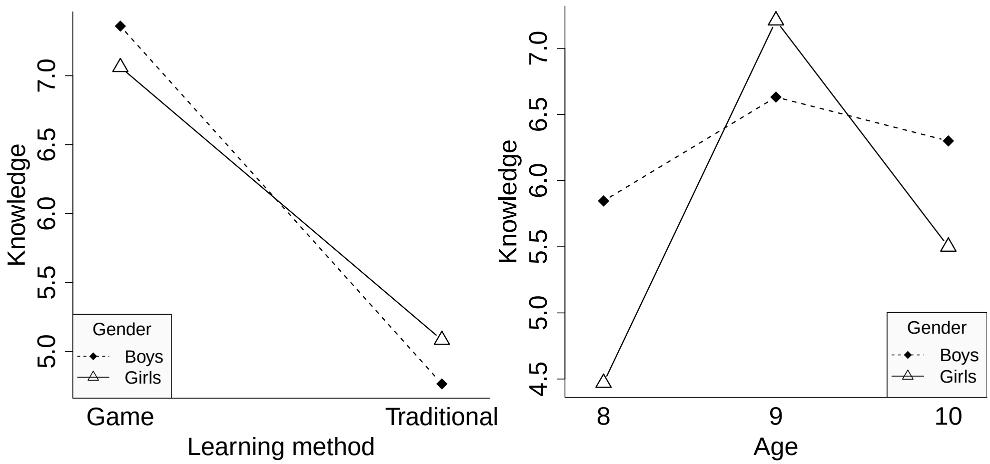


Figure 8.4: Scores of the knowledge variable before and after receiving either the game method (played in a large group or in pairs) or the traditional method

Large Group vs. Pairs

To check the homogeneity between the children that played the game in a large group and the children that played the game in pairs, an unpaired Mann-Whitney U test was performed for their initial knowledge. No statistically significant difference was found between PrePairs (3.5; 1.75) and Pre3A (3; 2) ($U = 285.5$, $Z = -1.419$, $p = 0.159$, $r = 0.185$). After playing the game, the results of the comparisons between PostPairs (8; 4.75) and Post3A (8; 5) indicated no statistically significant differences ($U = 372$, $Z = -0.098$, $p = 0.928$, $r = 0.013$). Figure 8.4 shows the box plot for the scores before and after playing for the children that played the game in a large group and the children that played in pairs. Therefore, our sixth hypothesis has not been corroborated.



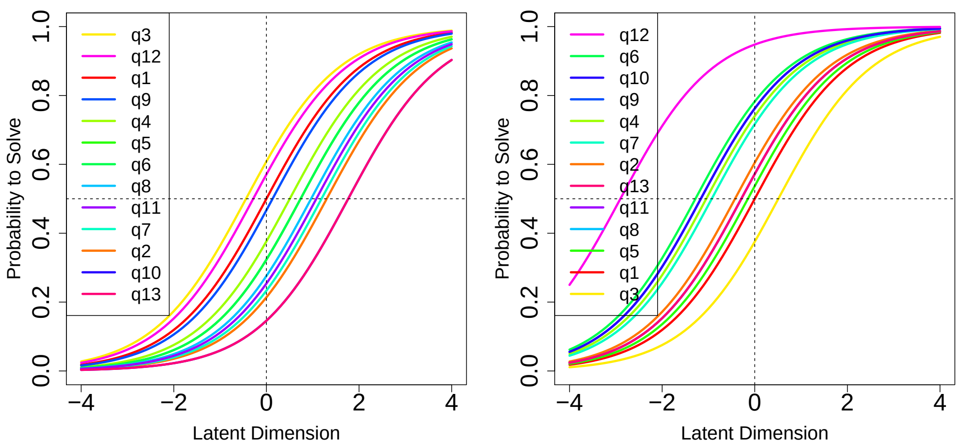
(a) Interaction by gender for each questionnaire (b) Interaction by gender for each age group

Figure 8.5: Interaction plots for the acquired knowledge

8.3.2 Rasch model analysis

In order to perform a qualitative analysis, as in the previous studies, the dichotomous Rasch model proposed by Rasch (1960) was used. Figure 8.6 shows the Item Characteristic Curve for each question. As mentioned in the previous study, the latent dimension shows the ability of the children measured in the interval [-4, 4], with 0 being a child with medium ability. The curve indicates the probability that a child of each ability has to correctly answer a question. The dotted lines represent the medium values for each axis (0 for ability and 0.5 for probability). Figure 8.6a shows the Item Characteristic Curve for the group of children who played the game first. It can be observed that, the hardest question for this group was Q13, where it was necessary for a child to have an ability value of 2 in order to have a probability of 0.5 to answer the question correctly. The easiest question was Q3, where a child with an ability value of -1 was enough to have a probability of 0.5 to answer correctly. The most balanced question in this group was Q1, which needed an ability value of 0 (the medium value) to have a probability of 0.5.

On the other hand, Figure 8.6b shows the group of children who received the traditional class first. It can be observed that the order of the questions had changed with respect to the other group. In this case, the hardest question was Q3 (which was the easiest question in the other group) and the easiest question was Q12. The most balanced question was again Q1. From these figures, it can be observed that the way the questions are distributed is not balanced in the traditional class group, which means that the contents were not assimilated as well as in the group that played with the game.

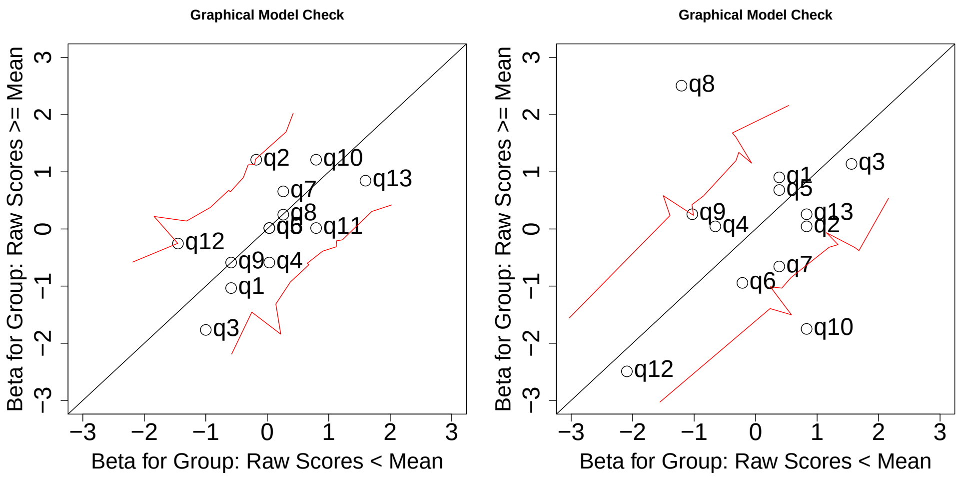


(a) Game group (b) Class group

Figure 8.6: Item Characteristic Curve for all questions

A graphical model check was also performed. The questions were grouped by raw scores and the ones that were higher than the mean were separated from the ones that were lower. The red lines represent the confidence bands. The results of the questions for the group of children who played the game first are shown in the graph in Figure 8.7a. For this group, every question was inside the confidence bands. In the case of the group of children who received the traditional class first, (shown in Figure 8.7b), the questions were more scattered; there are even some questions outside of the confidence bands (Q8 and Q10). This means that since the questions were the same for both groups, the group that played the game first was better prepared to answer them than the traditional class group.

In order to visually check the children and the questions, a Person-Item Map was plotted (Figure 8.8), where the estimated ability of the child and the question difficulty measures are placed side by side in one vertical dimension. The questions appear in

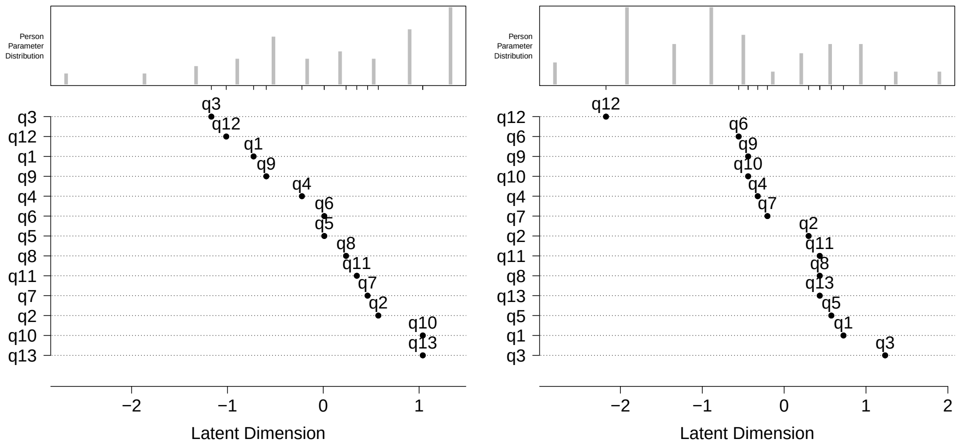


(a) Game group (b) Class groups

Figure 8.7: Graphical model check

order of difficulty. The Person-Parameter Distribution, which is at the top of the graph, is a distribution of the children's abilities. For the game group, Figure 8.8a shows the distribution of the children with respect to the questions, and the most of the children are located to the right of the most difficult questions. This means that these children were well prepared to answer those questions. In contrast, in the traditional class group, (shown in Figure 8.8b), most of the children are distributed between questions Q12 (the easiest) and Q6 (the second easiest). This means that most of the children were not as well prepared to answer the questions as the children from the other group.

Finally, the test proposed by Andersen (1973) was used in order to check the goodness of fit of the Rasch model. In this study, this test offered the values, LRvalue $= 35.274$, df $= 12$, $p = 0$, which fit the Chi-squared distribution. Therefore, the Rasch model is true in this study.



(a) Game group

(b) Class group

Figure 8.8: Person-Item Map for both groups

8.3.3 Satisfaction and usability outcomes

Large Group vs. Traditional Class

Several non-parametric statistical tests (the Wilcoxon Signed-rank sum test and the Mann-Whitney U test) were performed for the Likert questions to determine whether or not there were statistically significant differences for the satisfaction and usability questions (Table A.4), depending on which learning method they received first (Q14-Q16, Q18). The results of these tests for the unpaired data are shown in Tables 8.2-8.5. The results for the paired data are shown in Tables 8.6 and 8.7.

In the first place, we define a satisfaction variable that combines the answers of questions related to satisfaction (Table A.4). There were no statistically significant differences between the perceived satisfaction playing the game in large groups or the traditional class, ($U = 206$, $Z = 0.202$, $p = 0.853$, $r = 0.032$).

Second, we consider the individual questions and the unpaired data that show the differences between the different groups of children (i.e., the children who received the traditional class first compared with the children who played the game in a large group first). For all the combinations (LGroup first vs. TClass first; LGroup first vs. LGroup second; TClass first vs. TClass second; TClass second vs. LGroup second), there was only a statistically significant difference when the children who played the game in a large group first are compared with those who played the game in a large group second (LGroup first vs. LGroup second), and for the fun perceived in favour of the children who played the game after receiving the traditional classroom lesson, who had a lot more fun ($U = 619$, $Z = -3.078$, $p = 0.003^{**}$, $r = 0.342$).

Third, we consider the individual questions and the paired data that show the differences between the two learning methods (Large Groups vs. Traditional Class). This means that the answers from the children who played the game in a large group first were compared with their answers after receiving the traditional class and vice versa. From these results, no statistically significant differences were found for Q14-Q16, Q18.

Fourth, we consider the individual questions and the questions that are directly related to the game (Q17, Q19-Q24). This comparison includes the answers given by the children who played the game first and the children who played the game second. The results are shown in Table 8.8. Statistically significant differences were found only when asking about being able to touch the castle (Q17, $U = 483$, $Z = -3.282$, $p < 0.001^{**}$, $r = 0.367$) and when asking about recommending the game to friends (Q19, $U = 676$, $Z = -2.011$, $p = 0.045$, $r = 0.223$). In both cases, these differences were in favour of the children who had played the game second. This indicates that the children were more impressed by the game after receiving the same content from a traditional class.

Although few statistical significant differences were found for the satisfaction questions between the Large Group and the Traditional Class, if the means are considered, in almost all of the questions the means were higher in favor of the Large Group.

#	PosGame B	PosClass C	U	Z	p	r
Q14	5;1	5;0	729	−1.467	0.143	0.162
Q15	5;1	5;0	697.5	−1.713	0.089	0.189
Q16	5;1	5;1	914	0.784	0.444	0.087
Q18	10;1	10;0.75	653.5	−0.369	0.734	0.043

Table 8.2: Medians and Interquartile range of the B and C questionnaires, Mann-Whitney's U test analysis, and r effect size

The children were also asked which learning method they liked the most (Q26), why (Q27), and what they liked the most about the experience (Q28). The resulting percentages for Q26 are presented in Figure 8.9. Separated by three different groups

#	PosGame B	PosGame E	U	Z	p	r
Q14	5;1	5;0	619	−3.078	0.003**	0.342
Q15	5;1	5;0	710.5	−1.310	0.203	0.146
Q16	5;1	5;1	853.5	0.370	0.708	0.041
Q18	10;1	10;0	670.5	−1.786	0.070	0.200

Table 8.3: Medians and Interquartile range of the B and E questionnaires, Mann-Whitney's U test analysis, and r effect size

#	PosClass C	PosClass D	U	Z	p	r
Q14	5;0	5;0	373.5	−0.432	0.716	0.056
Q15	5;0	5;1	444.5	1.188	0.213	0.153
Q16	5;1	5;0.5	334.5	−1.008	0.339	0.130
Q18	10;0.75	10;0	243	−0.833	0.464	0.118

Table 8.4: Medians and Interquartile range of the C and D questionnaires, Mann-Whitney's U test analysis, and r effect size

#	PosClass D	PosGame E	U	Z	p	r
Q14	5;0	5;0	349.5	−1.300	0.240	0.169
Q15	5;1	5;0	340.5	−0.844	0.405	0.110
Q16	5;0	5;1	418	0.728	0.468	0.095
Q18	10;0	10;0	315	−0.158	0.676	0.021

Table 8.5: Medians and Interquartile range of the D and E questionnaires, Mann-Whitney's U test analysis, and r effect size

(group, grade and gender), it can be observed that in all cases, a high percentage of the children preferred the game over the traditional class. Some answers they gave when they were asked why were the following: "because there were funny things", "because I learned a lot", "because it has been more fun than in the classroom", "because we had to use electronic devices", "because it was entertaining", or "because with this game you can learn fast".

When they were asked about Q28 some of the answers were the following: "to touch the table", "the games", "to learn the colors the cavemen used", "the Roman

#	PosGame B	PosClass D	W	Z	p	r
Q14	5;0	5;0	3	0.513	1.000	0.083
Q15	5;0.5	5;1	15	0.864	0.531	0.140
Q16	5;1	5;0.5	23.5	−0.176	0.977	0.029
Q18	10;0	10;0	3	1.413	0.500	0.250

Table 8.6: Medians and Interquartile range of the B and D questionnaires, Wilcoxon Signed-rank sum test analysis, and r effect size

#	PosClass C	PosGame E	W	Z	p	r
Q14	5;0	5;0	3.5	−1.633	0.219	0.183
Q15	5;0	5;0	48	0.618	0.613	0.069
Q16	5;1	5;1	105	0.000	1.000	0.000
Q18	10;0	10;0	19	−1.480	0.202	0.182

Table 8.7: Medians and Interquartile range of the C and E questionnaires, Wilcoxon Signed-rank sum test analysis, and r effect size

#	PosGame B	PosGame E	U	Z	p	r
Q17	6;3	7;1	483	−3.282	<0.001**	0.367
Q19	5;1	5;0	676	−2.011	0.045**	0.223
Q20	5;1	5;0	684	−1.653	0.104	0.184
Q21	5;1	5;0	692	−1.605	0.118	0.178
Q22	5;1	5;1	798	−0.267	0.812	0.030
Q23	5;1	5;1	752	−0.778	0.446	0.086
Q24	5;1	5;1	776.5	−0.511	0.619	0.057

Table 8.8: Medians and Interquartile range of the B and E questionnaires, Mann-Whitney's U test analysis, and r effect size

city", "the cave paintings", "the projection", or "Prehistory". In the last question (Q29), the children were asked if they would like to use this game at school to learn other things. For the group of children who played the game first, all the children from 3rd grade voted "yes" (100%), while 81.81% of the children from 4th grade voted "yes" and 18.18% voted "no". For the group of children who played the game second,

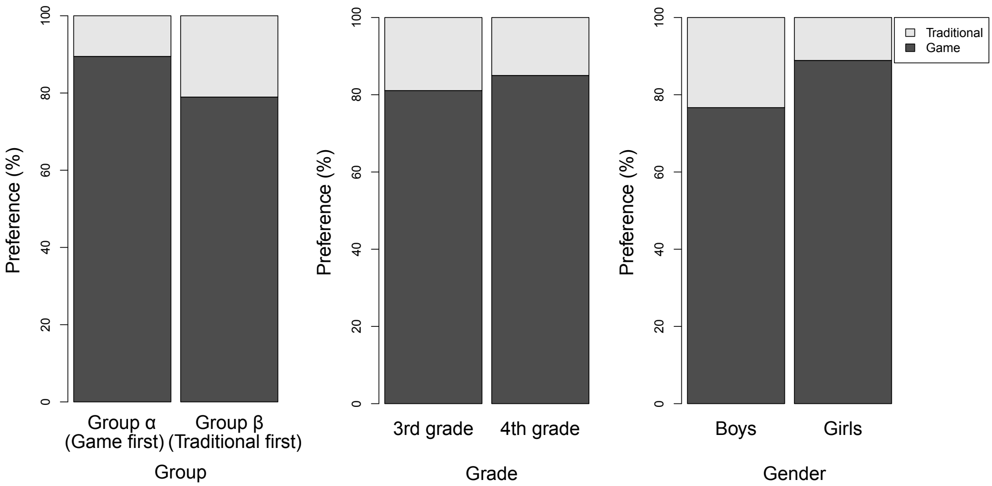


Figure 8.9: The children's vote for the preferred learning method from the different groups, grades, and gender

all the children (3rd grade and 4th grade) voted "yes" (100%). Figure 8.10 shows these results graphically.

Since there were no statistically significant differences between the Large Group and the Traditional Class, these results indicate that our seventh hypothesis has not been corroborated, even though there is considerable evidence of the satisfaction of the children with the game.

Large Groups vs. Pairs

As in the previous analysis, the data of the children that played the game in a large group were compared with those who played the game in pairs. The results for the individual questions are shown in Table 8.9. These results showed the following: the children who played in a large group found it easier to learn and scored the game higher; they would recommend the game to their friends; and they also appreciated the appearance of the game to a great extent than the children who played in pairs. Even though no statistically significant differences were found between the pairs and the large group for the satisfaction and usability questions, if the means are considered, they were higher in favour of the large group in all the questions. Moreover, if the satisfaction variable is considered, there was a statistically significant difference in

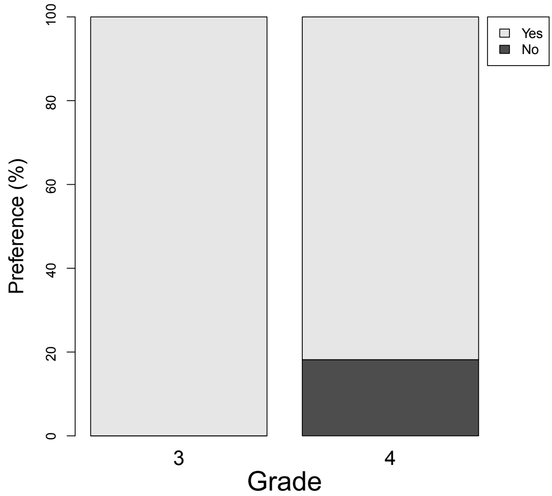


Figure 8.10: The children's vote for their preference about using the game at school when they played the game the first (questionnaire B)

favour of the large group ($U = 355.5$, $Z = 5.197$, $p < 0.001$**, $r = 0.843$). A similar trend in favor of the large group is observed when the usability variable is considered, ($U = 272$, $Z = 2.751$, $p = 0.006$**, $r = 0.446$). These results show that our eighth hypothesis has been corroborated.

