Sheffield Hallam University

Zombie Division : a methodological case study for the evaluation of game-based learning

HABGOOD, M. P. Jacob <http://orcid.org/0000-0003-4531-0507>

Available from Sheffield Hallam University Research Archive (SHURA) at:

http://shura.shu.ac.uk/10652/

This document is the author deposited version. You are advised to consult the publisher's version if you wish to cite from it.

Published version

HABGOOD, M. P. Jacob (2015). Zombie Division : a methodological case study for the evaluation of game-based learning. In: 9th European Conference on Games Based Learning : ECGBL 2015, Steinkjer, Norway, 8-9 October 2015. (In Press)

Copyright and re-use policy

See http://shura.shu.ac.uk/information.html

Sheffield Hallam University Research Archive
http://shura.shu.ac.uk

Zombie Division: A Methodological Case Study for the Evaluation of Game-Based Learning

M. P. Jacob Habgood
Sheffield Hallam University, Sheffield, UK
j.habgood@shu.ac.uk

Abstract: This paper discusses the methodological designs and technologies used to evaluate an educational videogame in order to support researchers in the design of their own evaluative research in the field of game-based learning. The Zombie Division videogame has been used to empirically evaluate the effectiveness of a more *intrinsically integrated* approach to creating educational games. It was specifically designed to deliver interventions as part of research studies examining differences in learning outcomes and motivation predicted by theoretical contrasts in educational design. The game was used in a series of evaluative studies, which employed experimental methodologies based around one or more treatment groups and a control. Multiple choice questions were used to measure knowledge and understanding before and after interventions (pre, post and delayed) and time-on-task was used as a measure of motivation and preference during interventions. Qualitative interview data was also collected and analysed as part of many of the studies in order to help support and explain the findings in more detail.

The experimental methodologies applied in these studies were augmented by a range of bespoke technology systems. This included an automated testing system which could randomly assign participants to treatment groups so that pre-test statistics were closely matched between groups. Large quantities of process data were recorded about players' interactions with the game in the form of time-stamped log files, and a stream of compressed controller data was saved allowing an entire playing session to be replayed in a video-like form. This rich set of process data was mined as part of a post-hoc analysis in order to identify evidence to help to enrich the understanding of users' interactions with the game.

This paper details the methodological design of both published and unpublished studies, as well as reflecting upon some of the potential pitfalls of classroom-based evaluations in order to illustrate successful and unsuccessful approaches for evaluating game-based learning.

Keywords: Game-Based Learning, Evaluation, Experimental Design, Learning Outcomes, Motivation

1. Introduction

Academic research into digital game-based learning now spans five decades (e.g. Cullingford, Mawdesley et al. 1979), yet relatively little has been established in terms of a common methodological practice for empirical research in this field (All, Castellar et al. 2014). Perhaps this should be expected given the inter-disciplinary nature of this area of study, which brings together computing, education and psychology, with any other discipline that seeks to apply game-based learning to its own teaching. It is natural for researchers to carry the epistemological assumptions of their own disciplines and seek to ask research questions which lend themselves to different methodological approaches. Nonetheless, the credibility of the field is hampered by its failure to produce a substantial body of empirical research demonstrating the effectiveness of game-based learning (Blunt 2007, O'Neil, Wainess et al. 2005, Abdul Jabbar, Felicia 2015). In order to facilitate more evaluative research within this field, this paper provides a practical insight into a research project which was able to empirically demonstrate an impact on motivation and learning.

Zombie Division is a third-person perspective videogame adventure which teaches mathematics to seven and eight year olds (figure 1). It was designed to examine the effectiveness of game-based learning in which the learning content is very tightly coupled with the gameplay (Habgood, Ainsworth et al. 2005). This approach was based on the observation that poor game-based learning often has a very loose coupling between learning and gameplay, allowing it to be characterised as "chocolate-covered broccoli" (Bruckman 1999). The term "intrinsic integration" was used to describe this tight coupling in keeping with earlier literature (Kafai 2001, Malone 1981) contrasting it with the "extrinsic integration" exhibited by typical 'edutainment' products (Papert 1998).

