Making games in the classroom: Benefits and gender concerns

Judy Robertson

Computer Science, Heriot-Watt University, Edinburgh EH14 4AS, United Kingdom

A R T I C L E   I N F O

Article history:
Received 12 September 2011
Received in revised form
13 December 2011
Accepted 18 December 2011

Keywords:
Game-based learning
Gender
Classroom
New media literacy

A B S T R A C T

This paper argues that making computer games as part of a classroom project can develop a range of new media storytelling, visual design and audience awareness skills. This claim is supported by data from the evaluation of a six week game making project in a state funded primary school in which 11–12 year old learners made their own computer games using software called Adventure Author. The paper reports on analysis of the games produced by the children and documents the range of new media storytelling skills used as well as examining how the pupils responded to peer reviews of their games. In light of concerns raised in the literature that girls may be disadvantaged by classroom games projects, it investigates whether there are gender differences in the game making skills displayed by the learners. The results of the study indicate that girls’ games score more highly than boys’, particularly on skills relating to storytelling.

© 2011 Elsevier Ltd. All rights reserved.

1. Introduction

Computer games are an integral part of many children’s lives, regularly attracting and sustaining their attention for long periods of time (Lenhart & Kahne, 2008). Educators are aware of the potential benefits of harnessing the motivational power of games for educational purposes; teachers are now putting in practice the recommendations of theorists such as Gee (2003) and are using games as part of their everyday classroom teaching (Futurelab, 2009; Groff, Howells, & Cranmer, 2010). Placing learners in the role of producers rather than merely consumers of technology, researchers have explored the benefits of enabling children to create their own games in the classroom (Carbonaro, Szafron, Cutumisu, & Schaeffer, 2010; Kafai, 1995; Meerbaum-salant & Ben-ari, 2010; Owston, Wideman, Ronda, & Brown, 2009; Robertson & Howells, 2008; Vos, van der Meijden, & Denessen, 2011). Creating games can be considered to be an important aspect of new media literacy: “games are a significant part of children’s cultural capital, and a potential expressive form which they can develop in the classroom...that should be more extensively valued and developed within education” (Buckingham & Burn, 2007, p.346). Buckingham and Burn (2007) write that the term “game literacy” implies that games make meaning through a kind of language, a language which can be gradually acquired and explicitly taught. If this is the case, what are the storytelling techniques for this language? What techniques can learners use to express themselves through the medium of game design? This paper addresses this question through the analysis of 3D role-playing games created by a class of 11–12 year old learners during a six week game making project in terms of new media storytelling techniques, visual design and audience awareness. In addition to analysis of the game products, the paper also examines an aspect of the game making process which is particularly relevant to new media literacy: peer evaluation. An aspect of game literacy is the critical reading of games (Buckingham & Burn, 2007). Do young learners demonstrate skills in this area when reviewing and offering comment on games made by their peers? Indeed, do young game designers capitalize on the advice offered to them by their peers or is the advice ignored?

Concerns have been raised about the implications of incorporating an activity which is apparently more attractive to boys into the classroom mainstream (Carr & Pelletier, 2008; Kelleher, 2008; Steiner, Kickmeier-rust, & Albert, 2009): will girls be at a disadvantage during game literacy lessons? If girls spend less time playing games at home, will their relative lack of game experience hold them back in class? Will girls be alienated by the introduction of a ‘boys’ toy’ into the classroom? If this was the case, then one might expect to find that girls would be less proficient at making games than boys. On the other hand, given that girls tend to outperform boys in traditional narrative tasks, such as writing stories (Beard & Burrell, 2010; Daly, 2003), it could be that girls’ lack of experience in gaming is mitigated by their...
competence in similar creative activities. This paper investigates whether this is indeed the case by comparing the storytelling techniques demonstrated in the learners’ games by gender.

The paper begins with a literature review of research into the opportunities for the development of new storytelling techniques offered by the task of designing computer games, as well as an assessment of the evidence that girls may be disadvantaged by game making activities in the classroom. A description of the Adventure Author software which the children used to make their games is given, followed by the results of a school based field study in which the storytelling techniques used by the children are analysed.

2. Background

2.1. Learning opportunities from game making

From a constructionist perspective, there are theoretical reasons for believing that making games can be educationally beneficial: Kafai (1995, 2006) has argued that when making games, learners also construct knowledge and their relationship to it (Kafai, 1995, 2006). She writes: “The learner is involved in all the design decisions and begins to develop technological fluency. Just as fluency in language means much more than knowing facts about the language, technological fluency involves not only knowing how to use new technological tools but also knowing how to make things of significance with those tools and most important, develop new ways of thinking based on use of those tools” (Kafai, 1995, p. 39). As technology has moved on to include 3D graphics since Kafai’s original work (for a review of current game making software for young learners see Good (2011)), opportunities have also arisen to develop new media literacy through game making.

In the last twenty years, notions of literacy have developed beyond the written word (Buckingham & Burn, 2007), taking a “turn to the visual” (Jewitt, 2008, p.9). Theorists have highlighted the differences between traditional alphabetic text and multimodal texts – which can contain arrangements of potentially non-linear visual, audio, and moving image elements as well as the written word – and discussed the additional challenges for both reader and writer when working in these more complex forms (Kress, 2003, p. xiii, 186 p.). A major challenge in literacy education is to encourage learner writers to anticipate the needs of their reader (Wyse, 1998); these audience awareness skills are equally important in non-linear multimodal texts whereby the player must often navigate their own path through the choices that they make. Compared with other multimodal texts, computer games offer added complexity for both player and designer, including the challenge that the player (anticipated by the designer) can move around inside the world of the text and experience it from more than one visual, spatial and textual perspective.

