

Chapter 2

A Meta-Analysis of Data Collection in Serious Games Research

Shamus P. Smith, Karen Blackmore, and Keith Nesbitt

Abstract Serious game analytics share many of the challenges of data analytics for computer systems involving human activity. Key challenges include how to collect data without influencing its generation, and more fundamentally, how to collect and validate data from humans where a primary emphasis is on what people are thinking and doing. This chapter presents a meta-analysis of data collection activities in serious games research. A systematic review was conducted to consider metrics and measures across the human–computer interaction, gaming, simulation, and virtual reality literature. The review focus was on the temporal aspect of data collection to identify if data is collected before, during, or after gameplay and if so what fundamental processes are used to collect data. The review found that the majority of data collection occurred post-game, then pre-game, and finally during gameplay. This reflects traditional difficulties of capturing gameplay data and highlights opportunities for new data capture approaches oriented towards data analytics. Also we identify how researchers gather data to answer fundamental questions about the efficacy of serious games and the design elements that might underlie their efficacy. We suggest that more standardized and better-validated data collection techniques, that allow comparing and contrasting outcomes between studies, would be beneficial.

Keywords Data collection • Serious games • Meta-review • Data analytics

S.P. Smith (✉)

School of Electrical Engineering and Computer Science, The University of Newcastle,
University Drive, Callaghan, NSW 2308, Australia
e-mail: shamus.smith@newcastle.edu.au

K. Blackmore • K. Nesbitt

School of Design, Communication and Information Technology,
The University of Newcastle, University Drive, Callaghan, NSW 2308, Australia
e-mail: karen.blackmore@newcastle.edu.au; keith.nesbitt@newcastle.edu.au

1 Introduction

This chapter is concerned with serious games and the data collected about serious games. It concerns itself with the process of turning such data into information and, through data analytics, drawing conclusions about that information. Thus, collected data is synthesized as information and ultimately into evidence.

There has been significant growth in research and industry attention, and public awareness, of the collection and analysis of the vast amounts of data that is available electronically. This has characterized the popularization of “Big Data” and the associated knowledge and value to be generated through both automated and perceptual techniques for data analytics. At one end of the data analytics pipeline is the notion of visual analytics, the use of visualization techniques to represent multidimensional data so that people can use their perceptual skills and incomplete heuristic knowledge to find useful patterns in the data. We might term these patterns information, and we note that these approaches rely on human skills that do not necessarily translate well to computers. More automated approaches often referred to as “data-mining” also exist. While the intention is the same, that is to identify useful patterns or information, the approach is complementary, relying on the strengths of computers to perform rapid, repetitive, error-free mathematical tasks that allow large amounts of data to be quickly processed.

Given that neither of these approaches to finding information in large data is discrete, good data analytics might well rely on combining the strengths of both approaches. However, information does not imply evidence. The accuracy of the information very much depends on the quality and validity of the data and the transformations that filter, abstract, and simplify the vast volumes of data to support analysis. If poor quality data is initially collected, then its progress through later stages of the analytics pipeline will be compromised and the validity of any identified patterns weakened.

Serious game analytics share many of the same challenges as data analytics in other computer systems that are focused on human activity. A typical challenge is how to collect data without influencing its generation and more fundamentally, how to collect and validate data from human participants where a primary focus is on what people are thinking and doing.

This chapter will explore data collection issues from serious games as the initial step to any serious gameplay analytics. We use a systematic review process to consider the metrics and measures across the human–computer interaction, gaming, simulation, and virtual reality literature. We identify how researchers gather data to try and answer fundamental questions about the efficacy of serious games and the design elements that might underlie their efficacy.

Data collection is interdisciplinary and a comprehensive literature review over computer science, psychology, and education, for instance, is outside the scope of this chapter. The focus here will be on the temporal aspect of data collection during serious game studies, namely how, and if, data is collected before, during, and after serious gameplay. The chapter uses a framework of traditional data collection methods to identify a core mapping to the serious game literature. The study is broad in that it covers diverse research from numerous disciplines, over a long time frame,

that have used a wide range of methods and been driven by different motivations. This diverse body of research is first found by systematically selecting eight relevant literature reviews from 2009 to 2014 related to serious game research. To provide depth to the study, each of the study papers identified in these literature reviews ($n=299$) are examined in terms of the way data is collected to assess the efficacy and usability of the games.

While the enthusiasm for serious games is unquestioned, the business case for serious games still requires more tangible evidence, both qualitative and quantitative. However, a first step to better evidence is a close examination of the data collected from serious games. It is this data that will be processed by any serious game analytics, and ultimately demonstrate the worth of the source serious games. This chapter provides a historical review of data collection as a resource for researchers in serious games, human–computer interaction, and anyone who is concerned about the collection and accuracy of gameplay data for future analytic purposes. Also, in the discussion section of this chapter, we will reflect upon the question of evidence and how well it relates to the two key issues of efficacy and usability in games that are used for serious purposes.

2 Study Method

To perform our study, we designed a systematic process that could be repeated or amended to accommodate both changes in scope and alternative research questions, and extended to incorporate future literature (see Fig. 2.1).

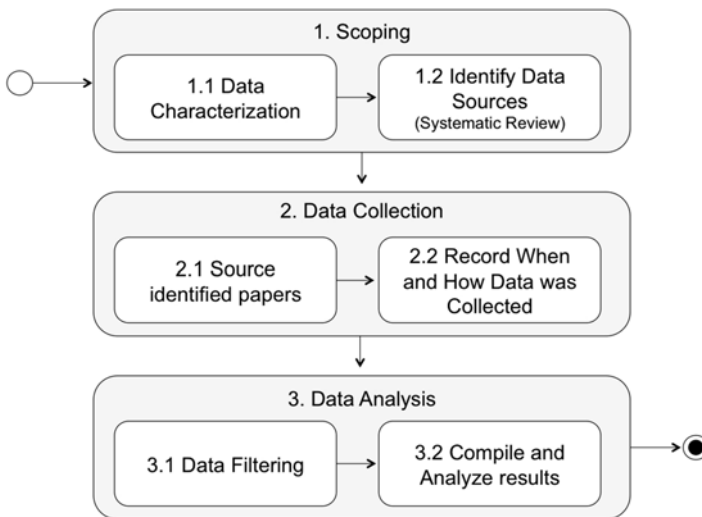


Fig. 2.1 Overview of the process used in this study

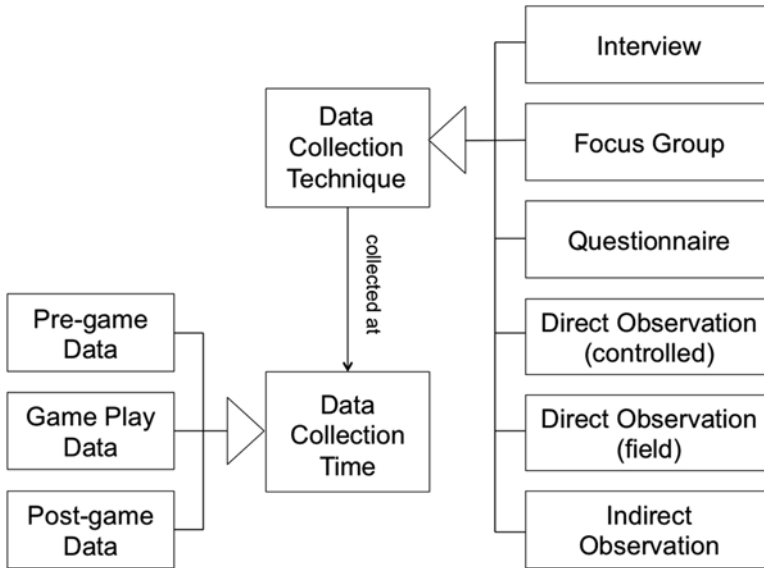


Fig. 2.2 Overview of the data characterization used in this study

2.1 Data Characterization

The process began with a data characterization activity where we identified the type of data we wished to collect about each individual game study (see 1.1 in Fig. 2.1). We focused on the temporal aspect of data collection during serious game studies, namely how, and if, data is collected before, during, and after serious gameplay (see Fig. 2.2). These stages are relatively standard across the range of methodological approaches used to evaluate games.

We decided to categorize the data collected during each of these stages based on common data gathering techniques (see Fig. 2.2). These frequently used techniques are taken from a list provided by Rogers, Sharp, and Preece (2011) in their popular text on human–computer interaction, and include:

- Interviews
- Focus Groups
- Questionnaires
- Direct observation in the field
- Direct observation in a controlled environment
- Indirect observation

Although the first three techniques are self-explanatory, the last three techniques require further clarification. *Direct observation* consists of observing actual user activity and typically involves the collection of qualitative data by capturing the details of what individual or groups are doing with a system, for example, with observers taking notes of user behaviors. This can be conducted *in the field*, where

users are interacting in the target environment for a system. Examples could include a normal teaching session in a classroom for an educational system or on a walking tour for a mobile application (Rogers et al., 2011, p. 262). Thus, the system is being used in a real-life situation, e.g., with high ecological validity (McMahan, Ragan, Leal, Beaton, & Bowman, 2011; Smith, Stibric, & Smithson, 2013). Direct observation *in a controlled environment* is typically a laboratory-based environment where conditions can be controlled and standardized between participants and sessions. This allows users to focus on a task without interruption. However, results from studies in such environment may not generalize as the conditions are, by default, artificial.