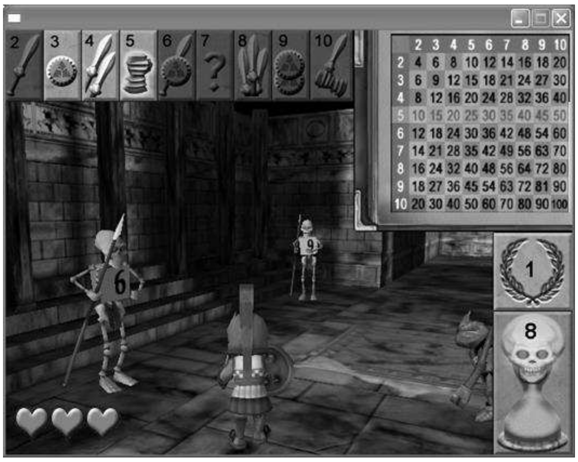

Figure 1: The intrinsic version of Zombie Division

2. Effective game-based learning

Any research design attempting to evaluate the effectiveness of game-based learning must first decide what it means to be effective. The Zombie Division research adopted a quantitative, quasi-experimental approach to evaluating effectiveness based on pre to post-test gains (as a measure of learning) and time-on-task (as a measure of motivation). In addition, a range of qualitative interview data was collected in order to enrich the interpretation of the experimental results. The combination of educational and motivational measures of effectiveness have intuitive face value for demonstrating the potential of game-based learning, but neither is without its complications.

2.1 Educationally effective

Learning gains are a common measure of effectiveness (O'Neil, Wainess et al. 2005, All, Castellar et al. 2014), but demonstrating knowledge acquisition doesn’t necessarily make a mode of learning effective—crawling hardly qualifies as an effective mode of travel, despite the ability to demonstrate progression! The concept of “effectiveness” naturally incorporates the idea of being (at least) comparable to the status quo, but finding the right comparison raises additional epistemological and methodological challenges.

There is often a desire to compare game-based learning to traditional modes of teaching (e.g. Randel, Morris et al. 1992), and frame research questions in terms of whether game-based learning is instructionally more effective than classroom teaching. However, the Zombie Division research was designed from the perspective that this kind of question realistically cannot be addressed in a typical, small to medium scale research study. The strong influence of an individual teacher’s performance during a short intervention (often at the end of a teaching year) presents a significant threat to the external validity of any results. The approach adopted within this research has tried to keep the teacher’s role as a common factor in any intervention, preferring to explore how classroom teaching can be augmented by game-based learning rather than replaced by it.

The Zombie Division studies were primarily concerned with the comparison between two different approaches to game-based learning: the loosely-coupled (extrinsic) approach offered by edutainment products (i.e. the status quo) and the tightly-coupled (intrinsic) alternative proposed by our research. The studies focussing on learning gains incorporated an additional control group condition which offered a baseline comparison with no learning (but identical gameplay). In this way it was possible to explore whether an extrinsic approach could produce any learning gains and whether the intrinsic approach was more or less effective than it.

2.2 Motivationally effective

Learning gains are not the only potential measure of an educational game’s effectiveness, and historically it is the motivational potential of game-based learning that has been the driving force for much of the interest into the field (e.g. Bowman 1982). The link between motivation, time-on-task and learning is supported by research outside of games (Vollmeyer, Rheinberg 2000) and time-on-task is an objective measure of motivation which has a long tradition of use in motivation research (e.g. Deci 1971). Self-reported measures of motivation can have a range of shortcomings (Fulmer, Frijters 2009), and lack the objectivity required to demonstrate efficacy in the same way. Consequently the Zombie Division studies recorded time-on-task as the principal measure of motivation and didn't pursue self-reported measures beyond collecting interview data to add depth to statistical findings.

Finding a meaningful comparison for motivation is equally problematic, but here the implicit comparison is not with the classroom, but the home. The 'utopian dream' of game-based learning research would be an educational game which could effectively compete for attention against blockbuster console games. However, it is unlikely that even the most commercially successful 'brain-training' franchises would perform very favourably in a direct comparison of this kind. In the home context researchers may have to consider an effective educational game as one which can compete for a slice of children's time amongst the complex picture of media consumption with in the home (Ofcom 2014). Classroom-based research allows the range of competing activities to be more carefully controlled, but it is not without its own complications.

3. Classroom research

The practical and ethical realities of undertaking research within a school setting has a significant influence on methodological design. Subject specialisation in secondary/high-school teaching makes research harder to accommodate within the school timetable, and the 7-11 age group is often preferred for the relative autonomy that school teachers have over the daily routine. Even so, any research programme carried out in schools will typically need the backing of both class teachers and the head of the institution in order to proceed. Our own experience running videogame related activities with schools (both research and outreach through the Games Britannia festival) is that videogames can still have strong negative associations amongst some staff, and consequently not all schools will engage with this kind of activity. It is particularly important therefore that game-based learning researchers are mindful of the responsibilities of using games for research.