In a series of case study illustrations, Howells and Robertson (Howells & Robertson, in press) outline the literacy opportunities offered by new media authoring environments. The paper illustrates new media storytelling techniques such as conveying plot through interactive dialogue, creating compelling visual design for locations in which significant story events take place, using “narrative vehicles” (a technique in which items which the player can carry or 3D props in the environments are tagged with plot related information) and designing reading pathways for the player to follow. The most effective games in the case study were the ones which integrated the visual design of the game with the storyline, and where the player’s spatial path through the game was anticipated in advance and supported by the designer both visually in the area design, and textually through conversations and narrative vehicles. The current paper takes this analysis further by looking at the techniques used by a larger set of learners and by rating them on a numerical scale for comparison between sub-skills and between genders.

The process used by learners to make their games is of importance because it can assist in understanding variations in the learners’ game products, and can be used to advise teachers who are considering incorporating game making into their practice about suitable approaches. Game making is a complex design task (Robertson & Nicholson, 2007) which requires a range of creative skills including problem finding, problem solving, evaluation and communication. Resnick (2007) offers a spiral model for creative thinking in children’s technology based projects such as game making or computational craft-work. He proposes an iterative process containing the stages of imagine, create, play, share and reflect. The children imagine a new idea of what they would like to make and then use the technology to create it. They play with their creation, and invite their friends to play with it too. By reflecting on their experiences of working on this project, they are in a position to imagine further projects, and so the cycle continues. He writes that “As students go through this process, over and over, they learn to develop their own ideas, try them out, test the boundaries, experiment with alternatives, get input from others, and generate new ideas based on their experiences.” (Resnick, 2007, p.18). Buckingham and Burn (2007) note that an important aspect of games literacy is the ability to critically review games. In the context of classroom game making, this involves peer and self assessment skills; the ability to constructively evaluate creative products made by oneself and one’s classmates. Self and peer assessment are part of a set of “learning how to learn” competencies (Black, McCormick, James, & Pedder, 2006) and have been developed widely in the UK through the Assessment is for Learning scheme. Assessing games requires a mixture of critical media analysis and technical software testing strategies (such as might be taught to novice computer science students). The learner must decide whether the story is compelling, whether the visual design of the game is attractive, whether the player’s abilities are a good match for the level of challenge, whether the player is given sufficient autonomy as well as clear goals, and also check that the behaviour of the game is as intended under all possible conditions.

2.2. Gender issues in classroom game making

Steiner et al. (2009) argue that a “gender sensitive approach [to educational gaming] is strongly needed, as the new learning technologies need to be appropriate and accepted for all students and ensure equal (learning) opportunities” (p. 5). Jenkins and Cassells (2008) warn: “It would be a big mistake to introduce games into the classroom if the results had the effect of further disadvantaging girls.” (p. 14). Kelleher (2008) believes that “using computer games as a primary end-goal to motivate students towards computer science is potentially dangerous” (p. 28) on the grounds that games do not appeal to boys and girls equally. Certainly it would be problematic if one gender were to be
disadvantaged by the enthusiastic adoption of a new technology. But how great a problem is this likely to be? In the current games market, where women gamers over the age of 25 are reputed to make up the largest block of PC gamers\(^2\) and 70% of children under eleven years old play video games with adult family members (Ulicsak & Cranmer, 2010), the longstanding stereotype of the computer gamer as a solitary teenage boy is fading. What evidence is there to support the concern that girls may be disadvantaged by games in the classroom? Further, what is known about gender differences in game making activities? The evidence from prior research is mixed, as indicated in the following literature review.

2.2.1. Gender differences in game playing habits

Firstly, consider the facts about game playing habits by gender. It is the case that boys spend more time playing games than girls. In a US study, Lenhart and Kahne (2008) found that almost all children (97%) play video games but that boys play more frequently and for longer durations. They found that 39% of boys play games daily, in comparison to 22% of girls, and that 34% of boys play for 2 h or more per day in comparison to 18% of girls. Heeter and Winn (2008) point out that this gender gap magnifies with age so that by the time a young male enrolls in college, he will have accumulated thousands more hours of game-play. With respect to game “modding” (using software tools to modify existing games) – an activity similar to creating games – Lenhart and Kahne found that 36% of boys had modded a game whereas only 20% of girl gamers had done so. Similarly, an earlier study by Hayes (2008) found that females were far less likely to take part in in-game content creation activities such as making new levels or characters. Possible factors which influence this, as cited by Hayes, include the facts that many games which are marketed at girls do not include the functionality which would make it possible to alter the game content, that girls tend not to take part in on-line game communities where they could learn the necessary skills, and that boys are more likely to value game modding skills in their peers.

2.2.2. Gender differences in game preferences

Recent findings relating to young learners’ attitudes to video games in the classroom indicate that the differences in expertise in computer games may indeed influence boys’ and girls’ preferences for using them for learning. In a recent study of students’ perceptions of using video games in the classroom with 858 secondary school pupils, Bourgonjon, Valcke, Soetaert, and Schellens (2010) found that male students favour computer games in the classroom more than girls, although both believe that games offer learning opportunities. Path analysis revealed that the gender differences were mediated by inequalities of gaming experience and therefore perceived ease of use. The authors argue that because there is such a diverse range of student experience with games, educators should not assume that learners will all be familiar with games, and that dedicated game literacy lessons would be of benefit. In fact, they suggest designing video games would be an appropriate activity to increase novice students’ confidence.

It also appears to be the case that males and females have different motivations for playing games and different preferences for content. In a study of women’s dislikes of video games, women considered themselves to be less competitive than men, and less confident about their ability to master competitive games. In addition, they disliked violence and gender stereotyped characters but enjoyed games with rich social interactions (Hartmann & Klimmt, 2006). Olson (2010) found that boys were significantly more likely to play games for fun, competition, challenge, enjoyment of in-game weapons and emotion-related reasons such as coping with anger or relaxing. Interestingly, boys and girls seemed equally motivated to “create their own world” within the game (p. 180), which again suggests that game making in class would not be off-putting to girls.