In contrast to direct observation, where the users can see that they are actively being observed, *indirect observation* involves gathering data where users are not distracted by the collection mechanism. This could include collecting qualitative data, for example, from a user diary, or quantitative data from automated event logging. The latter is particularly attractive for serious games as event logs can be tailored to collect any pertinent information; for example, task sequences, task completion times, and/or percentage of tasks accomplished. Loh (2009) details a number of logging examples including basic game event logs, After-Action Reports as graphical game logs, and biofeedback data to capture physiological reactions. Such *in-process* data collection is by its nature objective and can provide substantial volumes of data for further analytic treatment.

While these techniques cover a good range of the mixed methods used in game research, we also recognize that other categorizations could have been adopted. For example, an alternative and more detailed classification of 16 different data collection techniques used in games studies is provided by Mayer et al. (2014). While there is merit for more complex categorization, we recognized the difficulty of collecting our own data; game studies from various disciplines do not have a standardized approach to describing data collection methods. Since we intended to be reviewing a large and broad range of studies, we sought to keep our data classification as simple as possible. Thus focusing on specific techniques, for example, the use of telemetry or *Information Trails* (Loh, 2012) for indirect observation of gameplay, is outside the scope of the current review. However, we will revisit issues surrounding the data collection process in the discussion section.

2.2 Identify Data Sources (Systematic Review)

We adopted a systematic approach to identifying existing reviews of serious game research across domains (see 1.2 in Fig. 2.1). A systematic review is developed to gather, evaluate, and analyze all the available literature relevant to a particular research question or area of interest, based on a well-defined process (Bearman et al., 2012; González, Rubio, González, & Velthuis, 2010; Kitchenham et al., 2009). The systematic review methodology is extensively used in the healthcare domain (Bearman et al., 2012) and has been widely adopted in other areas including business (González et al., 2010), education (Bearman et al., 2012), and software engineering (Kitchenham et al., 2009; Šmite, Wohlin, Gorschek, & Feldt, 2010).

A systematic review methodology requires the identification of all published works relevant to the requirements. The search strategy adopted covers key term searches in relevant scholarly databases. We included the Web of Science, Scopus, EBSCOhost, and Wiley Interscience bibliographic databases in the search. The search was conducted over article titles to restrict results to primary studies, and includes journal articles, book chapters, and review papers in the results. Search results were restricted to papers published between 2009 and 2014 inclusive.

The objective of our systematic review was to identify all review articles of studies using serious games. To conduct a review that meets our objective, the search term used needs to accommodate two key purposes. The first purpose was to find published works relating to serious games. We expanded the term, *serious games*, to include references to studies of games for *applied, learning, teaching, or educational* purposes (Crookall, 2010). The second purpose is to find review or meta-review articles only as the basis for “drill-down” to individual studies. We therefore included the terms *review, meta-review, or meta-analysis* in the search term. Several preliminary searches were conducted to refine the individual and combined search terms to develop a search string that located articles of interest without too many false positives. The resulting Boolean search string that we used for the systematic review was:

((gam* AND (serious OR edutainment OR “applied gam*” OR learn* OR game-based learning OR educat* OR teach*) AND (review OR meta-review OR meta-analysis))

The search initially produced a total of 126 potential papers, of which 73 were found to be unique. These papers were then manually evaluated by title, abstract, and if necessary, by full text, based on the following inclusion criteria:

- Focused on the review of studies:
 - Using randomized control trials, experimental pretest/posttest control group design or quasi-experimental structure
 - Evaluating computer, console or mobile games
 - That was directed at achieving teaching and learning outcomes
- From any country
- Written in English

Papers not meeting the inclusion criteria were excluded from the systematic review. The review process is shown in Fig. 2.3 and identified ten papers for the analysis.

The evaluation process detailed in Fig. 2.3 shows the inclusion of an additional paper that did not appear in the initial 126 papers. Wattanasoontorn, Boada, García, and Sbert’s (2013) comprehensive study of serious games for health was not located using the Boolean search string due to the non-inclusion of the term *review* in the article title. It was, however, identified and noted in the preliminary searches that we used to refine the search terms. Wattanasoontorn et al. (2013) include 108 references in the broad health domain in their final review, making this a relevant and comprehensive piece of work for inclusion in our analysis. However, expanding the search string to ensure that this article was located results in an unwieldy number of

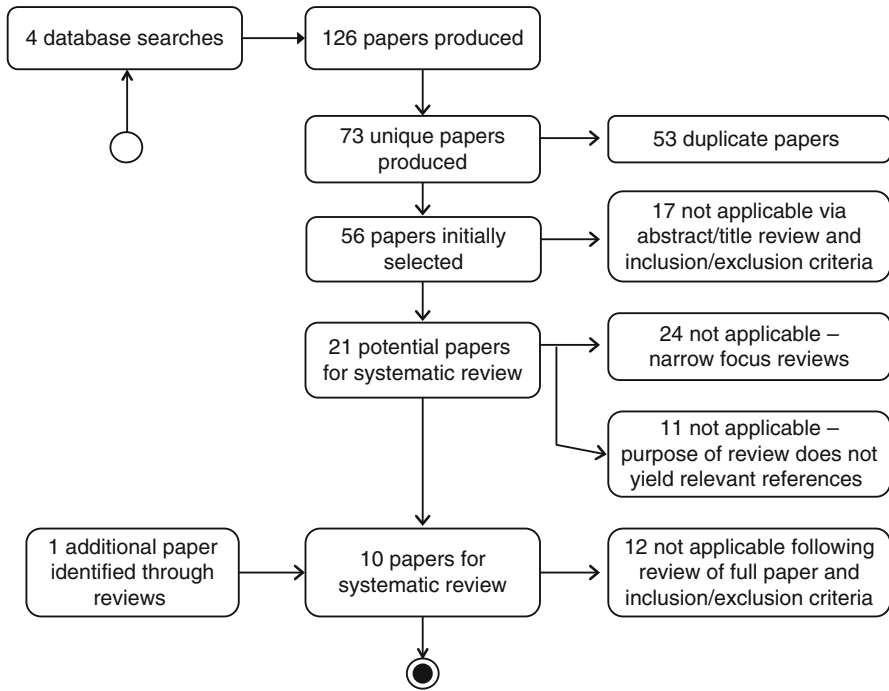


Fig. 2.3 Diagram of the selection process for the systematic review

irrelevant search results. More importantly, it results in a search term that does not meet the objective of the systematic review, which is to identify reviews or meta-reviews of serious games. Thus, the paper was simply added to the end results of the systematic review process.

2.3 Data Collection and Analysis

During the data collection process, we examined in detail the ten high level literature reviews identified in our systematic review. A description of each of these papers is provided in the next section of the chapter.

Two literature review papers, on analysis, were rejected during the data collection process. Hwang and Wu (2012) analyzed the research status and trends in digital game-based learning (DGBL) from 2001 to 2010. Specifically, they explored (1) whether the number of articles in this area is increasing or decreasing, (2) what the primary learning domains related to DGBL are, (3) whether there is a domain focus shift between the first 5 years (2001–2005) and the second 5 years (2006–2010), and (4) which are the major contributing countries of DGBL research. From an initial set of 4,548 papers, Hwang and Wu selected a total of 137 articles for review. However, their paper does not provide details of the specific 137 articles

selected, and we were therefore unable to identify data collection methods from these papers. Thus, we have excluded this review from our analysis. The Blakely, Skirton, Cooper, Allum, and Nelmes (2009) systematic review of educational games in the health sciences was also removed. They provide an analysis of the use of games to support classroom learning in the health sciences based on a review of 16 papers. However, it was deemed to be an earlier subset of the latter and more expansive review of serious games for health by Wattanasoontorn et al. (2013).

From the remaining eight literature review papers, we identified 299 referenced studies. Where possible, we then sourced each of the papers and recorded the data collection techniques used pre-game, during gameplay and post-game for each study. In a few cases, papers could not be sourced or the papers were not in English. These papers were excluded, as were all studies that were only reported in non-peer reviewed locations such as websites. We also excluded references to demonstrations or papers that only included a critical analysis of literature. Finally, we excluded duplicate studies so they were only included once in the analysis. This left a total of 188 referenced studies to be included in the final analysis.

3 Systematic Review Papers

The eight review papers identified from our systematic process that were used for data collection covered both general and domain-specific areas. Four of the papers are reviews of the general serious games area and were not focused on any specific area or application. However, two papers are focused on studies in the health domain, while the other two focused on medicine and the humanities. Each of these papers are described below and the number of contributory studies to our review are identified.

Connolly, Boyle, MacArthur, Hainey, and Boyle (2012) examined the “literature on computer games and serious games in regard to the potential positive impacts of gaming on users aged 14 years or above, especially with respect to learning, skill enhancement and engagement” (p. 1). This review paper focused on identifying empirical evidence and on categorizing games and their impacts and outputs. The majority of reviewed papers come under the serious games for education and training classification. We have reviewed data collection in these papers ($n=70$) across Connolly et al.’s categories of affective and motivational outcomes, behavioral change outcomes, knowledge acquisition/content understanding outcomes, motor skill outcomes, perceptual and cognitive skills outcomes, physiological arousal outcomes, and soft skill and social outcomes.