3.1. Ethical considerations

Classroom research has significant advantages in terms of participation, but careful consideration must be given to how long it is ethical to remove children from their normal studies. Any study which incorporates a control group is potentially putting one group of children at an educational disadvantage to their colleagues. Consultation with teachers in this research programme, concluded that around twenty minutes per day for seven days was an appropriate amount of time away from lessons. This may seem like a relatively low figure, and examples of much more time-consuming educational research does exist (e.g. Kafai 1996), but the limit takes into account the additional time (and disruption) involved in taking three separate groups of children out of their normal lessons and getting them to and from the school's IT suite. To compensate for the lack of educational content in the control group's intervention, these children were optionally offered the opportunity to play the intrinsic version of the game during their lunch hour in the weeks following the end of studies.

Age-ratings present another ethical issue which is specific to researchers using videogames. Most territories now have their own legal rating systems which determine what is considered appropriate content for games classified under different age categories. The PEGI rating system is used across most of Europe and it has been illegal for retailers to sell games to underage children in the UK since 2012 (Sweeny 2012). The formal age-rating process is only applicable to published products and not closed research prototypes, but details of the requirements are freely available and should be followed by researchers. The Zombie Division prototype was designed according to the PEGI rating for a 7+ game with "depictions of non-realistic violence towards fantasy characters", being the maximum level of violent content experienced in the game.

3.2 Participants and group sizes

Twenty participants is often considered the minimum group size for statistical reliability in quasi-experimental studies of this kind. That means that designs involving three treatment groups require more than sixty participants to accommodate some level of dropout. A typical primary school in the UK has two classes per year group, and most conform to a target class size of less than thirty pupils, making it impossible to run a three group classroom study without involving two cohorts of pupils. In the first classroom study (design 1) two complete year groups of 7-8 and 8-9 year olds were enlisted totalling sixty-four pupils. Fifty-eight pupils took part in the second classroom study (design 4), comprising of an entire 7-8 year group topped up with students picked by their class teacher from the year below. The class teacher's selection was based on combination of age, maturity and mathematical ability, and in mind of an ethical consideration for selecting students who would not suffer from being removed from their normal teaching.

The two studies examining motivation were both conducted outside of normal school hours and used fewer than three treatment groups. The first of these took place during the school holidays (design 2) and recruited forty-four, 7-8 year olds from different schools into just two treatment groups (no control). The second was run as an after school club (design 3) with just a single group of sixteen pupils. Throughout all four studies parental permission was always sought to take part in the studies, and notably never refused.

3.3 Matched, randomised assignment

Randomly assigning participants to treatment groups helps to avoid unintended differences between groups influencing the outcome of the intervention (confounding factors). However, differences between groups can still arise through random chance when the number of participants in each group is small (as they typically are in educational studies). Matched designs can help to avoid this by sorting participants based on a matching variable (often pre-test scores) and then randomly assigning similarly scoring participants between the treatment groups. Groups can be balanced for binary attributes such as gender at the same time by simply performing the same process separately for males and females. This process will generally ensure that the mean scores for each group are comparable and any difference observed at the end of the intervention can be more reliably attributed to the independent variable rather than any difference between treatment groups.

Matching participants based on a single pre-test variable is relatively easy to perform by hand, but usually requires a separation in time between the pre-test and main intervention (to perform the grouping). Unfortunately this separation often means that carefully matched pre-test scores will be distorted by an inevitable level of absenteeism at the actual intervention. Although the matching process still helps to minimise any potential distortion, lost participants can alter mean pre-test scores by a disappointing margin over the course of a longer study. Such differences can end up disguising any potential outcome of the study as ANOVA-based statistical analyses use the group variance to determine whether an outcome is statistically significant.

4. Designing a game for research

This paper is primarily concerned with discussing the methodological approaches which were successfully and unsuccessfully applied to the evaluation of Zombie Division. Both the design of the game and the main outcomes of the studies are described in detail elsewhere (Habgood, Ainsworth 2011, Habgood 2007) and it is beyond the scope of this paper to revisit those aspects of the work. Nonetheless, a brief summary is provided here to facilitate the ongoing discussion of methodological design.