8.4 Conclusions

Thanks to our frontal projection system, several children can play with the game at the same time, thereby saving resources and time and encouraging the children to act collaboratively, which links group members together (Johnson et al., 1998) and improves their learning outcomes (Neo, 2003). Moreover, the group size could be quite large (up to twelve). This group size is considerably larger than the group sizes used in previous studies (Sung and Hwang, 2013; Hwang et al., 2013; Cao et al., 2010; Jamil et al., 2011). In Sung and Hwang's work, three or four students work as a team. In Hwang et al.'s work, three students participated in each group. In other works (Cao et al., 2010; Jamil et al., 2011), the students played in groups of 4. In contrast, Rau and Heyl (1990) stated that it is difficult to ensure that all members in large groups participate. However, this study demonstrates that it is possible for all the members in a large group to participate in the learning activity and improve learning outcomes.

From the results, the children learned by playing collaboratively in large groups or in pairs. The statistically significant differences between the children's knowledge before and after playing the game proved that games of this kind are suitable for transmitting knowledge. These results are in line with other studies in this thesis.

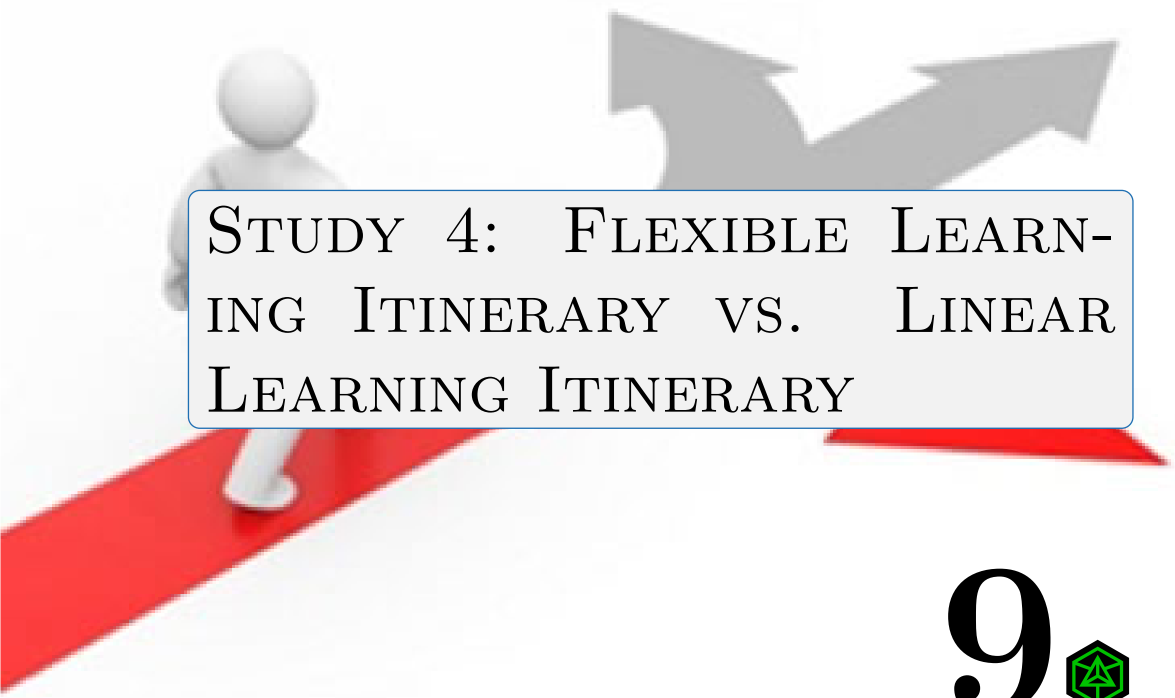
When the results (after receiving the learning methods) for playing in large groups were compared with attending the traditional class, statistically significant differences were found in favour of the game. If the Person Parameter Distribution from the Rasch model is taken into account, it can be observed that the game provides more ability to correctly answer the learning questions. This ability was higher in the group of children who played the game. Therefore, it can be concluded that our game facilitates the learning process when played in large groups, which corroborates our fifth hypothesis. However, no statistically significant differences were found between playing in large groups and playing in pairs, which does not corroborate our sixth hypothesis. Although unexpected, this is a good result because it implies that children can learn by playing in large groups or in pairs without having statistically significant differences in the acquired knowledge. This result offers the possibility of using the game for groups of different sizes without leading statistically significant differences in the acquired knowledge.

The children expressed their interest in incorporating systems of this type for learning different types of content. For the satisfaction between the large group and traditional class, if the means are considered, the means were higher in favour of the group for almost all the questions. However, there were no statistically significant differences. These results do not corroborate our seventh hypothesis. However, if the large group vs. the pairs is considered, there were statistically significant differences for the satisfaction in favour of the large group, corroborating our eighth hypothesis for satisfaction. Moreover, the means for the satisfaction variable were high in all the cases. Therefore, the children were satisfied with our game. Other authors have studied the positive relationship between satisfaction and learning outcomes (e.g., Lee et al. (2011); Shea et al. (2004)). Since the children were satisfied with our game, this satisfaction could positively influence their learning outcomes, results also obtained in other studies of this thesis.

Three of the questions (Table A.4) were related to usability and the means were very high (greater than 4.5 on a scale from 1 to 5 in the children that played in large groups) indicating that the children considered the game to be very easy to use. The children in the pairs group scored lower, existing statistically significant differences in favor of the large group, corroborating our eighth hypothesis for the usability. However, the means for the Pairs group were also quite high (almost 4 on a scale from 1 to 5). In addition, the people observing the participants during the game stated that a great majority of users did not have any problems interacting with the system. All of these arguments suggest that our game does help students focus their attention on the learning content, especially when they are playing in large groups.

#	PostPair	$\mu_{PP} \pm \sigma_{PP}$	PostLGroup B	$\mu_B \pm \sigma_B$	U	Z	p	r
Q14	5;0	4.78 ± 0.42	5;0	4.90 ± 0.44	151	-1.445	0.156	0.234
Q15	5;1	4.44 ± 0.68	5:0.25	4.75 ± 0.43	140	-1.406	0.165	0.228
Q16	4;1	3.89 ± 0.74	5;1	4.60 ± 0.58	89	-2.878	0.004**	0.467
Q18	10;1	9.28 ± 1.04	10;0	10.0 ± 0.00	104.5	-2.959	0.003**	0.486
Q19	4;1.75	3.89 ± 1.41	5;0	4.90 ± 0.44	90.5	-3.275	0.001**	0.531
Q20	4.5;1	4.17 ± 1.21	5;0.25	4.70 ± 0.56	131.5	-1.660	0.100	0.269
Q21	4.5;1	4.11 ± 1.24	5;1	4.55 ± 0.74	144	-1.203	0.236	0.195
Q22	5;1	4.50 ± 0.83	5;0	4.85 ± 0.36	144	-1.420	0.162	0.230
Q23	4;2.5	3.50 ± 1.34	5;0	4.95 ± 0.22	55.5	-4.227	<0.001**	0.686
Q24	4;1.75	4.00 ± 1.00	5;0	4.80 ± 0.51	94	-2.927	0.004**	0.475

Table 8.9: Medians, Interquartile range, Means and Standard deviations of the PostPair and PostLGroup B questionnaires, the Mann-Whitney U test analysis, and r effect size



Study 4: Flexible Learning Itinerary vs. Linear Learning Itinerary

9

"The measure of who we are is what we do with what we have."

— Vince Lombardi

9.1 Introduction

In this chapter a study that compares two different learning itineraries (Linear Learning and Flexible Learning) was carried out. The elaboration theory (Reigeluth, 1992) justifies the importance of sequencing content and teaching-learning activities in two fundamental analyses: the first is the representation of the organizing content, and, the second, the different development levels for structuring the learning sequence (Reigeluth, 1999). From this definition of learning itinerary, a first classification based on the freedom provided to the users leads to two different itineraries: Linear Learning Itineraries and Flexible Learning Itineraries. LLI are normally guided itineraries where the path is determined completely by the system. These types of itineraries can be seen as a collection of small pieces of information that are sequentially interconnected with each other as shown in Figure 9.1a. On the other hand, FLI offers users the chance to choose what the next step is in their learning process by changing the path, which is determined completely by the user. These itineraries may have different degrees of freedom depending on how they are designed. They can be viewed as a collection of small information units that are interconnected as a directed graph, which is shown in Figure 9.1b. In this chapter, we present two different versions of the game explained in section 3.2, one that uses a LLI and another that uses a FLI. For this study, the frontal projection system was used.

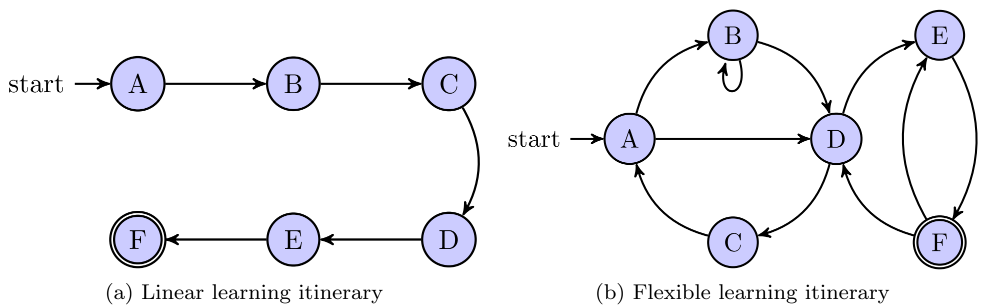


Figure 9.1: Different learning itineraries

The main differences of the FLI are the following:

- When playing, the children can choose the next historical age they want to learn.
- Each child decides the itinerary that he/she wishes to follow.
- The children can repeat the historical ages as many times as they want.
- If they do not want to learn about a historical age, they can choose to ignore it.

9.2 Design of the evaluations

The frontal projection system described in Chapter 5 was played by 29 children. This section explains in detail the participants, the measurements and the procedure carried out during the evaluations.

9.2.1 Description of the game

Two modes were defined for the game. In the first mode, the children could choose the order of the historical ages; they could repeat the historical ages twice or they could even omit some of them (Flexible Learning Itinerary). In the second mode, the children played the game in the real chronological order, from Prehistory to the Contemporary Period (Linear Learning Itinerary). The flowchart of the mini-games that made up the LLI is shown in Figure 3.3.

9.2.2 Participants

A total of 29 children participated in this study. There were 19 boys (65.52%) and 10 girls (34.48%). They were between seven and ten years old, and they had already finished their third grade academic course of primary education. The mean age was 8.62 ± 0.49 years old. The children were students from three different summer schools in Valencia.

9.2.3 Measurements

Two questionnaires in a web-based form were designed and used to determine the knowledge that the children acquired during the game (Table A.1) as well as their satisfaction and level of interaction (Table A.2). The knowledge questionnaire consisted of thirteen questions about the contents that the children learn while playing with the game. By comparing the answers given in the pretest by the children in the pretest with the answers given in the post-test, we were able to determine whether or not there had been an increase in knowledge. The satisfaction and interaction questionnaire consisted of eleven questions. The questionnaires were filled out individually for both playing modes (LLI and FLI).

9.2.4 Procedure

The games were played in pairs of one boy and one girl, two boys or two girls. The participants were assigned to one of the following two groups:

- Group A: Participants that played in pairs (collaborative mode). These pairs were made by one boy and one girl, two boys, or two girls.
- Group B: Participants that played solo (individual mode) with no more company than the monitor. The children of this group had to select the options that appeared in the two sides of the screen by using their two hands.

Figure 9.2 shows the procedure of the two groups graphically. The protocol used is given as follows:

1. A pair of children from Group A or a pair of children from Group B filled out the pre-test questionnaire (PreLinear for Group A and PreFlexible for Group B).
2. These children played the game.
3. They filled out the post-test questionnaire on-line (PosLinear for Group A and PosFlexible for Group B).

Since the questionnaires were filled out on-line, the answers were automatically stored in a remote database.

9.3 Results

As in the previous studies, the data were analyzed using the statistical open source toolkit *R* with the *RStudio* IDE.

9.3.1 Learning outcomes

To measure how much the children learned, the knowledge variable was analyzed. This was done by analyzing the answers to questions Q1 to Q13 in Table A.1 before playing (Pre-test) and after playing (Post-test). Several t-tests were performed to determine if there were statistically significant differences in the knowledge acquired. Figure 9.3 shows the box plot for the scores before and after playing. A high predominance of correct answers after playing the first game (PosLinear and PosFlexible) over the pre-test (PreLinear and PreFlexible) can be observed.

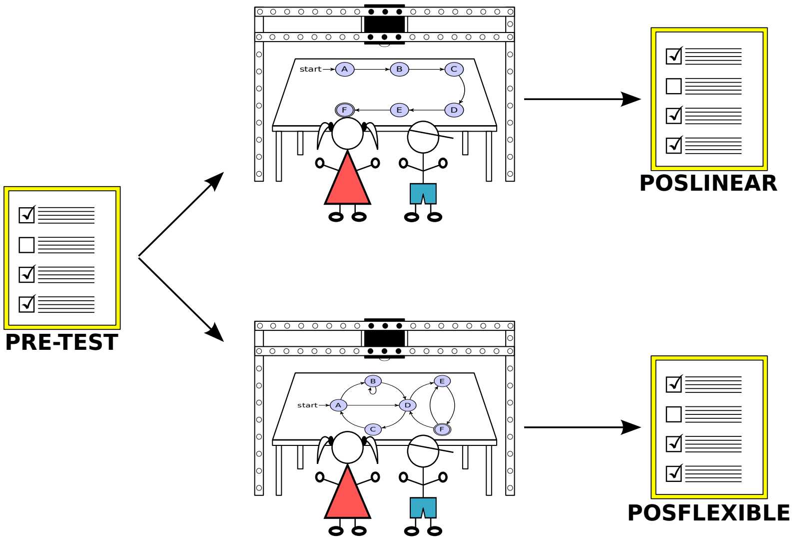


Figure 9.2: Study procedure

All t-tests are shown in the format: *(statistic [degrees of freedom], p-value, Cohen's d)*, and ** indicates some statistical significance at level $\alpha = 0.05$. First, to determine if there was a statistically significant difference between the initial knowledge in both pre-tests, an unpaired t-test was performed between PreLinear (3.61 $\pm$ 1.89) and PreFlexible (2.27 $\pm$ 1.21) ($t[27] = 2.03$, $p = 0.053$, Cohen's $d = 0.78$). No statistically significant differences were found. From a paired t-test, the scores of the knowledge variable between PreLinear (3.61 $\pm$ 1.89) and PosLinear (7.44 $\pm$ 3.22) showed statistically significant differences ($t[17] = -6.21$, $p < 0.001$**, Cohen's $d = -1.46$). Another paired t-test between the PreFlexible (2.27 $\pm$ 1.21) and the PosFlexible (6.18 $\pm$ 3.46) questionnaires revealed statistically significant differences ($t[10] = -3.67$, $p = 0.004$**, Cohen's $d = -1.11$). Finally, in order to determine whether or not there was a statistically significant difference between the acquired knowledge in the two groups, another unpaired t-test was performed between the knowledge in PosLinear (7.44 $\pm$ 3.22) and the knowledge in PosFlexible (6.18 $\pm$ 3.46) ($t[27] = 0.96$, $p = 0.345$, Cohen's $d = 0.37$). This showed that the knowledge retrieved while playing the linear mode was not significantly greater than the knowledge acquired while playing the flexible mode. To complete the analysis and to determine the questions that had statistically significant differences, several t-tests were per-

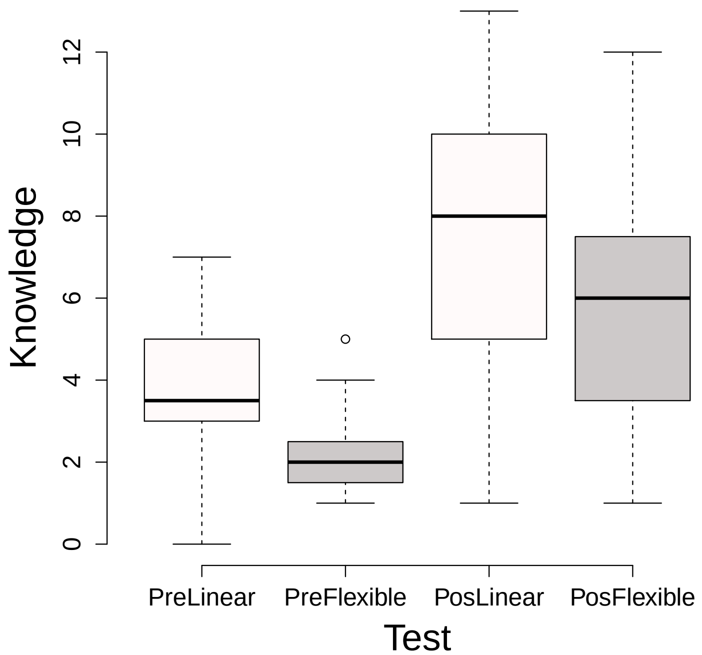


Figure 9.3: Scores of the knowledge variable in questionnaires before and after playing with the linear or flexible itineraries

formed for each question between PreLinear – PosLinear (Table 9.1), PreFlexible – PosFlexible (Table 9.2), and PosLinear – PosFlexible (Table 9.3). Table 9.1 shows that children who played the linear mode acquired significantly more knowledge in questions Q1, Q3, Q5, Q6, Q7, Q11, Q12, and Q13. This can be compared with the results in Table 9.2 provided by children who played the flexible mode. In this case, questions Q1, Q3, Q5, Q7, and Q12 were the ones with significant differences. Table 9.3 shows that none of the questions had statistically significant differences when the two modes were compared.