2.2.3. Gender, technology and narrative

Creating one’s own world in a game engine is a task which combines an interest in narrative with technology engagement. The combination of narrative and technology has been proposed both as a solution to poor attainment in writing by boys, and lack of interest in technology by girls. While researchers who study girl gamers are concerned that girls may be disadvantaged by the introduction of games into classrooms, literacy researchers have catalogued evidence that boys are disadvantaged by their teachers’ expectations in the narrative genre which may be at odds with their own reading preferences. Daly also discusses the ways in which boys’ writing attainment can be improved including by the use of ICT related activities such as multimedia authoring and the use of visual media.

On the other hand, girls’ interest and aptitude at narrative creation has caused some researchers to use it as a tool to motivate girls to become interested in technology. For example, Kelleher’s Storytelling Alice was designed to teach middle school girls how to create animated stories by learning computer programming. Girls using the version of the Alice software designed specifically for storytelling spent more time on task, were more inclined to share their programs with friends, and performed equally well in terms of programming skills as those using the generic version of Alice (Kelleher, 2008).

2.2.4. Studies of game making in the classroom

Evidence from two studies of the effect of game making activities on literacy skills suggests a lack of gender effect for interest or learning gains. Vos et al. (2011) conducted a study in which 235 elementary school students (10–12 years old) either played or constructed their own drag and drop memory games in the Dutch language. There was a significant difference in self reports of intrinsic motivation and deep learning strategy use; children who made their own games enjoyed it more and reported using deeper learning strategies than those who played an existing language game. There were no significant gender differences for perceived competence, interest or use of deep learning strategies. In a study of the effects of game construction on 9–10 year old children’s literacy, Owston et al., (2009) children were asked to create educational quiz questions which would be incorporated into an electronic version of a board game (Owston et al., 2009). The researchers report a significantly better performance on a logical sentence construction item on a battery of standardised tests after constructing the games. Interview data from teachers also suggested that the game development task helped the students with content

retention as well as a variety of research skills required to prepare the questions for their game. There were no gender differences reported on literacy scores or by teachers.

2.2.5. Studies in the role-play genre: Neverwinter Nights

Game genre is likely to make a difference to children’s preferences. Steiner et al. (2009) note that adventure and role-play games, which have been identified as appropriate for learning activities are appealing for both boys and girls. The game making activity reported in this paper was in the role-play genre using a toolkit based on the commercially available Neverwinter Nights 2 game. In considering the likely impact this environment might have on girls, discussion of the Neverwinter Nights (NWN) game world itself, and previous game making studies which utilised the same software are relevant.

NWN is a 3D fantasy role-playing game franchise based on the dungeon and dragons rule-set. Corneliusse and Mortense (2005) examine the issue of gender in this game: “The construction of gender in NWN is built on strongly stereotypical, dualistic and heteronormative understandings of gender, both in the player characters, the story and in the game environments. However, NWN does in some ways fulfill its promises to the female audience. Female characters in the game take on prominent positions, they display non-traditional positions confronted with men, and the game environment establishes a ‘natural’ battlefield for women” (p. 3). Although it was the NWN2 toolset which was used in the study reported here rather than the NWN game itself, the toolset is built on the same premise that gender “doesn’t matter” and that males and females can take on equal roles. It is then up to the user of the toolset to decide how to portray gender in the games that he or she creates. Jenkins and Cassell (2008) report the characteristics that are likely to appeal to girls in games including lead characters who are easy to relate to; an emphasis on exploration rather than hierarchical scoring; a focus on stories and character rather than fast paced action; real world settings and success through social relationships rather than combat. According to Corneliusse and Mortise’s analysis, the Neverwinter Nights world is likely to be appealing to girls on the grounds that it is strongly narrative based, with strong supporting characters and a large setting to explore. Elements which are likely to appeal less to girls are the fantasy setting and the combat aspects. However, when using the toolset, girls have the option to focus on the story aspects in their own games instead of combat if they choose. The fantasy setting is not possible to change.

Robertson and Howells (2008) consider the successful learner skills which 10 year old children developed by taking part in a project in which they constructed their own role-playing games using the NWN toolset: enthusiasm and motivation for learning, determination to reach high standards of achievement, independent and group learning, and linking and applying learning in new situations. There was no indication that girls were disadvantaged during this project.

Using Neverwinter Nights and a custom made programming toolkit called ScriptEase, Carbonaro et al. (2010) explored the computer science concepts which high school students learned by programming their own role-play games, concluding that the game authoring activity stimulated higher order thinking skills and that the games the learners made illustrated an understanding of the targeted computer science concepts. On the grounds that girls significantly outperformed boys on higher order thinking skills, that there were no significant gender differences for computer science skills, and that both groups were motivated by the activity, the authors claimed that their approach is “a viable gender-neutral approach to teaching Computer Science in particular and Science in general that may increase female participation in the discipline” (p. 1).

2.2.6. Summary

In summary, there are gender differences in the frequency and duration of boys’ and girls’ game-play sessions. Boys’ greater experience of gaming makes them more likely to look favourably on the use of games in the classroom, although both genders believe that games can bring educational benefits. Game making with an emphasis on storytelling has been used to interest girls in computer science, where visual and IT related tasks are recommended to address gender gaps in writing for boys which particularly manifest in the narrative genre. Studies of game making activities in the classroom have found no gender differences in learners’ interest or learning gains. Indeed, a study using the NWN toolset (similar to the software used in this study) suggests that girls may be better at some aspects of the game design task.