The biggest single influence on the methodological design of the Zombie Division studies was the limited amount of time which participants would have to play the game. Observing learning gains in just two hours of play is a significant challenge given the time required to become familiar with controls, avatars and objectives in any game. Nonetheless the concept was designed with this time constraint in mind and so a number of practical design decisions were made in order to try and facilitate its success as a research tool:

- The game and its control mechanisms were designed to mirror those of typical third-person action-adventure games familiar to children of the target age group (e.g. Spyro, Harry Potter, Zelda). Only the combat controls were unconventional as these required players to use the function keys to perform a corresponding division operation (e.g. F2 divided by two, F10 divided by ten).

- The educational content of the game was focussed on teaching a single mathematical concept (the inverse relationship between multiplication and division) and the game provided a mathematical representation which implicitly provided the correct division operations once this concept was understood. Thus a single mathematical 'epiphany' could potentially lead to a huge improvement in performance. The national curriculum at the time determined that the inverse relationship would be a relatively new concept for the year groups targeted by these studies.

In this way the intrinsic version of the game had a tight coupling between the gameplay and learning content where mathematical division was fundamental to the game's combat mechanic. The extrinsic version was exactly the same game but with the mathematical relationship removed from the combat mechanic. It then included identical mathematical content in the form of multiple choice questions at the end of each level (see figure 2). Both the intrinsic and extrinsic groups had access to the mathematical representation during the appropriate aspects of their respective games. The control version was the same as the extrinsic version without the multiple choice questions (i.e. no mathematical content at all).

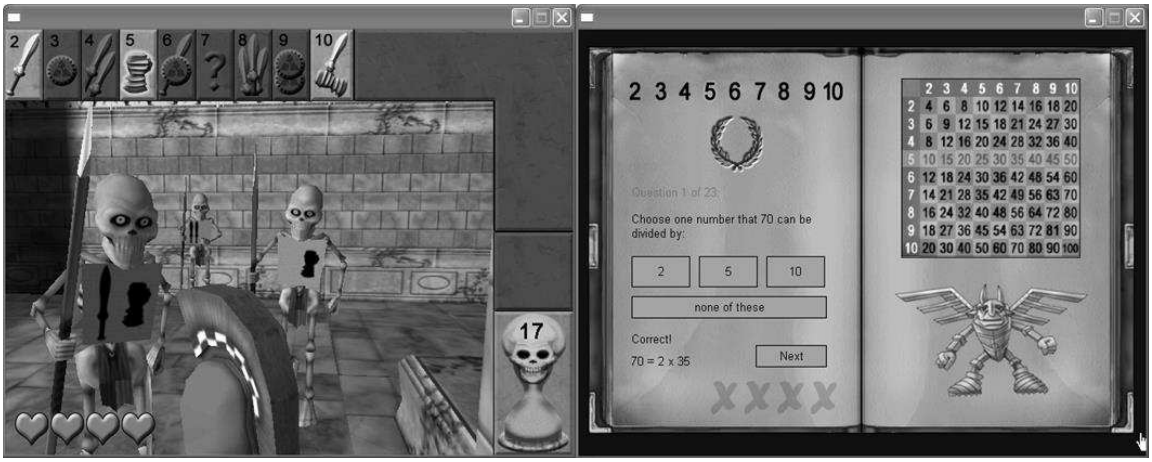

Figure 2: The extrinsic version of Zombie Division

5. Design 1: Learning gains

The design of the first study compared learning gains over five sessions, each lasting twenty minutes (once a day for a single week). Paper-based pre-tests were delivered on the Friday before the intervention week and used to perform a matched assignment of participants to each of the three conditions (intrinsic, extrinsic and control). For each session, children played the game together within their treatment groups inside the school's ICT suite. The children wore headphones and the game contained audio prompts which explained how to play the game without the need for additional direction by teachers or researchers. No classroom instruction was delivered relating to the game's content, and the teachers avoided covering the relevant mathematical concepts in class for the period of the intervention.

Statistical analysis of the results showed that learning had taken place, but there was no demonstrable difference in the learning between the three different conditions. The results looked as might be expected (improvements in the intrinsic and extrinsic conditions and almost no difference in the control), but large standard deviations meant that none of these differences were statistically significant. It was also noticeable that pre-test scores which were initially matched, had drifted upwards in the intrinsic and extrinsic groups as a result of absences, making it even more unlikely to detect any differences between mean group scores.

