Academic methods for usability evaluation of serious games: a systematic review

Rosa Yáñez-Gómez[1] · **Daniel Cascado-Caballero**[1] · **José-Luis Sevillano**[1]

Abstract In the last years, there has been an increasing interest in the design of video games as a tool for education, training, health promotion, socialization, etc. Usability, which is a key factor in any video game, becomes even more important in these so-called "serious games", where the users' special characteristics should be considered, and the game efficacy depends on the users' adherence and engagement. However, evaluation of the usability of this kind of games requires a redefinition of techniques, methods and even terminology. In this paper, we elicit six research questions and conduct a systematic review of the scientific literature, which resulted in the selection of 187 papers that contained the most relevant responses. The conclusions of this systematic review illustrate the general status of current academic usability evaluations of these games and the main trends in the selection of methodologies and how are they applied. This view may be a very valuable foundation for future research.

Keywords Human-computer interaction · Usability · Usability evaluation · Serious games · Interactive media · Systematic review

1 Introduction

Video games, which are hereafter referred to as games, are one of the most successful and representative examples of interactive media [108]. Furthermore, their interactive nature is one of their most important characteristics. As a result, their study from an Human-Computer Interaction (HCI) perspective is needed [13].

✉ Rosa Yáñez-Gómez
ryanez@us.es

[1] Department of Computer Technology and Architecture, ETS Ingenieria Informatica, Universidad de Sevilla, Avda. Reina Mercedes s/n., 41012 Seville, Spain

However, games "*differ in many respects from productivity software*" [122]. They are special software systems in which the interaction between the gamer and the user interface should be done in ways that do not demand the users' attention and do not interrupt the flow of work [59, 110]. Specifically, "*productivity applications are tools, and the design intention is to make tasks easier, efficient, less error-prone, and increase the quality of the results. Games, instead, are intended to be pleasurable to play and sufficiently challenging in order to provide a good gaming experience. The intention is to reduce the obstacles of fun, rather the obstacles of accomplishments*" [166].

This interaction is usually based on novel concepts, where user satisfaction and performance are specifically catered to as keys to attraction and sales success [110]. Furthermore, the different games genres in existence lead to the coexistence of different control schemes and interface layouts [139].

The usability of games in this scenario becomes crucial [59] and creates new challenges for traditional usability evaluation techniques [122, 139, 166, 172]. The first reported application of usability evaluation techniques to game development was in 1997 at Microsoft [59, 75, 129]. However, although usability of productivity software conforms to a known corpus, the usability of video games is a challenging matter of study.

All of these challenges dramatically increase when considering so-called serious games. Serious games are, basically, "games used for purposes other than mere entertainment" [224]. That is, serious games are "*applications of interactive technology that extend far beyond the traditional videogame market, including: training, policy exploration, analytics, visualization, simulation, education and health and therapy*" [181]. Additionally, many serious games (for instance, those with focus in healthcare) are usually played by users with special interaction needs, such as children or elderly users with different cognitive characteristics and sensory impaired people, as in the treatment of a wide range of pathologies, which can either directly or collaterally affect the interaction scenario. All of these constraints contribute to the special relevance of usability assurance as an unavoidable requirement, which crosses the design process.

Going even further, although designing games for skilled gamers is slightly simplified because there are known game languages, interface conventions and control schemes, the audience in the case of serious games includes non-gamers, which occasionally results in bad experiences because the target user 'does not get games' [150, 219]. These facts lead us to the questions of how to properly evaluate the usability of video games and what the most suitable evaluation techniques are. To shed light on these questions, a search of how these games have been evaluated in the literature is required.

In this regard, we present in this paper a systematic search of the academic literature using only scientific databases, as it is known that Universities and other Academic Institutions are one of the main actors of serious games research [224]. As a result, contributions from the Industry and other professional fora can be considered out of the scope of this paper.

The rest of the paper is organized as follows: in the next section, we explore some of the usability challenges that result from the special characteristics of serious games. This preliminary analysis results in a set of research questions exposed in Section 3, which we sought to answer through a systematic review of the scientific literature. The methodology of this systematic review is also detailed in Section 3, and the results are statistically described in Section 4. A non-systematic analysis of the results and extracted general conclusions are exposed in Section 5, as are the future lines of research opened by this study.

2 Usability evaluation of serious games

In this section, the usability challenges that result from the special characteristics of serious games are preliminarily discussed. In this discussion, a number of research questions arise and are indicated with the acronym RQ. In section 3.1, all of these research questions are recapitulated prior to describing the systematic review.

When evaluating the usability of these games, three approaches are possible: discarding traditional techniques and developing new evaluation techniques that were specially conceived for games; directly applying "off the shelf" techniques under the assumption that differences between games and productivity software are not critical; or adapting existing traditional techniques to overcome these differences [163] (this leads to RQ1). There is no consensus in the research literature regarding which approach should be preferred. Furthermore, some terms and concepts are occasionally confusing in the literature. Therefore, before a systematic search can be performed, an informal review of the literature is needed to categorize the different approaches, clearly define the terms and center our research.

Because an exhaustive description of every single existing evaluation technique, whether new or traditional, is necessarily outside the scope of this article, we will need a framework to refer to them; additionally, references are provided when a specific technique is named in case the reader desires more detailed information. Throughout this article, we refer to the classification shown in Fig. 1. Both traditional and novel usability evaluation techniques can be classified according to the scheme proposed [96, 107, 112, 220, 252, 257]. Over real systems or prototypes, the best alternatives are either evaluations conducted by experts, which are also known as Inspection Methods, or evaluations involving users, which are divided into Inquiry Methods and Testing Methods depending on the methodology adopted. With a more academic focus, Predictive Evaluation offers some predictions regarding the usability of a potential and not-yet-existent prototype.

According to the ISO 9241 definition, the term "usability" covers three different aspects: "effectiveness," "efficiency" and "satisfaction," with usability being *the extent to which a product can be used with effectiveness, efficiency, and satisfaction in a specified context of use* [105]. We will use these three terms as categories with which to classify the differences between games and productivity software, as well as a framework to analyze various usability evaluation techniques when they are applied to games. It is important to note the difference between effectiveness and efficiency: effectiveness is the capability to produce a desired result, whereas efficiency is the ability to produce the result while minimizing the effort, cost or time. According to this framework, usability evaluation techniques focus on task completion and functional adequacy (effectiveness), on performance in terms of the reduction of errors and

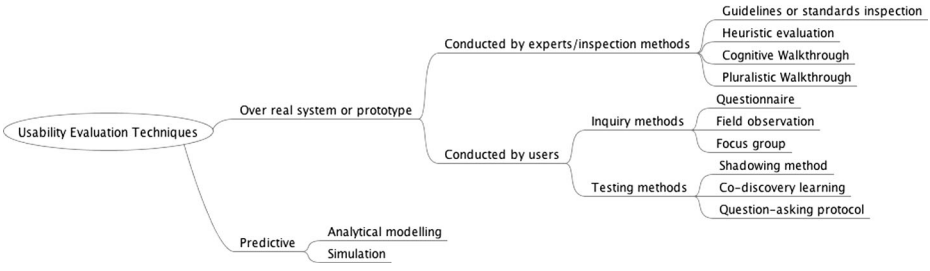

Fig. 1 Usability evaluation technique classification

maximization of interface aids (efficiency), and on the feelings, thoughts or aesthetic appeal and personal preferences of the users (satisfaction).

Starting with **effectiveness**, the differences between games and productivity software lead to difficulties when defining which tasks can be carried out using the game. Both Inspection Methods and user-based evaluations are, in some cases, task-driven: evaluators propose tasks that must be completed using the software being evaluated. Tasks are occasionally described step-by-step in a very detailed fashion, including how to use the tool being evaluated, as is the case with Cognitive Walkthrough [135, 175], Pluralistic Walkthrough [21, 22] and Task Analysis [8, 172, 215]. Pinelle, Wong and Stach [172] warn that "*many usability inspection techniques are not appropriate to games since they either rely on formal specifications of task sequences*" (RQ1.1).

This conflict can perhaps be solved by redefining "tasks" as in-game goals based on ludus rules (winner-loser) [7]. However, it is important to keep in mind that if the tasks are not well defined, the evaluator effect is higher [97], which can add bias to the results. Otherwise, however, the evaluator effect is a measure of reliability only and not of validity [97].

Regarding differences related to **efficiency**, performance in games is understood differently than in other domains (RQ1.2): "*users errors are usually undesirable in other domains, but are expected in many games since they are designed to challenge users and to force them to develop new skills so they can achieve in-game objectives*" [172]. Furthermore, in serious games, "*the pathways of exploration and trial and error loops to help the player acquire new knowledge and skills in the process*" [111, 150].

The definition of terms is again critical: the term "error" should be redefined to avoid confusion with erratic exploration of the interface or in-game narratives. It is "*imperative to differentiate hesitations and error due to a bad HI design from actual trial and errors from the exploratory nature of discovering gameplay elements*" [150]. Indeed, "*Games should be designed to be 'pleasantly frustrating experiences', challenging users beyond their skills*" [125, 150]. "*Usability metrics for serious games should distinguish in-game frustration from at-game frustration*" [81, 150]. According to a more careful definition of "error," game interfaces should be as error-free as productivity software interfaces: "*failure to design usable game interfaces can interfere with the larger goal of creating a compelling experience for users, and can have a negative effect on the overall quality and success of a game*" [172].

Delving deeply into the differences related to efficiency, it is noted that the instructor can interrupt the participant occasionally to ask questions in traditional user-based usability testing, whereas this is not always possible in games because the interruption can cause unnecessary difficulties for the participants [129] (RQ1.3); this is particularly true for some applications of serious games when the cognitive characteristics of the users are especially constrained.

On first consideration, the concept of the third aspect, **satisfaction**, appears to be similar in both games and productivity software. However, the use of the term is actually controversial.

To some authors, the focus of usability in games consists only of "*not presenting challenges not related to the game so that the player is focused only on having fun*" [49]. According to these authors, usability does not address questions about aesthetic appeal or personal preferences. "*Game usability does not address issues of entertainment, engagement, and storyline, which are strongly tied to both artistic issues (...) and technical issues (...)*" [172]. Alternately, if User Experience (UX) is defined as "*a person's perceptions and responses that result from the use and/or anticipated use of a product, system or service*" [64], it "*cannot be neglected in games*" because the user experience will consist "*of both functional and emotional sides, with greater focus on the emotional side, and it changes dynamically over the time*" [61, 192].

Indeed, some new expert review methods have focused primarily on anticipating satisfaction around playability and fun [44, 53, 65, 123, 145] and secondarily on pure effectiveness and efficiency in HCI [172].

New terms such as "player experience" and "playability" have been introduced to address questions of entertainment, engagement or appeal, thereby broadening the concept of usability but also increasing confusion in the terminology. Whereas some authors identify playability merely as "*the instantiation of the general concept of usability when applied to games*" [63], others, such as Olsen, Procci, & Bowers [44], distinguish between usability and playability, with usability referring to the effectiveness and efficiency categories defined previously. "*Usability is a more micro approach, focusing on the independent functionalities within individual components of a system*"; "*playability, on the other hand, focuses on a broader sense of overall functionality associates with the integration of several usable tools, allowing for successful and enjoyable interaction with a game*" [164].

How to distinguish between playability and usability is a key theme in the research literature: It seems that evaluation of both aspects could be based on the same techniques, and the problems detected in the evaluation of games cannot be easily identified as usability or playability issues [160].

For operative reasons, we will establish a terminology applicable to this paper where playability will be specifically identified with satisfaction as one aspect of usability, while the other two aspects, effectiveness and efficiency, will be separately considered as closer to HCI questions. This is only a useful convention of terminology and discards additional aspects, such as narration coherence and game narratives [7, 196], which are also closely related and could also be considered part of the usability concerns. However, our assumption is coherent with various proposals, such as Federoff's [65], who divided his heuristics into three areas: game interface, game mechanics and game playability. These three categories were proposed by Clanton [44] and can be understood as the game environment translation of the three aspects of usability. However, as noted, any attempt to maintain an agreed-upon terminology will encounter difficulties.

After these considerations and with regard to satisfaction, some authors believe that Heuristic Evaluation (HE) [154] could be suitably adapted to measure satisfaction (RQ1.4).

> "Heuristics as seen non relied on task sequences and flexible enough to be reused in new contexts" [139, 172].
> "The heuristic approach has been successfully used in previous game evaluations because it does not make assumptions about task structure, and it is flexible enough to be adapted to specialized domains" [171].