3. The adventure author software

Adventure Author is game authoring software aimed at 11–14 year old users. It was developed using a learner centred design process in consultation with pupils and teachers using the methodology outlined and has been iteratively evaluated in several classroom and informal education settings throughout the process. Adventure Author is a freely available plug-in for the commercially developed Neverwinter Nights 2 software which enables end users to make their own games. The components of Adventure Author are as follows: Fridge Magnets, which is designed to assist in recording creative ideas about what should happen in the game; Conversation Writer, which enables children to write interactive dialogues with conditions and consequences; and Comments Cards which supports structured peer review of the game once it is ready to be tested. The aspects of the tool which are of relevance to the analysis (the toolset and game engine, the Conversation Writer and Comments Cards) are described in more detail below.

The designer creates the game using the Neverwinter Nights 2 (NWN2) toolset (see Fig. 1) and tests it using the NWN2 game engine (see Fig. 2).

The toolset enables the designer to edit the terrain of their game area by sculpting 3D hills, valleys, lakes and rivers. The current state of the game area is shown in the centre of the screen in a 3D preview (see Fig. 1). Another major feature of the toolset is blueprint editing. Characters, buildings, furniture, trees, sounds and items are instances of blueprints, and can be chosen from extensive libraries on the right hand side of the screen. Characters act autonomously once the game is running and their behaviour is governed by automatic rules from the dungeons and dragons role-play rule-set implemented within NWN2. Their appearance is configurable by the designer and they can be made to interact with the player using the Conversation Writer.

The game engine is used to test a game which has been designed using the toolset. The designer can view and interact with the game world as a player would by exploring the 3D landscape, reading dialogue and making conversation choices, picking up and examining objects and fighting non-player characters.
Screenshots of Conversation Writer are shown in Fig. 3 and Fig. 4. The user can write dialogue between multiple characters in a branching format, shown in the format of a play script. Action sequences, such as triggers for attacks or exchanges of property can be initiated from conversation lines. Dialogue lines can also be specified to be presented only under certain conditions, such as “if the player is carrying the magic sword”. The user can also see an overview of the conversation structure in the form of a graphical dialogue tree (shown in Fig. 4).

Comments Cards is the peer and self evaluation tool in Adventure Author. It enables teachers to specify a set of criteria by which learners will evaluate each other’s games. A learner can review the game designed by a classmate by rating the game on each of the teacher proposed criteria. He can indicate overall quality by selecting stars on the interface which correspond to numerical ratings, and by justifying this quantitative rating with a textual explanation. The interface enables the learner whose game has been reviewed to type in a response to the review comments, and the teacher is also able to comment on the quality of reviewing. An excerpt of a partially completed evaluation is shown in Fig. 5. In this example, a child has reviewed the plot of the game which was created by one of his peers. He has rated different aspects of it with stars, and has written comments to justify his opinion. The game designer has responded by agreeing to take his reviewer’s advice, and the teacher has commented on the constructive criticism skills of the reviewer.

4. A school based field study

4.1. Participants and classroom environment

A field study was conducted to evaluate the Adventure Author software in a naturalistic setting and investigate the range of game making skills developed by the learners during the course of the project. The field study was conducted with a primary seven class in a state funded primary school in East Lothian, Scotland. The class teacher wished to try the Adventure Author software to support a cross-curricular project in fantasy literature themed around the novel The Hobbit by J.R.R. Tolkien. She chose to spend most of the class time for a six-week period on this project, although there were also regular timetabled slots for maths and physical education. The class consisted of 25 pupils (11–12 year old), 11 girls and 14 boys. A suite of ten Windows Vista Dell gaming laptops was lent to the school for the duration of the project.

4.2. Procedure

The pupils used the laptops individually, usually at the same time as their peers worked on other project related tasks with their teacher, including vocabulary exercises, discussion, brainstorming, story writing, drawing, craft-work, and individual reading of fantasy novels. For

![Fig. 1. A screenshot of the NWN2 toolset.](image-url)
example, in one vocabulary lesson, the children were asked to generate place and character names for the fantasy genre. They were asked to write a story containing these characters and places. They also made papier-mâché weapons and jewels to correspond to the props available in the toolset. As is appropriate in a naturalistic field study, the activities were planned and carried out using the teachers’ professional judgement rather than being prescribed by the researchers. Only the time spent using the laptop was recorded as this could be done automatically without placing an additional burden on the teacher. Adventure Author was used on eighteen days during the project. The users spent on average 5.8 h using the software (SD = 58 min due to absences and inevitable classroom scheduling variations).

Learners mostly had free choice of what game making activity to engage with when they used the software, particularly towards the end of the project. At the beginning of the project, the researchers and teacher ran one structured session on ideas generation using the Fridge Magnets and gave demos of how to use the Conversation Writer on the topics of creating branching conversations, creating conditions and specifying actions at the start of two sessions. The researchers also ran one structured session for peer review of the games using Comments Cards, which came two sessions before the end of the project. At the start of the peer review session, two of the researchers modelled the sort of interaction which should happen between reviewer and designer: the reviewer should play the game without intervention from the designer. Designers could answer specific questions from reviewers, but their focus was on looking for unanticipated behaviour on the part of the user while playing the game. In such circumstances, the reviewer will choose to take a different route through the area, or make a different choice in the conversation than the designer was expecting. Designers were asked to make notes on problems which occur and start thinking about how they might solve them during the testing. After testing was finished.
(perhaps after 10 min of play), the reviewer completed the Comments Cards form and the designer responded. The designer and reviewer
then swapped roles.