Wattanasoontorn et al. (2013) consider the use of serious games in the health domain area. They provide a survey of serious games for health and define a new classification, based on serious game, health, and player dimensions. For serious game subjects, they classify by game purpose and game functionality, for health, they classify by state of disease and finally for player, they consider two types of

player dimensions, player/non-player, and professional/nonprofessional. We have used Wattanasoontorn et al.'s (2013) classification and comparison of health games summary ($n=91$) which considers the following areas: detection (patient), treatment (patient), rehabilitation (patient), education (patient), health and wellness (non-patient), training for professional (non-patient), training for non-professional (non-patient).

Anderson et al. (2010) describe the use of serious games for cultural heritage; specifically, the use of games to support history teaching and learning and for enhancing museum visits. Their state-of-the-art review includes both a set of case studies and an overview of the methods and techniques used in entertainment games that can potentially be deployed in cultural heritage contexts. Here, we have focused on the former and reviewed data collection as noted in the case studies ($n=5$).

Girard, Ecalte, and Magnan (2013) review the results of experimental studies designed to examine the effectiveness of video games and serious games on players' learning and engagement. They have attempted to identify all the experimental studies from 2007 to 2011 that have used serious games for training or learning, and assessed their results in terms of both effectiveness and acceptability. Girard et al. (2013) had a two pass process for article inclusion/exclusion where the stricter second pass, only considering randomized controlled trial studies, resulted in only nine articles. Here, we have used the results from their first pass of the literature which resulted in 30 articles ($n=29$, we excluded one article written in French) published in scientific journals or in proceedings of conferences and symposia across the fields of cognitive science, psychology, human-computer interaction, education, medicine, and engineering where training has been performed using serious games or video games.

The systematic review of Graafland, Schraagen, and Schijven (2012) provides a comprehensive analysis of the use of serious games for medical training and surgical skills training. The authors focus on evaluating the validity testing evident in prior serious games research in the area and identify 25 articles through a systematic search process. Of these, 17 included games developed for specific educational purposes and 13 were commercial games evaluated for their usefulness in developing skills relevant to medical personnel. Of the 25 articles identified by Graafland et al. (2012), six were identified as having completed some validation process and none were found to have completed a full validation process. For the purpose of our study, we considered only articles explicitly identified by Graafland et al. (2012), that appeared in the supplementary information tables ($n=20$).

Papastergiou (2009) presents a review of published scientific literature on the use of computer and video games in Health Education (HE) and Physical Education (PE). The aim of the review is to identify the contribution of incorporating electronic games as educational tools into HE and PE programs, to provide a synthesis of empirical evidence on the educational effectiveness of electronic games in HE and PE, and to scope out future research opportunities in this area. Papastergiou (2009) notes that the empirical evidence to support the educational effectiveness of electronic games in HE and PE is limited, but that the findings presented in their review show a positive picture overall. We have reviewed data collection methods in the research articles featured in this review ($n=19$).

Vandercruysse, Vandewaetere, and Clarebout (2012) conducted a systematic literature review where the learning effects of educational games are studied in order to gain more insights into the conditions under which a game may be effective for learning. They noted that although some studies reported positive effects on learning and motivation, this was confounded by different learner variables and different context variables across the literature. Their review initially found 998 unique peer reviewed articles. After removing articles with (quasi) experimental research, only 22 journal articles were finally reviewed. It is these 22 articles that we have included in our data collection review.

Wilson et al. (2009) performed a literature review of 42 identified studies and examined relationships between key design components of games and representative learning outcomes expected from serious games for education. The key design components of fantasy, rules/goals, sensory stimuli, challenge, mystery, and control considered by Wilson et al. (2009) were identified as statistically significant for increasing the “game-like” feel of simulations (Garris & Ahlers, 2001) and key gaming features necessary for learning (Garris, Ahlers, & Driskell, 2002). These were examined in relation to both skill-based and affective learning outcomes. We included all 42 studies in our review.

4 Results

The total number of papers used in this study for data collection was 299. Table 2.1 provides a full list of the references examined, and the literature reviews from which those papers were sourced.

After examination of the 299 papers, and the filtering described in Sect. 2.3, we explored the data collection techniques described in 188 papers spanning 1981–2012. Eighty-four percent of the 188 papers were from the 10 year period of 2003–2012 (see Table 2.2). Also, 51 % of the papers were from the mid-region of this 10 year period, i.e., 2006–2009. However, this does not necessarily indicate a surge in serious game research but is more likely a consequence of publication time frames. Although the literature reviews determined by our search string were published up to 2014, the published research that they reported on was only up to 2012.

In total, 510 data collection techniques were used in the 188 studies. Of these, 33 % of data collection occurred pre-game, 21 % during gameplay, and 46 % in post-game evaluation phases (see Fig. 2.4). On average the total number of data collection methods used per study, across the three phases of pre-game, during gameplay, and post-game, was 2.71 (SD=1.2).

In terms of specific techniques for the pre-game phase ($n=169$), 52 % of the studies used questionnaires, 42 % of the studies used some form of test, 4 % of the studies used an interview, 2 % of the studies used an indirect observation, while only a single study employed a focus group in the pre-game phase (see Fig. 2.5).

For the post-game phase ($n=235$), 46 % of the studies used questionnaires, 37 % of the studies used some form of test, and 13 % of the studies used an interview.

Table 2.1 List and source of all references examined in the data collection process

<p>Anderson, E., F., McLoughlin, L., Liarokapis, F., Peters, C., Petridis, P., & de Freitas, S. (2010). Developing serious games for cultural heritage: A state-of-the-art review. <i>Virtual Reality, 14</i>(4), 255–275. (<i>n</i>=5)</p> <p>Arnold, D., Day, A., Glauert, J., Haegler, S., Jennings, V., Kevelham, B., Laycock, R., Magnost-Thalmann, N., Mam, J., Maupu, D., Papagiannakis, G., Thalmann, D., Yersin, B., & Rodriguez-Echavaria, K. (2008); Gaitatzes, A., Christopoulos, D., & Papaioannou, G. (2004); Jacobson, J., Handron, K., & Holden, L. (2009)</p>	<p>Connolly, T. M., Boyle, E. A., MacArthur, E., Hainey, T., Boyle, & J. M. (2012). A systematic literature review of empirical evidence on computer games and serious games. <i>Computers & Education, 59</i>(2), September 2012, 661–686. (<i>n</i>=70)</p> <p>Akkerman, S., Admiraal, W., & Huizenga, J. (2008); Anand, V. (2007); Assmann, J. J., & Gallenkamp, J. V. (2009); Aylon, Y., Glaser, C. B., Hall, J. I., Uribe, S., & Fried, M. P. (2005); Backlund, P., Engström, H., Johansson, M., Lebram, M., & Sjöden, B. (2008); Baldaro, B., Tuozzi, G., Codispoti, M., Montebrocci, O., Barbagli, F., Trombini, E., et al. (2004); Barlett, C. P., Vowels, C. L., Shanteau, J., Crow, J., & Miller, T. (2009); Beale, I. L., Kato, P. M., Marin-Bowling, V. M., Guthrie, N., & Cole, S. W. (2007); Boot, W. R., Kramer, A. F., Simons, D. J., Fabiani, M., & Gratton, G. (2008); Cameron, B., & Dwyer, F. (2005); Carvelho, T., Allison, R. S., Irving, E. L., & Herriot, C. (2008); Castelli, L., Corazzini, L. L., & Geminiani, C. G. (2008); Chiang, Y.-T., Cheng, C.-Y., & Lin, S. S. J. (2008); Chou, C., & Tsai, M.-J. (2007); Connolly, T. M., Boyle, E., & Hamey, T. (2007); Davidovitch, L., Parush, A., & Shuib, A. (2008); De Lucia, A., Francese, R., Passero, I., & Tortora, G. (2009); Eastin, M. S. (2006); Felicia, P., & Pitt, I. (2007); Feng, J., Spence, I., & Pratt, J. (2007); Fu, F.-L., Su, R.-C., & Yu, S.-C. (2009); Gentile, D. A., & Gentile, J. R. (2008); Green, C. S., & Bavelier, D. (2006); Halpern, D. F., & Wai, J. (2007); Hamalainen, R., Oksanen, K., & Hakkinen, P. (2008); Harr, R., Buch, T., & Hanghøj, T. (2008); Higuchi, S., Motohashi, Y., Liu, Y. L., & Maeda, A. (2005); Hogle, N. J., Widmann, W. D., Ude, A. O., Hardy, M. A., & Fowler, D. L. (2008); Houtkamp, J., Schuurink, E., & Toet, A. (2008); Hsu, C.-L., & Lu, H.-P. (2004); Huizenga, J., Admiraal, W., Akkerman, S., & ten Dam, G. (2007); Huizenga, J., Admiraal, W., Akkerman, S., & ten Dam, G. (2008); Ivory, J. D., & Kalyanaraman, S. (2007); Jennett et al. (2008); Jouriles, E. N., McDonald, R., Kullowatz, A., Rosenfield, D., Gomez, G. S., & Cuevas, A. (2008); Karakus, T., Inal, Y., & Cagiltay, K. (2008); Kiili, K., Ketamo, H., & Lainema, T. (2007); Kim, Y., & Ross, S. D. (2006); Lavender, T. (2008); Lee, M., & Faber, R. J. (2007); Lindh, J., Hrastinski, S., Bruhn, C., & Mozgira, L. (2008); Lucas, K., & Sherry, J. L. (2004); Mayer, I. S., Carton, L., de Jong, M., Leijten, M., & Dammers, E. (2004); Miller, M., & Hegelheimer, V. (2006); Nelson, M. R., Yaros, R. A., & Keum, H. (2006); Nomura, T., Miyashita, M., Shrestha, S., Makino, H., Nakamura, Y., Aso, R., et al. (2008); Nite, S., & Stephens, R. (2008); Orvis, K. A., Horn, D. B., & Belanich, J. (2008); Papastegiou (2009); Ravaja, N., Turpeinen, M., Saari, T., Puttonen, S., & Keltikangas-Jarvinen, L. (2008); Riegelsberger, J., Counts, S., Farnham, S. D., & Phillips, B. C. (2006); Rossio, E., & Papadakis, S. (2008); Russell, W. D., & Newton, M. (2008); Salminen, M., & Ravaja, N. (2008); Schneider, L.-P., & Cornwell, T. B. (2005); Schrader, P. G., & McCreery, M. (2007); Schwabe, G., Goth, C., & Frohberg, D. (2005); Stefanidis, D., Komdorffer, J. R., Jr., Sierra, R., Touchard, C., Dunne, J. B., & Scott, D. J. (2005); Stefanidis, D., Scerbo, M. W., Sechrist, C., Mostafavi, A., & Heniford, B. T. (2008); Steinkuehler, C., & Duncan, S. (2008); Sward, K. A., Richardson, S., Kendrick, J., & Maloney, C. (2008); Terlecki, M. S., & Newcombe, N. S. (2005); Vahed, A. (2008); van Reekum, C. M., Johnstone, T., Banse, R., Eitter, A., Wehrle, T., & Scherer, K. R. (2004); Wan, C.-S., & Chiou, W.-B. (2007); Weibel, D., Wissmath, B., Habegger, S., Steiner, Y., & Groner, R. (2008); Wijers, M., Jonker, V., & Kerstens, K. (2008); Yalon-Chamovitz, S., & Weiss, P. L. (2008); Yaman, M., Nerdel, C., & Bayrhuber, H. (2008); Yip, F. W. M., & Kwan, A. C. M. (2006)</p>
---	---