New heuristic guidelines appear to cover new considerations regarding playability and the new interaction concepts [53, 65, 123, 172]. Pinelle, Wong and Stach [172] defend the creation of new heuristic sets because traditional guidelines do not cover issues such as "*proper camera angles when displaying the game world or providing intuitive control mappings*". Pinelle, Wong, Stach, & Gutwin [171] introduce a new set of heuristics for evaluating multiplayer game usability by attending to the interaction primitives called "*mechanics of collaboration*".

Finally, and referring **to the three aspects** of usability (effectiveness, efficiency and satisfaction), user testing is widely recommended in the literature. Although some authors understand playtesting as part of some expert-based methods as well as user-based tests, we will follow the classification reported before in Fig. 1. Accordingly, the term playtesting is used in this article to refer to the application of traditional user testing techniques to games [74,

160, 194, 209] and it covers inquiry and testing methods shown in Fig. 1. Differences are explicitly noted when appropriate.

The previous discussion raises quite a few research questions regarding which techniques are effectively used in the research literature (RQ1), how they address specific differences of games compared with productivity software (RQ1.1, RQ1.2, RQ1.3 and RQ1.4) and how games are evaluated in the literature with regard to their characteristics and those of the evaluation environment (RQ2). All of these questions motivate the following systematic review of the literature.

3 Review method

To obtain answers to the questions presented in the previous section, we conducted a literature review, specifically focusing on methods for usability evaluations of serious games. To perform this review, we followed a systematic and structured method inspired by the guidelines of Kitchenham and Brereton [24, 118]. According to their recommendations, the review protocol validation included the execution of a pilot, and the data extraction was performed by one researcher and checked by the other two [24].

The research questions, the selection of source material and the exclusion and inclusion criteria are described in subsequent subsections.

3.1 Research questions

Our review is driven by the following research questions:

- RQ1: Which is the preferred approach when evaluating the usability of games: discarding traditional techniques and developing new evaluation techniques that are specially conceived for games; directly applying "off the shelf" techniques under the assumption that differences between games and productivity software are not critical; or adapting existing traditional techniques to overcome these differences?

  - RQ1.1: Are task-driven techniques used to evaluate games? Are they adapted in any way?
  - RQ1.2: Is performance in games understood differently than in other domains? Is the "error" concept redefined?
  - RQ1.3: Are instructor interruptions especially avoided during playtesting? Is the Think Aloud technique commonly applied?
  - RQ1.4: Is satisfaction measured with special attention? Are new heuristic guidelines used with this objective?

- RQ2: Which are the most remarkable characteristics of the tested games (e.g. area of study) and of the testing environment (e.g. number/age of users, type of interface, etc.) in the evaluations described in the literature?

3.2 Source material

Both manual and automated methods were applied to select candidate papers from leading journals, conferences and other related events. The knowledge databases selected were

IEEExplore, ACM Library and ISI Web of Knowledge. The date of the search tasks was May 1, 2015, but the results were updated on November 13, 2015, after a delay in applying the manual inclusion and exclusion criteria to the first sets of results.

The inclusion and exclusion criteria are described in the next subsection. The majority of these criteria were applied using the search tools provided by the knowledge databases (search strings are shown in Table 1), but a manual application of the rest of the criteria was necessary to obtain the final selection. The search strings are different because the search tools are based on the different models of the databases, but the manual application of the criteria overcomes these limitations. The quantitative results of the automatic filter are summarized in Table 1.

General terms like "games" and "usability" were used against more specific terms like "serious games" or "usability evaluation" to avoid the risk of missing references in an scenario where terminology used in literature is not uniform and sometimes not consistent.

After removing repeated items from the first 858 results, we obtained 706 items, with 78 of them unavailable to us. After a manual application of the rest of the inclusion and exclusion criteria, we found 441 not to be relevant according to those inclusion and exclusion criteria; the remaining 187 were considered relevant to our research objective [1–6, 9, 10, 12, 15–19, 23, 25, 27–43, 45–48, 50–52, 55–58, 60, 62, 66–73, 76–80, 82–94, 98–100, 102–104, 106, 109, 113–117, 119–121, 124, 126–128, 130–133, 136–138, 140–144, 146–149, 151–153, 159, 161, 162, 165, 167–169, 173, 174, 176, 178–180, 182–191, 193, 195, 197, 199–208, 210, 212–214, 216–218, 221–223, 225–235, 237–240, 242–245, 247–251, 253–256, 258, 259].

3.2.1 Inclusion and exclusion criteria

We included articles that were published in journals, magazines or conference publications and included the words "usability" and "game" in the body, the title or the abstract. The language selected was English, and we did not apply any type of publication date filter.

The use of the term "game" was controversial, but because a more specific term (such as "serious game") would exclude relevant results, we delegated the exclusion of non-relevant studies to further manual filtering. We had to exclude results about Game Theory that were accidentally included in the first set of results; they meet the inclusion criteria but obviously are outside the scope of this paper. Additionally, there were various game-like proposals, such as the simplification of captchas for authentication or systems based on game engines, which were also marked as not relevant for our study. We only include usability testing descriptions when the game is complete and offers an evaluable interface to the user. We also excluded analyses of particular aspects of the interface that excluded the overall interaction with the

Table 1 Number of search results

Knowledge database	Search string	Number of results
IEEExplore	*(("Abstract":usability) AND "Abstract":games)*	191
ACM Library	*(Abstract:usability and Abstract:games) and (PublishedAs:journal OR PublishedAs:proceeding OR PublishedAs:transaction)*	394
ISI Web of Knowledge	*Topic: (usability game)* *Filtered by: kind of document: (ARTICLE OR REPORT OR ABSTRACT OR REVIEW) AND Language: (ENGLISH)* *Period of time: All time*	273

game. Any type of gamification results, understanding gamification as the application of game mechanics to non-game contexts to engage users [54], were also included when the applications were designed using gamification approaches and their usability was evaluated. We also discarded usability tests based on physiological measures because we understood that they are outside the usual tools that are available in general usability laboratories and are instead typically developed ad hoc for a project. However, we included Brain-Computer Interfaces (BCI) because they are considered part of the game to be evaluated and not the evaluation tools.

Usability evaluations are occasionally part of a User-Centered Design approach, but we are interested in the specific usability evaluation sessions that are eventually carried out as part of the methodology, not the entire framework.

The recommended number of users involved in an evaluation is controversial in the literature: according to some authors, it depends on the specific technique applied [101]. We only included tests with five or more users because that seems to be the minimum number that the literature considers adequate to find usability bugs in an average user interface [157, 177, 241]. However, for expert analysis, we did not establish any filter regarding the number of experts running the test.

Additionally, we excluded proposals of new usability evaluation techniques to focus specifically on game usability evaluation. These new proposals occasionally include a test, but we considered them to not be sufficiently representative when they are example applications of the new proposal.

A summary of the inclusion and exclusion criteria is shown in Table 2. The majority of these criteria were applied automatically when using the search tools provided by the knowledge databases (the detailed search strings are provided in Table 1), but other criteria were applied manually, as described in this section.

4 Results of the review

In this section, we describe and analyze the 187 relevant studies that were finally selected from the systematic review.

Table 2 Summary of the inclusion and exclusion criteria

Inclusion criteria
• Terms “usability” and “game” in the Title, Body or Abstract.
• Articles published in journals, conferences (e.g., symposiums, workshops) and magazines.
• Articles written in English.
• Any publication date.
• The article includes at least one usability evaluation of a game.
• If the evaluation is based on user tests/playtesting, more than 4 users should be involved. If the evaluation is based on an Inspection Method, no filter is applied.
• Games whose interfaces are based on Brain-Computer Interaction.
Exclusion criteria
• Game-based solutions and game engine research when they are not referred to as a game usability evaluation.
• Evaluation of the performance and usability of a game when it is based on physiological measures.
• Description of proposals of new usability evaluation techniques specifically focused on games even when they include a usability evaluation example.

We start our analysis by classifying the items according to the area of study where the game being evaluated is to be applied (Fig. 2). We found four main categories: (1) Learning [3, 4, 12, 15–17, 19, 27, 28, 37–40, 47, 55, 60, 67, 68, 71, 76, 77, 85, 86, 93, 98, 100, 102, 104, 109, 115, 120, 128, 131, 137, 146, 147, 149, 167, 174, 176, 178, 179, 185, 188, 191, 197, 200, 201, 203–206, 210, 212, 217, 218, 223, 225, 227, 229, 239, 240, 242–244, 251, 253, 259], including formal or informal approaches; (2) Health [5, 6, 10, 18, 23, 29, 32, 45, 48, 52, 56–58, 66, 69, 72, 73, 79, 87, 91, 92, 94, 99, 103, 114, 119, 121, 127, 130, 140, 159, 162, 169, 173, 180, 183, 184, 186, 195, 213, 214, 216, 222, 226, 231–233, 238, 245, 247–249, 255], which can be divided into the categories shown in Fig. 3; (3) Accessibility [152, 161, 202, 207, 221], including any game that is not in the previous categories and focuses on the inclusion of people with sensory impairments; and (4) Others [1, 2, 9, 25, 30, 31, 33–36, 41–43, 46, 50, 51, 62, 70, 78, 80, 82–84, 88–90, 106, 113, 116, 117, 124, 126, 132, 133, 136, 138, 141–144, 148, 151, 153, 165, 168, 182, 187, 189, 190, 193, 199, 208, 228, 230, 234, 235, 237, 250, 254, 256, 258], which serves as a miscellanea category, where we included analysis of networked commercial games or mobile games that did not pay attention to any of the previous specific areas. This final group of tests included training tools for aircraft pilots, design tools for architects or engineers and analysis of famous commercial games, among others.

As defined in Section 2, the objectives of usability evaluations can be focused on different aspects of usability because the term usability groups different refinements of the human-interface interaction. We recorded the terms used in the selected studies to describe their objectives. The controversial use of terms produces a wide range of terms (see Fig. 4), but "satisfaction," "acceptance" and "engagement" are the three most common declared goals.

The trends in publications of studies regarding usability of games are shown in Fig. 5. The interest in this topic remains on an upward trend. Regarding the type of publication, 45 % of the articles studied were published in journals, and 55 % were published in conferences, workshops, or symposiums.

With regard to the type of interface used to interact with the games, 42 % of studies addressed very innovative designs, such as respiratory interfaces or interaction with the physical space where the game is to be played. The results of the categorization of interfaces are shown in Fig. 6. As said before, we included BCI interfaces when they are part of the game and not the evaluation tools. To address controversial terminology when classifying virtual reality solutions, we decided to divide games according to the manner by which they recognize

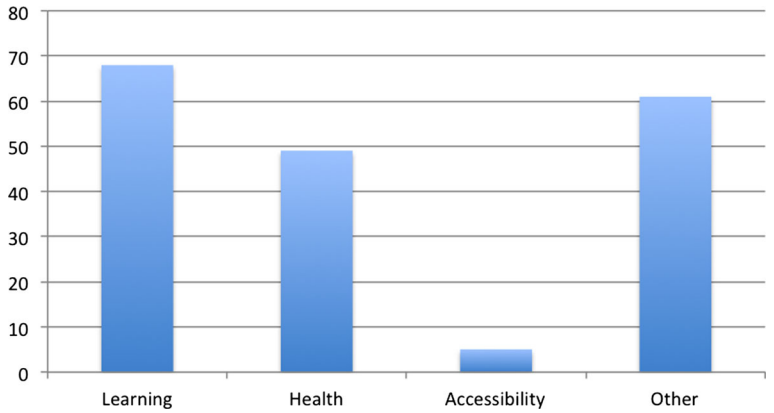

Fig. 2 Areas where usability evaluations are applied

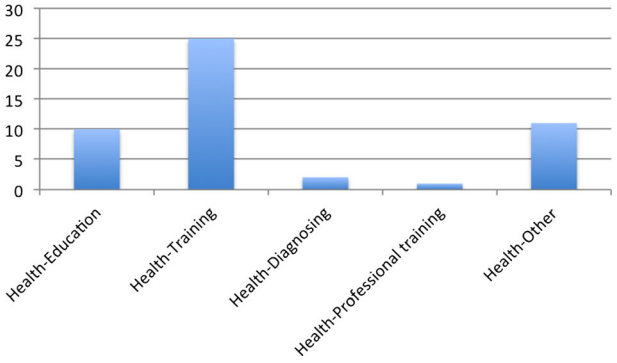

Fig. 3 Sub-categories of the application of video games in the health domain

the body of the player: through computer vision or using other sensors. We only considered sensors to be used indoors and established a specific category of wearable interfaces when the sensors are unplugged and able to be used outdoors. These last three categories (body recognition by computer vision, body recognition by other sensors and body recognition by wearable sensors) refer only to the way in which the gamer body is detected, and they do not make any assumptions about the degree of immersion in the game. Augmented reality interfaces were classified independently because their goal is recognition of the environment rather than the gamer's body. There is only one study of a game where environment recognition is achieved through sensors. It is placed in the Interactive Space category.