A multifactorial ANOVA test was also performed to take into account several factors simultaneously. The factors of gender, age, and game mode (linear or flexible) were between subjects. The effect size used was the partial eta-squared (η^2). This analysis is shown in Table 9.4, where the results show that there were statistically significant differences in the interactions of the gender and age factors, and the in-

#	PreLinear	PosLinear	t	p	Cohen's d
Q1	0.22 ± 0.42	0.56 ± 0.50	-2.92	0.010**	-0.69
Q2	0.28 ± 0.45	0.17 ± 0.37	1.00	0.331	0.24
Q3	0.17 ± 0.37	0.72 ± 0.45	-4.61	<0.001**	-1.09
Q4	0.22 ± 0.42	0.33 ± 0.47	-0.70	0.495	-0.16
Q5	0.28 ± 0.45	0.67 ± 0.47	-3.29	0.004**	-0.78
Q6	0.17 ± 0.37	0.44 ± 0.50	-2.56	0.020**	-0.60
Q7	0.06 ± 0.23	0.44 ± 0.50	-2.72	0.015**	-0.64
Q8	0.78 ± 0.42	0.61 ± 0.49	1.37	0.187	0.32
Q9	0.72 ± 0.45	0.83 ± 0.37	-1.00	0.331	-0.24
Q10	0.56 ± 0.50	0.78 ± 0.42	-1.17	0.260	-0.27
Q11	0.00 ± 0.00	0.67 ± 0.47	-5.83	<0.001**	-1.37
Q12	0.11 ± 0.31	0.67 ± 0.47	-4.61	<0.001**	-1.09
Q13	0.06 ± 0.23	0.56 ± 0.50	-4.12	<0.001**	-0.97

Table 9.1: Means and standard deviations of questions for PreLinear and PosLinear, t-test analysis, and Cohen's d, d.f. = 17

#	PreFlexible	PosFlexible	t	p	Cohen's d
Q1	0.09 ± 0.29	0.73 ± 0.45	-4.18	0.002**	-1.26
Q2	0.00 ± 0.00	0.09 ± 0.29	-1.00	0.341	-0.30
Q3	0.09 ± 0.29	0.64 ± 0.48	-3.46	0.006**	-1.04
Q4	0.09 ± 0.29	0.27 ± 0.45	-1.00	0.341	-0.30
Q5	0.00 ± 0.00	0.55 ± 0.50	-3.46	0.006**	-1.04
Q6	0.09 ± 0.29	0.18 ± 0.39	-0.56	0.588	-0.17
Q7	0.00 ± 0.00	0.45 ± 0.50	-2.89	0.016**	-0.87
Q8	0.82 ± 0.39	0.64 ± 0.48	1.00	0.341	0.30
Q9	0.64 ± 0.48	0.64 ± 0.48	0.00	1.000	0.00
Q10	0.36 ± 0.48	0.45 ± 0.50	-0.56	0.588	-0.17
Q11	0.09 ± 0.29	0.45 ± 0.50	-1.79	0.104	-0.54
Q12	0.00 ± 0.00	0.82 ± 0.39	-6.71	<0.001**	-2.02
Q13	0.00 ± 0.00	0.27 ± 0.45	-1.94	0.082	-0.58

Table 9.2: Means and standard deviations of questions for PreFlexible and PosFlexible questionnaires, t-test analysis, and Cohen's d, d.f. = 10

#	PosLinear	PosFlexible	t	p	Cohen's d
Q1	0.56 ± 0.50	0.73 ± 0.45	-0.91	0.373	-0.35
Q2	0.17 ± 0.37	0.09 ± 0.29	0.56	0.582	0.21
Q3	0.72 ± 0.45	0.64 ± 0.48	0.47	0.642	0.18
Q4	0.33 ± 0.47	0.27 ± 0.45	0.33	0.743	0.13
Q5	0.67 ± 0.47	0.55 ± 0.50	0.63	0.531	0.24
Q6	0.44 ± 0.50	0.18 ± 0.39	1.45	0.160	0.55
Q7	0.44 ± 0.50	0.45 ± 0.50	-0.05	0.960	-0.02
Q8	0.61 ± 0.49	0.64 ± 0.48	-0.13	0.897	-0.05
Q9	0.83 ± 0.37	0.64 ± 0.48	1.19	0.244	0.46
Q10	0.78 ± 0.42	0.45 ± 0.50	1.82	0.080	0.70
Q11	0.67 ± 0.47	0.45 ± 0.50	1.11	0.277	0.42
Q12	0.67 ± 0.47	0.82 ± 0.39	-0.87	0.394	-0.33
Q13	0.56 ± 0.50	0.27 ± 0.45	1.49	0.147	0.57

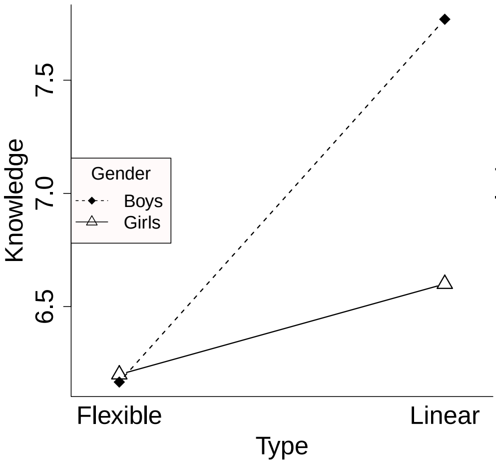
Table 9.3: Means and standard deviations of questions for PosLinear and PosFlexible questionnaires, t-test analysis, and Cohen's d, d.f. $= 27$

teraction of the mode and age factors, with p-values less than 0.07. The effect sizes revealed that the most influential factor was the gender and age interaction with a very large effect size, followed by the mode and age interaction which had a medium effect size.

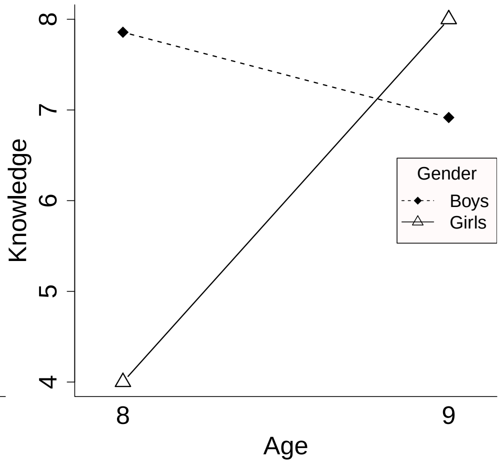
Factor	d.f.	F	p	partial η^2
Gender	1	0.03	0.858	0.001
Age	1	0.10	0.748	0.004
Mode (Linear/Flexible)	1	1.10	0.249	0.062
Gender:Age	1	5.93	0.023**	0.220
Mode:Age	1	3.40	0.079**	0.139
Other interactions	1	<0.90	>0.352	<0.042

Table 9.4: Multifactorial ANOVA for the knowledge variable, $N = 46$

Figure 9.4a shows the interaction plot between gender and the two modes of playing, indicating that boys acquired more knowledge than girls playing the linear mode and the same knowledge playing the flexible mode. Figure 9.4b shows the interaction plot between gender and age, indicating that 8-year-old boys acquired more knowledge than girls, but 9-year-old boys acquired less.



(a) By game mode



(b) By age

Figure 9.4: Interaction plots for the acquired knowledge

9.3.2 Satisfaction and interaction outcomes

To determine the satisfaction outcomes, the children were asked the questions Q14 to Q24 that are shown in Table A.3. The children's satisfaction was quite favorable, as the results in Table 9.5 show. Most of the children stated that they had a very good time playing the game (Q14). In addition, the children gave an average score of 9.14±1.83 when asked to score the game (Q23). We asked the children to rate the difficulty of the game from one to five (Q16), with one being great difficulty and five being the easiest. Of the 29 children who played the game, eleven children said that the game was easy by giving a score of four points, eight children said that the game was of normal difficulty (3 points), six children chose very easy as their option (5 points), three children chose "hard" (2 points), and only one child chose "very hard" (1 point). The average obtained with this data was 3.62±1.05 points. Therefore, this result indicates that the children's opinion was that the game was "easy". The children also liked the avatar. They gave it an average score of 3.90±1.29 points out of five (Q20). They also thought that the avatar was helpful during the game. They gave a score 4.07±1.07 points out of five (Q21). The rest of the results are shown in Table 9.5.

#	Bounds	Mean	Answer
Q14	[1-5]	4.83±0.37	Very good
Q15	[1-5]	4.44±0.97	To several friends
Q16	[1-5]	3.62±1.05	Easy
Q17	[1-5]	4.39±0.69	Good
Q18	[1-5]	4.22±0.85	Easy
Q19	[1-5]	4.52±0.74	A lot
Q20	[1-5]	3.90±1.29	Enough
Q21	[1-5]	4.07±1.07	A lot
Q22	[1-5]	4.34±0.94	A lot
Q23	[1-10]	9.14±1.83	Very good

Table 9.5: Satisfaction and interaction question results

To determine if there were statistically significant differences between the two game modes with respect to the satisfaction questions, an unpaired *t*-test was performed. Table 9.6 shows the results of the test. Statistically significant differences were only found for Q15 (where the children were asked about recommending the game to friends) and for Q20 (where the children were asked about liking the avatar character). For the two questions, the children who played the Flexible Learning Itinerary scored these questions higher.

To analyze the satisfaction variable, a multifactorial ANOVA test was performed. Table 9.7 shows the results. Statistically significant differences were found in the interaction between game mode and age factors, and in the interaction among gender, game mode, and age factors.

For Q24, which asked to children about the mini-games they liked the most, a vote count was performed. The children could vote for more than one mini-game. When the results between the children who played the FLI mode and the children who played the LLI mode are compared, some differences can be observed, as shown in the bar plots in Figure 9.5. Children from FLI preferred Prehistory, the Middle Ages, and the Contemporary Period the most (7 votes). However, children from LLI preferred the Middle Ages the most (12 votes). When all the votes are considered, the most preferred age was the Middle Ages (19 votes). Therefore, the three plots

#	Linear	Flexible	t	d.f.	p	Cohen's d
Q14	4.78 ± 0.42	4.91 ± 0.29	-0.89	27	0.382	-0.34
Q15	4.12 ± 1.08	5.00 ± 0.00	-2.49	25	0.020**	-0.99
Q16	3.89 ± 0.74	3.18 ± 1.27	1.83	27	0.078	0.70
Q17	4.41 ± 0.69	4.36 ± 0.64	0.18	26	0.860	0.07
Q18	4.35 ± 0.76	4.00 ± 0.89	1.05	25	0.305	0.42
Q19	4.50 ± 0.76	4.55 ± 0.50	-0.16	27	0.876	-0.06
Q20	3.50 ± 1.34	4.55 ± 0.78	-2.27	27	0.032**	-0.87
Q21	4.00 ± 1.00	4.18 ± 1.11	-0.44	27	0.664	-0.17
Q22	4.44 ± 0.68	4.18 ± 1.19	0.73	27	0.474	0.28
Q23	9.28 ± 1.04	8.91 ± 2.57	0.52	27	0.607	0.20

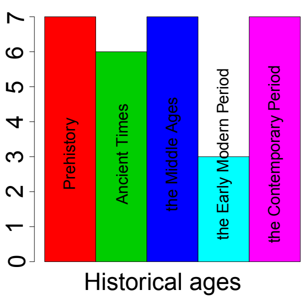
Table 9.6: Means and standard deviations for satisfaction questions, t-test analysis, and Cohen's d

Factor	d.f.	F	p	partial η^2
Gender	1	2.86	0.105	0.120
Age	1	0.23	0.629	0.011
Mode (Linear/Flexible)	1	0.08	0.780	0.003
Gender:Age	1	0.82	0.375	0.037
Mode:Age	1	3.62	0.070**	0.147
Gender:Mode	1	0.26	0.610	0.012
Gender:Mode:Age	1	3.28	0.084**	0.135

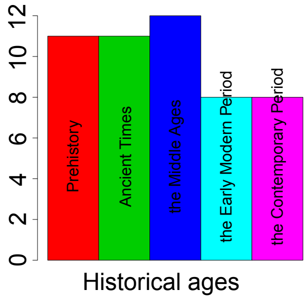
Table 9.7: Multifactorial ANOVA for the *satisfaction* variable, $N = 29$

show that the Middle Ages was the favorite.

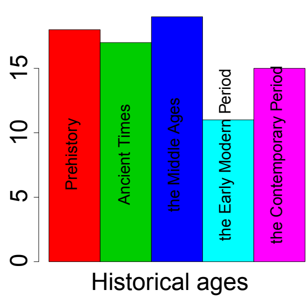
A Chi-squared test was performed to analyze the children's preferred mini-game. Table 9.8 shows the results of the test. These results show that there were no statistically significant differences between the two learning itineraries. The Prehistory, the Ancient Times, and the Middle Ages mini-games were the most voted in the two groups (LLI and FLI) for the preferred mini-game.



(a) FLI



(b) LLI



(c) Both learning itineraries taken into account

Figure 9.5: Comparative frequencies of the mini-games the children liked the most. The votes are shown on the y-axis (ordinate)

Age	Linear	Flexible	χ^2	p	V
Prehistory	1	1	0.00	1.000	0.03
Ancient Times	1	1	0.00	1.000	0.06
The Middle Ages	1	1	0.00	1.000	0.03
The Early Modern Period	0	0	0.28	0.596	0.17
The Contemporary Period	0	1	0.39	0.535	0.19

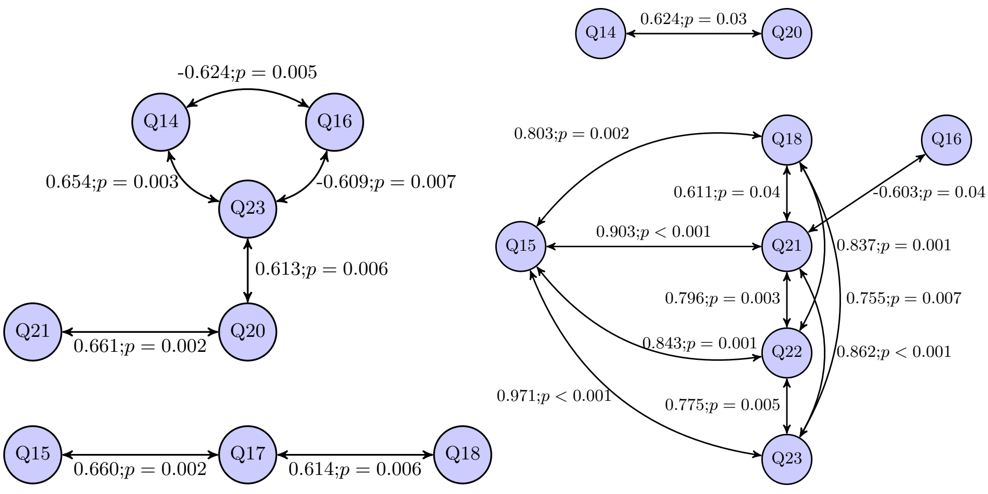
Table 9.8: Modes of the preferred mini-game, Chi-squared analysis, and Cramer's V, d.f. $= 1$, $N = 29$

9.3.3 Correlation analysis

A correlation analysis was performed to determine if there was a correlation among any of the questions. Some correlations were found when the two groups were analyzed together. There were correlations found between Q15 (Recommend the game to friends) and Q23 (Score the game) (0.639, $p < 0.001$), between Q18 (Difficulty of selecting the answers) and Q22 (How much the children learned) (0.691, $p < 0.001$), between Q21 (Did the avatar help) and Q22 (0.654, $p < 0.001$), between Q21 and Q23 (0.673, $p < 0.001$), and between Q22 and Q23 (0.681, $p < 0.001$). When the groups were analyzed separately, we found different correlations (Figure 9.6a Figure 9.6b).

When groups formed only by boys and groups formed only by girls were analyzed separately, there were different correlation results. In the group of boys, the correlations found were between Q14 and Q20 (0.614, $p = 0.005$), between Q15 and Q18 (0.648, $p = 0.002$), between Q15 and Q23 (0.664, $p = 0.001$), between Q18 and Q22 (0.809, $p < 0.001$), between Q21 and Q22 (0.714, $p < 0.001$), between Q21 and Q23 (0.737, $p < 0.001$), and between Q22 and Q23 (0.722, $p < 0.001$). In the group of girls, fewer correlations were found. These correlations were between Q17 and Q19 (-0.699,$p = 0.024$), and between Q17 and Q23 (0.620, $p = 0.05$).

When groups organized by age were analyzed separately, in the group of 8-year-old children, the correlations were between Q15 and Q20 (0.703, $p = 0.015$), between Q17 and Q18 (0.600, $p = 0.05$), and between Q20 and Q21 (0.679, $p = 0.02$). In the group of 9-year-old children, there were correlations between Q15 and Q23 (0.683, $p = 0.001$), between Q18 and Q22 (0.754, $p < 0.001$), between Q21 and Q22 (0.719, $p < 0.001$), between Q21 and Q23 (0.819, $p < 0.001$), and between Q22 and Q23 (0.755, $p < 0.001$).



(a) Linear Learning Itinerary correlations (b) Flexible Learning Itinerary correlations

Figure 9.6: Correlations among questions

9.4 Conclusions

From the results, it can be observed that the children learned while playing with the two modes (LLI and FLI), with statistically significant differences between the levels of knowledge before and after playing the game. This implies that the game is suitable for the transmission of knowledge. Result that is in line with the results obtained in the other studies of this thesis.

However, there were no statistically significant differences between playing with the flexible mode or the linear mode. In our ninth hypothesis, we predicted that the children would acquire more knowledge by playing with the Flexible learning Itinerary because they could decide how to learn the different historical ages. These results did not corroborate this hypothesis. Although unexpected, this is a good result because it implies that children can learn using the two modes without having statistically significant differences in the acquired knowledge. However, we still believe that the flexible mode is an alternative mode that some individuals might prefer over the linear mode, especially when other groups are considered such as young people or adults. Another factor that we also think has influenced this result is the limited time of use by each child and the pressure of participating in a study. Thus, we consider that a study with more freedom would benefit the flexible itinerary.

The analysis also revealed that boys improved their knowledge quite a bit more than girls, mainly while playing with the linear learning itinerary. From our point of view, and in line with previous studies (e.g., Bonanno and Kommers, 2005), this greater improvement in boys could be related to the fact that more boys play video games than girls, and for longer periods of time. For example, a study carried out by Bonanno and Kommers (2005) revealed that male junior college students were far more likely than females to play video games, and males also played for longer periods of time than females (more than double). However, more studies should be carried out to verify this trend.

In the analysis for satisfaction and interaction, there were only statistically significant differences for Q15 (recommending the game to friends) and for Q20 (liking the avatar) between the two modes. This result indicates that there were no statistically significant differences for the rest of the questions. If these questions are considered, the children perceived the game as easy to use. This is also a good result because one of the recommendations for a good interface is to be easy to use. As mentioned in the individual vs. collaborative study, several authors have considered usability or perceived ease of use as an important technical factor that affects educational effectiveness (Jones et al., 1999; Mayes and Fowler, 1999). Also, Sun et al. (2008) pointed out that learning systems that are easy to use help students to focus their attention on the learning content, and they are more motivated to learn. Therefore, according to the above-mentioned suggestions, our game in the two modes does help students focus their attention on the learning content.

With regard to the most preferred mini-games, the Middle Ages with the medieval castle was the most highly rated historical age. Prehistory was the second most highly rated. The least preferred mini-game was the Early Modern Period. Our explanation for this result is that the Middle Ages and Prehistory were the two periods that had the most animations, and the children had to interact with them. In addition, the rotating 3D model of the medieval castle greatly attracted the children's attention. The Early Modern Period and the Contemporary Period were the two periods in which the contents were mainly transmitted using videos and with little interaction. The children scored the Contemporary Period higher because it is the period that they are most familiar with. From these results, our recommendation is to transmit the content using more interactive material and using fewer videos.

Study 5: Text-only vs. Real Images vs. Fake Images

10

"You cannot have a positive life and a negative mind."

— Joyce Meyer

10.1 Introduction

In this chapter, we have focused on online formative assessment and multiple-choice questions. In this type of questionnaire, there is usually a question and several possible answers in which the student must select only one answer. It is very common for the answers to be just text. However, images could also be used. In this chapter, we have carried out a study to determine if an added image that represents/defines an object helps the children to choose the correct answer. The tenth hypothesis was that there would be significant differences between using only a text-only questionnaire and a questionnaire that, apart from the text also includes images. The eleventh hypothesis was that there would be significant differences between a questionnaire with images used during the learning process and a questionnaire with images that represent the item but that were not used during the learning process.

10.2 Design of the evaluations

The games described in Chapters 4 and 5 were played by a total of 94 children. This section presents the characteristics of the children that played the game, the measurements used during the experiment, and the steps followed.

10.2.1 Participants

A total of 94 children participated in this study. There were 46 boys (48.94%) and 48 girls (51.06%). They were between seven and eight years old, and they had already finished their second academic course of primary school. The mean age was 7.56 ± 0.50 years old. The children were students from three different summer schools in Valencia.

10.2.2 Measurements

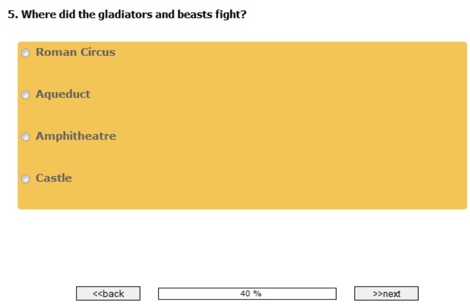
To retrieve the data for the analysis, we used three different web-based questionnaires:

1. A text-only questionnaire where all the questions were written in text-only and there were no images on it (Figure 10.1a).