4.3. Research questions

Qualitative and quantitative data was gathered during the project in the form of observation notes, log files, evaluation forms and copies
of the children’s games. Table 1 summarises the research questions which the analysis attempts to answer. The first question considers the
literacy skills which may be demonstrated in the learners’ games in terms of new media storytelling techniques (storyline, characters,
dialogue and imagination), visual design and audience awareness (spatial player guidance, player purpose, player choice and player
engagement). The second question relates to whether there will be gender differences in the techniques used, starting from the hypothesis
that girls will perform better. This hypothesis is based on recent evidence from a study in the same role-play genre (Carbonaro et al., 2010),
which found that games made by girls scored more highly on higher order thinking skills than boys’ games. Although other previous studies
found no gender effects as discussed in the literature review, the similarity of the software used in the Carbonaro et al.’s study (2010) makes
it the most useful comparison. These questions are answered using a game rating scheme devised by Howland (2011) which is described in
Section 4.1.1. Research question 3 is concerned with gender differences in the peer review process which may shed some light on any gender
differences which are found in the game products. This question is answered through analysis of the learners’ usage of the Comments Cards
software and time stamped data from the software log files. In particular, differences in the learners’ behaviour with respect to editing the
Table 1
Summary of research questions.

<table>
<thead>
<tr>
<th>Research questions</th>
<th>Data gathered</th>
<th>Analysis method</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 What are the characteristics of the games produced by learners in terms of storyline, visual design, spatial player guidance, player purpose, player choice, characters, dialogue, imagination and player engagement?</td>
<td>Game files</td>
<td>Rating of games on criteria specified by Howland (2011)</td>
</tr>
<tr>
<td>2 Are there gender differences in the games produced by learners?</td>
<td>Game files</td>
<td>Inferential statistics on ratings measured by scale Howland (2011)</td>
</tr>
<tr>
<td>3 Are there gender differences in the game making process?</td>
<td>Comments Cards files and log files</td>
<td>Qualitative analysis and inferential statistics on log file time data</td>
</tr>
</tbody>
</table>

Game following a peer review session are considered in the interests of deepening our understanding of young people’s skills in the critical analysis of games and the extent to which they can benefit from the suggestions of their peers.

4.4. Game analysis scheme

The games were analysed using an assessment scheme for games created by learners using the Neverwinter Nights 2 environment (Howland, 2011). This scheme was developed with input from an experienced teacher and literacy specialist and has been piloted on games produced by 112 learners in a previous unpublished study. The scheme requires the rater to score the games (on a scale between 0 and 10) on the dimensions of storyline; visual appearance of the areas; how well the designer guides the player around the area; the extent to which the player has a purpose in the game; the extent to which the player can make meaningful choices in the game; interesting characters and dialogue, imagination and challenge.

For each dimension, the scheme also asks the rater to indicate particular techniques which could potentially be used. These are described briefly here as useful indicators of the sorts of skills involved, but they were not used quantitatively in the analysis because a pilot test showed very low inter-rater reliability. For storytelling, these techniques include conveying the story in a visual way by using textures or terrain, the placement of objects such as trees, buildings or furniture and by custom character appearance. Story can also be conveyed textually in character, area or item names, text associated with objects and through character conversations. It is also possible to tell the story through the way the characters behave, such as by the animations they use or the actions they perform.

Techniques for making an area visually interesting include: raising or lowering ground to create a landscaped terrain; using textures to create paths or other features on the ground; and including trees, buildings, water and special effects. Other visual design skills include adding creatures to make a scene seem more lively and creating identifiable locations in which particular game episodes take place (story locations in the terminology of Howells & Robertson, in press).

The designer can use various techniques to guide the player around the landscape of the game including: writing directions in conversations; creating paths and signs; writing instructions in objects such as books (narrative vehicles in Howells & Robertson’s terminology, in press); and making landscape features for the player to navigate by (such as following the shore of a river).

The player can be informed of goals and objectives via conversations (either explicitly stated or implied), by the way a character behaves or through text associated with an object.

Possible techniques a designer can use to give the player choice in the game include enabling them to take different routes through the area and giving them choices in a conversation, particularly choices which have noticeable changes in game outcomes.

The scheme indicates potential conversational features including the use of humour, dialect, descriptive language, conversation which advances the plot, and the less positive feature of conversation which is merely utilitarian in that it is simply a trigger for a sequence of scripted behaviour to be invoked. For the dimensions of characters and imagination, the rater is asked to provide examples of interesting characters or imaginative aspects of the game under review. Finally, for the challenge dimension, the scheme considers whether the player has a challenge or quest to complete, and whether the battles are pitched at an appropriate level of difficulty for the player.

4.4.1. Game rating procedure

The rater first viewed the games in the toolset, inspected the characters, and then read the conversations and text on items. After rating all the games on dimensions 1–8 (storytelling to imagination) using the toolset, she played the games in order to rate them for challenge level. For consistency, when playing the games, a particular player character designed to be suitable for testing games was used. This player character was used by the children when they tested their own games during the project. At the start of each game, the player character was equipped with any weapons and armour which were supplied by the designer. The rater attempted to cooperate with the designer’s intentions, for example, by talking to characters rather than killing them and by collecting objects which might be useful. If the rater entered a fight and was killed, she would respawn the character once and then run away if possible in order to explore the rest of the level.

4.4.2. Inter-rater reliability

All games were rated by two experts (the author and a colleague with research experience in the area), and the average of these scores was used in all subsequent analysis. This is similar to the consensual agreement method used in the analysis of creative products (Baer & McKool, 2009). The researchers each independently rated a sample of six games before meeting for three hours to calibrate their scores. Minor changes were made to Howland’s scheme (2011) as a result of this. The raters then independently rated the remaining 19 games and inter-rater reliability statistics were computed for these games (not including the original six in the training set). Analysis of inter-rater agreement revealed that the challenge category had exceedingly low agreement, and so after further discussion this category was
revised to “Player Engagement” and the labels at the end points of the scale were disambiguated. The raters then revisited all the games to assess them for player engagement. Inter-rater reliability for player engagement was computed for all the twenty-five games because no particular games were used as a training set for the revised version of this category. The inter-rater reliability (Cronbach’s alpha) is shown in Table 2; these values are considered “respectable” (DeVellis, 2012, p. 203), although it is worth noting that the category of visual design is lower than the others.