(continued)

Table 2.1 (continued)

Girard, C., Ecalte, J., & Magnan, A. (2013). Serious games as new educational tools: How effective are they? A meta-analysis of recent studies, <i>Journal of Computer Assisted Learning</i> , 29(3), 207–219. (<i>n</i> = 30)
Annetta, L. A., Minogue, J., Holmes, S. Y., & Cheng, M.-T. (2009); Baker, R. S. J. D., D'Mello, M. M. T., & Graesser, A. C. (2010); Beale, I. L., Kato, P. M., Marin-Bowling, V. M., Guthrie, N., & Cole, S. W. (2007); Bloomfield, J., Roberts, J., & White, A. (2010); Boot, W. R., Kramer A. F., Simons, D. J., Fabiani, M., & Gratton G. (2008); Brom, C., Preuss, M., & Klement, D. (2011); Ciavarro, C., Dobson, M., & Goodman, D. (2008); Collier, B. D., & Scott, M. J. (2009); Duque, G., Fung, S., Mallet, L., Posel, N., & Fleischer, D. (2008); Echeverría, A., García-Campo, C., Nussbaum, M., Gil, F., Villalta, M., Améstica, M., & Echeverría, S. (2011); Hainey, T., Connolly, T., Stansfield, M., & Boyle, E. A. (2011); Ke, F. (2008); Kebrtchi, M., Hirumi, A., & Bai, H. (2010); Kim, B., Park, H., & Baek, Y. (2009); Knight, J. F., Carley, S., Tregunna, B., Jarvis, S., Smithies, R., de Freitas, S., Dunwell, I., & Mackway-Jones, K. (2010); Kobes, M., Helsloot, I., de Vries, B., & Post, J. (2010); Lim, C. P. (2008); Lindström, P., Gulz, A., Haake, M., & Sjöden, B. (2011); Liu, T.-Y., & Chu, Y.-L. (2010); Lorant-Royer, S., Munch, C., Mescélé, H., & Lieury, A. (2010); Lorant-Royer, S., Spiess, V., Goncalves, J., & Miller, A. (2008); Miller, L. M., Chang, C.-I., Wang, S., Beier, M. E., & Klisch, Y. (2011); Papastergiou (2009); Radillo, A. (2009); Robertson, D., & Miller, D. (2009); Sung, Y.-T., Chang, K.-E., & Huang, J.-S. (2008); Tanes, Z., & Cemalcilar, Z. (2010); Tuzun, H., Yilmaz-Soylu, M., Karakus, T., Inal, Y., & Kizilkaya G. (2009); Watson, W. R., Mong, C. J., & Harris, C. A. (2011); Wrzesien, M., & Raya, M. A. (2010)
Graafland, M., Schraagen, J. M., & Schijven, M. P. (2012). Systematic review of serious games for medical education and surgical skills training. <i>British Journal of Surgery</i> , 99(10), 1322–1330. (<i>n</i> = 20)
Andreatta, P. B., Maslowski, E., Petty, S., Shim, W., Marsh, M., Hall, T., et al. (2010); Badurdeen, S., Abdul-Samad, O., Story, G., Wilson, C., Down, S., & Harris, A. (2010); Bergeron, B. P. (2008); Bokhari, R., Bollman-McGregor, J., Kahol, K., Smith, M., Feinstein, A., Ferrara, J. (2010); Bowyer, M. W., Streete, K. A., Muniz, G. M., & Liu, A. V. (2008); Caudell, T. P., Summers, K. L., Holten, J., Hakamata, T., Mowafi, M., Jacobs, J., et al. (2003); Cowan, B., Sabri, H., Kapralos, B., Porte, M., Backstein, D., Cristancho, S., et al. (2010); Creutzfeldt, J., Hedman, L., Medin, C., Heinrichs, W. L., & Föllander-Tsai, L. (2010); Dev, P., Heinrichs, W. L., Youngblood, P., Kung, S., Cheng, R., Kusumoto, L., et al. (2007); Heinrichs, W. L., Youngblood, P., Harter, P., Kusumoto, L., & Dev, P. (2010); Knight, J. F., Carley, S., Tregunna, B., Jarvis, S., Smithies, R., de Freitas, S., et al. (2010); Kurenov, S. N., Cance, W. W., Noel, B., Mozingo, D. W. (2009); LeRoy, H. W., Youngblood, P., Harter, P. M., & Dev, P. (2008); Parvati, D. E. V., Heinrichs, W. L., & Patricia, Y. (2011); Rosenberg, B. H., Landsittel, D., & Averbch, T. D. (2005); Rosser, J. C., Jr, Lynch, P. J., Cuddihy, L., Gentile, D. A., Klonsky, J., & Merrell, R. (2007); Sadandanan, S., Dryfhout, V. L., & Sosnowski, J. P. (2008); Schlickum, M. K., Hedman, L., Enochsson, L., Kjellin, A., & Föllander-Tsai, L. (2009); Taekman, J. M., & Shelley, K. (2010); Youngblood, P., Harter, P. M., Srivastava, S., Moffett, S., Heinrichs, W. L., & Dev P. (2008)