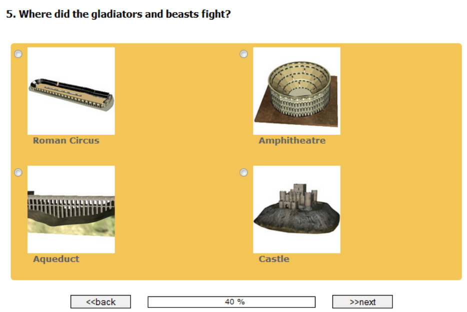
2. A questionnaire where all the questions had images taken from the game that was played. We refer to the images that appear in the game as real images (Figure 10.1b). The text was also included.

3. A questionnaire (similar to the previous one), where all the questions had images that did not appear in the game that was played but were representative images of the item specified in the text. We refer to these images as fake images (Figure 10.1c). The text was also included.

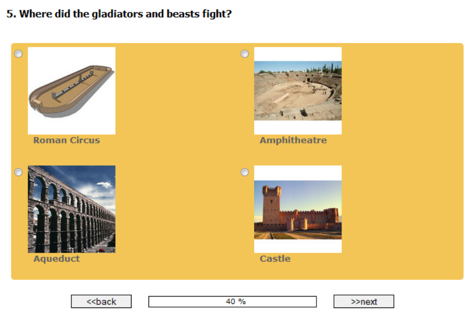
All three questionnaires contained thirteen knowledge questions about the contents of the game, shown in Table A.1. A pre-test and a post-test of these three questionnaires were used to carry out the study. We refer to the pre-tests as PreText, PreReal and PreFake; and we refer to the post-tests as PosText, PosReal and PosFake.



(a) Screenshot of Q5 of the text-only questionnaire



(b) Screenshot of Q5 of the questionnaire with real images



(c) Screenshot of Q5 of the questionnaire with fake images

Figure 10.1: Questionnaire screenshots of Q5

10.2.3 Procedure

The participants were assigned to one of the following three groups:

- Group A: The participants who filled out the text-only questionnaires before and after playing the game. There were 36 participants in this group (38.30%).

- Group B: The participants who filled out the questionnaires with real images before and after playing the game. There were 29 participants in this group (30.85%).

- Group C: The participants who filled out the questionnaires with fake images before and after playing the game. There were 29 participants in this group (30.85%).

Figure 10.2 shows graphically the procedure for the three groups. Since all the questionnaires were filled out online, the answers were automatically stored in a remote database. The questionnaires were filled out individually. Figure 10.3 shows a child filling out the text-only questionnaire.

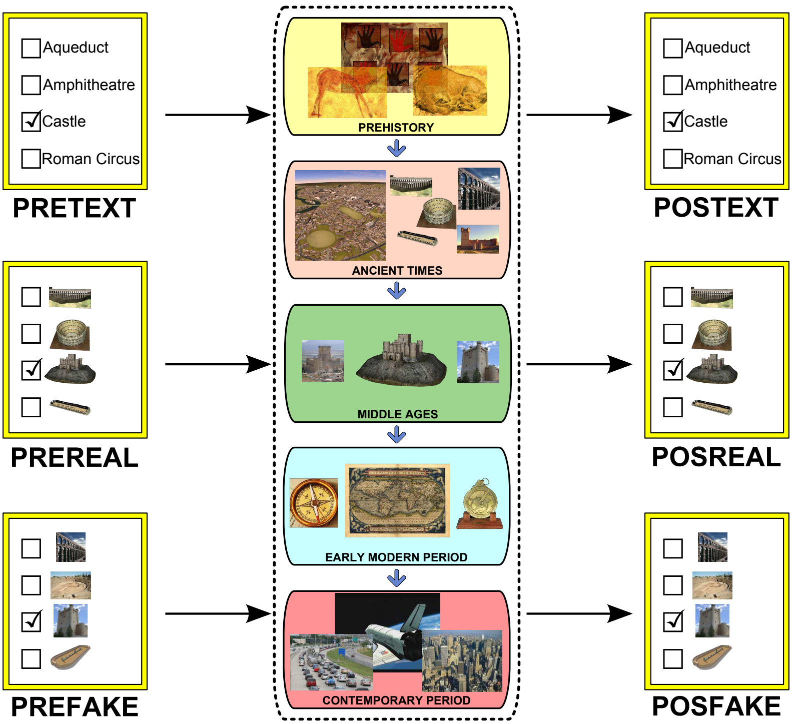


Figure 10.2: Study procedure

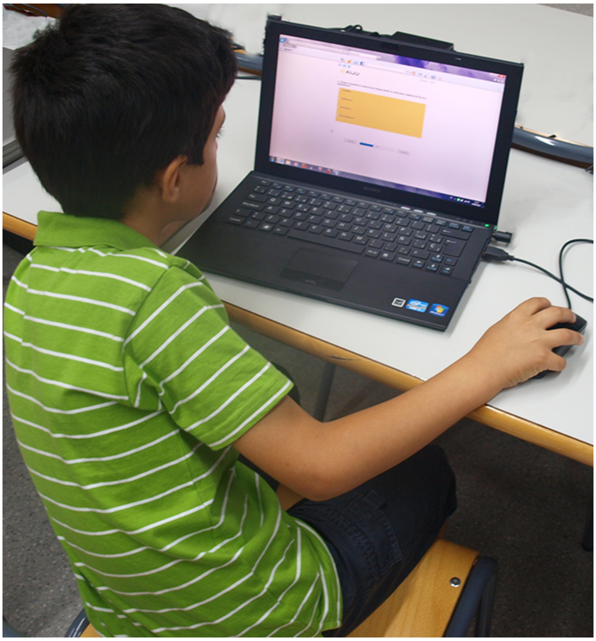


Figure 10.3: Child filling out the text-only questionnaire

10.3 Results

As in the previous studies, the data from this study were analyzed using the statistical open source toolkit *R* with the *RStudio* IDE.

10.3.1 Learning outcomes

To measure how much the children learned, the knowledge variable was analyzed. This was achieved by analyzing the answers to questions Q1 to Q13 in Table A.1 before playing (pre-test) and after playing (post-test). The knowledge value was obtained by summing up all the correct answers. Several t-tests were performed to determine if there were statistically significant differences in the knowledge acquired. Figure 10.4 shows the box plot for the scores before and after playing the game. As can be observed, there was a high dominance of correct answers after playing the game and using the two questionnaires that had images.

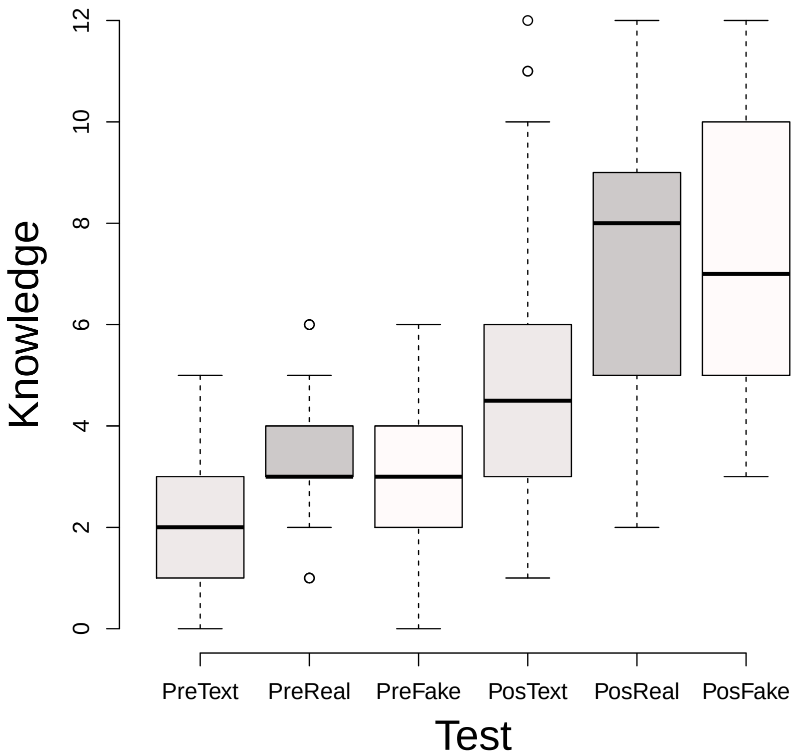


Figure 10.4: Scores of the knowledge variable before and after playing for each type of questionnaire used

All t-tests are shown in the format: (*statistic [degrees of freedom], p-value, Cohen's d*), and ** indicates the statistical significance at level $\alpha = 0.05$. First, to determine whether or not there were statistically significant differences between the initial knowledge in all types of pre-tests, some unpaired t-tests were performed. Statistically significant differences were found between PreText (2.20 ± 1.50) and PreReal (3.90 ± 1.90) ($t[63] = -3.94$, $p < 0.001$**, Cohen's $d = 0.98$); no statistically significant differences were found between PreFake (3.30 ± 1.60) and PreReal (3.90 ± 1.90) ($t[56] = -1.28$, $p = 0.20$, Cohen's $d = 0.34$). Finally, another unpaired t-test between PreFake (3.30 ± 1.60) and PreText (2.20 ± 1.50) ($t[63] = 2.74$,

$p = 0.008^{**}$, Cohen's $d = 0.68$) was performed, where statistically significant differences were found. This proved that children got a better score on the pre-test if it had images (Figure 10.4). In order to measure the knowledge acquired using each type of questionnaire, several t-tests were performed to compare each pre-test with its post-test. From a paired t-test, the scores of the knowledge variable between PreText (2.20 ± 1.50) and PosText (5.10 ± 2.90) showed statistically significant differences ($t[35] = -7.52$, $p < 0.001^{**}$, Cohen's $d = 1.25$). Another paired t-test between the PreReal (3.90 ± 1.90) and the PosReal (7.40 ± 2.80) questionnaires revealed statistically significant differences ($t[28] = -5.85$, $p < 0.001^{**}$, Cohen's $d = 1.09$). The last comparison between pre-test and post-test was performed between PreFake (3.30 ± 1.60) and PosFake (7.40 ± 2.70) with the results also showing statistically significant differences ($t[28] = -8.07$, $p < 0.001^{**}$, Cohen's $d = 1.50$). These results indicate that regardless of the method used for the assessment, the children acquired knowledge using the game. Finally, in order to determine whether or not there were statistically significant differences between the acquired knowledge in the three groups, further unpaired t-tests were performed between the knowledge in PosText (5.10 ± 2.90) and the knowledge in PosReal (7.40 ± 2.80) ($t[63] = -3.36$, $p = 0.001^{**}$, Cohen's $d = 0.84$) showing that the appearance of the real image helps in choosing the correct answer. When performing this same test using the questionnaire with fake images (7.40 ± 2.70), similar results were obtained ($t[63] = 3.35$, $p = 0.001^{**}$, Cohen's $d = 0.84$). These results showed statistically significant differences. When comparing the two questionnaires that had images, PosFake (7.41 ± 2.65) and PosReal (7.45 ± 2.71), the results showed that there were no statistically significant differences ($t[56] = -0.05$, $p = 0.962$, Cohen's $d = 0.01$). To complete the analysis and check the questions where there were statistically significant differences the following tests were performed. Since the value of the questions were dichotomous (`0`, wrong / `1`, right), several non-parametric McNemar's tests for paired data were performed for each question between PreText – PosText (Table 10.1), PreReal – PosReal (Table 10.2), and PosFake – PosReal (Table 10.3). Table 10.1 shows that the children who filled out the text-only questionnaire acquired more knowledge in seven questions. This can be compared with the results in Table 10.2 provided by the children who filled out the questionnaire with real images of the game. In this case, statistically significant differences were also obtained in seven questions, six of them the same as in the first case. For the children who filled out the questionnaire with fake images, Table 10.3 shows that there were nine questions with statistically significant differences (including the same six questions as in the previous analyses).

In order to compare the acquired knowledge for each question after playing the game, the results between the two post-tests with images were compared with several

#	PreText	PosText	χ^2	p	ϕ
Q1	0.05	0.50	14.06	<0.001**	0.62
Q2	0.08	0.31	6.12	0.013**	0.41
Q3	0.08	0.61	17.05	<0.001**	0.69
Q4	0.46	0.33	0.27	0.606	0.09
Q5	0.22	0.33	0.75	0.386	0.14
Q6	0.17	0.22	0.17	0.683	0.07
Q7	0.05	0.22	2.50	0.114	0.26
Q8	0.53	0.64	0.75	0.386	0.14
Q9	0.47	0.56	0.44	0.505	0.11
Q10	0.17	0.44	6.75	0.009**	0.43
Q11	0.08	0.36	8.10	0.004**	0.47
Q12	0.05	0.39	8.64	0.003**	0.49
Q13	0.00	0.17	4.17	0.041**	0.34

Table 10.1: Proportions for questions of the PreText and PosText questionnaires, McNemar's test analysis, and ϕ effect size, $N = 36$

#	PreReal	PosReal	χ^2	p	ϕ
Q1	0.34	0.72	7.69	<0.006**	0.52
Q2	0.59	0.38	3.12	0.077	0.33
Q3	0.17	0.72	14.06	<0.001**	0.70
Q4	0.17	0.34	1.23	0.267	0.21
Q5	0.55	0.72	1.45	0.228	0.22
Q6	0.10	0.17	0.17	0.683	0.08
Q7	0.34	0.69	5.79	0.016**	0.45
Q8	0.62	0.79	1.45	0.228	0.22
Q9	0.48	0.59	0.57	0.450	0.14
Q10	0.21	0.66	7.58	0.006**	0.51
Q11	0.14	0.66	10.32	0.001**	0.60
Q12	0.06	0.62	12.50	<0.001**	0.66
Q13	0.06	0.38	5.82	0.016**	0.45

Table 10.2: Proportions for questions of the PreReal and PosReal questionnaires, McNemar's test analysis, and ϕ effect size, $N = 28$

#	PreReal	PosReal	χ^2	p	ϕ
Q1	0.34	0.72	7.69	<0.006**	0.52
Q2	0.59	0.38	3.12	0.077	0.33
Q3	0.17	0.72	14.06	<0.001**	0.70
Q4	0.17	0.34	1.23	0.267	0.21
Q5	0.55	0.72	1.45	0.228	0.22
Q6	0.10	0.17	0.17	0.683	0.08
Q7	0.34	0.69	5.79	0.016**	0.45
Q8	0.62	0.79	1.45	0.228	0.22
Q9	0.48	0.59	0.57	0.450	0.14
Q10	0.21	0.66	7.58	0.006**	0.51
Q11	0.14	0.66	10.32	0.001**	0.60
Q12	0.06	0.62	12.50	<0.001**	0.66
Q13	0.06	0.38	5.82	0.016**	0.45

Table 10.3: Proportions for questions of the PreFake and PosFake questionnaires, McNemar's test analysis, and ϕ effect size, $N = 28$

Fisher exact tests for unpaired data. In this case, only Q12 had statistically significant differences $p = 0.007$**) in favor of the questionnaire with real images (proportions 0.62 vs. 0.24).

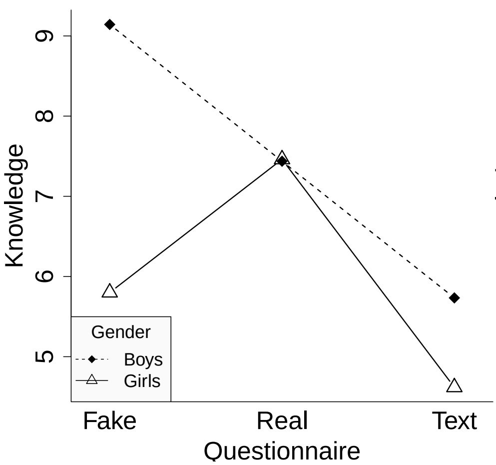
A multifactorial ANOVA test was also performed to take into consideration several factors simultaneously. The factors were gender, age, and questionnaire. The effect size used was the partial eta-squared (η^2). The results of the analysis shown in Table 10.4 indicate that there were statistically significant differences in the gender and questionnaire factors. The effect sizes revealed that the most influential factor was the questionnaire with large size, followed by gender which had a medium size. No statistically significant differences were found in the interactions between factors. A Tukey's post-hoc pairwise comparison revealed statistically significant differences between PosFake and PosText ($p = 0.009$**) and between PosReal and PosText ($p = 0.007$**), which corroborate previous analyses.

Figure 10.5a shows the interaction plot between gender and the three types of questionnaires. Boys acquired more knowledge than girls using the Text and Fake-image questionnaires. For the questionnaire with real images, both genders obtained the same score. Figure 10.5b shows the interaction plot between gender and age, where the older children had higher scores than the younger children. However, this

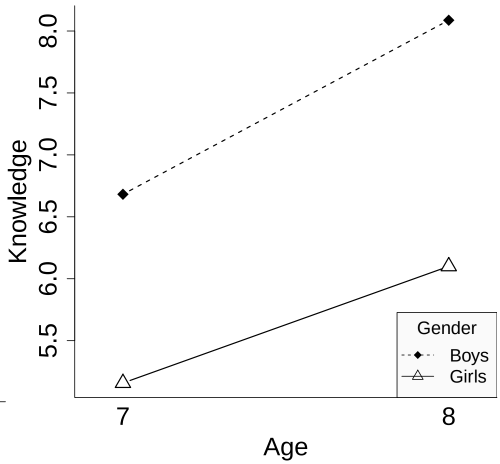
Factor	d.f.	F	p	partial η^2
Gender	1	6.99	0.009**	0.078
Age	1	2.87	0.093	0.033
Questionnaire	2	6.91	0.001**	0.144
Interactions	≤ 2	<2.64	>0.076	<0.061

Table 10.4: Multifactorial ANOVA for the knowledge variable, $N = 94$

difference between the two ages was not statistically significant.



(a) Interaction by gender of each questionnaire

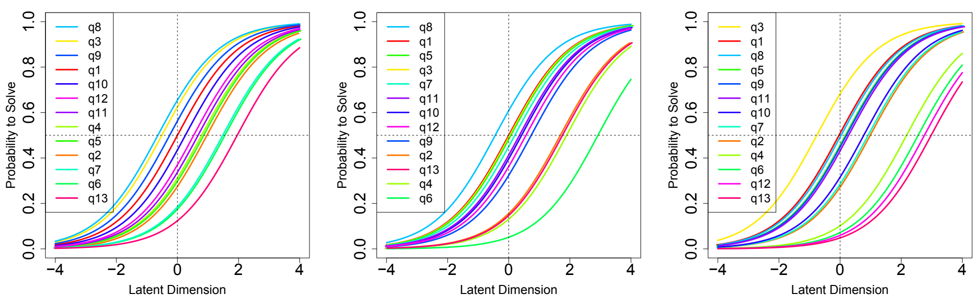


(b) Interaction by gender of each age

Figure 10.5: Interaction plots for the acquired knowledge

10.3.2 Rasch model analysis

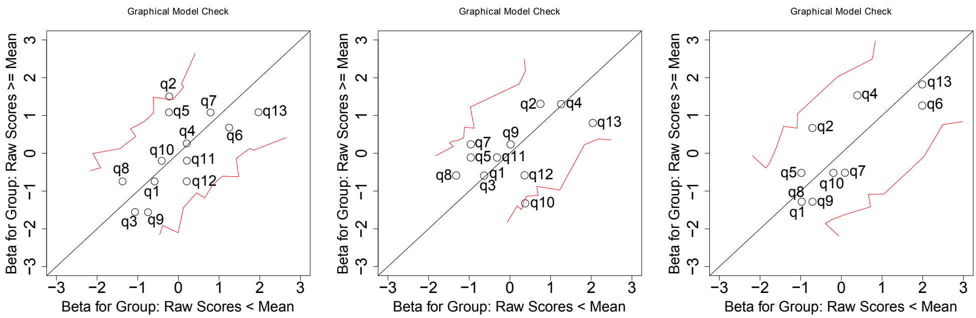
To complete the statistical analysis, as in previous studies, the dichotomous Rasch model was used (Rasch, 1960). Figure 10.6 shows the Item Characteristic Curve for every question. All the questions in the graph appear ordered by probability to answer the question correctly. Figure 10.6a shows the Item Characteristic Curve for the group of children who used the text questionnaire. It can be observed that, in this group, the hardest question was Q13, where it was necessary for a child to have an ability value of 2 in order to have a probability of 0.5 to answer this question correctly. The easiest question was Q8, where a child with an ability value of -1 was enough to have a probability of 0.5 in order to answer the question correctly. The most balanced question of this group was Q1, which needed an ability of 0 (the medium value) to have a probability of 0.5. Figure 10.6b shows the Item Characteristic Curve for the group of children who used the real-image questionnaire. The order of the questions changed with respect to the previous group. In this group, the most difficult question was Q6 and the easiest was Q8. The most balanced questions for this group were Q1 and Q5 which share the most balanced position. Figure 10.6c shows the Item Characteristic Curve for the group of children who used the fake-image questionnaire. Here the order of the questions also changed. The hardest question was Q13 and the easiest question was Q3. The most balanced questions in this group were Q1 and Q8. In summary, it can be observed from these graphs that, in the text group, the questions are grouped in one cluster. This means that even though the questions have different latent dimensions, they have the same level of magnitude. In contrast, in the other two types of questionnaires, the questions are grouped as two clustered sets of data. This means that the difference between the easier questions and the more difficult questions is more distinguishable in these two types of questionnaires. Although they seem more difficult, the latent dimensions of the children from these groups are enough to solve these questions satisfactorily (Figure 10.8 completes this information).