Paying attention to the type of evaluation techniques applied in the different tests, we found that only 9 % of the studies used some form of Inspection Methods; in more than half of these cases (5 % of the total), they were complementary user tests that were carried out together and described in the same article. Further, 13 % of the studies include more than one user-based evaluation.

In the case of items including users tests, we can classify them according to the age of the users taking part in the tests, as shown in Fig. 7. The category "students" covers a wide and imprecise range of ages from preadolescents to university students including PhD students. "Elderly" includes those papers declared to work with this population but age ranges vary from different authors: 55, 60 or 65 are common low limits. "Others" includes analysis over adults of different ages, professionals of any kind and heterogeneous groups of users. However, it is also relevant to distinguish healthy users from users who are suffering from an illness or impairment; the latter can especially interfere with or modify the way they interact

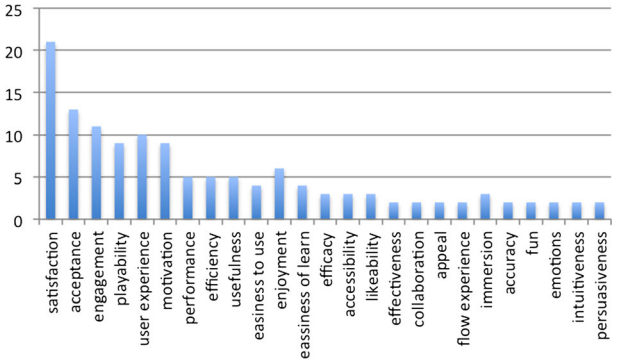

Fig. 4 Different aspects of usability declared as goals of tests with a frequency of appearance of at least two in the review

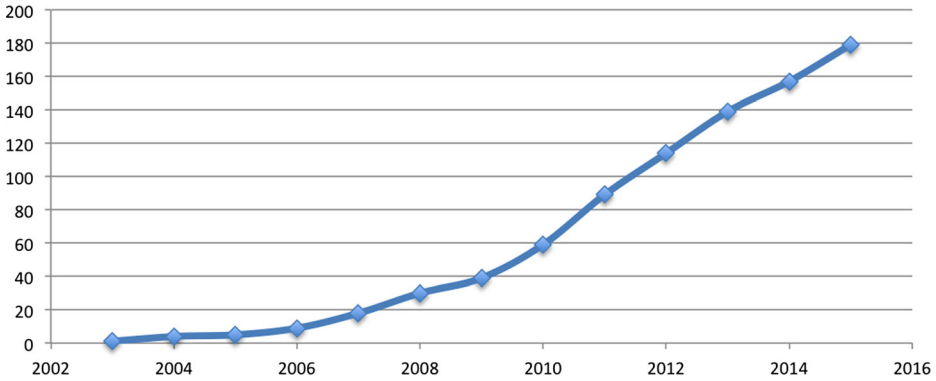

Fig. 5 Trends of publications

with the game. It is important to remember that, in this case, games are usually applied as a therapeutic tool, and it is important that they perfectly meet gamers' needs. Eighty-three percent of studies analyzed were carried out on healthy users, and 17 % of studies involved patients or a combination of users with different capabilities.

The recommended number of users involved in an evaluation is controversial in the literature: according to some authors, it depends on the specific technique applied [101, 236]. As noted, we only included tests with five or more users given that the literature considers this to be an adequate number of individuals to find usability bugs in an average user interface [157, 177, 241]. However, for expert analysis, we did not establish any filter regarding the number of experts running the test.

Given this filter, Fig. 8 shows the frequency histogram of the number of users who participated in the tests in the relevant studies. It can be seen that the most common sets of users include 10, 15 or 20 users. The same analysis of expert-based tests yields that the most frequent numbers of experts involved in a given evaluation are three, four and five, as shown in Fig. 9.

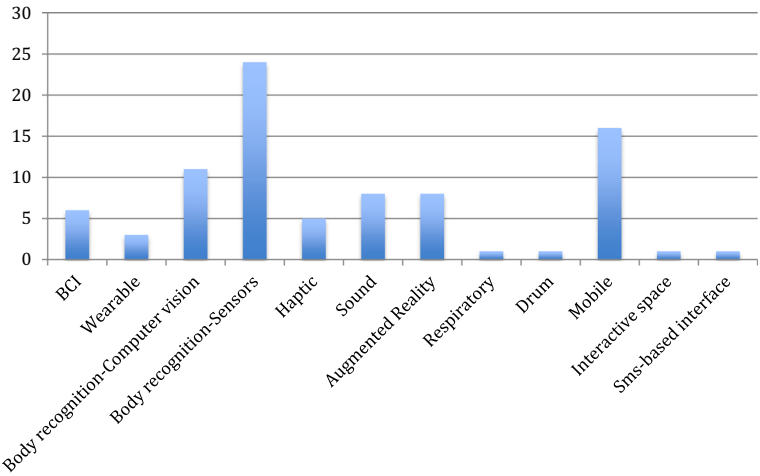

Fig. 6 Innovative interfaces used in 42 % of the tests

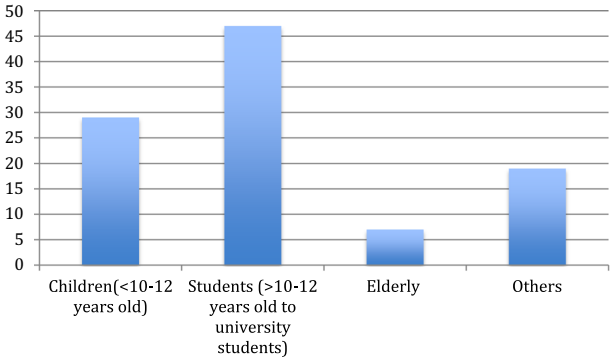

Fig. 7 Classifications of users involved in user tests according to their ages

Additional information about the number of individuals involved in any type of test is shown in Table 3.

Regarding information about individuals involved in the tests, 48 % of studies provide information about gender and 61 % give information about the ages of the users; 91 % specify the range of ages, and, simultaneously or not, 41 % provide the average age of users. Only 3 % of studies specify that users where paid in any form.

Statistical analysis of the results is conducted in 14 % of the studies. These statistical validations are especially prevalent when the objective of the study is to compare different interfaces.

With regard to the techniques applied, user-testing techniques are divided into those based on playtesting (93 %) and those based on a post-use analysis of user satisfaction or performance (7 %), with the main difference being the time taken by the user to explore the interface; if the time exceeds a unique session, it is considered a test post-use and not a playtest. The frequency histograms of techniques for both types of user evaluations are shown in Figs. 10 and 11, respectively. Note that these figures show the number of papers using these techniques, which are applied either in isolation or combined into the different tests. In both cases it can be seen that the most frequent technique applied is the completion of ad hoc questionnaires by the users, followed by interviews, standard questionnaires, direct observation, logs or performance measures, Think Aloud method [158], analysis of recorded sessions, adaptation of standard questionnaires, Focus Group [156], SEEM (Structured Expert Evaluation Method) [11], Heuristic Evaluation [154], Cognitive Walkthrough [246] and Wizard of Oz [20]. Among standard questionnaires, the second most frequent technique, the most frequently used are

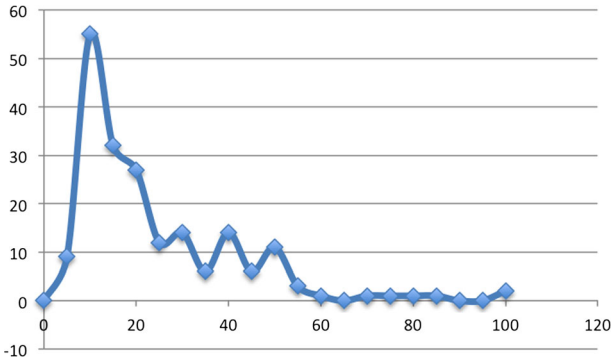

Fig. 8 Frequency histogram of the number of users involved in the tests

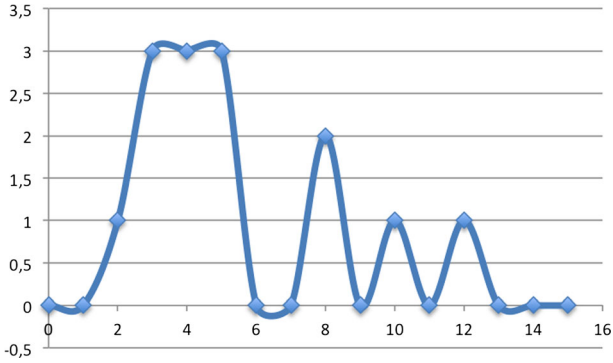

Fig. 9 Frequency histogram of the number of experts involved in the tests

System Usability Scale (SUS) [26] in 32% of cases, End-User Usability Questionnaire) [198] in 5% of cases and the Nasa Task Load Index questionnaire (NASA-TLX) [95] in 7% of cases.

If we filter the results according to the area of study where the game is being evaluated, the differences are not very relevant. In the Health domain, ad hoc questionnaires are also the most frequent technique (42 %) followed by standard questionnaires (17 %) and interviews (10 %). On the other hand, in the Learning domain ad hoc questionnaires are again the most frequent technique (46 %), followed by observation (12,5 %), standard questionnaires (11 %) and heuristic evaluation (11 %).

The equivalent analysis for experts shows that heuristic evaluation is the inspection method more frequently used (see Fig. 12). Other methods such as Cognitive Walkthrough [246], Focus Group [156] or PSSUQ (Post-Study System Usability Questionnaire) [134] are only marginally used. Among the heuristic evaluations, the heuristic checklist by Nielsen [155] is the most commonly used followed by Schneiderman's proposal [211], Bastien and Scapin's [14], ad hoc checklists and Pierotti's [170] (see Fig. 13). Note that Fig. 13 counts the number of times a given heuristic test is used, even if it is applied several times in a single paper.

It is also interesting to analyze the evolution of the preferred techniques along the years. Figure 14 shows the evolution of the six most frequent techniques in the last years. Again, the use of ad hoc questionnaires is in good shape, with standard questionnaires being the second most used in the last years. Observation was not unusual for a few years but it has not been used recently.

Table 3 Summary of the number of users and experts involved in the evaluations analyzed

Expert tests		
	Mode	4
	Minimum	2
	Maximum	58
	Median	5
	Mean	8.3
User tests		
	Mode	10
	Minimum	5[a]
	Maximum	735
	Median	19
	Mean	36.1

[a]tests with fewer users were excluded from the analysis even if they were realized as a complement to other evaluations

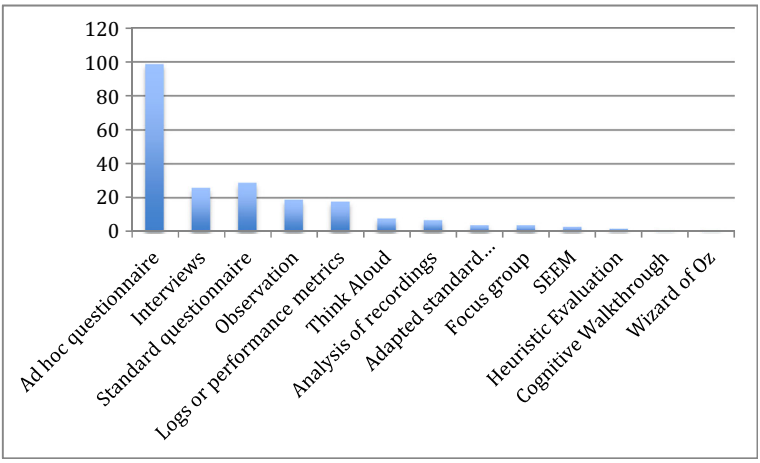

Fig. 10 Evaluation techniques applied in playtesting studies

5 Discussion

Returning to the research questions posed in Section 3 and based on the data collected from the systematic review, we found a wide dispersion of responses.

Focusing on RQ1 –*Which is the preferred approach when evaluating usability of games?*–, as thoroughly explained in Section 2, there are three potential approaches for evaluating the usability of games: the direct application of traditional methodologies, the adaptation of them, or the use of new proposals. According to the data collected from the systematic review, we can conclude that traditional usability evaluation techniques are more prevalent when evaluating games, not only regarding playtesting and related tools but also when using expert-based tests.