- Papastergiou, M. (2009).** Exploring the potential of computer and video games for health and physical education: A literature review. *Computers & Education*, 53(3), November 2009, 603–622. (*n*=19)
- Bartholomew, L., Gold, R., Parcel, G., Cyzewski, D., Sockrider, M., Fernandez, M., et al. (2000); Beale, I., Kato, P., Marin-Bowling, V., Guthrie, N., & Cole, S. (2007); Chin, M., Paw, A., Jacobs, W., Vaessen, E., Titze, S., & van Mechelen, W. (2008); Ciavarro, C., Meanley, J., Bizzocchi, J., & Goodman, D. (2005); Cullen, K., Watson, K., Baranowski, T., Baranowski, J., & Zakeri, I. (2005); Fery, Y., & Ponserre, S. (2001); Goodman, D., Bradley, N., Paras, B., Williamson, I., & Bizzocchi, J. (2006); Hewitt, M., Denman, S., Hayes, L., Pearson, J., & Wallbanks, C. (2001); Hornung, R., Lennon, P., Garrett, J., DeVellis, R., Weinberg, P., & Strecher, V. (2000); Lieberman, D. (2001); Munguba, M., Valdes, M., & da Silva, C. (2008); Russell, W. (2007); Sell, K., Lillie, T., & Taylor, J. (2007); Silk, K., Sherry, J., Winn, B., Keecker, N., Horodynski, M., & Sayir, A. (2008); Tan, B., Aziz, A., Chua, K., & Teh, K. (2002); Turnin, M., Tauber, M., Couvaras, O., Jourret, B., Bolzonella, C., Bourgeois, O., et al. (2001); Tuzun, H. (2007); Umnithan, V., Houser, W., & Fernhall, B. (2006); Yawn, B., Algatt-Bergstrom, P., Yawn, R., Wollan, P., Greco, M., Gleason, M., et al. (2000)
- Vandercruyse, S., Vandewaele, M., & Clarebout, G. (2012).** Game based learning: A review on the effectiveness of educational games. In M. M. Cruz-Cunha (Ed.), *Handbook of research on serious games as educational, business, and research tools* (pp. 628–647). Hershey, PA: IGI Global. (*n*=22)
- Annetta, L. A., Mangrum, J., Holmes, S., Collazo, K., & Cheng, M.-T. (2009a); Annetta, L. A., Mimogue, J., Holmes, S. Y., & Cheng, M.-T. (2009b); Beale, I. L., Kato, P. M., Marin-Bowling, V. M., Guthrie, N., & Cole, S. W. (2007); Brown, S. J., Lieberman, D. A., Gemeny, B. A., Fan, Y. C., Wilson, D. M., & Pasta, D. J. (1997); Chuang, T.-Y., & Chen, W.-F. (2009); Din, F. S., & Calao, J. (2001); Ebner, M., & Holzinger, A. (2005); Fontana, L., & Beckerman, A. (2004); Goodman, D., Bradley, N. L., Paras, B., Williamson, I. J., & Bizzocchi, J. (2006); Huitzenga, J., Admiraal, W., Akkerman, S., & ten Dam, G. (2009); Ke, F. (2008); Ke, F., & Grabowski, B. (2007); Miller, D. J., & Robertson, D. P. (2010); Moreno, R., & Mayer, R. (2005); Papastergiou (2009); Prendinger, H., Mori, J., & Ishizuka, M. (2005); Richards, D., Fassbender, E., Bilgin, A., & Thompson, W. F. (2008); Virvou, M., Katsionis, G., & Manos, K. (2005); Warren, S. J., Dondlinger, M. J., & Barab, S. A. (2008); Wrzesnien, M., & Raya, M. A. (2010); Yip, F. W. M., & Kwan, A. C. M. (2006); Yu, F.-Y. (2003)
- Wattanasoontorn, V., Boada, I., García, R., Sbert, M. (2013)** Serious games for health. *Entertainment Computing*, 4(4), December 2013, 231–247. (*n*=91)

(continued)

Table 2.1 (continued)

- Anchor Bay Entertainment (2008); Anderson, C. (2008); Applied Research Associates Inc (2012); Archimage Inc (2006); Association RMC/BFM (2011); Atkinson, S., & Narasimhan, V. (2010); Bartolome, N., Zorrilla, A., & Zapirain, B. (2010); BBG Entertainment (2009); Blitz Games (2009); Botella, C., Breton-Lopez, J., Quero, S., Baos, R., Garcia-Palacios, A., Zaragoza, I., & Alcaniz, A. (2011); BreakAway Ltd (2010); Burke, J. W., McNeill, M. D. J., Charles, D. K., Morrow, P. J., Crosbie, J. H., McDonough, S. M. (2009); Cagatay, M., Ege, P., Tokdemir, G., & Cagiltay, N. (2012); Chan, W., Qin, J., Chui, Y., & Heng, P. (2012); Clawson, J., Patel, N., & Starnet, T. (2010); Collision Studios (2009); De Bortoli A., & Gaggi, O (2011); de Urturi, Z., Zorrilla, A., & Zapirain, B. (2011); Deponiti, D., Maggiorini, D., & Palazzi, C. (2009); Diehl, L., Lehmann, E., Souza, R., Alves, J., Esteves, R., & Gordan, P. (2011); EAD Nintendo (2007); Edheads (2007); e-Learning Studios (2012); Electronic Arts (2010); EMCO3 (2012); Faculty of Medicine, Imperial College London (2008); fatworld.org (2007); Finkelstein, J., Wood, J., Cha, E., Orlov, A., & Dennison, C. (2010); Fishing Cactus (2010); Fuchslöcher, A., Niesenhaus, J., & Kirner, N. (2011); Gago, J., Barreira, T., Carrasosa, R., & Segovia, P. (2010); Gameloft (2006); GENIOUS Interactive (2012); Glasgow Caledonian University (2008); Grau, S., Tost, D., Campeny, R., Moya, S., & Ruiz, M. (2010); Hatfield, D. (2008); Hopelab (2010); Imbeault, F., Bouchard, B., & Bouzouane, A. (2011); Inhalotherapie Kinsithrapie Respiratory Association for Research and Education (2012); Innovation in Learning Inc (2010); Intelligent Systems (2007); Janomedia (2006); Johnston, E., & Duskin, B. (2004); K.T.M. Advance (2010); Kim, J. A., Kang, K. K., Yang, H. R., & Kim, D. (2009); Laikari, A. (2009); Lakeside Center for Autism (2011); LearningGames Lab (2009); Lightning Fish Games (2009); Lin, J. K., Cheng, P. H., Su, Y., Wang, S. Y., Lin, H. W., Hou, H. C., Chiang, W. C., Wu, S. W., Luh, J. J., & Su, M. J. (2011); LudoMedic 2011 (2011); Mc G.III. University (2006); Mckanna, J. A., Jimison, H., & Pavel, M. (2009); Mili, F., Barr, J., Harris, M., & Pittiglio, L. (2008); Miller, J. (2010); Milo Foundation (2012); MIRROR project (2012); Montreal Science Centre (2004); Moya, S., Grau, S., Tost, D., Campeny, R., & Ruiz, M. (2011); Nauta, H., & Spil, T. (2011); Nike Inc (2012); Nintendo (2005); Nordic Innovation Centre (2007); Pervasive Games (2009); QOVEO (2009); Queiros, S., Vilaca, J., Rodrigues, N., Neves, S., Teixeira, P., & Correia-Pinto, J. (2011); RANJ Serious Games (2009); Raylight Srl (2009); Red Hill Studios (2012); ResponDesign (2004); Sabri, H., Cowan, B., Kapralos, B., Porte, M., Backstein, D., & Dubrowskie, A. (2010); Scarle, S., Dunwell, I., Bashford-Rogers, T., Selmanovic, E., DeBattista, K., Chalmers, A., Powell, J., & Robertson W. (2011); Schnauer, C., Pintaric, T., Kosterink, S. J., & Vollenbroek, M. (2011); Skills2Learn Ltd (2010); Sliney, A., & Murphy, D. (2008); Succubus Interactive (2009); The Diablotines (2011); The Partnership for Food Safety Education (2008); TruSim, Blitz Games Studios (2008); Ubisoft Divertissements (2012); Van Loon, E., Peper, C., van de Rijt, A., & Salverda, A. (2011); Vazquez, M., Santana-Lopez, V., Skodova, M., Ferrero-Alvarez-Rementeria, J., & Torres-Olivera A. (2011); Verduin, M. L., LaRowe, S. D., Myrick, H., Cannon-Bowers, J., & Bowers, C. (2012); Vermont Department of Health (2008); Vermont department of health, Khemia (2008); Vidani, A., Chittaro, L., & Carchietti, E. (2010); Virtual Heroes Inc (2009); Visual Imagination Software (2010); Vtmen, A., & Leikas, J. (2009); Wang, Q., Sourina, O., & Nguyen, M. K. (2010); Warner Bros Entertainment (2008)

- Wilson, K. A., Bedwell, W. L., Lazzara, E. H., Salas, E., Burke, C. S., Estock, J. L., Orvis, K. L., & Conke, C. (2009).** Relationships between game attributes and learning outcomes: Review and research proposals. *Simulation Gaming, 40*(2), 217–266. (n=42)
- Adams (1998); Belanich, J., Sibley, D. E., & Orvis, K. L. (2004); Blunt (2007); Bowers & Jentsch (2001); Crown (2001); Day, Arthur, & Gettman (2001); Dennis & Harris (1998); Driskell, J. E., & Dwyer, J. D. (1984); Gander, S. (2002); Garris & Ahlers (2001); Gopher, D., Weil, M., & Bareket, T. (1994); Gremmen, H., & Potters, H. (1995); Habgood, M. P. J. (2005); Habgood, M. P. J., Ainsworth, S. E., & Benford, S. (2005); Lepper, M. R. (1985); Lepper, M. R., & Chabay, R. W. (1985); Malone, T. W. (1981); Marks (2000); Mayer, R. E., Mautone, P., & Prothero, W. (2002); McFarlane, Sparrowhawk, & Heald (2002); Merzenich, M. M., Jenkins, W. M., Johnston, J. P., Schreiner, C., Miller, S. L., & Tallal, P. (1996); Noble, A., Best, D., Sidell, C., & Strang, J. (2000); Orvis, K. A., Orvis, K. L., Belanich, J., & Mullin, L. N. (2008); Pange, J. (2003); Parker & Lepper (1992); Parry, S. B. (1971); Prince, C., & Jentsch, F. (2001); Proctor, M. D., Panko, M., & Donovan, S. J. (2004); Ricci, K. E., Salas, E., & Cannon-Bowers, J. A. (1996); Rieber, L. P. (1996); Rieber, L. P., & Noah, D. (1997); Ronen, M., & Eliahu, M. (2000); Rosas, R., Nussbaum, M., Cumsille, P., Marianov, V., Correa, M., & Flores, P., et al. (2003); Serrano, E. L., & Anderson, J. E. (2004); Squire, Giovanetto, Devane, & Durga (2005); Thomas, R., Cahill, J., & Santilli, L. (1997); Van Eck (2006); Van Eck, R., & Dempsey, J. (2002); Veale, T. K. (1999); Virvou, M., Katsionis, G., & Manos, K. (2005); Westrom, M., & Shaban, A. (1992); Woodman, M. D. (2006)