5. Results

5.1. Research question 1: characteristics of children’s games

Table 3 shows the means and standard deviations of scores for each dimension of the rating scale. The mean total percentage score for the games was 42% with a standard deviation of 20. The lowest score was 5% for a very short game with very little conversation or indication of the player’s purpose, and the highest was 69% for an imaginative detective story which linked the visual design of the area with the story.

The mean scores for each dimension are fairly similar, slightly below the middle of the scale, suggesting that the children mastered a range of basic techniques, but required practice to use them in more sophisticated ways. Visual design tended to be rated most highly, whereas player guidance scored least well (but with a wider variation). No game scored the maximum score of 10 in any dimension, but several received the minimum score of 0 for dimension in cases where the designer had not included conversations, or the game which they created was not playable.

Story was most commonly conveyed through text in the conversations with characters. It was also apparent from character names, objects in the area (for example embedded in the text of a magic book) and the terrain or textures in the level (for example an island in a land which has been flooded by an evil witch).

All of the games were decorated with trees and vegetation, most contained buildings or other structures, and a large proportion showed mastery of terrain and texturing skills and the inclusion of water features. Some of the learners made identifiable locations within the area in which a story event would take place, for example a clearing in the forest where the player could find the blacksmith’s workshop.

The games were less likely to incorporate techniques for guiding the player in the level: around half of the games contained instructions in conversation, paths or clues in the landscape. Textual guidance in the form of sign posts or instructions attached to items were relatively rare. The designers’ games illustrate that they did have the technical skills to create conversations or create story locations; the fact that these were not more often used to guide the player suggests that the designer lacked audience awareness skills. The designers perhaps did not anticipate that a player who was new to the game would find it hard to know where to go next.

In terms of techniques for notifying the player of his objectives, the designers most frequently directly stated explanations in written dialogue. Subtle hints in conversations and text in objects were relatively rare. Some games had no indication at all of player objectives which made them difficult to understand.

Player choice was mostly apparent through conversational choices, and a large proportion of these had notable consequences in terms of action sequences which ensued. A common pattern was to offer the player a choice of a polite or rude dialogue option, with politeness rewarded with a gift or a new ally, and rudeness penalized with an attack. A relatively low proportion of games had a coherent choice of routes mapped through the area, which suggests that the designers were not anticipating the spatial route planning of the player. The best games contained dialogue with humour or characterization or attempts at dialect and poetic language.

In terms of player engagement, games with low scores did not offer an opportunity for the player to do anything in the game such as exploration, combat or undertake a quest. This was sometimes because the area was very sparsely populated, and sometimes because the designer had misjudged the difficulty of the level and the player got killed almost immediately after starting the game. Higher scoring games gave the player interesting quests, attractive areas to explore and balanced combat in which the player was neither instantly defeated nor invincible.

To summarise: the designers showed strengths in portraying a storyline in the game, particularly through written dialogues but also through visual features of the area design in special locations where story events could happen. The games could be improved by generally increasing a set of skills relating to player (audience) awareness such as guiding the player spatially round the environment, giving objectives in other ways than through conversations, giving the player a choice of spatial routes through the area and creating battles at a level of difficulty appropriate to the player. The use of techniques for making dialogue more compelling (such as humour, poetic language and dialect) would also benefit the games.

\[\text{Cronbach's alpha}\]

<table>
<thead>
<tr>
<th>Category</th>
<th>Cronbach's alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td>Story</td>
<td>.931</td>
</tr>
<tr>
<td>Visual design</td>
<td>.740</td>
</tr>
<tr>
<td>Player Guidance</td>
<td>.804</td>
</tr>
<tr>
<td>Purpose</td>
<td>.907</td>
</tr>
<tr>
<td>Choice</td>
<td>.925</td>
</tr>
<tr>
<td>Characters</td>
<td>.924</td>
</tr>
<tr>
<td>Dialogue</td>
<td>.931</td>
</tr>
<tr>
<td>Imagination</td>
<td>.788</td>
</tr>
<tr>
<td>Player engagement</td>
<td>.812</td>
</tr>
<tr>
<td>Total</td>
<td>.939</td>
</tr>
</tbody>
</table>

\[\text{Table 2: Inter-rater reliability statistics.}\]

---

\[\text{Footnote: The revised version of the scale is available at REMOVED FOR REVIEW.}\]
5.2. Research question 2: gender differences in children’s games

Descriptive statistics for each dimension of the games made by boys and girls are shown in Table 4. Girls tended to score more highly than boys overall, and in all dimensions. There was a higher variance in the quality of boys’ games.

Given the findings of a recent study in which girls performed better in a game making task (Carbonaro et al., 2010), one tailed t-tests were conducted to test the hypothesis that the games produced by girls were rated more highly. All of the dependent variables had a normal distribution. The total score, story, choice, characters and dialogue categories had unequal variances; in these cases Welch’s t-test was used. Alpha was set at .05, but a Bonferroni correction was used to correct for multiple tests (corrected alpha = .005). Effect size calculations (Cohen’s d) were also performed. The results are shown in Table 5.4

5.3. Research question 3 – gender differences in the game making process

One way to investigate the reasons for the gender differences in the game product is to examine the game making process. The time spent in different types of game making activities between genders was therefore investigated. As the learners’ overall time using the software was related to classroom circumstances, one would not expect to see gender differences unless all the boys or girls as a group were given another activity instead of game making (such as taking part in a girls only sporting event). Differences relating to the proportion of time spent in different sorts of game making activities, however, might indicate that boys or girls have different preferences about sub-activities or value them differently. Furthermore, some insight can be gained by analyzing gender patterns in the learners’ use of the Comment Card tool which was used towards the end of the project. It is anticipated that more effective use of this tool would lead to a better game.