Traditional evaluation techniques are preferred, even with 42 % of interfaces being very innovative and defining new interactions with users (Fig. 6). Overall, 93 % of the tests were developed with user-based testing or playtesting (equivalent terms in our terminology), and data were collected in the majority of cases using traditional tools, such as questionnaires, interviews, and observations (Figs. 10 and 11). Only in 9 % of evaluations was some form of traditional expert-based test applied (Fig. 12). The marginal use of expert-based tests is likely due to the recommendation of the application in the early phases of design and to the

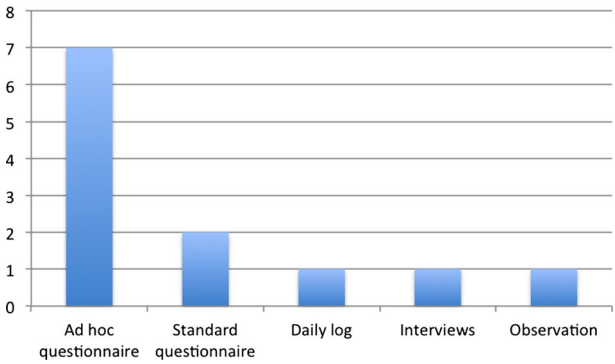

Fig. 11 Evaluation techniques applied in post-use analysis

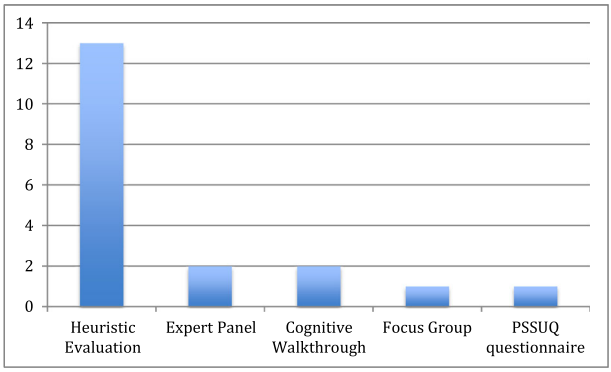

Fig. 12 Evaluation techniques applied in expert-based tests

development of the majority of evaluations over completely functional prototypes or completed versions of the games. Avoiding expert-based tests removes an important tool in usability assurance: they are a lower-cost method than user-based tests and facilitate the detection of the majority of significant usability issues in every phase of software development [220]. It is impossible to conclude whether the application of expert-based methods could have improved the results of user-based game tests, but it could be a good hypothesis to evaluate in future research.

However, a second approach was possible: adapting traditional methodologies to games. This leads us to the sub-question of RQ1.1: *Are task-driven techniques used for game evaluation? Are they adapted in any way?* In Section 2, some of the substantial differences noted in the literature when comparing the evaluation of games with that of productivity software were exposed. The first one addresses the task descriptions required in task-driven methodologies. When this type of methodology is used in the set of articles studied in the review, the tasks selected were very small and clear (such as "complete login" and "buy a new item for the character"). These definitions are far from complex use cases and are described step-by-step when applying these techniques to productivity software evaluations. However, the majority of playtesting sessions were not task-driven, and exploratory navigation was the preferred interaction in the studied articles.

Regarding to RQ1.2 *–Is performance in games understood differently than in other domains? Is the "error" concept redefined?–*, 10 % of the playtesting sessions involved the collection of logs or performance metrics. However, due to the reasons explained in Section 2,

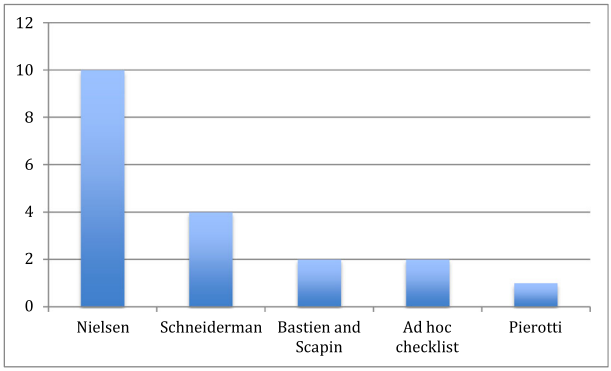

Fig. 13 Heuristic checklists used in HE-based tests

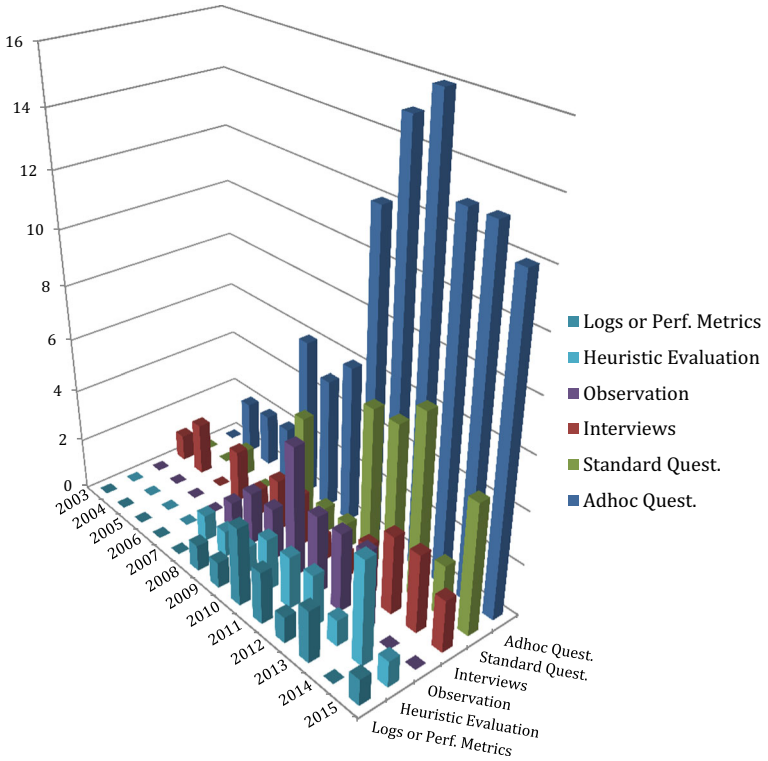

Fig. 14 Evolution of the use of the six most frequent evaluation techniques in the last years

errors are redefined, and the performance of small tasks is measured to test the clarity of the interface and the adequacy of the metaphors, but not to evaluate the ease of completion of complex use cases as in productivity software. This type of metric is especially useful in tests whose objective is to compare alternative interfaces.

Concerning RQ1.3 –*Are specially avoided instructor's interruptions during playtesting? Is Think Aloud technique commonly applied?–,* although the interruption of gamers when using the game is avoided and interviews and questionnaires are posed when the playing phase is ended, the Think Aloud Method is used in 5 % of cases, either in combination or not with observation or recordings, among other activities. However, the convenience of using this method is directly related to the type of users targeted, their cognitive capabilities and the environment of the test.

With respect to RQ1.4 –*Is satisfaction measured with special attention? Are new heuristic guidelines used with this objective?–*, the most prevalent method to measure the satisfaction of users is the posing of questionnaires to the users after playtesting sessions. However, standard questionnaires (the only ones statistically validated) were used in only 17 % of the cases, compared with the 58 % of the playtesting sessions in which ad hoc questionnaires were used (also 17 % of standard questionnaires compared with 58 % of ad hoc questionnaires in the case of testing based on the use of games during a longer period of time). A summary of the methods used is shown in Figs. 10 and 11. Ad hoc questionnaires prevail in all domains (health, learning) and their use has not decreased in the last years, as shown in Fig. 14.

It seems that, despite their statistical validity, standard questionnaires are not capable of capturing specific information of the game being tested, and researchers tend to add personalized questions about what they consider the most controversial aspects of their design. There is a gap between the focus of standard tests and the information required in practice. This results in the conclusions of studies being particular results and not generalizable because the tools used are not statistically validated.

As exposed in Section 2, many new heuristic guidelines designed especially for games attempt to cover aspects close to satisfaction. However, despite the relatively high number of new heuristic checklists proposed for games [53, 65, 123, 172], traditional heuristics guidelines are preferred. In particular, the heuristic checklist by Nielsen [155] prevails over all other checklists (Fig. 13).

The research question RQ2 –*Which are the most remarkable characteristics of the tested games and the testing environment in evaluation described in the literature?*–, when applied over the set of articles from the systematic review, draws very interesting conclusions. As shown in Fig. 2, the areas most interested in usability of games are learning and health. The use of serious games to improve learning effectiveness and therapeutic effect (the evaluation of these impacts is usually developed in parallel with usability evaluations) is a matter of study because the special characteristics of the targeted users need to be addressed. In the case of learning, children are commonly the targeted users of games, and extracting information about usability and satisfaction from them is challenging. Observation or interviews with their teachers and families are usually applied as complement to the direct testing with children. Simplification and adaptation of questionnaires is also needed when capturing information from children.

In the case of serious games applied to health interventions, the typical users are people with sensory impairments, children (as in the previous case) and elderly, with their own specific constraints on sensory and cognitive capabilities. These special types of users also make paying attention to usability particularly necessary.

Regarding the users involved in the tests of any types of games (Fig. 7), we can detect that students are the most frequently recruited, either because they are the target users or because they are more accessible in schools or universities.

As shown in Fig. 4, the most prevalent objective when evaluating games is to measure the satisfaction, acceptance and engagement of users. Usually, this interest is caused by the known improvement of the learning effectiveness and therapeutic effect when users are motivated.

As thoroughly explained, the number of users or experts involved in tests is a controversial point in the literature [101, 157, 177, 236, 241]. In practice, there is a wide variability in the number of individuals recruited for the tests, as shown in Figs. 8 and 9. However, the most frequent values are coherent with the most cited recommendations in the literature. Future studies about the effects of different numbers of individuals in tests could draw some conclusions, but thus far only some approaches have been conducted.

6 Conclusions and future work

In summary, this paper presents a systematic review of the academic and scientific research literature on usability evaluation methods in video games, with a particular focus on the so-called “serious games”. Based on this review, we formulate some research questions and try to answer them in section 5. With regard to some of the most controversial aspects of usability

evaluation, it is difficult to extract definite conclusions from the evaluations in the articles studied: the number of individuals needed to be involved, the reliability or validity of new methodologies, etcetera; nor about other questions, such as which demographic information is necessary to capture from the population of the test (e.g., genre, range of age, average age) or if statistical analysis is indispensable in proving the validity of results. However, this systematic review illustrates the general status of current usability evaluations and the main trends in the selection of methodologies and how are they applied. An in depth description or analysis of the pros and cons of each of these methods is out of the scope of the paper but this view of the current status of usability of games is a valuable foundation upon which to develop future research not only on video games, but also on other interactive media. An exhaustive list of references is available to future game developers and enables them to check what usability evaluation methods have been used previously in literature and in what context, always with the possibility of query the original sources.

As shown in Fig. 5, interest in usability of games grows on an upward trend. Despite the lack of consensus in the use of terminology or the corpus of techniques to be used, the application of games on health and learning areas leads to the need of an special attention on usability. Our results show that the ad hoc questionnaires prevail in all areas (health, learning), and this has not changed in the last years. There is not any standardized technique or methodology being embraced by researchers. We believe that there is a need for a new evaluation methodology that can capture the special features of usability of serious games and can be accepted as a standard for the scientific community.

As mentioned in the Introduction, our main interest is in research in the academic community, as Universities and other Academic Institutions are one of the main actors of serious games research [224]. For this reason, our review uses scientific databases and therefore some relevant contributions from the Industry may be missed. A new search from industry/professional fora, conferences and reports should be made, and could also be the topic for further work.

During the systematic review, we marked as not relevant to our research questions quite a few new proposals, which likely try to advance the area of knowledge and improve the methodologies and tools. An interesting future work is to analyze these new proposals to determine why their use is not popular in the literature. Additionally, we find indispensable a consensus on the use of terms to form a solid base for future contributions to this area of knowledge.

Acknowledgments This work has been supported by project PROCUR@-IPT-2011-1038-900000, funded by the program INNPACTO of MINECO and FEDER funds; and by the Telefonica Chair “Intelligence in Networks” of the Universidad de Sevilla, Spain.