(a) Text-only (b) Real images (c) Fake images

Figure 10.6: Item Characteristic Curve for all questions

A graphical model check test was also performed. In this test, the questions were grouped by raw scores and the ones which were higher than the mean were separated from the ones which were lower. The red lines represent the confidence bands. The results for the questions are shown in the graphs in Figure 10.7. For the group of children who used the text questionnaires (Figure 10.7a), it can be observed that only Q2 is narrowly out of the confidence bands; for the group of children who used the real-image questionnaire (Figure 10.7b), Q10 is touching the confidence bands; finally, for the group of children who used the fake-image questionnaire (Figure 10.7c), every question is inside the confidence bands. Therefore, the questions are appropriate for the assessment of the acquired knowledge for the three types of questionnaires.



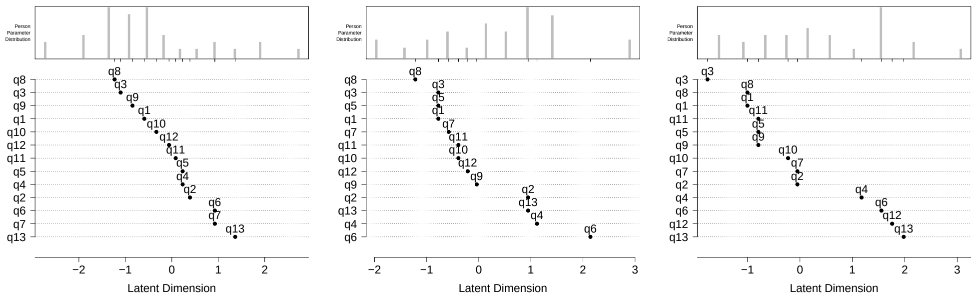
(a) Text-only (b) Real images (c) Fake images

Figure 10.7: Graphical model check

In order to visually check the children and the questions, a Person-Item Map was plotted, where the estimated ability of the child and the question difficulty measures are placed side by side in one vertical dimension. The questions appear in order of difficulty. The Person-Parameter Distribution (which is at the top of the graph) is a distribution of the children's abilities.

The Person-Item Map for each group of children is shown in Figure 10.8. For the text questionnaire group (Figure 10.8a), the hardest question (Q13) was easier than the ability of 8.33% of the children, and the easiest question (Q8) was more difficult than the ability of 33.33% of the children. For the real-image questionnaire group (Figure 10.8b), the hardest question (Q6) was easier than the ability of 6.89% of the children, and the easiest question (Q8) was more difficult than the ability of 10.34% of the children. For the fake-image questionnaire group (Figure 10.8c), the most difficult question (Q13) was easier than the ability of 10.34% of the children, and the easiest question (Q3) was harder than the ability of 0% of the children.

It can be observed in Figure 10.8 that the distributions of children who used a questionnaire with images are moved to the right, which means that most of the children were able to correctly answer most of the questions. In the case of the text-only questionnaire, the distribution shows that most of the children were in the lower levels and near the easiest questions. The questions were grouped in the same way they were distributed in Figure 10.6. In the text-only group, the questions were grouped in one cluster, and in the other two questionnaires, there were two differentiate clusters of questions. In summary, it can be concluded from these graphs that the children who used the questionnaires with images acquired greater latent ability for answering the questions than the children who used the text-only questionnaire.



(a) Text-only (b) Real images (c) Fake images

Figure 10.8: Person-Item Map for both groups

To check the goodness of fit of the Rasch model, the test proposed by Andersen (1973) was used. In this study, this test offered the following values that fit the Chi-squared distribution: LRvalue = 14.44, df = 12, p = 0.274. Therefore, the Rasch model is true in this study.

10.4 Conclusions

Three different types of questionnaires were designed and tested for assessment purposes. We carried out a study with children to determine whether the use of images that accompany an item to be identified in a question affects the selection of the correct answer in any way. We compared the text-only questionnaires with questionnaires that had images that appeared in the game that was played (real images) and questionnaires that had representative images of the item specified (fake images) that had not appeared in the game played.

From the initial knowledge of the three groups, statistically significant differences were found when comparing the text-only questionnaire with either of the two questionnaires with images. No statistically significant differences were found between the questionnaires with real and fake images. This was the result that we expected since, before playing the game, the two types of images (real or fake) represent the same concept. This result implies that the images gave an additional clue in selecting the correct answer.

Even though it was not the primary objective of this study, the acquired knowledge variable was analysed to assure that the learning method used is effective when it comes to transmitting knowledge in the short-term. The results indicated that regardless of the questionnaire used for the assessment (text-only, real, or fake), the children acquired statistically significant improvement in knowledge using the game. This result is in line with other studies in this thesis.

From the knowledge scores obtained after playing the game, statistically significant differences were found only when comparing the text-only questionnaire with either of the two questionnaires with images regardless of whether or not the images were exactly the same as the ones used in the game. These results corroborate the tenth hypothesis (the questionnaire with images are better than the text-only questionnaire) but do not support the eleventh hypothesis (real images are better than fake images). Even though we expected both hypotheses to be corroborated, it is still an excellent result because it means that images (real or fake) help the students to choose the right answer. As in the pre-test, the students did not choose the right answer with only text. However, when the associated image was included, they were able to choose the right answer. Therefore, it can be concluded that, to a great extent, the use of images in the questionnaires helps student to select the correct answer. This conclusion is in line with the work of Saß et al. (2012). Moreover, this study also demonstrates that it does not matter whether or not images are used during the instruction of the material or whether the images used were the same or different as those used during instruction. In formative assessment, the training does not finish until the end of the course and the assessment is part of the training process. If images are added to answers, the children can relate an image with its definition during the assessment, which contributes to completing their training. Therefore, based on these results, images added to answers could be used in formative assessment as a reinforcement of the knowledge that the children have while performing the tests.

Based on our own studies and those of other authors mentioned in this work, we can conclude that computer-based assessment offers different advantages, which they can be that it helps in the increasing of the engagement of the students (Anderson et al., 2005); it reduces the costs of paper, time, and processing, and is less vulnerable to the influence of the faculty (Dommeyer et al., 2002b); and it facilitates self-assessment. Self-assessment has advantages for both teachers and students (McConnell, 2006). It provides immediate feedback and helps to eliminate the distance between teachers and students. Moreover, students are more independent, which can promote self-confidence. If the assessment is online, it helps in overcoming the problems that traditional learning environments have (restrictions of teaching schedules and large numbers of concurrent students) (Wang, 2011; Wang et al., 2007).

VI

Discussion

Concluding remarks

11

"Live as if you were to die tomorrow. Learn as if you were to live forever."

— Mahatma Gandhi

11.1 Conclusions

In this thesis, we have studied the impact on the learning of children when playing video games with an educational background by using Natural User Interfaces, autostereoscopy and frontal projection. Psychologists and philosophers have studied the influence that playing games has on the learning process of children and they have concluded that entertainment is an important factor that helps to improve learning (Albert and Mori, 2001; Taran, 2005). An educational video game was designed and developed, based on the topic of the passage of time, learning about different historical periods. To our knowledge, this is the first work in which autostereoscopy and

NUI have been used for the development of an educational game. For this game, two different configurations were developed (detailed in Chapters 4 and 5). Moreover, it is also the first work in which autostereoscopic display + NUI are compared with large-size tabletop display + NUI.

The design of the game was performed by following the theories of the Constructivism (Dewey, 1963), the approach for computer-supported group-based learning systems proposed by Strijbos et al. (2004), the design guidelines for classroom collaborative games proposed by Villalta et al. (2011), and the DPE framework (Winn, 2008). By using this video game, the children acquired a significant knowledge about the historical ages. This significance in knowledge was found in the five studies. Another general significance found in all the studies is about satisfaction and usability. The children were really engaged with the game and they had fun. The general satisfaction of the game was highly scored, which means that the children learned the contents of the game while having a great time.

Table 11.1 shows a summary of the studies carried out throughout this thesis and the more relevant characteristics of each one of them.

Study	Individual	Pairs	Large Groups	Different systems	Different methods
Autostereoscopy vs. Frontal Projection		•		•	
Individual vs. Collaborative	•	•	•		•
LLI vs. FLI	•	•			•
Text-Only vs. Real Images vs. Fake Images		•		•	•

Table 11.1: Summary of the studies carried out

The conclusions for each study are detailed as follows:

- **Autostereoscopy vs. frontal projection:** In the first study, the children who evaluated the systems increase their knowledge about the topic of the game, and the increment of knowledge was no statistically significant different; and, the 3D perception was totally perceived by the children. Besides, the usability and satisfaction outcomes were highly rated, which means that the children liked our systems and they found them easy to use.

- **Individual vs. collaborative:** In the second study, the children also increase their knowledge, but in this case, the children who played with the game in the collaborative mode (in pairs) increased their knowledge more than the children who played with the game in the individual mode. This indicates that collaboration helps to improve the children's skills (Gokhale, 1995; Li, 2002) and they can obtain a better score in the tests (Neo, 2003).

- **Traditional learning vs. large groups or pairs:** In the third study, we have compared two learning methods (playing a game collaboratively on a large-size tabletop display and attending a traditional classroom lesson). The collaborative system have been splitted into two groups: one group of children playing in pairs and another group of children playing in large groups. The results showed that there were statistically significant differences between the traditional method and the game played in a large group in favour of children who played the game in the large group. There were no statistically significant differences between learning in large groups or learning in pairs. Therefore, with our system, several children can play with the game at the same time, thereby saving resources and time and encouraging the children to act collaboratively, which links group members together (Johnson et al., 1998). Moreover, the group size could be quite large (up to twelve). This group size is considerably larger than the group sizes used in previous studies (Sung and Hwang, 2013; Hwang et al., 2013; Cao et al., 2010; Jamil et al., 2011). In contrast, Rau and Heyl (1990) stated that it is difficult to ensure that all members in large groups participate. However, this study demonstrates that it is possible for all the members in a large group to participate in the learning activity and improve learning outcomes.

- **LLI vs. FLI:** In the fourth study, we tried to determine whether playing with the linear or free itinerary affected several aspects. From the learning outcomes, we would like to highlight that, even though there were no statistical significant differences between the learning itineraries, the children acquired new knowledge and that the free and the linear modes facilitated an increase in knowledge. Therefore, this study reveals the potential of computer-based learning games as a tool in the learning process, in both free and linear itinerary modes. Moreover, the free itinerary allows the children to personalize their learning and to increase their autonomy.

- **Text-Only vs. Real Images vs. Fake Images:** In the fifth study, three different types of questionnaires were designed and tested for assessment purposes. We carried out a study with children to determine whether the use of images that accompany an item to be identified in a question affects the selection of the correct answer in any way. We compared the text-only questionnaires with questionnaires that had images that appeared in the game that was played (real images) and questionnaires that had representative images of the item specified (fake images) that had not appeared in the game played. The results indicated that the children that filled out the questionnaires with images scored higher than the children that filled out the text-only questionnaires; however, there were no differences between real images and fake images. This indicates that the images improves the score that the children obtain when filling out the questionnaires. These results are in line with previous studies (Dindar et al., 2013; Saß et al., 2012).

From the studies performed we can present the following general conclusions:

- The new technologies are appropriated for developing educational games and autostereoscopy is a technology to exploit in their development.

- With applications like the ones presented in this thesis, children can learn using new technologies and, at the same time, they can have a good time playing the game, which promotes their learning.

- To play collaboratively facilitates the effectiveness of games that have educational purposes. Children can learn a wide variety of educational topics by using new technologies and having fun with other students at the same time they are playing with them.

- Children can perceive the 3D sensation of an autostereoscopic display correctly, and this feature engages them into the game, turning it more attractive and pleasant.

11.2 Scientific contributions

The publications deriving from this thesis are the following:

11.2.1 JCR Journals

- Martín-SanJosé, J.F., Juan, M.C., Seguí, I., García-García, I. (2015). The effects of computer-based games and collaboration in large groups vs. collaboration in pairs or traditional methods. *Computers & Education* 87 42–54, DOI:10.1016/j.compedu.2015.03.018 Impact Factor = 2.630 Q1

- Martín-SanJosé, J.F., Juan, M.C., Mollá, R., Vivó, R. (2015). Advanced displays and Natural User Interfaces to support learning. *Interactive Learning Environments.* Accepted, in press. DOI:10.1080/10494820.2015.1090455 Impact Factor = 0.750 Q2

- Martín-SanJosé, J.F., Juan, M.C., Torres, E., Vicent, M.J. (2014). Playful interaction for learning collaboratively and individually. *Journal of Ambient Intelligence and Smart Environments* 6(3) 295–311. DOI:10.3233/AIS-140257 Impact Factor = 1.082 Q2

- Martín-SanJosé, J.F., Juan, M.C., Gil-Gómez, J.A., Rando, N. (2014). Flexible learning itinerary vs. linear learning itinerary. *Science of Computer Programming* 88 3–21. DOI:10.1016/j.scico.2013.12.009 Impact Factor = 0.548 Q4

11.2.2 Latindex journals

- Martín-SanJosé, J.F., Juan, M.C., Mollá, R., Abad, F. (2015). The Effects of Images on Multiple-choice Questions in Computer-based Formative Assessment. *Digital Education.* Accepted, in press.

11.2.3 Conferences

- Martín-SanJosé, J.F. (2015). Aprender etapas históricas con gestos y visualización 3D. *II Encuentro de estudiantes de doctorado. Universidad Politécnica de Valencia*, poster 115.

- Martín-SanJosé, J.F. (2014). Dispositivos de visualización avanzados e interacción natural para que los niños puedan aprender jugando. *I Encuentro de estudiantes de doctorado. Universidad Politécnica de Valencia*, poster 119.

- Martín-SanJosé, J.F., Juan, M.C., Giménez, M., Cano, J. (2013). A Computer-Based Learning Game for Studying History. *XXIII Congreso Español de Informática Gráfica, CEIG 2013*, pp. 177-186.

11.2.4 Books

- Martín-SanJosé, J.F., Juan, M.C., Martín, A., Bonet, S. (2012). *Manual de buenas prácticas de aplicación de la Realidad Aumentada en empresas.* CreateSpace Independent Publishing Platform. ISBN: 978-1-47-830783-9

11.2.5 Other diffusions

- Juan, M.C., Martín-SanJosé, J.F., Cárdenas, S., Loachamín, M., Rodríguez, D. (2014). Aplicaciones de la Realidad Aumentada, Autoestereoscopía e Interfaces Naturales en Educación y Psicología, *I Jornada de Aplicaciones Industriales de Investigación*, Valencia.

11.3 Future work

With regard to future work, we can separate it into two different sections. On one hand, the improvements that can be made to the systems; and, on the other hand, additional comparatives using the same or the improved systems.

11.3.1 Possible improvements to the systems

The autostereoscopic system could be improved by displaying the video image in 3D and not just the virtual objects, this can be done by using several cameras to capture the real-world image. Augmented Reality could also be considered (Arino et al., 2014). The interaction could also be improved, adding the recognition of new gestures. The game could be enhanced by adding more interactive material for the periods in which the content is now mainly transmitted using videos. The system could be used in a mobile device such as a tablet, or even mobile devices capables of generating 3D visualization for the user.

Another challenge could be to make the game less linear and predictable. One extension could be to add activities that have a gradual increase in difficulty (Villalta et al., 2011); for example, adapting the game difficulty to students in different academic grades and also allowing teachers to be more involved in the game. Another possible extension could be to add open skills such as learning objectives and ill-structured tasks (Strijbos et al., 2004).

One way of guiding students in their learning that is normally used by teachers is to use instructional prompts when students give an incorrect answer. If the feedback arrives via a graduated prompt approach, it facilitates the students' thinking and gives correct answers step by step (Campione and Brown, 1987). Images could be incorporated in systems that already include feedback in order to determine to what extent the inclusion of images improves self-assessment. The same idea could be applied to systems that include personalized assessment. According to Wang (2014), learners are likely to experience better e-Learning effectiveness when they conduct self-evaluation via Web-based dynamic assessment.

Since the evaluation was achieved by filling out some on-line questionnaires, making these questionnaires more interactive by using the same devices on which they played the games would make the children more willing to fill them out, and it would provide immediate feedback to both students and teachers.

11.3.2 Additional comparatives

Taking into account the collaborative background of the game, it could be enhanced by adding other play modes that are not only collaborative but also competitive. According to Bachour et al. (2010), if participation is not balanced in a collaborative learning situation, the children who participated less would get worse scores than those who participated more. We tried to facilitate this balance. However, a formal study could be carried out to determine the influence of the most active child when they play in pairs/groups. Considering the free itinerary, another study could be allow the children more freedom, e.g., to learn a period of history in different sessions and whenever the children wanted. Another study could be carried out to determine the benefits of the static system and a mobile system.

With regard to the factors that influence the behavioural intention to use a computer-based assessment, Terzis and Economides (2011) conducted a study to investigate these factors. From their results, they concluded that perceived ease of use and perceived playfulness have a direct effect on the use of computer-based assessment. According to those in charge of our studies who were supervising the activities, the children had no problems using the questionnaires. Informal questions and the children's comments indicate that it was easy to use. However, a formal study could confirm this assertion and also take into account the playfulness aspect. We have only studied the increase in knowledge in the short-term, but possible future work could study long-term learning.