Table 2.2 Number of serious game papers for each year in the 10 year period from 2003 to 2012

Year	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
No. of papers	2	7	15	10	18	49	19	22	12	3

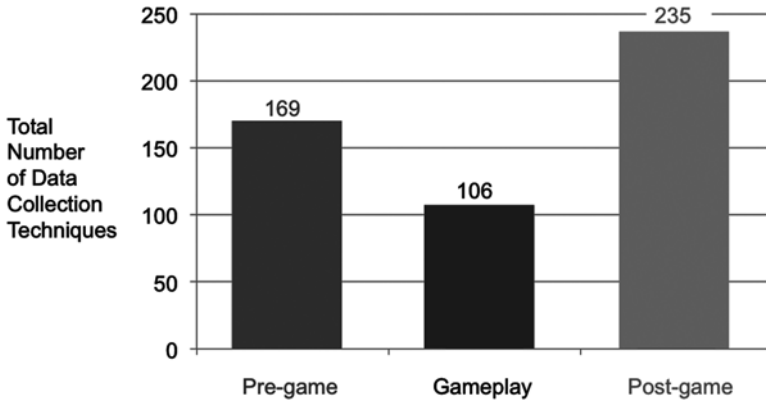


Fig. 2.4 Number of data collection techniques used per phase of study

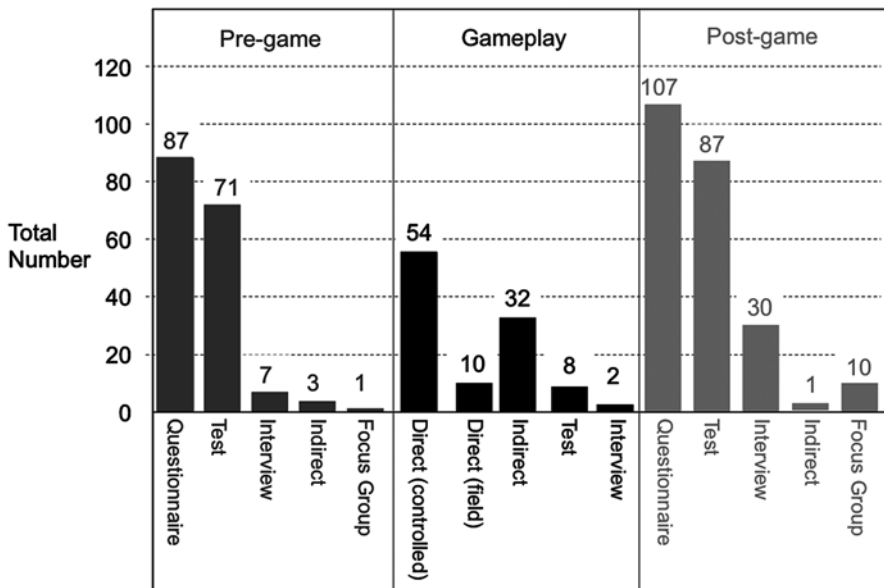


Fig. 2.5 Number of specific data collection techniques used per phase of study

A single study used an indirect observation, while 4 % of the studies employed a focus group in the post-game phase (see Fig. 2.5).

In the context of the specific techniques used during gameplay ($n=106$), 46 % of the studies used some form of direct observation in a controlled environment, 9 % of the studies used some form of direct observation in the field, 30 % of the

studies used an indirect observation method, around 8 % used a test, while two of the studies employed an interview during the gameplay phase of evaluation (see Fig. 2.5).

5 Discussion

5.1 Issues Highlighted Within Our Study Outcomes

The majority of the studies we reviewed used multiple data collection methods (80.3 %, Fig. 2.6). Surveys and questionnaires are good at getting shallow data from a large number of people but are not good at getting, deep, detailed data; participants may try to impress interviewers during interviews (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003); and duration logging systems may not take into account participant thinking time (Lazar, Feng, & Hochheiser, 2010). The studies included in our review were dominated by the use of questionnaires and formal tests (see Fig. 2.5). Also, the majority of data collection occurred post-game, followed by pre-game collection and finally captured during gameplay. Although these results may have been biased by our own sampling techniques, it may also highlight a need for integrating more focus groups and indirect observation techniques into game evaluations.

Even allowing for sampling errors, it was notable that objective techniques such as biometrics or psychometrics, as well as newer techniques such as path tracking or crowd sourcing, were largely absent. This reflects traditional difficulties of capturing gameplay data, that is, if such data collection is not part of the game design, and it also highlights opportunities for new data capture approaches oriented towards data

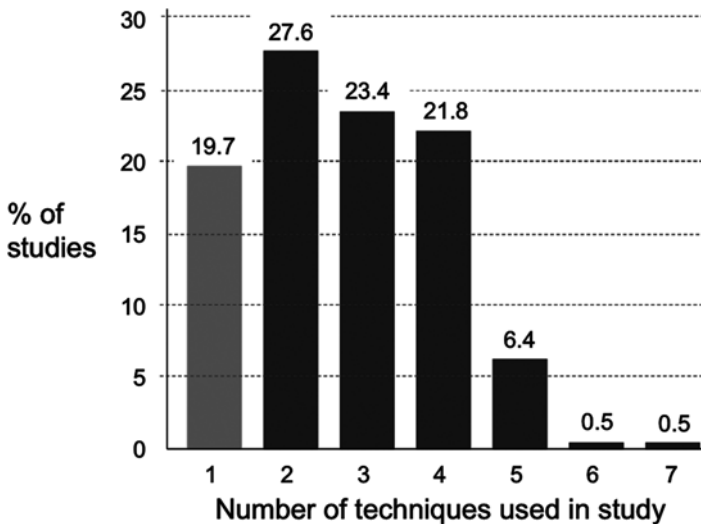


Fig. 2.6 Percentage of studies that used multiple data collection techniques

analytics. Loh (2012) notes that "... few [commercial] developers would actually be interested in 'in-process' data collection unless it somehow contributed to the usability of their games ..." and goes on to consider alternative approaches for empirical data gathering via telemetry and psychophysiological measures.

Also, different data collection techniques have inherent biases (Podsakoff et al., 2003). Thus, it is important to consider multiple data collection methods. As Rogers et al. (2011) observe, it is "important not to focus on just one data gathering technique but to use them flexibly and in combination so as to avoid biases" (p. 223). The framework used by Mayer et al. (2014) provides a good examination of the breadth of approaches that can be used in data collection.

Most of the studies were collecting data to demonstrate the use of serious games as an intervention tool, for instance, to demonstrate the impact of a serious game in an educational setting. Thus, a minimum expectation could be for a pretest and post-test, and it would also be desirable to obtain some in game data, e.g., score or duration metric. As seen in Fig. 2.6, only 52.7 % of the studies reviewed used three or more data collection techniques. Exploring this further is outside the scope of this chapter, but is an important area for future research if serious game evaluations and experimental designs are to be considered demonstrably robust.

Another feature highlighted in the data collection that occurred during gameplay was the lack of direct observation in the field (10 %) compared to observations that were made in a controlled environment such as a computer laboratory (54 %). This is understandable, as research by nature tends to occur in university environments, and controlled environments allow contextual variations associated with data collection to be controlled in traditional experimental designs. Again, our own data sampling methods make it difficult to argue the significance of this finding but it still needs to be considered that experimental serious game research might need to be extended to include more situated case studies and perhaps participatory methods.

When reflecting further on the content of the various studies we encountered during our process, as well as some of the problems encountered in the data collection process, a number of generic issues of serious games research were highlighted. These generic issues, which we discuss next, include:

- What data is being collected?
- When data is being collected?
- Where data is being collected?
- Who is involved in data collection?
- Why data is being collected?

5.2 What Data Is Being Collected?

In our study, we found there was a tendency to collect certain types of data during the different phases. During the pre-game, this data tends to include demographic information such as gender, age, nationality, and culture. It was also common to gather data surrounding previous experience and skills with computers, games, and related technology such as simulations and virtual reality. Less common is the

collection of data of a participant's attitudes, their intrinsic and extrinsic motivation, learning and personality styles, etc. In many cases, both pre- and posttests were used to generate skill or knowledge performance metrics directly related to intended serious outcomes of the game.

During the gameplay, measures tend to focus on issues of performance. Types of data include game metrics, such as time to complete, number of errors or levels of progress. Less common were measures that examined player approaches to completing the game and measures of experience such as flow, immersion, presence, and the general affective states of the participant (see, for example, Jennett et al., 2008).