References

1. Abdulhak SA, Cha KH, Kang DK (2011) Evaluation of online social network games usability using verbal protocol analysis. In: 2011 6th International Conference on Computer Sciences and Convergence Information Technology (ICCIT), pp 7–11. IEEE
2. Abeele VV, Zaman B, Abeele MV (2008) The unlikeability of a cuddly toy interface: an experimental study of preschoolers’ likeability and usability of a 3D Game played with a cuddly toy versus a keyboard. In Fun and Games, pp 118–131. Springer, Berlin
3. Adamo-Villani N, Wright K (2007) SMILE: an immersive learning game for deaf and hearing children. In: ACM SIGGRAPH 2007 educators program. ACM, p 17

4. Ahmad WFW, Shaarani ARS, Afrizal S (2012) Mobile language translation game. In: Computer & Information Science (ICCIS), 2012 International Conference on (Vol. 2), pp 1099–1104. IEEE
5. Albu M, Atack L, Srivastava I (2015) Simulation and gaming to promote health education: results of a usability test. Health Educ J 74(2):244–254
6. Ancker JS, Kukafka R (2007) A combined qualitative method for testing an interactive risk communication tool. In: AMIA Annual Symposium Proceedings (Vol. 2007), p 16. American Medical Informatics Association
7. Ang CS, Avni E, Zaphiris P (2008) Linking pedagogical theory of computer games to their usability. Int J E-Learning 7(3):533–558
8. Annett J, Duncan KD (1967) Task analysis and training design. Occup Psychol 12:211–221
9. Arroyo-Palacios J, Romano DM (2009) Exploring the use of a respiratory-computer interface for game interaction. In Games Innovations Conference, 2009. ICE-GIC 2009. International IEEE Consumer Electronics Society's, pp 154–159. IEEE
10. Aung YM, Al-Jumaily A, Anam K (2014) A novel upper limb rehabilitation system with self-driven virtual arm illusion. In: 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp 3614–3617. IEEE
11. Baauw E, Bekker MM, Barendregt W (2005) A structured expert evaluation method for the evaluation of children's computer games. Proceedings of human–computer interaction INTERACT 2005, Rome, 12–16 September 2005, pp 457–469
12. Baauw E, Bekker MM, Markopoulos P (2006) Assessing the applicability of the structured expert evaluation method (SEEM) for a wider age group. In: Proceedings of the 2006 conference on Interaction design and children, pp 73–80. ACM
13. Barr P (2008) Video game values: Play as human-computer interaction. PhD Thesis, Victoria University of Wellington
14. Bastien J, Scapin D (1991) Ergonomic criteria for the evaluation of human-computer interfaces, Rapport de Recherche N. 156. INRIA, France
15. Bekker MM, Baauw E, Barendregt W (2008) A comparison of two analytical evaluation methods for educational computer games for young children. Cogn Tech Work 10(2):129–140
16. Bellotti F, Berta R, De Gloria A, D'ursi A, Fiore V (2012) A serious game model for cultural heritage. J Comput Cult Herit 5(4):17
17. Bellotti F, Ferretti E, De Gloria A (2005) Discovering the European heritage through the ChiKho educational web game. In: Intelligent Technologies for Interactive Entertainment, pp 13–22. Springer, Berlin
18. Benveniste S, Jouvelot P, Péquignot R (2010) The MINWii Project: Renarcissization of patients suffering from Alzheimer's Disease through video game-based music therapy. In: Entertainment Computing-ICEC 2010, pp 79–90. Springer, Berlin
19. Bernhaupt R, Schwaiger D, Riegler S, Enthaler D (2007) Evaluating children's gaming experiences. In: Proceedings of the international conference on Advances in computer entertainment technology, pp 224–225. ACM
20. Bernsen NO, Dybkjær H, Dybkjær L (1994) Wizard of oz prototyping: how and when. Proc. CCI Working Papers Cognit. Sci./HCI, Roskilde, Denmark
21. Bias R (1991) Interface-walkthroughs: efficient collaborative testing. IEEE Softw 8(5):94–95
22. Bias RG (1994) The pluralistic usability walkthrough: coordinated empathies. In: Usability inspection methods. John Wiley & Sons, Inc., p 63–76
23. Boulay M, Benveniste S, Boespflug S, Jouvelot P, Rigaud AS (2011) A pilot usability study of MINWii, a music therapy game for demented patients. Technol Health Care 19(4):233–246
24. Brereton P, Kitchenham BA, Budgen D, Turner M, Khalil M (2007) Lessons from applying the systematic literature review process within the software engineering domain. J Syst Softw 80(4):571–583
25. Bronner S, Pinsker R, Noah JA (2015) Physiological and psychophysiological responses in experienced players while playing different dance exer-games. Comput Hum Behav 51:34–41
26. Brooke J (1996) SUS-A quick and dirty usability scale. Usability Eval Ind 189(194):4–7
27. Brown DJ, McHugh D, Standen P, Evett L, Shopland N, Battersby S (2011) Designing location-based learning experiences for people with intellectual disabilities and additional sensory impairments. Comput Educ 56(1):11–20
28. Browne K, Anand C (2013) Gamification and serious game approaches for introductory computer science tablet software. In: Proceedings of the First International Conference on Gameful Design, Research, and Applications, pp 50–57. ACM
29. Brown-Johnson CG, Berrean B, Cataldo JK (2015) Development and usability evaluation of the mHealth Tool for Lung Cancer (mHealth TLC): a virtual world health game for lung cancer patients. Patient Educ Couns 98(4):506–511

30. Burns M, Colbert M (2013) Web analytics: the new purpose towards predictive mobile games. In: Advances in computer entertainment, pp 1–13. Springer International Publishing
31. Cabral C, Dehanov J, Dias JMS, Bastos R (2005) Developing games with Magic Playground: a gesture-based game engine. In: Proceedings of the 2005 ACM SIGCHI International Conference on Advances in computer entertainment technology, pp 361–362. ACM
32. Cameirão MS, Badia SB, Oller ED, Verschure PFMJ (2010) Neurorehabilitation using the virtual reality based Rehabilitation Gaming System: methodology, design, psychometrics, usability and validation. J Neuroeng Rehabil 7(1):48
33. Cao Y, McGill WL (2013) LinkIT: a Ludic elicitation game for eliciting risk perceptions. Risk Anal 33(6): 1066–1082
34. Caputo FM, Giachetti A (2015) Evaluation of basic object manipulation modes for low-cost immersive Virtual Reality. In: Proceedings of the 11th Biannual Conference on Italian SIGCHI Chapter, pp 74–77. ACM
35. Carrigy T, Naliuka K, Paterson N, Haahr M (2010) Design and evaluation of player experience of a location-based mobile game. In: Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries, pp 92–101. ACM
36. Carter C, Mehdi Q, Hartley T (2012) Navigational techniques to improve usability and user experience in RPG games. In: 2012 17th International Conference on Computer Games (CGAMES), pp 86–93. IEEE
37. Centieiro P, Romão T, Dias AE (2011) A location-based multiplayer mobile game to encourage pro-environmental behaviours. In: Proceedings of the 8th International Conference on Advances in Computer Entertainment Technology, pp 31. ACM
38. Chammas A, Moraes A, Teixeira EA (2011) Ergonomics and usability of children interfaces: spore®, the case study. Work (Reading, Mass) 41:1024–1031
39. Chang SB, Deng YC, Cheng HN, Liao HC, Yu FY, Chan TW (2007) Implementation and evaluation of EduBingo for arithmetic drill. In: The First IEEE International Workshop on Digital Game and Intelligent Toy Enhanced Learning. DIGITEL'07. pp 99–103. IEEE
40. Chee YS, Tan EM, Liu Q (2010) Statecraft X: enacting citizenship education using a mobile learning game played on Apple iPhones. In: 2010 6th IEEE International Conference on Wireless, Mobile and Ubiquitous Technologies in Education (WMUTE), pp 222–224. IEEE
41. Chen KH, Shen KS, Ma MY (2012) The functional and usable appeal of Facebook SNS games. Internet Res 22(4):467–481
42. Choi YJ (2009) Providing novel and useful data for game development using usability expert evaluation and testing. In: Sixth International Conference on Computer Graphics, Imaging and Visualization, 2009. CGIV'09. pp 129–132. IEEE
43. Chung KM, Shin DH (2015) Effect of elastic touchscreen and input devices with different softness on user task performance and subjective satisfaction. Int J Hum Comput Stud 83:12–26
44. Clanton C (1998) An interpreted demonstration of computer game design. In: CHI 98 Conference Summary on Human Factors in Computing Systems. ACM, p 1–2
45. Clawson J, Patel N, Starner T (2010) Dancing in the streets: the design and evaluation of a wearable health game. In: 2010 International Symposium on Wearable Computers (ISWC), pp 1–4. IEEE
46. Cornett S (2004) The usability of massively multiplayer online roleplaying games: designing for new users. In: Proceedings of the SIGCHI conference on Human factors in computing systems, pp 703–710. ACM
47. Couceiro RM, Papastergiou M, Kordaki M, Veloso AI (2013) Design and evaluation of a computer game for the learning of Information and Communication Technologies (ICT) concepts by physical education and sport science students. Educ Inf Technol 18(3):531–554
48. Cruz VT, Pais J, Bento V, Mateus C, Colunas, M., Alves I, … Rocha NP (2013) A rehabilitation tool designed for intensive web-based cognitive training: description and usability study. JMIR Res Protocols, 2(2)
49. Cybis W, Holtz A, Faust R (2010) Ergonomia e usabilidade. Novatec, São Paulo
50. Dahlan Bin Abdul Ghani (2015) Wayang fighters: preserving the art of Wayang Kulit (shadow puppets) through game application. In: Proceedings of the 9th International Conference on Ubiquitous Information Management and Communication (IMCOM '15). ACM, New York, Article 26:7
51. de Carvalho BJA, Soares MM, das Neves AMM, Medeiros RP (2013) Interactive doodles: a comparative analysis of the usability and playability of google trademark games between 2010 and 2012. In: Design, user experience, and usability. Health, learning, playing, cultural, and cross-cultural user experience, pp 508–517. Springer, Berlin
52. DeShazo J, Harris L, Turner A, Pratt W (2010) Designing and remotely testing mobile diabetes video games. J Telemed Telecare 16(7):378–382
53. Desurvire H, Caplan M, Toth JA (2004) Using heuristics to evaluate the playability of games. In: CHI'04 extended abstracts on Human factors in computing systems. ACM, p 1509–1512

54. Deterding S, Dixon D, Khaled R, Nacke L (2011) From game design elements to gamefulness: defining gamification. In: Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments, pp 9–15. ACM
55. Diah NM, Ismail M, Ahmad S, Dahari MKM (2010) Usability testing for educational computer game using observation method. In: 2010 International Conference on Information Retrieval & Knowledge Management, (CAMP), pp 157–161. IEEE
56. Diehl LA, Souza RM, Alves JB, Gordan PA, Esteves RZ, Jorge MLSG, Coelho ICM (2013) InsuOnline, a serious game to teach insulin therapy to primary care physicians: design of the game and a randomized controlled trial for educational validation. JMIR Res Protocols, 2(1)
57. Direito A, Jiang Y, Whittaker R, Maddison R (2015) Smartphone apps to improve fitness and increase physical activity among young people: protocol of the Apps for IMproving FITness (AIMFIT) randomized controlled trial. BMC Public Health 15(1):635
58. Dores AR, Carvalho IP, Barbosa F, Almeida I, Guerreiro S, Leitão M, … Castro-Caldas A (2011) Serious games: are they part of the solution in the domain of cognitive rehabilitation?. In: Serious Games Development and Applications, pp 95–105. Springer, Berlin
59. Dyck J, Pinelle D, Brown B, Gutwin C (2003) Learning from games: HCI design innovations in entertainment software. In: Graphics interface, p 237–246
60. Ebner M, Holzinger A (2007) Successful implementation of user-centered game based learning in higher education: an example from civil engineering. Comput Educ 49(3):873–890
61. Eladhari MP, Ollila EM (2012) Design for research results experimental prototyping and play testing. Simul Gaming 43(3):391–412
62. El-Tayeh A, Gil N, Freeman J (2008) A methodology to evaluate the usability of digital socialization in 'virtual' engineering design. Res Eng Des 19(1):29–45
63. Fabricatore C, Nussbaum M, Rosas R (2002) Playability in action videogames: a qualitative design model. Hum Comput Interact 17(4):311–368
64. FDIS, I. 9241-210 (2009) Ergonomics of human system interaction-Part 210: human-centered design for interactive systems
65. Federoff MA (2002) Heuristics and usability guidelines for the creation and evaluation of fun in video games (Doctoral dissertation, Indiana University)
66. Ferracani A, Pezzatini D, Seidenari L, Del Bimbo A (2014) Natural and virtual environments for the training of emergency medicine personnel. Univ Access Inf Soc, 1–12
67. Ferraz M, Romão T, Câmara A (2010) The Alberto's Gravimente Toys: children's fiction on technological design. In: Proceedings of the 7th International Conference on Advances in Computer Entertainment Technology, pp 7–10. ACM
68. Ferrer V, Perdomo A, Rashed-Ali H, Fies C, Quarles J (2013) How does usability impact motivation in augmented reality serious games for education?. In: 2013 5th International Conference on Games and virtual worlds for serious applications (VS-GAMES), pp 1–8. IEEE
69. Fitzgerald D, Trakarnratanakul N, Dunne L, Smyth B, Caulfield B (2008) Development and user evaluation of a virtual rehabilitation system for wobble board balance training. In: Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE, pp 4194–4198. IEEE
70. Fitz-Walter Z, Tjondronegoro D, Koh D, Zrobok M (2012) Mystery at the library: encouraging library exploration using a pervasive mobile game. In: Proceedings of the 24th Australian Computer-Human Interaction Conference, pp 142–145. ACM
71. Fitz-Walter Z, Wyeth P, Tjondronegoro D, Scott-Parker B (2013) Driven to drive: Designing gamification for a learner logbook smartphone application. In: Proceedings of the First International Conference on Gameful Design, Research, and Applications, pp 42–49. ACM
72. Franck LS, Jones M (2003) Computer-taught coping techniques for venepuncture: preliminary findings from usability testing with children, parents and staff. J Child Health Care 7(1):41–54
73. Frutos-Pascual M, Zapirain BG, Zorrilla AM (2014) Adaptive tele-therapies based on serious games for health for people with time-management and organisational problems: preliminary results. Int J Environ Res Public Health 11(1):749–772
74. Fullerton T, Swain C, Hoffman S (2004) Game design workshop: designing, prototyping, & playtesting games. CRC Press
75. Fulton B, Romero R (2004) User-testing in a hostile environment: overcoming resistance and Apathy in your game company. In: Game Developer's Conference, San Jose CA
76. Furió D, González-Gancedo S, Juan MC, Seguí I, Rando N (2013) Evaluation of learning outcomes using an educational iPhone game vs. traditional game. Comput Educ 64:1–23
77. Furió D, Juan MC, Seguí I, Vivó R (2015) Mobile learning vs. traditional classroom lessons: a comparative study. J Comput Assist Learn 31(3):189–201