Bibliography

Al-Qahtani, A.A. and Higgins, S.E. (2013). Effects of traditional, blended and e-learning on students' achievement in higher education. *Journal of Computer Assisted Learning*, 29(3):220–234, doi:10.1111/j.1365-2729.2012.00490.x. [Cited on page 17]

Albert, D. and Mori, T. (2001). Contributions of cognitive psychology to the future of e-learning. *Bulletin of the Graduate School of Education*, 50:25–34. [Cited on page 155]

Anacleto, J.C., Fels, S., and Silvestre, R. (2012). Transforming a paper based process to a natural user interfaces process in a chronic care hospital. *Procedia Computer Science*, 14:173–180, doi:10.1016/j.procs.2012.10.020. [Cited on page 25]

Andersen, E.B. (1973). A goodness of fit test for the rasch model. *Psychometrika*, 38(1):123–140, doi:10.1007/BF02291180. [Cited on pages 96, 110 and 149]

Anderson, H.M., Cain, J., and Bird, E. (2005). Online Student Course Evaluations: Review of Literature and a Pilot Study. *American Journal of Pharmaceutical Education*, 69(1):34–43, article no. 5, doi:10.5688/aj690105. [Cited on pages 23 and 151]

Anneta, L.A., Minogue, J., Holmes, S.Y., and Cheng, M.T. (2009). Investigating the impact of video games on high school students' engagement and learning about genetics. *Computers & Education*, 53(1):74–85, doi:10.1016/j.compedu.2008.12.020. [Cited on page 78]

Arbaugh, J.B. (2002). Managing the on-line classroom: a study of technological and behavioral characteristics of web-based MBA courses. *The Journal of High Technology Management*, 13(2):203–223, doi:10.1016/S1047-8310(02)00049-4. [Cited on page 16]

Arino, J.J., Juan, M.C., Gil-Gómez, J.A., and Mollá, R. (2014). A comparative study using an autostereoscopic display with augmented and virtual reality. *Behaviour &*

Information technology, 33(6):646–655, doi:10.1080/0144929X.2013.815277. [Cited on pages 4, 26 and 160]

Arreola, R.A. (1995). *Developing a Comprehensive Faculty Evaluation System*. Jossey-Bass, Boston. [Cited on page 21]

Avery, R.J., Bryant, W.K., Mathios, A., Kang, H., and Bell, D. (2006). Electronic Course Evaluations: Does an Online Delivery System Influence Student Evaluations? *The Journal of Economic Education*, 37(1):21–37, doi:10.3200/JECE.37.1.21-37. [Cited on page 24]

Bachour, K., Kaplan, F., and Dillenbourg, P. (2010). An Interactive Table for Supporting Participation Balance in Face-to-Face Collaborative Learning. *IEEE Transactions on Learning Technologies*, 3(3):203–213, doi:10.1109/TLT.2010.18. [Cited on page 161]

Barron, B. (2000). Achieving Coordination in Collaborative Problem-Solving Groups. *Journal of the Learning Sciences*, 9(4):403–436, doi:10.1207/S15327809JLS0904_2. [Cited on page 19]

Beale, I.L., Kato, P.M., Marin-Bowling, V.M., Guthrie, N., and Cole, S.W. (2007). Improvement in cancer-related knowledge following use of a psychoeducational video game for adolescents and young adults with cancer. *Journal of Adolescent Health*, 41(3):263–270, doi:10.1016/j.jadohealth.2007.04.006. [Cited on page 15]

Beck, J. and Wade, M. (2004). *Got Game: How a New Gamer Generation is Reshaping Business Forever*. Boston: Harvard Business School Press. [Cited on page 13]

Beck, J. and Wade, M. (2006a). *The Kids are Alright: How the Gamer Generation is Changing the Workplace*. Boston: Harvard Business School Press. [Cited on page 13]

Beck, J. and Wade, M. (2006b). *The kids are alright: How the gamer generation is changing the workplace*. Harvard Business School Press, Boston. [Cited on page 15]

Bekebrede, G., Warmelink, H., and Mayer, I. (2011). Reviewing the need for gaming in education to accommodate the net generation. *Computers & Education*, 57(2):1521–1529, doi:10.1016/j.compedu.2011.02.010. [Cited on page 13]

Blecic, I., Cecchini, A., Rizzi, P., and A. Trunfio, G. (2002). Playing with Automata. An Innovative Perspective for Gaming Simulation. In Bandini, S., Chopard, B., and Tomassini, M., editors, *Cellular Automata*, volume 2493 of *Lecture Notes in Computer Science*, pages 337–348. Springer Berlin Heidelberg, Berlin. [Cited on page 4]

Bonanno, P. and Kommers, P.A.M. (2005). Gender differences and styles in the use of digital games. *Educational Psychology*, 25(1):13–41, doi:10.1080/0144341042000294877. [Cited on page 134]

Bourgonjon, J., Grove, F.D., Smet, C.D., Looy, J.V., Soetaert, R., and Valcke, M. (2013). Acceptance of game-based learning by secondary school teachers. *Computers & Education*, 67:21–35, doi:10.1016/j.compedu.2013.02.010. [Cited on page 15]

Bransford, J., Brown, A., and Cocking, R. (2000). *How people learn: brain, mind, experience, and school.* National Academy Press, Washington, DC, expanded edition. [Cited on page 21]

Brooke, J. (1996). SUS - A quick and dirty usability scale. In Jordan, P.W., Thomas, B., McClelland, I.L., and Weerdmeester, B., editors, *Usability evaluation in industry*, pages 189–194. Taylor&Francis, London. [Cited on page 187]

Brown, J., Hinze, S., and Pellegrino, J. (2008). Technology and formative assessment. In Good, T., editor, *Technology*, pages 245–255. Sage Publications, Thousand Oaks, California. [Cited on page 22]

Brusilovsky, P. (2001). Adaptive hypermedia. *User Modeling and User-Adapted Interaction*, 11(1–2):87–110, doi:10.1023/A:1011143116306. [Cited on page 20]

Brusilovsky, P., Eklund, J., and Schwarz, E. (1998). Web-based education for all: A tool for development adaptive courseware. *Computer Networks and ISDN Systems*, 30(1–7):291–300, doi:10.1016/S0169-7552(98)00082-8. [Cited on page 20]

Brusilovsky, P., Schwarz, E., and Weber, G. (1996a). Elm-art: An intelligent tutoring system on world wide web. In C. Frasson, G.G. and Lesgold, A., editors, *Proceedings of Third International Conference on Intelligent Tutoring Systems, ITS-96. Lecture Notes in Computer Science, Vol. 1086, Berlin: Springer Verlag*, pages 261–169. [Cited on page 20]

Brusilovsky, P., Schwarz, E., and Weber, G. (1996b). A tool for developing adaptive electronic textbooks on www. In *Proceedings of WebNet'96, World Conference of the Web Society*, pages 64–69, San Francisco, California, USA. [Cited on page 20]

Busschots, B., Raeside, L., Keating, J.G., and Waddington, S. (2006). The VTIE Collaborative Writing Environment. In *WEBIST conference on web information systems and technologies*, pages 221–228. [Cited on page 24]

Campione, J. and Brown, A. (1987). Linking dynamic assessment with school achievement. In Lidz, C., editor, *Dynamic assessment: An international approach to evaluating learning potential*, pages 82–115. The Guilford Press, New York. [Cited on page 161]

Cao, X., Lindley, S., Helmes, J., and Sellen, A. (2010). Telling the whole story: anticipation, inspiration and reputation in a field deployment of TellTable. In *Proceedings of the 2010 ACM conference on computer supported cooperative work*, pages 251–260, Savannah, Georgia, USA. [Cited on pages 116 and 157]

Carvalho, D., Bessa, M., Oliveira, L., Guedes, C., Peres, E., and Magalhães, L. (2012). New interaction paradigms to fight the digital divide: A pilot case study regarding multi-touch technology. *Procedia Computer Science*, 14:128–137, doi:10.1016/j.procs.2012.10.015. [Cited on page 25]

Challis, D. (2005). Committing to quality learning through adaptive online assessment. *Assessment & Evaluation in Higher Education*, 30(5):519–527, doi:10.1080/02602930500187030. [Cited on page 21]

Chan, L.K.Y. and Lau, H.Y.K. (2012). MagicPad: the projection based 3D user interface. *International Journal on Interactive Design and Manufacturing (IJIDeM)*, 6(2):75–81, doi:10.1007/s12008-012-0141-4. [Cited on page 27]

Chang, Y.J., Chen, S.F., and Chuang, A. (2011a). A gesture recognition system to transition autonomously through vocational tasks for individuals with cognitive impairments. *Research in Developmental Disabilities*, 32(6):2064–2068, doi:10.1016/j.ridd.2011.08.010. [Cited on page 25]

Chang, Y.J., Chen, S.F., and Huang, J.D. (2011b). A kinect-based system for physical rehabilitation: A pilot study for young adults with motor disabilities. *Research in Developmental Disabilities*, 32(6):2566–2570, doi:10.1016/j.ridd.2011.07.002. [Cited on pages 4 and 25]

Chin, J.P., Diehl, V.A., and Norman, L.K. (1988). Development of an instrument measuring user satisfaction of the human-computer interface. In *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '88*, pages 213–218, New York, New York, USA. ACM Press. [Cited on page 187]

Chuang, C.H., Chen, Y.N., Tsai, L.W., Lee, C.C., and Tsai, H.C. (2014). Improving Learning Performance with Happiness by Interactive Scenarios. *The Scientific*

World Journal, 2014(Article ID 807347):12 pages, doi:10.1155/2014/807347. [Cited on page 18]

Cisco, Intel, and Microsoft (2009). Transforming education: Assessing and teaching 21st century skills. [Cited on page 22]

Clark, A. (1997). *Being There: Putting Brain, Body, and World Together Again*. MIT Press, Cambridge, MA. [Cited on page 4]

Cohé, A. and Hachet, M. (2012). Beyond the mouse: Understanding user gestures for manipulating 3D objects from touchscreen inputs. *Computers & Graphics*, 36(8):1119–1131, doi:10.1016/j.cag.2012.09.004. [Cited on page 25]

Conejo, R., Barros, B., Guzmán, E., and Garcia-Viñas, J.I. (2013). A web based collaborative testing environment. *Computers & Education*, 68:440–457, doi:10.1016/j.compedu.2013.06.001. [Cited on page 23]

Connolly, T.M., Boyle, E.A., MacArthur, E., Hainey, T., and Boyle, J.M. (2012). A systematic literature review of empirical evidence on computer games and serious games. *Computers & Education*, 59(2):661–686, doi:10.1016/j.compedu.2012.03.004. [Cited on page 15]

Connolly, T.M., Stansfield, M., and Hainey, T. (2007). An application of games-based learning within software engineering. *British Journal of Educational Technology*, 38(3):416–428, doi:10.1111/j.1467-8535.2007.00706.x. [Cited on page 14]

Connolly, T.M., Stansfield, M., and Hainey, T. (2011). An alternate reality game for language learning: ARGuing for multilingual motivation. *Computers & Education*, 57(1):1389–1415, doi:10.1016/j.compedu.2011.01.009. [Cited on page 14]

Couse, L.J. and Chen, D.W. (2010). A Tablet Computer for Young Children? Exploring Its Viability for Early Childhood Education. *Journal of Research on Technology in Education*, 43(1):75–98, doi:10.1080/15391523.2010.10782562. [Cited on page 14]

Csíkszentmihályi, M. (1990). *Flow: The Psychology of Optimal Experience*. Harper & Row, New York. [Cited on page 40]

De Bra, P. (1996). Teaching hypertext and hypermedia through the web. *Journal of Universal Computer Science*, 2(12):797–804. [Cited on page 20]

De Bra, P. and Calvi, L. (1998). AHA! An open Adaptive Hypermedia Architecture. *The New Review of Hypermedia and Multimedia*, 4(1):115–140, doi:10.1080/13614569808914698. [Cited on page 20]

De Freitas, S. and Oliver, M. (2006). How can exploratory learning with games and simulations within the curriculum be most effectively evaluated? *Computers & Education*, 46(3):249–264, doi:10.1016/j.compedu.2005.11.007. [Cited on page 34]

DeVincenzi, A., Yao, L., Ishii, H., and Raskar, R. (2011). Kinected conference: augmenting video imaging with calibrated depth and audio. In *Proceedings of the ACM 2011 conference on Computer supported cooperative work*, pages 621–624. [Cited on page 25]

Dewey, J. (1963). *Experience and Education*. New York: Collier. [Cited on pages 35 and 156]

Dietz, P. and Leigh, D. (2001). Diamondtouch: a multi-user touch technology. In *Proc. of the 14th Annual ACM Symposium on User Interface Software and Technology*, pages 219–226, New York. [Cited on page 28]

Dindar, M., Yurdakul, I.K., and Dönmez, F.I. (2013). Multimedia in Test Items: Animated Questions vs. Static Graphics Questions. *Procedia - Social and Behavioral Sciences*, 106:1876–1882, doi:10.1016/j.sbspro.2013.12.213. [Cited on pages 24 and 158]

Do-Lenh, S., Kaplan, F., and Dillenbourg, P. (2009). Paper-based concept map: the effects of tabletop on an expressive collaborative learning task. In *Proceedings of the 23rd British HCI Group Annual Conference on People and Computers: Celebrating People and Technology*, pages 149–158, Swinton, UK. [Cited on page 29]

Dommeyer, C.J., Baum, P., Chapman, K.S., and Hanna, R.W. (2002a). Attitudes of Business Faculty Towards Two Methods of Collecting Teaching Evaluations: Paper vs. Online. *Assessment & Evaluation in Higher Education*, 27(5):455–462, doi:10.1080/0260293022000009320. [Cited on page 23]

Dommeyer, C.J., Baum, P., and Hanna, R.W. (2002b). College Students' Attitudes Toward Methods of Collecting Teaching Evaluations: In-Class Versus On-Line. *Journal of Education for Business*, 78(1):11–15, doi:10.1080/08832320209599691. [Cited on pages 22 and 151]

Dommeyer, C.J., Baum, P., Hanna, R.W., and Chapman, K.S. (2004). Gathering faculty teaching evaluations by in class and online surveys: their effects on response rates and evaluations. *Assessment & Evaluation in Higher Education*, 29(5):611–623, doi:10.1080/02602930410001689171. [Cited on pages 22 and 23]

Dube, T. and Ma, M. (2010). A flexible e-assessment system inspired by design methodology management to accommodate diverse learning styles. *International Journal of Digital Society*, 1(1):44–52. [Cited on page 22]

Durkin, K. and Barber, B. (2002). Not so doomed: Computer game play and positive adolescent development. *Journal of Applied Developmental Psychology*, 23(4):373–392, doi:10.1016/S0193-3973(02)00124-7. [Cited on page 78]

Ebner, M. and Holzinger, A. (2007). Succesful implementation of user-centered game based on learning in higher education: an example from civil engineering. *Computers & Education*, 49:873–890, doi:10.1016/j.compedu.2005.11.026. [Cited on page 5]

Egenfeldt-Nielsen, S. (2007). *Beyond Edutainment: Exploring the Educational Potential of Computer Games*. Continuum Press. [Cited on page 78]

Elices, J.A., Del Caño, M., and Verdugo, M.A. (2002). Interacción entre iguales y aprendizaje. Una perspectiva de investigación. *Revista de Psicología General y Aplicaciones*, 55(3):421–438. [Cited on page 19]

Esenther, A. and Wittenburg, K. (2005). Multi-user multi-touch games on diamond-touch with the dtflash toolkit. In Maybury, M., Stock, O., and Wahlster, W., editors, *Intelligent Technologies for Interactive Entertainment.*, pages 315–319. Lecture Notes in Computer Science. Springer Berlin Heidelberg. [Cited on page 29]

Facer, K., Joiner, R., Stanton, D., Reid, J., Hull, R., and Kirk, D. (2004). Savannah: mobile gaming and learning? *Journal of Computer Assisted Learning*, 20(6):399–409, doi:10.1111/j.1365-2729.2004.00105.x. [Cited on page 15]

Falcão, T.P. and Price, S. (2009). What have you done! The role of 'interference' in tangible environments for supporting collaborative learning. In *Proceedings of Computer Supported Collaborative Learning 2009*, pages 325–334, Rhodes, Greece. International Society of the Learning Sciences. [Cited on pages 28 and 29]

Fischer, G.H. (2006). Rasch models. In Rao, C. and Sinharay, S., editors, *Psychometrics*, volume 26 of *Handbook of Statistics*, pages 515 – 585. Elsevier. [Cited on page 93]

Fishkin, K.P. (2004). A taxonomy for and analysis of tangible interfaces. *Personal Ubiquitous Computing*, 8(5):347–358, doi:10.1007/s00779-004-0297-4. [Cited on page 25]

Fjeld, M., Fredriksson, J., Ejdestig, M., Duca, F., Bötschi, K., Voegtli, B., and Juchli, P. (2007). Tangible user interface for chemistry education: Comparative evaluation and re-design. In *Proceedings of the International Conference on Human Factors in Computing Systems (CHI'07)*, pages 805–808. [Cited on page 16]

Freitas, R. and Campos, P. (2008). Smart: a system of augmented reality for teaching 2nd grade students. In *Proceedings of the 22nd British HCI Group Annual Conference on People and Computers: Culture, Creativity, Interaction - Volume 2*, pages 27–30. [Cited on page 14]

Furió, D., González-Gancedo, S., Juan, M.C., Seguí, I., and Costa, M. (2013a). The effects of the size and weight of a mobile device on an educational game. *Computers & Education*, 64:24–41, doi:10.1016/j.compedu.2012.12.015. [Cited on pages 16 and 187]

Furió, D., González-Gancedo, S., Juan, M.C., Seguí, I., and Rando, N. (2013b). Evaluation of learning outcomes using an educational iphone game vs. traditional game. *Computers & Education*, 64:1–23, doi:10.1016/j.compedu.2012.12.001. [Cited on pages 6, 16, 17, 34 and 187]

Furió, D., Juan, M.C., Seguí, I., and Vivó, R. (2015). Mobile learning vs. traditional classroom lessons: a comparative study. *Journal of Computer Assisted Learning*, 31(3):189–201, doi:10.1111/jcal.12071. [Cited on pages 18 and 187]

Garris, R., Ahlers, R., and Driskell, J.E. (2002). Games, motivation and learning: A research and practice model. *Simulation and gaming*, 33(4):441–467, doi:10.1177/1046878102238607. [Cited on page 15]

Gathy, P., Denef, J.F., and Haumont, S. (1991). Computer-assisted self-assessment (CASA) in histology. *Computers & Education*, 17(2):109–116, doi:10.1016/0360-1315(91)90089-A. [Cited on page 23]

Gikandi, J.W., Morrow, D., and Davis, N.E. (2011). Online formative assessment in higher education: A review of the literature. *Computers & Education*, 57(4):2333–2351, doi:10.1016/j.compedu.2011.06.004. [Cited on page 21]

Gimbert, B. and Cristol, D. (2004). Teaching curriculum with technology: Enhancing children's technological competence during early childhood. *Early Childhood Education*, 31(3):207–216, doi:10.1023/B:ECEJ.0000012315.64687.ee. [Cited on page 14]

Girard, R., Ecalle, J., and Magnan, A. (2013). Serious games as new educational tools: how effective are they? A meta-analysis of recent studies. *Journal of Computer*

Assisted Learning, 29(3):207–219, doi:10.1111/j.1365-2729.2012.00489.x. [Cited on pages 15 and 17]

Gokhale, A.A. (1995). *Collaborative Learning Enhances Critical Thinking*. Digital Library and Archives of the Virginia Tech University Libraries. [Cited on pages 19 and 157]

Grabinger, S. (1996). Screen design. In Kommers, P., Grabinger, S., and Dunlap, J., editors, *Hypermedia Learning Environments: Instructional Design and Integration*. Lawrence Erlbaum Associates, New Jersey. [Cited on page 20]

Guillén-Nieto, V. and Aleson-Carbonell, M. (2012). Serious games and learning effectiveness: The case of it's a deal! *Computers & Education*, 58(1):435–448, doi:10.1016/j.compedu.2011.07.015. [Cited on page 15]

Hargreaves, E. (2008). Assessment. In McCulloch, G. and Crook, D., editors, *The Routledge international encyclopedia of education*, pages 37–38. Routledge, New York. [Cited on page 21]

Harris, A., Rick, J., Bonnett, V., Yuill, N., Fleck, R., Marshall, P., and Rogers, Y. (2009). Around the table: are multiple-touch surfaces better than single-touch for children's collaborative interactions? In *Proceedings of Computer Supported Collaborative Learning 2009*, pages 335–344. [Cited on page 28]

Hatch, A., Higgins, S., Joyce-Gibbons, A., and Mercier, E. (2011). Numbernet: using multi-touch technology to support within and between group mathematics learning. In *Proceedings of computer-supported collaborative learning conference*, pages 176–183. [Cited on page 29]

Hattie, J. and Timperley, H. (2007). The Power of Feedback. *Review of Educational Research*, 77(1):81–112, doi:10.3102/003465430298487. [Cited on page 21]

Heeter, C., Winn, B., and Greene, D. (2005). Theories meet realities: designing a learning game for girls. In *Proceedings of the 2005 conference on Designing for User eXperience*, New York. [Cited on page 39]