This first evaluation attempted a very clean approach to the (quasi) experimental design by trying to control for any influence that teachers or researchers might have on the intervention. While methodologically desirable, later designs would acknowledge the important role that teachers play in providing the link between game-based learning and children's wider understanding (Sandford, Williamson 2005). At this stage the game contained relatively little support for scaffolding children's understanding of the mathematical concepts involved, and for some children it was clear that applying the chosen mathematical representation (a number square) to division tasks was well outside of their "zone of proximal development". The children's class teachers recommended using a multiplication grid as a more familiar representation they might readily associate with multiplication and division and so this was adopted in subsequent studies (see figure 2).

Further investigation of the pre-test results showed that the testing instrument had introduced a significant flaw in the design of the study. Despite trialling the test in advance on children the same age, it was too easy for the children who took part in the actual study, as one quarter of them achieved a score of over 70% in the pre-test. This creates a “ceiling-effect” in which the higher scoring children have less scope to demonstrate improvement as a result of the intervention. This was supported by the statistical analysis which showed that the improvement between pre and post-test was only observable in the most difficult questions.

6. Design 2: Time-on-task vs. alternative

The design of the second study compared time-on-task over three sessions lasting forty-five minutes during a single day. A computer-based pre-test was delivered at the start of the study and used to perform a matched assignment of participants to the two conditions (intrinsic and extrinsic). For the first forty-five minutes of the study, participants were asked to play their assigned versions of Zombie Division with no alternative activity provided. Following a short break, they were given access to a “home screen” which allowed them to switch freely between playing their assigned version of Zombie Division and a range of non-educational games from the BBC website. This continued for the two remaining forty-five minute sessions, with a short break in-between.

Analysis of the results showed that there was no statistical difference between the amounts of time spent playing the two different versions of the game. However, boys across both groups did play Zombie Division for longer than girls (74% of their optional playing time compared to 51%).

Undertaking an entire intervention in a single day avoids the problem of absentees, but requires grouping to take place almost instantaneously. Even though this study was primarily concerned with motivation it is likely that mathematical ability could have a direct effect on children’s motivation to play a mathematical game, so matched grouping based on pre-test score was still desirable. The ceiling-effects observed in the first study had already prompted the implementation of a computer-based testing system which could deliver a (practically) endless number of questions in a fixed amount of time in order to prevent this problem re-occurring. For this study a client-server system was added in which each participant’s data was sent back to a server program to be matched. Once all the pre-tests were received, the groups were assigned and broadcast back to the client machines so that participants could immediately begin playing the correct version of the game.

While the system itself worked flawlessly during the study, it proved to be methodologically problematic as it introduced contamination between groups. Many participants sitting adjacent to each other found themselves assigned to different treatment groups, and it became clear that individuals in different conditions were directly competing to see who could progress the furthest. This competition was a particular issue for the way this study was examining motivation and the system was not used again. Nonetheless it could be effective in a situation in which it was possible to relocate the students between the pre-test and intervention.

7. Design 3: Time-on-task with free switching

The design of the third study compared time-on-task over three weekly sessions lasting forty-five minutes each as part of an after-school computer club. Throughout the study participants had access to a “home-screen” which allowed them to freely switch between the intrinsic and extrinsic versions of the game without suffering any overall loss of progress. The order of the selection buttons was randomised each time they returned to the home screen. Participants were initially shown the two versions of the game running side-by-side and asked to make sure that they tried playing both versions during the club. They were also free to choose to do their usual club activities instead (just about any activity available on the school PCs). In a fourth club session the pupils were interviewed about their experiences and asked to describe the differences between the two versions.

Statistical analysis of the results showed a huge preference for the intrinsic version of the game, which children played for over seven times longer than the extrinsic version.

The direct comparison between the two versions of the game demonstrated that children had a clear preference for the intrinsic version of the game, but the interview data showed a more complicated set of motivations behind this choice. Competition within the group was clearly a strong motivating factor and some children felt that it was quicker to progress through levels in the intrinsic version of the game (as the quiz took additional time). Nonetheless it was clear from some of the other comments that they possessed a surprisingly deep understanding of the design choices used in the two versions and one even astutely described the intrinsic version of the game as “like subliminal learning with maths” (Habgood, Ainsworth 2011, p.195).