78. Gamberini L, Corradi N, Zamboni L, Perotti M, Cadenazzi C, Mandressi S, … Aman P (2011) Saving is fun: designing a persuasive game for power conservation. In: Proceedings of the 8th International Conference on Advances in Computer Entertainment Technology, p 16. ACM
79. Ganesan S, Anthony L (2012) Using the kinect to encourage older adults to exercise: a prototype. In: CHI'12 Extended Abstracts on Human Factors in Computing Systems, pp 2297–2302. ACM
80. Gerling KM, Klauser M, Niesenhaus J (2011) Measuring the impact of game controllers on player experience in FPS games. In: Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments, pp 83–86. ACM
81. Gilleade KM, Dix A (2004) Using frustration in the design of adaptive videogames. In: Proceedings of the 2004 ACM SIGCHI International Conference on Advances in computer entertainment technology. ACM, p 228–232
82. Goh DHL, Lee CS, Chua AYK (2010) Do games motivate mobile content sharing? Springer, Berlin, pp 61–70
83. Goh DHL, Razikin K, Chua AY, Lee CS, Tan KT (2011) Understanding location-based information sharing in a mobile human computation game. In: Internet of things (iThings/CPSCom), 2011 International Conference on and 4th International Conference on Cyber, Physical and Social Computing, pp 209–216. IEEE
84. Gonçalves D, Jesus R, Grangeiro F, Romão T, Correia N (2008) Tag around: a 3D gesture game for image annotation. In: Proceedings of the 2008 International Conference on Advances in Computer Entertainment Technology, pp 259–262. ACM
85. González C, Muñoz V, Toledo P, Mora A, Moreno L (2014) EMATIC: an inclusive educational application for tablets. In: Proceedings of the XV International Conference on Human Computer Interaction, p. 94. ACM
86. Goodman D, Bradley NL, Paras B, Williamson IJ, Bizzochi J (2006) Video gaming promotes concussion knowledge acquisition in youth hockey players. J Adolesc 29(3):351–360
87. Guidali M, Duschau-Wicke A, Broggi S, Klamroth-Marganska V, Nef T, Riener R (2011) A robotic system to train activities of daily living in a virtual environment. Med Biol Eng Comput 49(10):1213–1223
88. Gürkök H, Hakvoort G, Poel M, Nijholt A (2011) User expectations and experiences of a speech and thought controlled computer game. In: Proceedings of the 8th International Conference on Advances in Computer Entertainment Technology, p 53. ACM
89. Gürkök H, Hakvoort G, Poel M (2011) Evaluating user experience in a selection based brain-computer interface game A comparative study. In: Entertainment computing-ICEC 2011, pp 77–88. Springer, Berlin
90. Ha J, Park S (2013) A comparative study by age in evaluating web-based 3D model house. Archit Res 15(4):175–182
91. Haggerty K, Yadamsuren B, Kibaru F, Basic J (2013) Children's information needs of a patient engagement system in a hospital room: insights from usability testing in a simulation laboratory. Proc Am Soc Inf Sci Technol 50(1):1–5
92. Halic T, Kockara S, Demirel D, Willey M, Eichelberger K (2014) MoMiReS: Mobile mixed reality system for physical & occupational therapies for hand and wrist ailments. In: Innovations in Technology Conference (InnoTek), 2014 IEEE, pp 1–6. IEEE
93. Hannig A, Kuth N, Özman M, Jonas S, Spreckelsen C (2012) eMedOffice: a web-based collaborative serious game for teaching optimal design of a medical practice. BMC Med Educ 12(1):104
94. Hart P, Bierwirth R, Fulk G, Sazonov E (2014) The design and evaluation of an activity monitoring user interface for people with stroke. In Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE, pp 5908–5911. IEEE
95. Hart SG, Staveland LE (1988) Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. Adv Psychol 52:139–183
96. Heo J, Ham DH, Park S, Song C, Yoon WC (2009) A framework for evaluating the usability of mobile phones based on multi-level, hierarchical model of usability factors. Interact Comput 21(4):263–275
97. Hertzum M, Jacobsen NE (2001) The evaluator effect: a chilling fact about usability evaluation methods. Int J Hum Comput Interact 13(4):421–443
98. Hou HT, Li MC (2014) Evaluating multiple aspects of a digital educational problem-solving-based adventure game. Comput Hum Behav 30:29–38
99. Höysniemi J, Aula A, Auvinen P, Hännikäinen J, Hämäläinen P (2004) Shadow boxer: a physically interactive fitness game. In: Proceedings of the third Nordic conference on Human-computer interaction, pp 389–392. ACM
100. Hwang MY, Hong JC, Jong JT, Lee CK, Chang HY (2009) From fingers to embodiment: A study on the relations of the usability, dependability of the embodied interactive video games and the elders, Äô flow experience. In: Learning by Playing. Game-based Education System Design and Development, pp 464–472. Springer, Berlin
101. Hwang W, Salvendy G (2010) Number of people required for usability evaluation: the 10¬±2 rule. Commun ACM 53(5):130–133
102. Infante C, Weitz J, Reyes T, Nussbaum M, Gómez F, Radovic D (2010) Co-located collaborative learning video game with single display groupware. Interact Learn Environ 18(2):177–195

103. Isleyen F, Gulkesen KH, Cinemre B, Samur MK, Zayim N, Sen KS (2013) Evaluation of the usability of a serious game aiming to teach facial expressions to schizophrenic patients. Stud Health Technol Inform 205: 662–666
104. Ismail M, Diah NM, Ahmad S, Kamal NAM, Dahari MKM (2011) Measuring usability of educational computer games based on the user success rate. In: 2011 International Symposium on Humanities, Science & Engineering Research (SHUSER), pp 56–60. IEEE
105. ISO, S. 9241-11 (1998) Ergonomic requirements for office work with visual display terminals (VDTs)-Part II guidance on usability
106. Istance H, Hyrskykari A, Vickers S, Chaves T (2009) For your eyes only: controlling 3d online games by eye-gaze. In: Human-Computer Interaction-INTERACT 2009, pp 314–327. Springer, Berlin
107. Ivory MY, Hearst MA (2001) The state of the art in automating usability evaluation of user interfaces. ACM Comput Surv 33(4):470–516
108. Jett J, Sacchi S, Lee JH, Clarke RI (2016) A conceptual model for video games and interactive media. J Assoc Inf Sci Technol 67(3):505–517
109. Jimenez-Diaz G, Gonzalez-Calero PA, Gomez-Albarran M (2012) Role-play virtual worlds for teaching object-oriented design: the ViRPlay development experience. Softw: Pract Exp 42(2):235–253
110. Jørgensen AH (2004) Marrying HCI/Usability and computer games: a preliminary look. In: Proceedings of the third Nordic conference on Human-computer interaction. ACM, p 393–396
111. Ju E, Wagner C (1997) Personal computer adventure games: their structure, principles, and applicability for training. SIGMIS Database 28(2):78–92
112. Karat J (1997) User-centered software evaluation methodologies. Handb Hum Comput Interact 2:689–704
113. Karouzaki E, Savidis A (2012) A framework for adaptive game presenters with emotions and social comments. Int J Comput Games Technol 2012:6
114. Khan DU, Ananthanarayan S, Le AT, Schaefbauer CL, Siek K (2012) Designing mobile snack application for low socioeconomic status families. In: 2012 6th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth), pp 57–64. IEEE
115. Kim YS, Collins M, Bulmer W, Sharma S, Mayrose J (2013) Haptics Assisted Training (HAT) System for children's handwriting. In World Haptics Conference (WHC), 2013, pp 559–564. IEEE
116. Kim SL, Suk HJ, Kang JH, Jung JM, Laine TH, Westlin J (2014) Using unity 3D to facilitate mobile augmented reality game development. In: 2014 I.E. World Forum on Internet of Things (WF-IoT), pp 21–26. IEEE
117. Kistler F, André E (2015) How can I Interact?: comparing full body gesture visualizations. In: Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play, pp 583–588. ACM
118. Kitchenham B, Brereton P (2013) A systematic review of systematic review process research in software engineering. Inf Softw Technol 55(12):2049–2075
119. Klisch Y, Miller LM, Beier ME, Wang S (2012) Teaching the biological consequences of alcohol abuse through an online game: impacts among secondary students. CBE Life Sci Educ 11(1):94–102
120. Klisch Y, Miller LM, Wang S, Epstein J (2012) The impact of a science education game on students' learning and perception of inhalants as body pollutants. J Sci Educ Technol 21(2):295–303
121. Konstantinidis EI, Billis AS, Mouzakidis CA, Zilidou VI, Antoniou PE, Bamidis PD (2016). Design, implementation, and wide pilot deployment of fit for all: an easy to use exergaming platform improving physical fitness and life quality of senior citizens. IEEE J Biomed Health Inform. 20(1):189–200
122. Korhonen H (2010) Comparison of playtesting and expert review methods in mobile game evaluation. In: Proceedings of the 3rd International Conference on Fun and Games. ACM, p 18–27
123. Korhonen H, Koivisto EM (2006) Playability heuristics for mobile games. In: Proceedings of the 8th conference on Human-computer interaction with mobile devices and services. ACM, p 9–16
124. Kos'myna N, Tarpin-Bernard F (2013) Evaluation and Comparison of a Multimodal Combination of BCI Paradigms with Consumer-Grade Hardware and Eye Tracking. IEEE Transactions on Computational Intelligence and AI in Games (in print, 2013)
125. Koster R (2013) Theory of fun for game design. O'Reilly Media, Inc
126. Kothandapani S, Goh DHL, Razikin K (2012) Influence of playability and usability in a mobile human computation game. In: 2012 32nd International Conference on Distributed Computing Systems Workshops (ICDCSW), pp 75–78. IEEE
127. Krebs P, Burkhalter JE, Snow B, Fiske J, Ostroff JS (2013) Development and alpha testing of QuitIT: an interactive video game to enhance skills for coping with smoking urges. JMIR Res Protocols, 2(2)
128. Laine TH, Sedano CAI, Joy M, Sutinen E (2010) Critical factors for technology integration in game-based pervasive learning spaces. IEEE Trans Learn Technol 3(4):294–306
129. Laitinen S (2006) Do usability expert evaluation and test provide novel and useful data for game development? J Usability Stud 2(1):64–75
130. Lange B, Flynn S, Proffitt R, Chang CY, Rizzo AS (2010) Development of an interactive game-based rehabilitation tool for dynamic balance training. Top Stroke Rehabil 17(5):345–352