Heinrich, E. and Lawn, A. (2004). Onscreen marking support for formative assessment. In *Proceedings from Ed-Media 2004 world conference on educational multimedia, hypermedia & telecommunications*, pages 1985–1992. [Cited on page 22]

Hernández, S.R. (2008). The constructivist model and the new technologies, applied to the learning process. *Revista de Universidad y Sociedad del Conocimiento*, 5(2):26–35, doi:10.7238/rusc.v5i2.335. [Cited on page 35]

Hornecker, E., Marshall, P., Dalton, N., and Rogers, Y. (2008). Collaboration and interference: awareness with mice or touch input. In *Proceedings of the 2008 ACM conference on Computer supported cooperative work*, pages 167–176, New York, USA. [Cited on page 29]

Hunicke, R., LeBlanc, M., and Zubek, R. (2004). Mda: A formal approach to game design and game research. In *Proceedings of the Challenges in Game AI Workshop, Nineteenth National Conference on Artificial Intelligence*, pages 1–5. [Cited on pages 38, 39 and 40]

Hwang, G.J. and Chang, H.F. (2011). A formative assessment-based mobile learning approach to improving the learning attitudes and achievements of students. *Computers & Education*, 56(4):1023–1031, doi:10.1016/j.compedu.2010.12.002. [Cited on page 22]

Hwang, W.Y., Shadiev, R., Huang, Y.M., Cai, Y.T., Yang, Y.S., and Su, J.H. (2013). Effects of drag-and-response interaction mechanism of multi-touch operated tabletop technology on users' awareness and collaborative performance. *Computers and Education*, 67:130–141, doi:10.1016/j.compedu.2013.03.004. [Cited on pages 116 and 157]

IMS Global Learning Consortium (2008). IMS question and test interoperability specification. [Cited on page 22]

Jamil, I., O'Hara, K., Perry, M., Karnik, A., and Subramanian, S. (2011). The effects of interaction techniques on talk patterns in collaborative peer learning around interactive tables. In *Proceedings of Conference on Human Factors in Computing Systems*, pages 3043–3052. ACM Press. [Cited on pages 28, 116 and 157]

Johnson, D., Johnson, R., and Holubec, E. (1998). *Cooperation in the classroom*. Allyn & Bacon, Boston. [Cited on pages 19, 116 and 157]

Johnson, D.W. (1981). Student-student interaction: the neglected variable in education. *Educational Researcher*, 10(1):5–10, doi:10.3102/0013189X010001005. [Cited on page 36]

Jones, A., Scanlon, E., Tosunoglu, C., Morris, E., Ross, S., and Butcher, P. (1999). Contexts for evaluating educational software. *Interacting with Computers*, 11(5):499–516, doi:10.1016/S0953-5438(98)00064-2. [Cited on pages 98 and 134]

Jordà, S., Kaltenbrunner, M., Geiger, G., and Alonso, M. (2006). The reactable: a tangible tabletop musical instrument and collaborative workbench. In *ACM SIGGRAPH 2006 Sketches*, page 91, New York. [Cited on page 28]

Jordà, S., Geiger, G., Alonso, M., and Kaltenbrunner, M. (2007). The reactable: exploring the synergy between live music performance and tabletop tangible interfaces. In *Proceedings of the 1st international conference on Tangible and embedded interaction*, TEI '07, pages 139–146, New York, NY, USA. ACM. [Cited on page 28]

Juan, M.C., Beatrice, F., and Cano, J. (2008a). An Augmented Reality System for Learning the Interior of the Human Body. In *The 8th IEEE International Conference on Advanced Learning Technologies*, pages 186–188. [Cited on page 16]

Juan, M.C. and Calatrava, J. (2011). An Augmented Reality System for the Treatment of Phobia to Small Animals Viewed Via an Optical See-Through HMD: Comparison With a Similar System Viewed Via a Video See-Through HMD. *International Journal of Human-Computer Interaction*, 27(5):436–449, doi:10.1080/10447318.2011.552059. [Cited on page 5]

Juan, M.C., Canu, R., and Giménez, M. (2008b). Augmented Reality Interactive Storytelling Systems Using Tangible Cubes for Edutainment. In *Proceedings of the 2008 Eighth IEEE International Conference on Advanced Learning Technologies*, ICALT '08, pages 233–235, Washington, DC, USA. IEEE Computer Society. [Cited on page 16]

Juan, M.C., Carrizo, M., Giménez, M., and Abad, F. (2011a). Using an augmented reality game to find matching pairs. In *Proceedings of the 19th International Conference on Computer Graphics, Visualization and Computer Vision*, pages 59–66. [Cited on pages 16 and 34]

Juan, M.C., Furió, D., Alem, L., Ashworth, P., and Cano, J. (2011b). ARGreenet and BasicGreenet: Two mobile games for learning how to recycle. In *Proceedings of the 19th International Conference on Computer Graphics, Visualization and Computer Vision*, pages 25–32. [Cited on page 16]

Juan, M.C., Llop, E., Abad, F., and Lluch, J. (2010a). Learning words using augmented reality. In *Proceedings of the 10th IEEE International Conference on Advanced Learning Technologies (ICALT)*, pages 422–426. [Cited on page 16]

Juan, M.C., Toffetti, G., Abad, F., and Cano, J. (2010b). Tangible Cubes Used as the User Interface in an Augmented Reality Game for Edutainment. In *Proceedings of* *the 2010 10th IEEE International Conference on Advanced Learning Technologies*, ICALT '10, pages 599–603, Washington, DC, USA. IEEE Computer Society. [Cited on page 16]

Kagan, J. and Kogan, N. (1970). Individual variation in cognitive processes. In Mussen, P., editor, *Carmichaell's manual of child psychology*, pages 1273–1365. Wiley. [Cited on page 20]

Kakehi, Y., Iida, M., Naemura, T., Shirai, Y., Matsushita, M., and Ohguro, T. (2005). Lumisight table: An interactive view-dependent tabletop display. *IEEE Computer Graphics and Applications*, 25(1):48–53, doi:10.1109/MCG.2005.14. [Cited on page 27]

Kaltenbrunner, M. (2009). reactivision and tuio: a tangible tabletop toolkit. In *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces*, ITS '09, pages 9–16, New York, NY, USA. ACM. [Cited on page 28]

Kaltenbrunner, M. and Bencina, R. (2007). reactivision: a computer-vision framework for table-based tangible interaction. In *Proceedings of the 1st international conference on Tangible and embedded interaction*, TEI '07, pages 69–74, New York, NY, USA. ACM. [Cited on page 28]

Kamel Boulos, M.N., Blanchard, B.J., Walker, C., Montero, J., Tripathy, A., and Gutierrez-Osuna, R. (2011). Web GIS in practice X: a Microsoft Kinect natural user interface for Google Earth navigation. *International Journal of Health Geographics*, 10(45):1–14, doi:10.1186/1476-072X-10-45. [Cited on page 25]

Kaptelinin, V. (1999). Learning Together: Educational Benefits and Prospects for Computer Support. *Journal of the Learning Sciences*, 8(3-4):499–508, doi:10.1080/10508406.1999.9672078. [Cited on page 19]

Kaufmann, H. (2004). *Geometry Education with Augmented Reality*. PhD thesis, Institut für Softwartechnik und Interaktive Systeme. [Cited on page 16]

Kebritchi, M. and Hirumi, A. (2008). Examining the pedagogical foundations of modern educational computer games. *Computers & Education*, 51(4):1729–1743, doi:10.1016/j.compedu.2008.05.004. [Cited on page 4]

Keeves, J. (1994). Assessment in schools, methods of assessment. In Husen, Torsten, Postlethwaite, and Neville, editors, *The international encyclopedia of education*, pages 362–370. Pergamon Press, Oxford, 2 edition. [Cited on page 21]

Kharrufa, A., Leat, D., and Olivier, P. (2010). Digital mysteries: designing for learning at the tabletop. In *ACM International Conference on Interactive Tabletops and Surfaces ITS '10*, pages 197–206, New York, USA. [Cited on page 29]

Kim, H., Lee, G., Yang, U., Kwak, T., and Kim, K.H. (2012). Dual Autostereoscopic Display Platform for Multi-user Collaboration with Natural Interaction. *ETRI Journal*, 34(3):466–469, doi:10.4218/etrij.12.0211.0331. [Cited on page 26]

Kirschner, F., Paas, F., and Kirschner, P. (2009). Individual and group-based learning from complex cognitive tasks: Effects on retention and transfer efficiency. *Computers in Human Behavior*, 25(2):306–314, doi:10.1016/j.chb.2008.12.008. [Cited on page 19]

Kobourov, S.G., Pavlou, K., Cappos, J., Stepp, M., Miles, M., and Wixted, A. (2005). Collaboration with diamondtouch. In Costabile, M. and Paternò, F., editors, *Human-Computer Interaction - INTERACT 2005*, pages 986–989. Lecture Notes in Computer Science. Springer Berlin Heidelberg. [Cited on page 29]

Kubicki, S., Lebrun, Y., Lepreux, S., Adam, E., Kolski, C., and Mandiau, R. (2013). Simulation in contexts involving an interactive table and tangible objects. *Simulation Modelling Practice and Theory*, 31:116–131, doi:10.1016/j.simpat.2012.10.012. [Cited on page 27]

Kusek, J.Z. and Rist, R.C. (2004). *Ten Steps to a Results-Based Monitoring and Evaluation System: A Handbook for Development Practitioners*. World Bank Publications, Washington, DC. [Cited on page 24]

Lam, A., Chow, K., Yau, E., and Lyu, M. (2006). ART: augmented reality table for interactive trading card game. In *Proceedings of the 2006 ACM international conference on Virtual Reality Continuum and its Applications in Industry*, pages 357–360. [Cited on page 28]

Lange, B., Chang, C.Y., Suma, E., Newman, B., Rizzo, A.S., and Bolas, M. (2011). Development and evaluation of low cost game-based balance rehabilitation tool using the microsoft kinect sensor. In *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, volume 2011, pages 1831–1834. [Cited on page 25]

Largillier, G. (2011). Tablet interfaces should evolve with the user. *Electronic Design*, 59(6):19. [Cited on page 14]

Larsen, L.B., Jensen, R.B., Jensen, K.L., and Larsen, S. (2005). Development of an automatic pool trainer. In *Proceedings of the 2005 ACM SIGCHI International Conference on Advances in computer entertainment technology*, ACE '05, pages 83–87, New York, NY, USA. ACM. [Cited on page 16]

Lave, J. and Wenger, E. (1991). *Situated Learning: Legitimate Peripheral Participation*. Cambridge University Press, Cambridge, UK. [Cited on page 4]

Law, E.L.C., Kickmeier-Rust, M.D., Albert, D., and Holzinger, A. (2008). Challenges in the Development and Evaluation of Immersive Digital Educational Games. In *Proceedings of the 4th Symposium of the Workgroup Human-Computer Interaction and Usability Engineering of the Austrian Computer Society on HCI and Usability for Education and Work*, USAB '08, pages 19–30, Berlin, Heidelberg. Springer-Verlag. [Cited on page 34]

Layne, B.H., Decristoforo, J.R., and Mcginty, D. (1999). Electronic versus traditional student ratings of instruction. *Research in Higher Education*, 40(2):221–232, doi:10.1023/A:1018738731032. [Cited on page 23]

Lazarinis, F., Green, S., and Pearson, E. (2010). Creating personalized assessments based on learner knowledge and objectives in a hypermedia Web testing application. *Computers & Education*, 55(4):1732–1743, doi:10.1016/j.compedu.2010.07.019. [Cited on page 23]

Leclercq, D. and Gilles, J.L. (1993). Hypermedia: Teaching through assessment. In *Item Banking: Interactive Testing and Self-Assessment, NATO ASI Series - Series F: Computer and Systems Sciences, Vol. 112*, pages 31–47. [Cited on page 20]

Leclercq, D. and Pierret, D. (1989). A computerized open learning environment to study intrapersonal variations in learning styles: Delin. In *Proceedings of the 6th International Conference on Technology and Education (ICTE)*, pages 268–272, Florida, Orlando. [Cited on page 20]

Lee, S.J., Srinivasan, S., Trail, T., Lewis, D., and Lopez, S. (2011). Examining the relationship among student perception of support, course satisfaction, and learning outcomes in online learning. *Internet and Higher Education*, 14(3):158–163, doi:10.1016/j.iheduc.2011.04.001. [Cited on page 117]

Lewis, J.R. (1995). IBM computer usability satisfaction questionnaires: Psychometric evaluation and instructions for use. *International Journal of Human-Computer Interaction*, 7(1):57–78, doi:10.1080/10447319509526110. [Cited on page 187]

Li, Q. (2002). Exploration of collaborative learning and communication in an educational environment using computer-mediated communication. *Journal of Research on Technology in Education*, 34(4):503–516, doi:10.1080/15391523.2002.10782364. [Cited on pages 19 and 157]

Lien, C.L., Huang, C.Y., Wang, C.Y., and Chen, G.D. (2012). Using Kinect to track learning behavior of students in the classroom as video Portfolio to enhance reflection learning. In *Proceedings of the 20th International Conference on Computers in Education (ICCE '12)*. [Cited on page 25]

Llamas-Nistal, M., Caeiro, M., Castro, M., Plaza, I., and Tovar, E. (2011). Use of LMS functionalities in engineering education. In *Proceedings from 41st ASEE/IEEE frontiers in education conference (FIE 2011)*, pages S1G–1–S1G–6, 12–15. [Cited on page 23]

Llamas-Nistal, M., Fernández-Iglesias, M.J., González-Tato, J., and Mikic-Fonte, F.A. (2013). Blended e-assessment: Migrating classical exams to the digital world. *Computers & Education*, 62:72–87, doi:10.1016/j.compedu.2012.10.021. [Cited on page 23]

Maimone, A., Bidwell, J., Peng, K., and Fuchs, H. (2012). Enhanced personal autostereoscopic telepresence system using commodity depth cameras. *Computers & Graphics*, 36(7):791–807, doi:10.1016/j.cag.2012.04.011. [Cited on page 26]

Malone, T. and Lepper, M. (1987). Making Learning Fun: A Taxonomy of Intrinsic Motivations for Learning. In Snow, R. and Farr, M., editors, *Aptitude, Learning, and Instruction Volume 3: Conative and Affective Process Analyses*, pages 223–253. Hillsdale, New Jersey. [Cited on page 15]

Mansor, E., Angeli, A.D., and Bruijn, O. (2009). The fantasy table. In *Proceedings of the 8th International Conference on Interaction Design and Children*, pages 70–79, New York. [Cited on page 29]

Marco, J., Cerezo, E., Baldasarri, S., Mazzone, E., and Read, J. (2009). Bringing tabletop technologies to kindergarten children. In *Proceedings of the 23rd British HCI Group Annual Conference on People and Computers: Celebrating People and Technology. British Computer Society*, pages 103–111, Swinton, UK. [Cited on page 29]

Mayer, I.S., Carton, L., de Jong, M., Leijten, M., and Dammers, E. (2004). Gaming the future of an urban network. *Futures*, 36(3):311–333, doi:10.1016/S0016-3287(03)00159-9. [Cited on page 15]

Mayes, J.T. and Fowler, C.J. (1999). Learning technology and usability: A framework for understanding courseware. *Interacting with Computers*, 11(5):485–497, doi:10.1016/S0953-5438(98)00065-4. [Cited on pages 98 and 134]

McConnell, D. (2006). *E-Learning groups and communities: Imaging learning in the age of the internet*. OU Press. [Cited on page 151]

Monahan, T., McArdle, G., and Bertolotto, M. (2008). Virtual reality for collaborative e-learning. *Computers & Education*, 50(4):1339–1353, doi:10.1016/j.compedu.2006.12.008. [Cited on page 16]

Mumtaz, S. (2001). Children's enjoyment and perception of computer use in the home and the school. *Computers & Education*, 36(4):347–362, doi:10.1016/S0360-1315(01)00023-9. [Cited on page 5]

Neo, M. (2003). Developing a collaborative learning environment using a web-based design. *Journal of Computer Assisted Learning*, 19(4):462–473, doi:10.1046/j.0266-4909.2003.00050.x. [Cited on pages 19, 116 and 157]

Nicol, D.J. and Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: a model and seven principles of good feedback practice. *Studies in Higher Education*, 31(2):199–218, doi:10.1080/03075070600572090. [Cited on page 21]

Nocent, O., Piotin, S., Benassarou, A., Jaisson, M., and Lucas, L. (2012). Toward an immersion platform for the world wide web using autostereoscopic displays and tracking devices. In *Proceedings of the 17th International Conference on 3D Web Technology*, Web3D '12, pages 69–72, New York, NY, USA. ACM. [Cited on page 26]

Oblinger, D.G. (2004). The Next Generation of Educational Engagement. *Journal of Interactive Media in Education*, 8:1–18, doi:10.5334/2004-8-oblinger. [Cited on page 15]

O'Donnell, A. and Dansereau, D. (1995). Scripted cooperation in student dyads: a method for analyzing and enhancing academic learning and performance. In Hertz-Lazarowitz, R., editor, *Interaction in cooperative groups: The theoretical anatomy of group learning*, pages 120–141. Cambridge University Press. [Cited on page 36]

Omura, K., Shiwa, S., and Miyasato, T. (1998). Lenticular autostereoscopic display system: multiple images for multiple viewers. *Journal of the Society for Information Display*, 6(4):313–324, doi:10.1889/1.1985257. [Cited on page 49]

O'Neil, H.F., Wainess, R., and Baker, E.L. (2005). Classification of learning outcomes: Evidence from the computer games literature. *The Curriculum Journal*, 16(4):455–474, doi:10.1080/09585170500384529. [Cited on page 14]

Oosterhof, A., Conrad, R., and Ely, D. (2008). *Assessing learners online*. Pearson, New Yersey. [Cited on page 21]

Papastergiou, M. (2009). Digital game-based learning in high school computer science education: Impact on educational effectiveness and student motivation. *Computers & Education*, 52(1):1–12, doi:10.1016/j.compedu.2008.06.004. [Cited on pages 15 and 78]

Pask, G. (1976). Styles and strategies of learning. *British Journal of Educational Psychology*, 46(2):128–148, doi:10.1111/j.2044-8279.1976.tb02305.x. [Cited on page 20]

Patton, M.Q. (1997). *Utilization-Focused Evaluation*. Sage Publications, Thousand Oaks, California, 3 edition. [Cited on page 24]

Periverzov, F. and Ilieş, H.T. (2012). 3D Imaging for hand gesture recognition: Exploring the software-hardware interaction of current technologies. *3D Research*, 3(3):1–15, doi:10.1007/3DRes.03(2012)1. [Cited on page 25]

Piper, A. and Hollan, J. (2009). Tabletop displays for small group study: affordances of paper and digital materials. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems CHI '09*, pages 1227–1236, New York, USA. [Cited on page 29]

Prensky, M. (2001). *Digital game-based learning*. New York: McGraw-Hill. [Cited on pages 13, 14 and 15]

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. University of Chicago Press, Chicago. [Cited on pages 93, 108 and 146]

Rashad, A., Youssif, A., Abdel-Ghafar, R., and Labib, A.E. (2008). E-assessment tool: a course assessment tool integrated into knowledge assessment. In Iskander, M., editor, *Innovative techniques in instruction technology, e-learning, e-assessment, and education*, pages 7–12. Springer, Netherlands. [Cited on page 22]

Rau, W. and Heyl, B.S. (1990). Humanizing the college classroom: Collaborative learning and social organization among students. *Teaching Sociology*, 18(2):141–155. [Cited on pages 19, 116 and 157]

Ravelli, B. (2000). Anonymous Online Teaching Assessments: Preliminary Findings. In *Annual National Conference of the American Association for Higher Education*, Charlotte, North Carolina. [Cited on page 23]