8. Design 4: Learning gains with teacher-led reflection

The design of the final study is meticulously documented elsewhere (Habgood, Ainsworth 2011), but mirrored that of the first study with the addition of teacher-led reflection activities half-way through the intervention week. This consisted of a thirty five-minute session away from the game in which a classroom teacher delivered mathematical reflection activities on division. For each group the content was tailored to the context of their gaming experience, but contained identical learning content (including the numerical examples). For the control group the same learning content was delivered without reference to the game. The final difference was the inclusion of another short playing session two weeks later followed by a delayed post-test.

Statistical analysis of the results showed that children in all conditions demonstrated learning gains (as would be expected given the inclusion of the teacher-led reflection), but children who played the intrinsic game scored significantly higher than children in either of the other two groups.

The methodological design of the final study was a culmination of the hard-won experience of the entire programme of research. As well as including the teacher-led reflection, additional instructional scaffolding had been added to the game itself (mirrored in both the intrinsic and extrinsic versions). This was trialled in the earlier motivational studies to ensure that it was functioning correctly and appeared to benefit children's understanding of the mathematical concepts involved. The problems with ceiling effects observed in the first study were solved by the computer-based test and provided an additional set of process data which could be compared with that already produced by the game.

9. Process Data

Zombie Division contained two separate systems for generating detailed process data from the player's interactions with the game. The first created a time-stamped log file recording key game events such as entering a room, attacking a skeleton or collecting a key. These log files were mined in a post-hoc analysis for additional information about the player's learning and behaviour (Baker, Habgood et al. 2007). This process of data-mining is comparable to attempts to record and analyse 'game analytics' in commercial games (Sifa, Drachen et al. 2013). For example it was possible to use this data to observe a strategy in which children ran into a room containing skeletons, quickly ran out again, and waited before running back in again to attack. This suggested that they were deliberately creating time and space to work out the answers from a safe position. In fact this behaviour was replaced over time as pupils realised that pausing the game had the same effect. The increase in players' use of this pausing behaviour (and the difference with the extrinsic group) can be seen in figure 3.

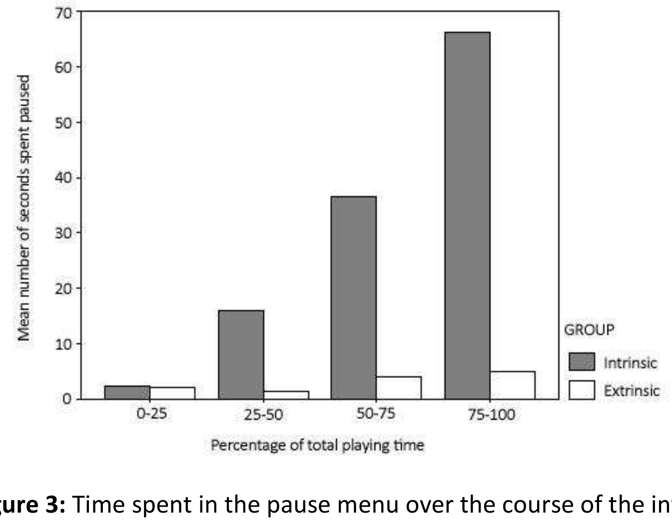

Figure 3: Time spent in the pause menu over the course of the intervention

9.1 Input streaming

The videogaming industry has long used the deterministic properties of games to record and replay in-game action using controller input streams. This relies on the principle that code should produce identical output each time it receives the same input. Even random number generation is deterministic and will produce the same output starting from the same "seed". This makes it possible to record a stream of input data (information about which keys are pressed for every frame of the game) and use it to replay the game in a video-like form at a later point in time. However, it takes just a tiny fraction of the storage required by a video and will compress very effectively using a simple run-length-encoded (RLE) algorithm. In the case of Zombie Division this means that the 130 hours of playing time from the final study takes up less than 8 MB of storage! The principal danger with this method of recording is that it is reliant on keeping the exact same version of the game used to record the data and the smallest change in the game itself results in wholly inaccurate replays.

10. Conclusion

Game-based learning is not as young a field as many might believe, but it has yet to produce a substantial body of empirical evidence which establishes the potential of learning games as effective educational tools. Researchers have approached this field from a diverse range of research interests and with a broad range of methodological approaches—all of which contribute to the depth and richness of understanding necessary to explain such complex artefacts. Nonetheless, the advancement of the field requires more empirical studies which are designed to collect evidence which directly supports or rejects the fundamental assumption that game-based learning can make an effective contribution to education.