Table 6 shows the time spent by learners in different game making activities by gender. The times are calculated from software log files based on the learner’s use of the user interface. The “ideas generation” entry in the table refers to the amount of time which the learners spent adding, viewing, editing rearranging or deleting Fridge Magnets, but obviously does not capture the amount of time the learners spent thinking about ideas. “Adding blueprints” refers to inserting blueprints representing characters, buildings, props and other decoration into the game level. “Sculpting terrain” includes editing the landscape of the game level to make features such as rivers, hills, valleys and changing textures to indicate terrain features such as snow. “Peer review” indicates time spent with the Comments Cards feature of Adventure Author open (during which time the learners were instructed to verbally discuss their games), but not the time playing the game under review. “Conversation writing” refers to the activities of typing, editing and deleting words in lines of conversation, as well as time spent on structuring a branching conversation and specifying conditions and branches. “Games testing” is the activity of the learner playing their own game, usually followed by game editing in the authoring software. The table entries of adding, editing and deleting content give a higher level overview of time spent in different stages of the editing process, and are summed across the other authoring activities in the table. For example, the activity of “adding new content” includes adding new blueprint, typing words in the conversation editor and creating new Fridge Magnets.

It can be seen that the learners spent a mean of around 6 h using the software. There is a standard deviation of around an hour due to naturally occurring variations in the classroom routine. As would be expected on a creative task, the learners spent the highest proportion of time adding new content, rather than deleting or editing content. They spent most time adding blueprints to populate the game level (around an hour and forty minutes), followed by around an hour and a half spent in testing their games. There is a high standard deviation for most of the activities, reflecting the learners’ personal preferences, apart from the activities of generating ideas and peer review, where the teacher and researchers ran the sessions to a set schedule. In terms of gender differences, a t-test reveals no significant difference in the total time spent in game making between boys and girls ($t = .590, n.s.$). Given the finding from the previous section that the quality of girls’ dialogue is higher, a post-hoc t-test on the time spent in writing conversations was justified ($t = .011, p < .01$), indicating that girls spent longer than boys writing conversations.

Although the time spent in the peer review session was the same for boys and girls because the session was structured by the teacher and researchers, it may be the case that their patterns of behaviour in response to the advice they received from their peers were different. Full analysis of the learners’ use of the Comments Cards tool is reported elsewhere, (Robertson, 2011, p. 26) but the issue of whether there is a gender difference in improving one’s game based on peer advice is most pertinent here. Following an action plan for improving a game would be likely make the game better, and so gender differences of this sort in the game making process could offer an explanation for

---

4 The analysis was conducted in R using the script documented here: http://www.unt.edu/rss/class/Jon/Benchmarks/BayesFactors_JDS_Mar2011.pdf. Full results, including Bayes factors can be found here: REMOVED FOR REVIEW.
differences in the quality of the game product. Previous analysis in indicated that the learners were mature about generating and responding to constructive criticism, and that they could formulate plans based on the advice given by teachers or reviewers (Robertson, 2011, p. 26). Did they actually make the changes they proposed? The analysis performed to answer this question is as follows: for each designer who got specific feedback in the Comments Card peer review session, and had enough time to respond (21 designers\(^5\)), the games and log files were checked by hand to see if the designers had taken the advice. It was recorded whether they took the reviewer’s advice, took it partially, followed their own plan, or followed their own plan partially, or did neither. Eight out of 21 followed their plan completely, while 11 (52%) did some of their plan, and 2 did not carry out the plan at all. Ten out of 21 designers followed the reviewer’s advice completely while a further four did so partially. In total, 76% of the pupils took action in future sessions which corresponded to actions noted on the Comments Cards. There was a significant gender difference here; girls were more likely to take some action in response to comments than boys (Mann–Whitney \(U = 27.5; p < .05\)). The fact that girls were more likely to attempt to improve their games based on peer review is a possible contributing factor to why their final products were rated more highly as described in Section 5.2.

6. Discussion

Observations and discussions with the teacher during the school field study indicated that Adventure Author was well received by pupils and teachers. Discussions with the children’s parents who were invited to the final session of the project revealed that the parents appreciated that the children had learned a lot from taking part and were impressed by what they created. The children spent considerable time working on their games, and were motivated by this task. They enjoyed working on other creative activities such as story writing and crafts as part of the project.

A summary of answers to the research questions in shown in Table 7.

6.1. Game characteristics: the need for player awareness

The designers’ games showed that they had mastered basic technical skills such as those used to write interactive dialogue or create a landscape. The games were often visually appealing, and some contained story locations (special areas of the landscape in which plot events could unfold). There were some imaginative ideas in the games, particularly for the storylines, and some designers attempted to use humour and poetic language to give their dialogue impact. However, most designers could improve their game by structuring the game around the needs of the player by explaining the game objectives, giving visual clues for spatial navigation around the environments and ensuring that the player will have the opportunity to engage in the game by carrying out a quest, exploring an interesting area, or taking part in a battle of appropriate difficulty. Lack of player awareness for game designers is perhaps analogous to the common problem of lack audience awareness for novice writers. It can be addressed in writing tasks by giving the writer a real audience and encouraging a writing process composed of multiple drafts (Wyse, 1998). Analogously, player awareness can be encouraged through iterative testing and peer evaluation. It would perhaps be beneficial to include it earlier in the game design process than in this study, or at least to make more time for the designer to make changes based on feedback.

6.2. Game making process: the benefits of peer review as part of the game making process

When using Comments Cards, the pupils showed themselves to be accomplished peer reviewers, and demonstrated their critical analysis skills in the games domain (Robertson, 2011, p. 26). It is encouraging that 76% of the designers followed up on the feedback given by reviewers because it suggests that the exercise of peer reviewing is not only beneficial from the perspective of the reviewer developing critical analysis skills, but also is useful for assisting the designer in improving her work. Such evaluation criteria could also be used to develop critical analysis skills using off the shelf games.