131. Lee TS, Goh SJA, Quek SY, Phillips R, Guan C, Cheung YB, … Krishnan KRR (2013) A brain-computer interface based cognitive training system for healthy elderly: a randomized control pilot study for usability and preliminary efficacy
132. Lee PW, Wang HY, Tung YC, Lin JW, Valstar A (2015) TranSection: hand-based interaction for playing a game within a virtual reality game. In: Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems, pp 73–76. ACM
133. Lee PW, Wang HY, Tung YC, Lin JW, Valstar A (2015) TranSection: hand-based interaction for playing a game within a virtual reality game. In: Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '15). ACM, New York, p 73–76
134. Lewis JR (1995) IBM computer usability satisfaction questionnaires: psychometric evaluation and instructions for use. Int J Hum Comput Inter 7(1):57–78
135. Lewis C, Polson PG, Wharton C, Rieman J (1990) Testing a walkthrough methodology for theory-based design of walk-up-and-use interfaces. In: Proceedings of the SIGCHI conference on Human factors in computing systems. ACM, p 235–242
136. Liao LD, Chen CY, Wang IJ, Chen SF, Li SY, Chen BW, … Lin CT (2012) Gaming control using a wearable and wireless EEG-based brain-computer interface device with novel dry foam-based sensors. J Neuroeng Rehabil 9(1): 5
137. Lin HCK, Hsieh MC, Wang CH, Sie ZY, Chang SH (2011) Establishment and usability evaluation of an interactive AR learning system on conservation of fish. Turk Online J Educ Technol 10(4):181–187
138. Liu B, Ji Y, Zhang C (2010) A Study and Realization of Mobile Social Network System. In: 2010 Proceedings of the 5th International Conference on Ubiquitous Information Technologies and Applications (CUTE), pp 1–4. IEEE
139. Livingston IJ, Mandryk RL, Stanley KG (2010) Critic-proofing: how using critic reviews and game genres can refine heuristic evaluations. In: Proceedings of the International Academic Conference on the Future of Game Design and Technology. ACM, p 48–55
140. Llorens R, Alcañiz M, Navarro MD, Ferri J, Noe E (2013) Self-awareness rehabilitation through a multi-touch virtual game board after acquired brain injury. In: 2013 International Conference on Virtual Rehabilitation (ICVR), pp 134–138. IEEE
141. Lu Z, Lal Khan MS, Ur Réhman S (2013) Hand and foot gesture interaction for handheld devices. In: Proceedings of the 21st ACM International conference on Multimedia, pp 621–624. ACM
142. Lugrin JL, Cavazza M, Charles F, Le Renard M, Freeman J, Lessiter J (2013) Immersive fps games: user experience and performance. In: Proceedings of the 2013 ACM international workshop on Immersive media experiences, pp 7–12. ACM
143. Luojus P, Koskela J, Ollila K, Mäki SM, Kulpa-Bogossia R, Heikkinen T, Ojala T (2013) Wordster: collaborative versus competitive gaming using interactive public displays and mobile phones. In: Proceedings of the 2nd ACM International Symposium on Pervasive Displays, pp 109–114. ACM
144. Lv Z, Halawani A, Fen S, Li H (2015) Touch-less interactive augmented reality game on vision based wearable device. arXiv preprint arXiv:1504.06359
145. Malone TW (1982) Heuristics for designing enjoyable user interfaces: lessons from computer games. In: Proceedings of the 1982 conference on Human factors in computing systems. ACM, p 63–68
146. Marco J, Baldassarri S, Cerezo E (2013) NIKVision: developing a tangible application for and with children. J UCS 19(15):2266–2291
147. Marco J, Cerezo E, Baldassarri S (2013) Bringing tabletop technology to all: evaluating a tangible farm game with kindergarten and special needs children. Pers Ubiquit Comput 17(8):1577–1591
148. Martins T, Romão T, Sommerer C, Mignonneau L, Correia N (2008) Towards an interface for untethered ubiquitous gaming. In: Proceedings of the 2008 International Conference on Advances in Computer Entertainment Technology, pp 26–33. ACM
149. Miller LM, Chang CI, Wang S, Beier ME, Klisch Y (2011) Learning and motivational impacts of a multimedia science game. Comput Educ 57(1):1425–1433
150. Moreno-Ger P, Torrente J, Hsieh YG, Lester WT (2012) Usability testing for serious games: making informed design decisions with user data. Adv Hum Comput Interact 2012:4
151. Morón MJ, Yáñez R, Cascado D, Suárez-Mejías C, Sevillano JL (2014) A mobile memory game for patients with Acquired Brain Damage: a preliminary usability study. In: 2014 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI), pp 302–305. IEEE
152. Moustakas K, Tzovaras D, Dybkjær L, Bernsen NO (2009) A modality replacement framework for the communication between blind and hearing impaired people. In: Universal access in human-computer interaction. Applications and services, pp 226–235. Springer, Berlin
153. Nam CS, Johnson S, Li Y, Seong Y (2009) Evaluation of human-agent user interfaces in multi-agent systems. Int J Ind Ergon 39(1):192–201

154. Nielsen J (1994) Heuristic evaluation. Usability Inspection Methods 17(1):25–62
155. Nielsen J (1994) Usability inspection methods. In Conference companion on Human factors in computing systems, pp 413–414. ACM
156. Nielsen J (1997) The use and misuse of focus groups. IEEE Softw 14(1):94–95
157. Nielsen J (2000) Why you only need to test with 5 users. Current Issues in Web Usability, Alertbox
158. Nielsen J, Clemmensen T, Yssing C (2002) Getting access to what goes on in people's heads?: reflections on the think-aloud technique. In: Proceedings of the second Nordic conference on Human-computer interaction, pp 101–110. ACM
159. Nijenhuis SM, Prange GB, Amirabdollahian F, Sale P, Infarinato F, Nasr N, … Rietman JS (2015) Feasibility study into self-administered training at home using an arm and hand device with motivational gaming environment in chronic stroke. J Neuroeng Rehabil 12(1): 1
160. Novick D, Vicario J, Santaella B, Gris I (2014) Empirical analysis of playability vs. usability in a computer game. In: Design, user experience, and usability. User experience design for diverse interaction platforms and environments. Springer International Publishing, p 720–731
161. Ntoa S, Adami I, Prokopiou G, Antona M, Stephanidis C (2011) A-cross: an accessible crossword puzzle for visually impaired users. In: Universal Access in Human-Computer Interaction. Users Diversity, pp 342–351. Springer, Berlin
162. Nystrom D, Nalbandian A, Reichenbach M, O'Keefe B (2012) From exercise to socialize: Improving peer relationships in the fitness center. In: 2012 I.E. International Games Innovation Conference (IGIC), pp 1–5. IEEE
163. Ollila E (2009) Using prototyping and evaluation methods in iterative design of innovative mobile games. Unpublished doctoral dissertation). Tampere University of Technology, Finland
164. Olsen T, Procci K, Bowers C (2011) Serious games usability testing: how to ensure proper usability, playability, and effectiveness. In: Design, user experience, and usability. Theory, methods, tools and practice. Springer, Berlin, p 625–634
165. Paavilainen J, Korhonen HJ, Alha K (2014) Common playability problems in social network games. In: CHI'14 Extended Abstracts on Human Factors in Computing Systems, pp 1519–1524. ACM
166. Pagulayan RJ, Keeker K, Wixon D, Romero RL, Fuller T (2002) User-centered design in games. In: Jacko JA, Sears A (eds) The human-computer interaction handbook. L. Erlbaum Associates Inc., Hillsdale, pp 883–906
167. Paliokas I, Arapidis C, Mpimpitsos M (2011) Playlogo 3d: A 3d interactive video game for early programming education: Let logo be a game. In: 2011 Third International Conference on Games and Virtual Worlds for Serious Applications (VS-GAMES), pp 24–31. IEEE
168. Papaloukas S, Stoli C, Patriarcheas K, Xenos M (2010) A survey on how online games usability affects the retention of information. In: 2010 14th Panhellenic Conference on Informatics (PCI), pp 209–213. IEEE
169. Pastor I, Hayes HA, Bamberg SJ (2012) A feasibility study of an upper limb rehabilitation system using kinect and computer games. In: 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp 1286–1289. IEEE
170. Pierotti D (1995) Heuristic evaluation-a system checklist. Xerox Corporation
171. Pinelle D, Wong N, Stach T, Gutwin C (2009) Usability heuristics for networked multiplayer games. In: Proceedings of the ACM 2009 international conference on Supporting group work. ACM, p 169–178
172. Pinelle D, Wong N, Stach T (2008) Heuristic evaluation for games: usability principles for video game design. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, p 1453–1462
173. Plow M, Finlayson M (2014) A qualitative study exploring the usability of Nintendo Wii Fit among persons with multiple sclerosis. Occup Ther Int 21(1):21–32
174. Pol MH, Lagro J, Fluit LR, Lagro, ÄêJanssen TL, Rikkert O, Marcel GM (2014) Teaching geriatrics using an innovative, individual-centered educational game: students and educators win. A proof-of-concept study. J Am Geriatr Soc 62(10):1943–1949
175. Polson PG, Lewis C, Rieman J, Wharton C (1992) Cognitive walkthroughs: a method for theory-based evaluation of user interfaces. Int J Man Mach Stud 36(5):741–773
176. Pranantha D, Bellotti F, Berta R, De Gloria A (2012) A format of serious games for higher technology education topics: a case study in a digital electronic system course. In: 2012 I.E. 12th International Conference on Advanced Learning Technologies (ICALT), pp 13–17. IEEE
177. Preece J, Sharp H, Rogers Y (2015) Interaction design-beyond human-computer interaction. Wiley
178. Proctor MD, Panko M, Donovan SJ (2004) Considerations for training team situation awareness and task performance through PC-gamer simulated multiship helicopter operations. Int J Aviat Psychol 14(2):191–205
179. Raisamo R, Hippula A, Tuominen E, Pasto V, Hasu M (2006) Testing usability of multimodal applications with visually impaired children. IEEE MultiMedia 3:70–76
180. Rand D, Kizony R, Weiss PTL (2008) The Sony PlayStation II EyeToy: low-cost virtual reality for use in rehabilitation. J Neurol Phys Ther 32(4):155–163

181. Raybourn EM, Bos N (2005) Design and evaluation challenges of serious games. In: CHI'05 Extended Abstracts on Human Factors in Computing Systems. ACM, p 2049–2050
182. Rebolledo-Mendez G, Dunwell I, Martínez-Mirón EA, Vargas-Cerdán MD, De Freitas S, Liarokapis F, García-Gaona AR (2009) Assessing Neurosky's usability to detect attention levels in an assessment exercise. In: Human-computer interaction. New Trends, pp 149–158. Springer, Berlin
183. Regenbrecht H, Hoermann S, McGregor G, Dixon B, Franz E, Ott C, Hale L, Schubert T, Hoermann J (2012) Visual manipulations for motor rehabilitation. Comput Graph 36(7):819–834, ISSN 0097–8493
184. Regenbrecht H, McGregor G, Ott C, Hoermann S, Schubert T, Hale L, … Franz E (2011) Out of reach?-A novel AR interface approach for motor rehabilitation. In: 2011 10th IEEE International Symposium on Mixed and augmented reality (ISMAR), pp 219–228. IEEE
185. Rehm M, Leichtenstern K (2012) Gesture-based mobile training of intercultural behavior. Multimedia Systems 18(1):33–51
186. Reichlin L, Mani N, McArthur K, Harris AM, Rajan N, Dacso CC (2011) Assessing the acceptability and usability of an interactive serious game in aiding treatment decisions for patients with localized prostate cancer. J Med Internet Res 13(1)
187. Reβin M, Haffner C (2007) Human-centered design of interactive TV games with SMS backchannel. In: Interactive TV: a shared experience, pp 37–46. Springer, Berlin
188. Rias RM, Abdullah M, Salam NHA (2011) A multimedia learning application for undergraduates in Computer Science: The design, development and usability. In: 2011 3rd International Congress on Engineering Education (ICEED), pp 169–175. IEEE
189. Rice M, Wan M, Foo MH, Ng J, Wai Z, Kwok J, … Teo L (2011) Evaluating gesture-based games with older adults on a large screen display. In: Proceedings of the 2011 ACM SIGGRAPH Symposium on Video Games, pp 17–24. ACM
190. Rogers K, Frommel J, Breier L, Celik S, Kramer H, Kreidel S, Brich J, Riemer V, Schrader C (2015) Mobile augmented reality as an orientation aid: a scavenger hunt prototype. In: 2015 International Conference on Intelligent environments (IE), pp 172–175
191. Romero M, Usart M, Popescu M, Boyle E (2012) Interdisciplinary and international adaption and personalization of the metavals serious games. In: Serious Games Development and Applications, pp 59–73. Springer, Berlin
192. Roto V (2006) Web browsing on mobile phones: characteristics of user experience. Helsinki University of Technology
193. Rouanet P, Danieau F, Oudeyer PY (2011) A robotic game to evaluate interfaces used to show and teach visual objects to a robot in real world condition. In: Proceedings of the 6th international conference on Human-robot interaction, pp 313–320. ACM
194. Rouse III R (2010) Game design: theory and practice. Jones & Bartlett Learning
195. Rubio Ballester B, Bermúdez i Badia S, Verschure PF (2012) Including social interaction in stroke VR-based motor rehabilitation enhances performance: a pilot study. Presence Teleop Virt 21(4):490–501
196. Ryan ML (Ed) (2004) Narrative across media: the languages of storytelling. U of Nebraska Press
197. Salvador-Herranz G, Perez-Lopez D, Ortega M, Soto E, Alcaniz M, Contero M (2013) Manipulating virtual objects with your hands: a case study on applying desktop augmented reality at the primary school. In: 2013 46th Hawaii International Conference on System Sciences (HICSS), pp 31–39. IEEE
198. Sánchez J (2003) End-user and facilitator questionnaire for Software Usability. Usability evaluation test. University of Chile, Santiago, Chile
199. Sánchez J (2012) Development of navigation skills through audio haptic videogaming in learners who are blind. Proc Comput Sci 14:102–110
200. Sánchez J, Espinoza M (2011) Audio haptic videogaming for navigation skills in learners who are blind. In: The proceedings of the 13th international ACM SIGACCESS conference on Computers and accessibility, pp 227–228. ACM
201. Sánchez J, Flores H, Sáenz M (2008) Mobile science learning for the blind. In: CHI'08 extended abstracts on Human factors in computing systems, pp 3201–3206. ACM
202. Sánchez J, Hassler T (2007) AudioMUD: a multiuser virtual environment for blind people. IEEE Trans Neural Syst Rehabil Eng 15(1):16–22
203. Sánchez J, Sáenz M (2006) 3D sound interactive environments for blind children problem solving skills. Behav Inform Technol 25(4):367–378
204. Sánchez J, Sáenz M, Garrido JM (2010) Usability of a multimodal video game to improve navigation skills for blind children. ACM Trans Access Comput (TACCESS) 3(2):7
205. Sánchez J, Salinas A, Sáenz M (2007) Mobile game-based methodology for science learning. In: Human-computer interaction. HCI applications and services, pp 322–331. Springer, Berlin
206. Santos B, Romão T, Dias AE, Centieiro P (2013) eVision: a mobile game to improve environmental awareness. In: Advances in computer entertainment, pp 380–391. Springer International Publishing