References

ABDUL JABBAR, A.I. and FELICIA, P., 2015. Gameplay Engagement and Learning in Game-Based Learning: A Systematic Review. *Review of Educational Research,* **in press**.

ALL, A., CASTELLAR, E.N.P., VAN LOOY, J. and IMINDS, M., 2014. Defining best practices for assessing the effectiveness of digital game-based learning, *64rd International Communication Association (ICA) Annual Conference* 2014.

BAKER, R.S., HABGOOD, M.J., AINSWORTH, S.E. and CORBETT, A.T., 2007. Modeling the acquisition of fluent skill in educational action games. *User Modeling 2007.* Springer, pp. 17-26.

BLUNT, R., 2007. Does game-based learning work? Results from three recent studies, *Proceedings of the Interservice/Industry Training, Simulation, & Education Conference* 2007, pp. 945-955.

BOWMAN, R.F., 1982. A "Pac-Man" theory of motivation: Tactical implications for classroom instruction. *Educational Technology,* **22**(9), pp. 14-16.

BRUCKMAN, A., 1999. Can educational be fun? *Game Developers Conference '99*, March 1999.

CULLINGFORD, G., MAWDESLEY, M.J. and DAVIES, P., 1979. Some experiences with computer based games in civil engineering teaching. *Computers and Education,* **3**, pp. 159-164.

DECI, E.L., 1971. Effects of externally mediated rewards on intrinsic motivation. *Journal of Personality and Social Psychology,* **18**(1), pp. 105-115.

FULMER, S.M. and FRIJTERS, J.C., 2009. A Review of Self-Report and Alternative Approaches in the Measurement of Student Motivation. *Educational Psychology Review,* **21**(3), pp. 219-246.

HABGOOD, M.P.J., 2007. *The Effective Integration of Digital Games and Learning Content.* PhD edn. University of Nottingham.

HABGOOD, M.J. and AINSWORTH, S.E., 2011. Motivating children to learn effectively: Exploring the value of intrinsic integration in educational games. *The Journal of the Learning Sciences,* **20**(2), pp. 169-206.

HABGOOD, M., AINSWORTH, S. and BENFORD, S., 2005. Endogenous fantasy and learning in digital games. *Simulation & Gaming,* **36**(4), pp. 483-498.

KAFAI, Y.B., 2001. *The educational potential of electronic games: From games-to-teach to games-to-learn.*

KAFAI, Y.B., 1996. Learning design by making games: Children's development of strategies in the creation of a complex computational artifact. In: Y.B. KAFAI and M. RESNICK, eds, *Constructionism in practice: Designing, thinking and learning in a digital world.* Mahwah, NJ: Lawrence Erlbaum Associates, pp. 71-96; 4.

MALONE, T.W., 1981. Toward a theory of intrinsically motivating instruction. *Cognitive Science,* **5**(4), pp. 333-369.

OFCOM, 2014. *Children and Parents: Media Use and Attitudes Report.* Independent regulator and competition authority for the UK communications industries.

O'NEIL, H.F., WAINESS, R. and BAKER, E.L., 2005. Classification of learning outcomes: Evidence from the computer games literature. *The Curriculum Journal,* **16**(4), pp. 455-474.

PAPERT, S., 1998. Does easy do it? Children, games and learning. *Game Developer,* **June 1998**, pp. 87-88.

RANDEL, J.M., MORRIS, B.A., WETZEL, C.D. and WHITEHILL, B.V., 1992. The effectiveness of games for educational purposes: A review of recent research. *Simulation & gaming,* **23**(3), pp. 261-276.

SANDFORD, R. and WILLIAMSON, B., 2005. *Handbook on games and learning.* Bristol: Futurelab.

SIFA, R., DRACHEN, A., BAUCKHAGE, C., THURAU, C. and CANOSSA, A., 2013. Behavior evolution in tomb raider underworld, *Computational Intelligence in Games (CIG), 2013 IEEE Conference on* 2013, IEEE, pp. 1-8.

SWEENY, M., 2012. PEGI video game ratings become law. *The Guardian Online*.

VOLLMEYER, R. and RHEINBERG, F., 2000. Does motivation affect performance via persistence? *Learning and Instruction,* **10**(4), pp. 293-309.