Significantly more girls than boys took action as a result of a peer review sessions. It is not surprising that willingness to take on board advice from peers would result in a more polished product, so this could be a partial explanation of why the girls’ games were rated more highly.

---

\(^5\) Four authors were either absent at the peer review session or in subsequent sessions and did not have the opportunity to make improvements based on peer review.
6.3. Gender differences: interpretations of effect size

The gender differences in the game products can be explained to some extent by the differences in the game making process. Girls chose to spend more time writing dialogue than boys, and so scored more highly in the dialogue and storyline dimensions. They also devoted more time to following up actions plans for improving their games based on advice from peer reviewers, with the result that their overall game quality was higher.

According to Cohen’s heuristics (1992) for interpreting effect sizes, the effect sizes for those categories with significant differences (i.e. total score, story, player guidance, player purpose, choice, character, dialogue and imagination categories) are large. In order to put these effect sizes in the context of game making in education, an effect size was calculated using the descriptive statistics reported in the study of high school game makers by Carbonaro et al. (2010). In this previous study Cohen’s $d = -.68$. The present study has larger effects in the same direction as this previous work. In the field of education in general, Hattie (2008), p. 392 states that an effect size of $d = 1.0$ (one standard deviation above the mean) “is typically associated with advancing children’s achievement by two to three years...” (p. 8). In the present study, the total score, story, player purpose, choice, characters, dialogue and imagination all had effect sizes of over 1.

How can we account for such large effects? Firstly, the results might be affected by methodological weaknesses. By chance, this may be a class with unusually large differences in baseline achievement between boys and girls. Unfortunately, the researchers do not have access to the pupils’ national curriculum scores for comparison. It is worth noting that both the researchers who assessed the game are female, as are the class teacher and the researcher who devised the rating scale. It may be the case that male teachers or researchers would have emphasized different aspects of game making. Additionally, the games were not blinded for analysis as both the raters could associate the pupils with their designers from face to face tuition in the classroom. We did not have the resources to recruit additional expert raters. In order to address these issues in the future, a larger study could be conducted in which other measures of achievement, such as story writing grades, are collected and the games are analysed blind.

An explanation for the direction of the effects could be that the rating scale measures aspects of the game which favour girls’ pre-existing strengths. The aspects of game making which are assessed in Howland’s rating scale (2011) were developed in conjunction with an expert literacy educator and were designed to be consistent with the Scottish Curriculum for Excellence literacy outcomes. It could be argued that they are more heavily weighted towards skills which are already valued within the classroom than the more visual, interactive aspects of game design. Findings from the literature on gender differences in children’s writing would suggest that girls are more likely to do well on activities associated with fictional writing (Beard & Burrell, 2010). This could explain their higher performance on the story, characters, dialogue and possibly imagination categories in this study. As the pupils often used dialogue to convey player choice, and the purpose of the game was often revealed to the player through dialogue, the differences in these categories could be explained by the girls’ higher performance in dialogue writing. The categories for which significant differences were not detected (visual design and player engagement) are less dependent on textual narrative skills, and so girls’ prior level of achievement in fictional writing would be less relevant. That being

---

6 A mapping of literacy outcomes to Adventure Author skills can be found under the Experiences and Outcomes section at http://www.adventureauthor.org/teaching-materials.html.
said, the relationship between new media literacy skills and traditional story writing is not entirely straightforward, as discussed in the case study analysis by Howells & Robertson (in press). Conventions from traditional story writing do not always transfer to the game medium. For example, heavy use of dialogue may be interesting in written stories, but may not be tolerated in a game by an impatient player. In addition, inexperienced gamers may underestimate the importance of the 3D spatial game world, and omit to test the game from the player’s point of view. Self-review by proof reading and editing dialogue is not enough – the designer must also interact with the game to check that the player’s needs are met.

The choice of game engine (Neverwinter Nights 2) is also biased towards creating narrative based games because it enables the creation of games in the role-play genre. Different choices of game engines might produce different results. Certainly, two recent studies of game making in an educational quiz genre found no gender differences (Owston et al., 2009; Vos et al., 2011) whereas Carbonaro et al. (2010) found similar results to the present study using very similar role-playing game software.

7. Conclusions

In conclusion, this paper has argued that making computer games as part of a classroom project has the potential to develop a range of new media storytelling, visual design and audience awareness skills. Evidence from a six week field study in a classroom demonstrates that young game designers can use game making as a medium for expressing imaginative narrative ideas, creating games with interactive dialogue and aesthetically pleasing visual designs. However, the designers required more development of player awareness skills such as how to effectively guide the player spatially around the environment and making them aware of the game objectives. Iterative peer evaluation is a possible approach to addressing this shortcoming, as the results of peer review late in the creative process were promising.

In spite of concerns published in the literature that girls could be disadvantaged by their lack of prior knowledge of computer games during classroom based gaming activities, the opposite effect was found in this study. Girls performed considerably higher in the educational game making activity overall, and for the dimensions of story, player guidance, player purpose, choice, character, dialogue and imagination. To an extent this can be explained in terms of differences in the game making process: girls spent more time writing conversations and were more likely to follow advice from peer reviewers to improve their games. Another possible explanation for this is that girls’ higher prior knowledge and enjoyment of fictional writing outweighed the disadvantages of their lower prior gaming knowledge. An implication for educators is that they should be aware of possible gender effects in choosing game making software and assessment schemes. It would be unfortunate if the introduction of game making technology in the classroom merely reinforced existing patterns of attainment around gender. To this end, a focus on non-textual aspects of game design such as visual design and player engagement is recommended alongside narrative. In the best case, girls may be encouraged to engage with technology as a result of their enjoyments of storytelling while boys may be motivated to develop their storytelling skills through their interest in gaming technology.

Acknowledgements

The author would like to thank Katy Howland, Cathrin Howells and Andrew Macvean for their contributions to this work. The project was funded by EPSRC.

References