207. Scardovelli TA, Frère AF (2015) The design and evaluation of a peripheral device for use with a computer game intended for children with motor disabilities. Comput Methods Prog Biomed 118(1):44–58
208. Scheible J, Tuulos VH, Ojala T (2007) Story Mashup: design and evaluation of novel interactive storytelling game for mobile and web users. In: Proceedings of the 6th international conference on Mobile and ubiquitous multimedia, pp 139–148. ACM
209. Schell J (2008) The art of game design. Morgan Kaufmann
210. Schmitz B, Klemke R, Specht M, Hoffmann M, Klamma R (2012) Developing a mobile game environment to support disadvantaged learners. In: 2012 I.E. 12th International Conference on Advanced Learning Technologies (ICALT), pp 223–227. IEEE
211. Schneiderman B, Plaisant C (2004) Designing the user interface: strategies for effective human-computer interaction. Addison Wesley, New York, p 652
212. Schwebel DC, McClure LA, Severson J (2014) Usability and feasibility of an internet-based virtual pedestrian environment to teach children to cross streets safely. Virtual Reality 18(1):5–11
213. Sharma SV, Shegog R, Chow J, Finley C, Pomeroy M, Smith C, Hoelscher DM (2015) Effects of the quest to lava mountain computer game on dietary and physical activity behaviors of elementary school children: a pilot group-randomized controlled trial. J Acad Nutr Diet
214. Shegog R, Markham C, Peskin M, Dancel M, Coton C, Tortolero S (2007) 'It's Your Game': An innovative multimedia virtual world to prevent HIV/STI and pregnancy in middle school youth. In Medinfo 2007: Proceedings of the 12th World Congress on Health (Medical) Informatics; Building Sustainable Health Systems, IOS Press, p 983
215. Shepherd A (1989) Analysis and training in information technology tasks. Task analysis for human-computer interaction, pp 15–55
216. Shin JH, Ryu H, Jang SH (2014) A task-specific interactive game-based virtual reality rehabilitation system for patients with stroke: a usability test and two clinical experiments. J Neuroeng Rehabil 11:32
217. Song CR (2006) Design and Usability of Web-Based Serious Game for Taiwanese EFL Undergraduate Students. In: International Conference on Networking, International Conference on Systems and International Conference on Mobile Communications and Learning Technologies. ICN/ICONS/MCL 2006, pp 220–220. IEEE
218. Song D, Karimi A, Kim P (2011) Toward designing mobile games for visually challenged children. In: 2011 International Conference on e-Education, Entertainment and e-Management (ICEEE), pp 234–238. IEEE
219. Squire K (2005) Changing the game: what happens when video games enter the classroom? Innov: J Online Educ 1(6):5
220. Suárez M (2011) SIRIUS: Sistema de Evaluación de la Usabilidad Web Orientado al Usuario y basado en la Determinación de Tareas Críticas. Universidad de Oviedo, Oviedo
221. Suárez CO, Marco J, Baldassarri S, Cerezo E (2011) Children with special needs: comparing tactile and tangible interaction. In: Human-Computer Interaction-INTERACT 2011, pp 495–498. Springer, Berlin
222. Sunwoo J, Yuen W, Lutteroth C, Wünsche B (2010) Mobile games for elderly healthcare. In: Proceedings of the 11th International Conference of the NZ Chapter of the ACM Special Interest Group on Human-Computer Interaction, pp 73–76. ACM
223. Susaeta H, Jimenez F, Nussbaum M, Gajardo I, Andreu JJ, Villalta M (2010) From MMORPG to a classroom multiplayer presential role playing game. J Educ Technol Soc 13(3):257–269
224. Susi T, Johannesson M, Backlund P (2007) Serious games: an overview. Technical Report HS- IKI -TR-07-001. School of Humanities and Informatics, University of Skövde, Sweden
225. Sutcliffe A, Alrayes A (2012) Investigating user experience in Second Life for collaborative learning. Int J Hum Comput Stud 70(7):508–525
226. Tadayon R, Panchanathan S, McDaniel T, Fakhri B, Laff M (2014) A toolkit for motion authoring and motor skill learning in serious games. In: 2014 I.E. International Symposium on Haptic, Audio and Visual Environments and Games (HAVE), pp 59–64. IEEE
227. Tan JL, Goh DHL, Ang RP, Huan VS (2013) Participatory evaluation of an educational game for social skills acquisition. Comput Educ 64:70–80
228. Tang TY, Man CY, Hang CP, Cheuk LS, Kwong CW, Chi YC, … Kam S (2008) A study of interaction patterns and awareness design elements in a massively multiplayer online game. Int J Comput Games Technol 2008: 5
229. Taylor GS, Barnett JS (2012) Evaluation of wearable simulation interface for military training. Hum Factors: J Hum Factors Ergon Soc, 0018720812466892
230. Thiele S, Meyer L, Geiger C, Drochtert D, Woldecke B (2013) Virtual archery with tangible interaction. In: 2013 I.E. Symposium on 3D User Interfaces (3DUI), pp 67–70. IEEE
231. Tolentino GP, Battaglini C, Pereira ACV, de Oliveria RJ, de Paula MGM (2011) Usability of Serious Games for Health. In: 2011 Third International Conference on Games and Virtual Worlds for Serious Applications (VS-GAMES), pp 172–175. IEEE
232. Tong T, Chignell M (2013) Designing game-based cognitive assessments for elderly adults. In: Proceedings of the First International Conference on Gameful design, research, and applications, pp 127–130. ACM

233. Tong T, Chignell M (2014) Developing a serious game for cognitive assessment: choosing settings and measuring performance. In: Proceedings of the Second International Symposium of Chinese CHI, pp 70–79. ACM
234. Torrente J (2012) Reusable game interfaces for people with disabilities. In: Proceedings of the 14th international ACM SIGACCESS conference on Computers and accessibility, pp 301–302. ACM
235. Trewin S, Laff M, Hanson V, Cavender A (2009) Exploring visual and motor accessibility in navigating a virtual world. ACM Trans Access Comput (TACCESS) 2(2):11
236. Tullis TS, Stetson JN (2004) A comparison of questionnaires for assessing website usability. In: Usability Professional Association Conference, pp 1–12
237. Vajda P, Ivanov I, Lee JS, Ebrahimi T (2011) Epitomize your photos. Int J Comput Games Technol 2011:9
238. Valdes B, Hilderman CG, Hung CT, Shirzad N, Van der Loos HF (2014) Usability testing of gaming and social media applications for stroke and cerebral palsy upper limb rehabilitation. In: 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp 3602–3605. IEEE
239. Villalta M, Gajardo I, Nussbaum M, Andreu JJ, Echeverría A, Plass JL (2011) Design guidelines for classroom multiplayer presential games (CMPG). Comput Educ 57(3):2039–2053
240. Virvou M, Katsionis G (2008) On the usability and likeability of virtual reality games for education: the case of VR-ENGAGE. Comput Educ 50(1):154–178
241. Virzi RA (1992) Refining the test phase of usability evaluation: how many subjects is enough? Hum Factors: J Hum Factors Ergon Soc 34(4):457–468
242. Wallner G, Kriglstein S (2011) Design and evaluation of the educational game DOG eometry: a case study. In: Proceedings of the 8th International Conference on Advances in Computer Entertainment Technology, p 14. ACM
243. Wang AI, Morch-Storstein OK (2008) An evaluation of a mobile game concept for lectures. In: IEEE 21st Conference on Software Engineering Education and Training. CSEET'08, pp 197–204. IEEE
244. Wang AI, Wu B, Bakken SK (2010) Experiences from implementing a face-to-face educational game for iPhone/iPod Touch. InGames Innovations Conference (ICE-GIC), 2010 International IEEE Consumer Electronics Society's, pp 1–8. IEEE
245. Weightman APH, Preston N, Holt R, Allsop M, Levesley M, Bhakta B (2010) Engaging children in healthcare technology design: developing rehabilitation technology for children with cerebral palsy. J Eng Des 21(5):579–600
246. Wharton C, Rieman J, Lewis C, Polson P (1994) The cognitive walkthrough method: a practitioner's guide. In: Usability inspection methods, pp 105–140. John Wiley & Sons, Inc.
247. Whitlock LA, McLaughlin AC, Leidheiser W, Gandy M, Allaire JC (2014) Know before you go: feelings of flow for older players depends on game and player characteristics. In: Proceedings of the first ACM SIGCHI annual symposium on Computer-human interaction in play, pp 277–286. ACM
248. Wise K, Kobayashi K, Searchfield GD (2015) Feasibility study of a game integrating assessment and therapy of tinnitus. J Neurosci Methods 249:1–7
249. Wrońska N, Garcia-Zapirain B, Mendez-Zorrilla A (2015) An iPad-Based tool for improving the skills of children with attention deficit disorder. Int J Environ Res Public Health 12(6):6261–6280
250. Wu W, Rank S (2015) Responsive environmental Diegetic audio feedback for hand gestures in audio-only games. In: Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play, pp 739–744. ACM
251. Yamabe T, Asuma H, Kiyono S, Nakajima T (2011) Feedback design in augmented musical instruments: A case study with an ar drum kit. In: 2011 I.E. 17th International Conference on Embedded and Real-Time Computing Systems and Applications (RTCSA) (Vol. 2), pp 126–129. IEEE
252. Yáñez Gómez R, Cascado Caballero D, Sevillano JL (2014) Heuristic evaluation on mobile interfaces: a new checklist. Sci World J 2014
253. Yang HC, Zapata-Rivera D (2010) Interlanguage pragmatics with a pedagogical agent: the request game. Comput Assist Lang Learn 23(5):395–412
254. Yoon JW, Yang SI, Cho SB (2012) Adaptive mixture-of-experts models for data glove interface with multiple users. Expert Syst Appl 39(5):4898–4907
255. Youm S (2015) The development and user satisfaction evaluation of Internet-Based N-screen healthcare walking content to increase continuous usage motivation. Telemedicine and e-Health
256. Yun C, Shastri D, Pavlidis I, Deng Z (2009) O'game, can you feel my frustration?: improving user's gaming experience via stresscam. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp 2195–2204. ACM

257. Zhang D. Overview of usability evaluation methods. http://www.usabilityhome.com. Retrieved in March 2014
258. Zhang M, Xu M, Han L, Liu Y, Lv P, He G (2012) Virtual Network Marathon with immersion, scientificalness, competitiveness, adaptability and learning. Comput Graph 36(3):185–192
259. Zin NAM, Nasir NYM (2008) Evaluation of an edutainment animated folktales software to motivate socio-cultural awareness among children. In: Third International Conference on Convergence and Hybrid Information Technology, 2008. ICCIT'08. (Vol. 1), pp 315–319. IEEE