Loh (2011) considers gameplay measures with the analogue of Black box and open box approaches in the context of game assessment metrics. Specifically, he defines *ex situ* data collection where the game environment is a Black box and data is collected without access to internal details. This could be the pre- and post-game collections metrics noted in this chapter (see Fig. 2.5) or psychophysiological measures collected during game sessions. The open box approach supports *in situ* data collection, for example, log files, game events, or user-generated action data, e.g., *Information Trails* (Loh, 2012). In contrast to psychophysiological measures, such *in situ* data would have no external noise as the data collection occurred within a closed environment. This could be of significant interest for serious game analytic approaches as a way to triangulate data across collection sources, similar to the use of *immersidata* to collect and index user-player behaviors from gameplay logs and video clips (Marsh, Smith, Yang, & Shahabi, 2006). Also, as noted in the previous section, there have traditionally been difficulties in capturing and using *in situ* data as it requires access to the internal processes of a serious game (e.g., to collect telemetry data), and it can be problematic to efficiently process the large volumes of data generated. Both topics are prominent in the other chapters of this book and are a focus of ongoing serious game analytics research.

During post-game evaluations, it was more common to obtain subjective feedback concerning game experience and issues surrounding fun and engagement. This phase was also when measures of player satisfaction with the game, such as clarity, realism, aesthetics, and ease of use, as well as perceived suitability were usually made.

Other types of data that might be useful to collect within studies include the quality or experience of any facilitators involved, the general context of field studies such as the interaction with others and their roles in the study, and potential organizational impacts such as management structure and culture (Mayer et al., 2014).

Some of the variation in data collection is related to the intention of studies and whether they relate to measuring the efficacy of serious outcomes, or the usability and quality of the game itself. Many studies address both efficacy and usability issues as they are related. One issue that needs to be considered in relation to what data is collected surrounds player profiling. The importance of this is highlighted in one study that used the specialized "Ravens advanced progressive matrices" to examine the relationship between general cognitive ability and any measured knowledge outcomes from the game (Day, Arthur, & Gettman, 2001). The inference is that underlying individual traits such as cognitive ability might be a good indicator of player performance in learning tasks. This suggests other psychological tests that might assist in measuring player traits such as risk-taking, general personality

traits, performance under stress, learning styles, teamwork, and other factors that might be relevant in some applications.

A benefit of adopting these traditional instruments is that they have been validated and are well understood, at least under laboratory conditions. Otherwise, we might also need to question the validity of questionnaires, surveys, and other measuring instruments currently being used in game studies (Boyle, Connolly, & Hainey, 2011; Slater, 2004).

5.3 When Data Is Being Collected?

Our own study identified a variety of data being collected in the post-game, game-play, and post-game phases of evaluation. Most of the data collected in our study occurred post-game, while the least occurred during the gameplay phase. Arguably there is an opportunity to improve levels of data collection occurring during the gameplay to support a better understanding of how specific game elements relate to the intended serious outcomes.

There are also other aspects of timing that should be considered in data collection. This may partly be related to the whether the study is intent on measuring aspects of the process or purely outcomes (Bowers & Jentsch, 2001). Thus, the relevance of process evaluation versus game efficacy or usability measures may impact on when data is collected.

In terms of learning applications, it may also be important to consider interactions between other forms of instruction that occur before, during, or after the game intervention (Van Eck, 2006). This might also apply to application of serious games for health, where additional treatments may occur in conjunction with game use.

This highlights the issue of deciding when and how often to collect data for evaluating games. Although many studies used mixed methods, data was not necessarily collected over the life of the study. By contrast, in the study by Squire, Giovanetto, Devane, and Durga (2005) games were played over 5 weeks and data was collected over this entire time frame. The time frame of data collection may be influenced by the intent and domain of the study, for example, whether the research is concerned with the direct and immediate influence of playing the game versus the indirect or long-term impact of the game. It is probably important to get short-term feedback involving gaining self-reported, subjective feedback from participants, for example, regarding participant satisfaction, or self-perceived learning as well as immediate changes to attitudes, skills, or knowledge. Medium or longer term data might be required to understand aspects of team or organizational change especially related to social issues.

We also found that the timing of outcome measures varied depending on domain. For example, in some learning applications there may be a greater tendency to measure longer term learning factors such as the time required to transfer or regain knowledge (Day et al., 2001; Dennis & Harris, 1998; Parker & Lepper, 1992). This implies testing skills, not just immediately after completing the game, but also at later intervals such as a few days, weeks, or months to measure the permanence of any immediate outcomes and issues of retention and reacquisition of knowledge.

5.4 Where Data Is Being Collected?

There was a tendency for evaluation to occur in controlled rather than field situations. Data was collected in a variety of contexts including primary schools, secondary schools, universities, and industry settings. One potential issue of study context is what else is happening during the study that might impact on outcomes and yet is not necessarily being reported (Van Eck, 2006).

The location of data collection also indirectly raises issues of cost. For example, onsite studies should be fast and efficient to ensure they do not unnecessarily impact on the time or resources of participating partners (Mayer et al., 2014). It also confirms the need for unobtrusive and perhaps covert data collection techniques (Mayer et al., 2014), not just to improve data validity, but to minimize impact on the standard workflow of participants involved in case studies, for example, the use of in situ methods (Loh, 2011). Stakeholders may also need to be persuaded that more extensive contextual data as well as extended longitudinal data gathering needs to occur beyond the obvious and minimal (Mayer et al., 2014).

5.5 Who Is Involved in Data Collection?

In our study, we identified a range of stakeholders involved in projects including students, teachers, researchers, game developers, and industry partners. All of these various stakeholders are candidates to be involved in evaluation. Such evaluations may need to bear in mind influences related to the motivations of stakeholders surrounding the process and outcomes. For example, the game designer may be enthusiastic to measure the aesthetics, the software engineer the usability, the scientist, the efficacy, and the manager the cost. All stakeholders may also be keen to find positive outcomes whether the motivation is for publication, ongoing employment, or other personal gains. Thus, it may be worthwhile to consider collecting data related to the exact role of various participants in the project and any intrinsic and extrinsic motivations of the parties (Mayer et al., 2014).

Marks (2000) highlights the obvious sampling issues in one project where university students were used to evaluate a game intended to teach military staff. It was not clear that measured effects on such a population would transfer to the intended group. By contrast, in another study three different questionnaires are used for pupils (players), parents, and teachers (McFarlane, Sparrowhawk, & Heald, 2002).

While most studies focus on individuals playing games, there is also interest in evaluating the efficacy of learning team-based, rather than individual, skills. Marks (2000), in considering some of the pros and cons of using computer simulations for team research, highlights the need for measuring the longitudinal impact of skills related to teamwork. Data may need to be collected that considers team performance rather than individual performance where games are designed to teach teamwork (Bowers & Jentsch, 2001). While a number of evaluation models exist that

focus on the learning of individuals, there has been less attention given to the data required to assess learning in teams or in larger collectives such as organizations and informal networks (Mayer et al., 2014).

5.6 *Why Data Is Being Collected?*

During our meta-review, we encountered the use of games across a wide variety of different domains and not surprisingly, we found a variety of expectations about the most appropriate methods of data collection and the types of data collected. For example, in one study participants were partly evaluated on the basis of an essay that reflected on their experience using the game (Adams, 1998). This is in contrast to another learning study that directly measured changes to student knowledge using tests as well as surveying the students and seeking feedback from external stakeholders such as parents and teachers (Crown, 2001).

While it is easy to understand the reasons for such differences, the variations make it harder to compare and contrast data results from different game studies. The usefulness to the serious game community of adopting standardized testing approaches that allow for comparison has been highlighted previously (Blunt, 2007; Mayer et al., 2014).

Despite some good work in the area of relating game design features to serious outcomes (Wilson et al., 2009), most studies focus on collecting data to support the message of efficacy rather than data that helps explain why and how they are effective or indeed how to apply design rules that lead to the required efficacy (Garris & Ahlers, 2001; Van Eck, 2006).

6 Conclusions

A complication of data collection for games is that not all games are created equal (Loh, 2009). Van Eck (2006) makes the key point that any taxonomy of games is as complex as learning taxonomies. While not all games are the same, the situation is complicated by the overlap of simulations, virtual reality, and partial gamification of traditional approaches. There is also wide variety in the types of games being used in studies. Some are small in scope and custom built by individuals while others are constructed in multi-discipline projects that involve discipline experts and professional game developers. Other studies simply make use of off-the-shelf games. This range of projects means that the data collection techniques need to be flexible.

In this study, we developed a review process for performing a meta-analysis on data collection techniques used in serious game research. We found that while many studies used a variety of methods, they were not necessarily intended to triangulate findings. The number of data collection techniques also varied considerably, with a number of studies using only a single measure. Our study also highlighted a number

of variations and subsequently raised questions around what data is being collected, when data is being collected, how data is being collected, where data is being collected, and why data is being collected.

Our systematic review approach identified a number of significant literature reviews that allowed us to examine data collection processes across broad domain and temporal spaces. However, not all bibliographic sources were included in the search parameters and thus relevant literature has potentially been missed. Despite this, the results provide a representative sample of serious game research that allows us to draw valid conclusions about approaches and issues in data collection.