Reeves, T. and Hedberg, J. (2009). Evaluation strategies for open and distributed learning environments. In Spratt, C. and Lajbcygier, P., editors, *E-Learning technologies and evidence based assessment approaches*, pages 234–253. Information Science Reference, New York. [Cited on page 21]

Reigeluth, C.M. (1992). Elaborating the elaboration theory. *Educational Technology Research & Development*, 40(3):80–86, doi:10.1007/BF02296844. [Cited on page 121]

Reigeluth, C.M. (1999). The elaboration theory: Guidance for scope and secuence decisions. In Reigeluth, C.M., editor, *Instructional design theories and models: Vol. II, A new paradigm of instruction theory*. Erlbaum. [Cited on page 122]

Rick, J., Harris, A., Marshall, P., Fleck, R., Yuill, N., and Rogers, Y. (2009a). Children designing together on a multi-touch tabletop: an analysis of spatial orientation and user interactions. In *Proceedings of the 8th International Conference on Interaction Design and Children*, IDC '09, pages 106–114, New York, NY, USA. ACM. [Cited on page 28]

Rick, J., Rogers, Y., Haig, C., and Yuill, N. (2009b). Learning by doing with shareable interfaces. *Children, Youth and Environments*, 19(1):321–342. [Cited on page 29]

Roman, D. (2010). Interact naturally. *Communications of the ACM*, 53(6):12, doi:10.1145/1743546.1743552. [Cited on page 14]

Rosas, R., Nussbaum, M., Cumsille, P., Marianov, V., Correa, M., Flores, P., Grau, V., Lagos, F., López, X., López, V., Rodriguez, P., and Salinas, M. (2003). Beyond nintendo: design and assessment of educational video games for first and second grade students. *Computers & Education*, 40(1):71–94, doi:10.1016/S0360-1315(02)00099-4. [Cited on page 78]

Rota, D. (1981). *Computer assisted instruction, lecture instruction, and combined computer-assisted/lecture instruction: A comparative experiment*. PhD thesis, Unpublished doctoral dissertation . University of Pittsburgh. [Cited on page 21]

Rudd, T. (2007). *Interactive whiteboards in the classroom*. Futurelab Report: IWBs 2007, Bristol. [Cited on page 25]

Sandor, C. and Klinker, G. (2005). A rapid prototyping software infrastructure for user interfaces in ubiquitous augmented reality. *Personal and Ubiquitous Computing*, 9(3):169–185, doi:10.1007/s00779-004-0328-1. [Cited on page 4]

Saß S., Wittwer, J., Senkbeil, M., and Köller, O. (2012). Pictures in Test Items: Effects on Response Time and Response Correctness. *Applied Cognitive Psychology*, 26(1):70–81, doi:10.1002/acp.1798. [Cited on pages 24, 150 and 158]

Selim, H. (2007). Critical success factors for e-learning acceptance: confirmatory factor models. *Computers & Education*, 49(2):396–413, doi:10.1016/j.compedu.2005.09.004. [Cited on page 16]

Shaffer, D.W., Squire, K.R., Halverson, R., and Gee, J.P. (2004). Video Games and the Future of Learning. *Phi Delta Kappan*, 87(2):104–111. [Cited on page 14]

Shea, P., Fredericksen, E., Pickett, A., and Pelz, W. (2004). Faculty development, student satisfaction, and reported learning in the SUNY learning network. In Duffy, T. and Kirkley, J., editors, *Learner-centered theory and practice in distance education*, pages 343–377. Lawrence Elrbaum Associates, Mahway, NJ. [Cited on page 117]

Shi, H., Revithis, S., and Chen, S. (2002). An agent enabling personalized learning in e-learning environments. In *Proceedings of the First International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS'02)*, pages 847–848. [Cited on page 16]

Slater, M., Usoh, M., and Steed, A. (1994). Depth of presence in virtual environments. *Presence-Teleoperators and Virtual Environments*, 3(2):130–144. [Cited on page 188]

Slavin, R.E. (1980). Cooperative learning in teams: State of the art. *Educational Psychologist*, 15(2):93–111, doi:10.1080/00461528009529219. [Cited on page 36]

Smeaton, A.F., Foley, C., Gurrin, C., Lee, H., and McGivney, S. (2006). Collaborative searching for video using the fischlar system and a diamondtouch table. In *Proceedings of the First IEEE International Workshop on Horizontal Interactive Human-Computer Systems*, pages 151–159. [Cited on page 29]

Sorenson, D.L. and Johnson, T.D. (2003). *Online Student Ratings of Instruction: New Directions for Teaching and Learning*. Jossey-Bass, San Francisco, CA. [Cited on page 23]

Squires, D. (1999). Usability and educational software design. *Special issue of interacting with computers*, 11(5):463–466, doi:10.1016/S0953-5438(98)00062-9. [Cited on page 98]

Stankov, S., Grubišić, A., and Žitko, B. (2004). E-learning paradigm and Intelligent tutoring systems. In Kniewald, Z., editor, *Annual 2004 of the Croatian academy of engineering*, pages 21–31. Croatian Academy of Engineering, Zagreb. [Cited on page 16]

Stockley, D. (2013). E-learning definition and explanation (e-learning, online training, online learning. Available at `http://derekstockley.com.au/elearning-definition.html`. [Cited on page 16]

Strijbos, J.W., Martens, R.L., and Jochems, W.M.G. (2004). Designing for interaction: Six steps to designing computer-supported group-based learning. *Computers & Education*, 42(4):403–424, doi:10.1016/j.compedu.2003.10.004. [Cited on pages 35, 156 and 160]

Suh, S., Kim, S.W., and Kim, N.J. (2010). Effectiveness of MMORPG-based instruction in elementary English education in Korea. *Journal of Computer Assisted Learning*, 26(5):370–378, doi:10.1111/j.1365-2729.2010.00353.x. [Cited on page 29]

Sun, P.C., Tsai, R.J., Finger, G., Chen, Y.Y., and Yeh, D. (2008). What drives a successful e-Learning? An empirical investigation of the critical factors influencing learner satisfaction. *Computers & Education*, 50(4):1183–1202, doi:10.1016/j.compedu.2006.11.007. [Cited on pages 98 and 134]

Sung, H.Y. and Hwang, G.J. (2013). A collaborative game-based learning approach to improving students' learning performance in science courses. *Computers & Education*, 63:43–51, doi:10.1016/j.compedu.2012.11.019. [Cited on pages 116 and 157]

Taherkhani, R. and Kia, M. (2012). Designing a high accuracy 3D auto stereoscopic eye tracking display, using a common LCD monitor. *3D Research*, 3(3):1–7, doi:10.1007/3DRes.03(2012)2. [Cited on page 26]

Tan, K.T.W., Lewis, E.M., Avis, N.J., and Withers, P.J. (2008). Using augmented reality to promote an understanding of materials science to school children. In *ACM SIGGRAPH ASIA 2008 educators programme on - SIGGRAPH Asia '08*, pages 1–8, New York, USA. ACM Press. [Cited on page 5]

Tapscott, D. (1998). *Growing up digital. The rise of the net generation*. New York: McGraw-Hill. [Cited on page 13]

Taran, C. (2005). Motivation techniques in elearning. In *Proceedings of the Fifth IEEE International Conference on Advanced Learning Technologies*, ICALT '05, pages 617–619, Washington, DC, USA. IEEE Computer Society. [Cited on page 155]

Terzis, V. and Economides, A.A. (2011). The acceptance and use of computer based assessment. *Computers & Education*, 56(4):1032–1044, doi:10.1016/j.compedu.2010.11.017. [Cited on page 161]

Torres, R.T., Preskill, H.S., and Piontek, M.E. (2004). *Evaluation Strategies for Communicating and Reporting: Enhancing Learning in Organizations*. Sage Publications, Thousand Oaks, California, 2 edition. [Cited on page 24]

Tufte, E.R. (1989). *The Visual Display of Quantitative Information*. Graphics Press, Cheshire, Connecticut. [Cited on page 24]

Villalta, M., Gajardo, I., Nussbaum, M., Andreu, J., Echeverría, A., and Plass, J. (2011). Design guidelines for Classroom Multiplayer Presential Games (CMPG). *Computers & Education*, 57(3):2039–2053, doi:10.1016/j.compedu.2011.05.003. [Cited on pages 36, 37, 98, 156 and 160]

Virvou, M., Katsionis, G., and Manos, K. (2005). Combining software games with education: Evaluation of its educational effectiveness. *Educational Technology & Society*, 8(2):54–65. [Cited on page 5]

Vonderwell, S., Liang, X., and Alderman, K. (2007). Asynchronous discussions and assessment in online learning. *Journal of Research on Technology in Education*, 39(3):309–328. [Cited on page 21]

Wang, K.H., Chen, L.C., Chu, P.Y., and Cheng, Y.M. (2009). A Study on the Design of Augmented Reality User Interfaces for Mobile Learning Systems in Heritage Temples. In *Proceedings of the Virtual and Mixed Reality*, volume 5622 of *Lecture Notes in Computer Science*, pages 282–290. [Cited on page 16]

Wang, T.H. (2010). Web-based dynamic assessment: Taking assessment as teaching and learning strategy for improving students' e-Learning effectiveness. *Computers & Education*, 54(4):1157–1166, doi:10.1016/j.compedu.2009.11.001. [Cited on page 22]

Wang, T.H. (2011). Developing Web-based assessment strategies for facilitating junior high school students to perform self-regulated learning in an e-Learning environment. *Computers & Education*, 57(2):1801–1812, doi:10.1016/j.compedu.2011.01.003. [Cited on page 151]

Wang, T.H. (2014). Developing an assessment-centered e-Learning system for improving student learning effectiveness. *Computers & Education*, 73:189–203, doi:10.1016/j.compedu.2013.12.002. [Cited on pages 23 and 161]

Wang, Y.S., Wang, H.Y., and Shee, D.Y. (2007). Measuring e-learning systems success in an organizational context: Scale development and validation. *Computers in Human Behavior*, 23(4):1792–1808, doi:10.1016/j.chb.2005.10.006. [Cited on pages 13, 21 and 151]

Warwick, P., Mercer, N., Kershner, R., and Staarman, J.K. (2010). In the mind and in the technology: The vicarious presence of the teacher in pupil's learning of science in collaborative group activity at the interactive whiteboard. *Computers & Education*, 55(1):350–362, doi:10.1016/j.compedu.2010.02.001. [Cited on page 25]

Weber, G. and Specht, M. (1997). User modeling and adaptive navigation support in www-based tutoring systems. In *Proceedings of 6th International Conference on User Modeling. Wien: SpringerWienNewYork*, pages 289–300. [Cited on page 20]

Weiss, M., Schwarz, F., Jakubowski, S., and Borchers, J. (2010). Madgets: actuating widgets on interactive tabletops. In *Proceedings of the 23nd annual ACM symposium on User interface software and technology*, pages 293–302, New York. [Cited on page 29]

Wilson, K., Boyd, C., Chen, L., and Jamal, S. (2011). Improving student performance in a first-year geography course: Examining the importance of computer-assisted formative assessment. *Computers & Education*, 57(2):1493–1500, doi:10.1016/j.compedu.2011.02.011. [Cited on page 23]

Winn, B. (2008). The design, play, and experience framework. In Ferdig, R., editor, *Handbook of research on effective electronic gaming in education*, pages 1010–1024. IGI Global. [Cited on pages 38 and 156]

Witkin, H.A. and Goodenough, D.R. (1977). Field dependence and interpersonal behavior. *Psychological Bulletin*, 84(4):661–689, doi:10.1037/0033-2909.84.4.661. [Cited on page 20]

VII

Appendices

Questionnaires

A

In this appendix, all the questionnaires that had been used in these studies are presented. The design of the questionnaires is explained. The column labeled with '#' shows the question numbering, and the columns labeled as Pre, Pos1 and Pos2 show the questionnaire that every question appears in. The possible choices to select as answers are placed below the questions.

A.1 Design

For designing the questionnaires, we count with the collaboration of the pedagogues of AIJU and two school teachers that used to collaborate with AIJU. Our experience in the design of previous questionnaires for similar studies significantly influenced this design (e.g., Furió et al. (2015); Furió et al. (2013a,b)). For our previous designs and also for this design, already designed questionnaires were considered. For usability and user experience, the SUS (System Usability Scale (Brooke, 1996)), the CSUQ (Computer System Usability Questionnaire (Lewis, 1995)) or the QUIS (Questionnaire for User Interaction Satisfaction (Chin et al., 1988)) were revised. The presence questions were based on the Slater et al.'s questionnaire (1994). The knowledge questions were extracted from the learning content and they were established by the pedagogues and school teachers involved in the APRENDRA project.

A.2 Learning questions

To measure the knowledge acquired by the children, before and after playing with the game, they were asked with the following questions. These questions were used in every study, as explained in Chapters 6 - 10.

#	Pre	Pos1	Question
Q1	•	•	Which of the following figures did the cavemen paint in the caves? a) Houses b) Deers c) Bisons d) Boats e) Hands f) Carts
Q2	•	•	Tell the name of a cave with cave paintings. a) Bajamira cave b) Miradentro cave d) Altamira cave e) Cave paintings cave
Q3	•	•	Which of the following colours were used for painting in Prehistory? a) Green b) Red c) Violet d) Blue e) Ochre f) Black
Q4	•	•	Ancient Times started with the: a) Invention of the wheel b) Invention of writing c) Discovery of America d) Fall of the Roman Empire e) Invention of the compass
Q5	•	•	Where did the gladiators and beasts fight? a) Roman circus b) Aqueduct c) Amphitheatre d) Castle
Q6	•	•	Which of the following characteristics correspond to Ancient Times? a) Some people lived in castles b) There were aqueducts and amphitheatres c) Mankind started to paint in caves d) The compass was used to navigate

#	Pre	Pos1	Question
Q7	•	•	What is the name of the fortification in front of the walls of the castle that protected the main door from enemies? a) Moat b) Keep c) Barbican d) Defensive tower
Q8	•	•	Which structure surrounds the castle and can be full of water? a) Barbican b) Moat c) Road d) Keep
Q9	•	•	What part of the castle did the Castle's Lord and his family live in? a) Keep b) Barbican c) Wall d) Defensive tower
Q10	•	•	Which event marked the start of the Early Modern Period? a) The invention of writing b) The discovery of America c) The invention of the mobile phone d) The trip to the moon
Q11	•	•	Select the inventions used for sailing in the Early Modern Period. a) Compass b) Television c) Astrolabe d) Map e) Mobile phone f) Spaceship
Q12	•	•	Place the historical ages in the correct order. a) Ancient Times b) the Contemporary Period c) Prehistory d) the Early Modern Period e) the Middle Ages
Q13	•	•	Place each invention in the correct age. a) Map b) Mobile phone c) Cave paintings d) Aqueduct e) Castle

Table A.1: Learning questions (numbered as in the questionnaires)

A.3 System comparison questions

To determine the differences between the two systems the children were asked with the following questions. Each study had different questions for this questionnaire, detailed list is as follows.

A.3.1 Study 1

These questions were used for the study explained in Chapter 6.

#	Pos1	Pos2	Question
Q14	•	•	How much fun did you have? [1-5]
Q15	•	•	Would you recommend this game to friends? [1-5]
Q16	•	•	What was the difficulty of the game? [1. Very difficult / 2. Difficult / 3. Regular / 4. Easy / 5. Very easy]
Q17	•	•	Did you understand the rules of the game? [1-5]
Q18	•	•	Selecting the answers was: [1. Very difficult / 2. Difficult / 3. Regular / 4. Easy / 5. Very easy]
Q19	•	•	How much did you like the images in the game? [1-5]
Q20	•		How much did you like the Clock Avatar (Mr. Tic-Tac)? [1-5]
Q21	•		How much did Mr. Tic-Tac help you during the game? [1-5]
Q22	•	•	How much did you learn during the game? [1-5]
Q23	•	•	Evaluate on a 1-7 scale the sensation of viewing the castle. Did it look like coming out of the screen? [1-7]
Q24	•	•	Did you think you were able to touch the castle? [1-7]

#	Pos1	Pos2	Question
Q25	•	•	Score the game from 1 to 10. [1-10]
Q26	•		Which of all the mini-games did you like the most? [Prehistory / Ancient Times / the Middle Ages / the Early Modern Period / the Contemporary Period]
Q27		•	Which system did you enjoy the most? Why? [Auto / Frontal / Both / None]
Q28		•	Which system was the easiest to use? [Auto / Frontal / Both / None]
Q29		•	Which system was the most comfortable? [Auto / Frontal / Both / None]
Q30		•	Which system did you control better? [Auto / Frontal / Both / None]
Q31		•	In which system were the images viewed better? [Auto / Frontal / Both / None]
Q32		•	Would you recommend any of these systems to friends? [Auto / Frontal / Both / None]
Q33		•	Which system was the most fun? [Auto / Frontal / Both / None]
Q34		•	Would you like to use any of these systems at school? [Auto / Frontal / Both / None]
Q35		•	Would you change anything about the game?

Table A.2: System comparison questions for the study 1 (numbered as in the questionnaires)

A.3.2 Studies 2 and 4

These questions were used for the studies explained in Chapters 7 and 9.

#	Pre	Pos	Question
Q14		•	How much fun did you have? [1-5]

#	Pre	Pos	Question
Q15		•	Would you recommend this game to friends? [1-5]
Q16		•	What was the difficulty of the game? [1. Very difficult / 2. Difficult / 3. Regular / 4. Easy / 5. Very easy]
Q17		•	Did you understand the rules of the game? [1-5]
Q18		•	Selecting the answers was: [1. Very difficult / 2. Difficult / 3. Regular / 4. Easy / 5. Very easy]
Q19		•	How much did you like the images in the game? [1-5]
Q20		•	How much did you like the Clock Avatar (Mr. Tic-Tac)? [1-5]
Q21		•	How much did Mr. Tic-Tac help you during the game? [1-5]
Q22		•	How much did you learn during the game? [1-5]
Q23		•	Score the game from 1 to 10. [1-10]
Q24		•	Which of all the mini-games did you like the most? [Prehistory / Ancient Times / the Middle Ages / the Early Modern Period / the Contemporary Period]

Table A.3: System comparison questions for the studies 2 and 4 (numbered as in the questionnaires)

A.3.3 Study 3

These questions were used for the study explained in Chapter 8. The TQ column shows the type of questions: SA indicates Satisfaction, US indicates usability, and PRE indicates preference.

#	TQ	B	C	D	E	Question
Q14	SA	•	•	•	•	How much fun did you have? [1-5]

#	TQ	B	C	D	E	Question
Q15	SA	•	•	•	•	How much did you learn? [1-5]
Q16	US	•	•	•	•	How difficult was it to learn "The timeline"? [1. Very difficult / 2. Difficult / 3. Regular / 4. Easy / 5. Very easy]
Q17	SA	•			•	Did you think you were able to touch the castle or the bridge? [1-7]
Q18	SA	•	•	•	•	Score the game/activity from 1 to 10. [1-10]
Q19	SA	•			•	Would you recommend this game to friends? [1-5]
Q20	US	•			•	Did you understand the rules of the game? [1-5]
Q21	US	•			•	Selecting the answers was: [1. Very difficult / 2. Difficult / 3. Regular / 4. Easy / 5. Very easy]
Q22	SA	•			•	How much did you like the images in the game? [1-5]
Q23	SA	•			•	How much did you like the Clock Avatar (Mr. Tic-Tac)? [1-5]
Q24	SA	•			•	How much did Mr. Tic-Tac help you during the game? [1-5]
Q25	PRE	•			•	Which of all the mini-games did you like the most? [Prehistory / Ancient Times / the Middle Ages / the Early Modern Period / the Contemporary Period]
Q26	PRE			•	•	Which of the following learning methods did you like the most? [a) Frontal projection (the game). b) Traditional class.]
Q27	PRE			•	•	Why? (Referring to Q26)
Q28	PRE			•	•	What did you like the most about the experience?
Q29	SA	•			•	Would you like to use this game at school to learn other things?

Table A.4: System comparison questions for the study 3

"It's kind of fun to do the impossible."

— Walt Disney