In summary, the data collected for serious game research is broad in scope, measuring both targeted performance skills, behavioral factors related to both the process and outcomes. For example, the data may be designed to measure changes in knowledge, attitudes, skills, or behavior. The data collected can also be multi-level in scope, designed to measure fine grain individual skills or large-scale organization attitudes. Data is collected using a wide range of objective and subjective methods that may fall across a range of longitudinal scales. The currently used data collection techniques might align more with discipline traditions than necessarily intentions of evaluations. Even though single studies often incorporate a variety of techniques, the data is not necessarily triangulated as might be expected in a true mixed method approach. The review also found that the majority of data collection occurred post-game, then pre-game, and finally during gameplay. This, perhaps, reflects traditional difficulties of capturing gameplay data and highlights opportunities for new data capture (i.e., in situ collection) and analysis approaches oriented towards data analytics. We suggest that more standardized and better-validated data collection techniques, that allow comparing and contrasting outcomes between studies, would be beneficial to the broader serious games community and specifically to those interested in serious game analytics.

References

- Adams, P. C. (1998). Teaching and learning with SimCity 2000. *Journal of Geography*, 97(2), 47–55.
- Anderson, E., McLoughlin, L., Liarokapis, F., Peters, C., Petridis, P., & Freitas, S. (2010). Developing serious games for cultural heritage: a state-of-the-art review. *Virtual Reality*, 14(4), 255–275. doi:10.1007/s10055-010-0177-3.
- Bearman, M., Smith, C. D., Carbone, A., Slade, S., Baik, C., Hughes-Warrington, M., et al. (2012). Systematic review methodology in higher education. *Higher Education Research & Development*, 31(5), 625–640.
- Blakely, G., Skirton, H., Cooper, S., Allum, P., & Nelmes, P. (2009). Educational gaming in the health sciences: Systematic review. *Journal of Advanced Nursing*, 65(2), 259–269. doi:10.1111/j.1365-2648.2008.04843.x.
- Blunt, R. (2007). Does game-based learning work? Results from three recent studies. In *Proceedings of the Interservice/Industry Training, Simulation, & Education Conference* (pp. 945–955). Orlando, FL: National Defense Industrial Association.
- Bowers, C. A., & Jentsch, F. (2001). Use of commercial off-the-shelf, simulations for team research. In C. A. Bowers & E. Salas (Eds.), *Advances in human performance and cognitive engineering research* (pp. 293–317). Mahwah, NJ: Lawrence Erlbaum.

- Boyle, E. A., Connolly, T. M., & Hainey, T. (2011). The role of psychology in understanding the impact of computer games. *Entertainment Computing*, 2(2), 69–74.
- Connolly, T. M., Boyle, E. A., MacArthur, E., Hainey, T., & Boyle, J. M. (2012). A systematic literature review of empirical evidence on computer games and serious games. *Computers & Education*, 59(2), 661–686. doi:10.1016/j.compedu.2012.03.004.
- Crookall, D. (2010). Serious games, debriefing, and simulation/gaming as a discipline. *Simulation & Gaming*, 41(6), 898–920.
- Crown, S. W. (2001). Improving visualization skills of engineering graphics students using simple JavaScript web based games. *Journal of Engineering Education*, 90(3), 347–355.
- Day, E. A., Arthur, W., Jr., & Gettman, D. (2001). Knowledge structures and the acquisition of a complex skill. *Journal of Applied Psychology*, 86(5), 1022–1033.
- Dennis, K. A., & Harris, D. (1998). Computer-based simulation as an adjunct to ab initio flight training. *International Journal of Aviation Psychology*, 8(3), 261–276.
- Garris, R., & Ahlers, R. (2001, December). *A game-based training model: Development, application, and evaluation*. Paper presented at the Interservice/Industry Training, Simulation & Education Conference, Orlando, FL.
- Garris, R., Ahlers, R., & Driskell, J. E. (2002). Games, motivation and learning: A research and practice model. *Simulation & Gaming*, 33(4), 441–467.
- Girard, C., Ecalle, J., & Magnan, A. (2013). Serious games as new educational tools: How effective are they? A meta-analysis of recent studies. *Journal of Computer Assisted Learning*, 29(3), 207–219. doi:10.1111/j.1365-2729.2012.00489.x.
- González, L. S., Rubio, F. G., González, F. R., & Velthuis, M. P. (2010). Measurement in business processes: A systematic review. *Business Process Management Journal*, 16(1), 114–134.
- Graafland, M., Schraagen, J. M., & Schijven, M. P. (2012). Systematic review of serious games for medical education and surgical skills training. *British Journal of Surgery*, 99(10), 1322–1330.
- Hwang, G.-J., & Wu, P.-H. (2012). Advancements and trends in digital game-based learning research: A review of publications in selected journals from 2001 to 2010. *Journal of Educational Technology*, 43(1), E6–E10. doi:10.1111/j.1467-8535.2011.01242.x. Blackwell.
- Jennett, C., Cox, A. L., Cairns, P., Dhoparee, S., Epps, A., Tijs, T., et al. (2008). Measuring and defining the experience of immersion in games. *International Journal of Human Computer Studies*, 66(9), 641–661.
- Kitchenham, B., Brereton, O. P., Budgen, D., Turner, M., Bailey, J., & Linkman, S. (2009). Systematic literature reviews in software engineering—A systematic literature review. *Information and Software Technology*, 51(1), 7–15.
- Lazar, J., Feng, J. H., & Hochheiser, H. (2010). *Research methods in human-computer interaction*. Chichester, England: Wiley.
- Loh, C. S. (2009). Research and developing serious games as interactive learning instructions. *International Journal of Gaming and Computer-Mediated Simulations*, 1(4), 1–19. IGI Global.
- Loh, C. S. (2011). Using in situ data collection to improve the impact and return of investment of game-based learning. In *Proceedings of ICEM-SIIE 2011, the 61st International Council for Educational Media (ICEM) and the XIII International Symposium on Computers in Education (SIIE) Joint Conference* (pp. 801–811). ICEM-SIIE.
- Loh, C. S. (2012). Information trails: In-process assessment for game-based learning. In D. Ifenthaler, D. Eseryel, & X. Ge (Eds.), *Assessment in game-based learning: Foundations, innovations, and perspectives* (pp. 123–144). New York: Springer.
- Marks, M. A. (2000). A critical analysis of computer simulations for conducting team research. *Small Group Research*, 31(6), 653–675.
- Marsh, T., Smith, S. P., Yang, K., & Shahabi, C. (2006) Continuous and unobtrusive capture of user-player behaviour and experience to assess and inform game design and development. In *1st World Conference for Fun'n Games (FNG 2006)* (pp. 79–86). Preston, England: University of Central Lancaster.
- Mayer, I., Bekebrede, G., Hartevelde, C., Warmelink, H., Zhou, Q., Ruijven, T., et al. (2014). The research and evaluation of serious games: Toward a comprehensive methodology. *British Journal of Educational Technology*, 45(3), 502–527.

- McFarlane, A., Sparrowhawk, A., & Heald, Y. (2002). *Report on the educational use of games: An exploration by TEEM of the contribution which games can make to the education process*. Last Access on 11th April, from https://pantherfile.uwm.edu/tjoosten/LTC/Gaming/teem_gamesined_full.pdf
- McMahan, R. P., Ragan, E. D., Leal, A., Beaton, R. J., & Bowman, D. A. (2011). Considerations for the use of commercial video games in controlled experiments. *Entertainment Computing*, 2(1), 3–9.
- Papastergiou, M. (2009). Exploring the potential of computer and video games for health and physical education: A literature review. *Computers & Education*, 53(3), 603–622. doi:10.1016/j.compedu.2009.04.001.
- Parker, L. E., & Lepper, M. R. (1992). Effects of fantasy contexts on children's learning and motivation: Making learning more fun. *Journal of Personality and Social Psychology*, 62(4), 625–633.
- Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: a critical review of the literature and recommended remedies. *Journal of Applied Psychology*, 88(5), 879.
- Rogers, Y., Sharp, H., & Preece, J. (2011). *Interaction design: Beyond human-computer interaction* (3rd ed.). Hoboken, NJ: Wiley.
- Slater, M. (2004). How colorful was your day? Why questionnaires cannot assess presence in virtual environments. *Presence*, 13(4), 484–493.
- Šmite, D., Wohlin, C., Gorschek, T., & Feldt, R. (2010). Empirical evidence in global software engineering: A systematic review. *Empirical Software Engineering*, 15(1), 91–118.
- Smith, S. P., Stibric, M., & Smithson, D. (2013). Exploring the effectiveness of commercial and custom-built games for cognitive training. *Computers in Human Behavior*, 29(6), 2388–2393.
- Squire, K., Giovanetto, L., Devane, B., & Durga, S. (2005). From users to designers: Building a self organizing game-based learning environment. *TechTrends*, 49(5), 34–42.
- Van Eck, R. (2006). Digital game based learning: It's not just the digital natives who are restless. *Educause Review*, 41(2), 17–30.
- Vandercruyssen, S., Vandewaetere, M., & Clarebout, G. (2012). Game based learning: A review on the effectiveness of educational games. In M. M. Cruz-Cunha (Ed.), *Handbook of research on serious games as educational, business, and research tools* (pp. 628–647). Hershey, PA: IGI Global.
- Wattanasoontorn, V., Boada, I., García, R., & Sbert, M. (2013). Serious games for health. *Entertainment Computing*, 4(4), 231–247.
- Wilson, K. A., Bedwell, W. L., Lazzara, E. H., Salas, E., Burke, C. S., Estock, J. L., et al. (2009). Relationships between game attributes and learning outcomes review and research proposals. *Simulation & Gaming*, 40(2), 217–266.