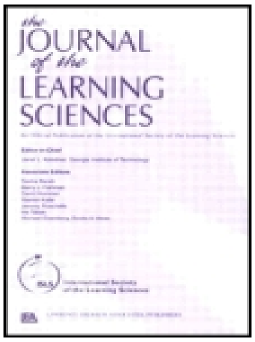

Journal of the Learning Sciences

ISSN: 1050-8406 (Print) 1532-7809 (Online) Journal homepage: http://www.tandfonline.com/loi/hlns20

# Motivating Children to Learn Effectively: Exploring the Value of Intrinsic Integration in Educational Games

M. P. Jacob Habgood & Shaaron E. Ainsworth

**To cite this article:** M. P. Jacob Habgood & Shaaron E. Ainsworth (2011) Motivating Children to Learn Effectively: Exploring the Value of Intrinsic Integration in Educational Games, Journal of the Learning Sciences, 20:2, 169-206, DOI: 10.1080/10508406.2010.508029

**To link to this article:** http://dx.doi.org/10.1080/10508406.2010.508029

Published online: 20 Apr 2011.

THE JOURNAL OF THE LEARNING SCIENCES, 20: 169–206, 2011

ISSN: 1050-8406 print / 1532-7809 online
DOI: 10.1080/10508406.2010.508029

# Motivating Children to Learn Effectively: Exploring the Value of Intrinsic Integration in Educational Games

M. P. Jacob Habgood
*School of Psychology and Learning Sciences Research Institute*
*University of Nottingham*
*and*
*Department of Computing*
*Sheffield Hallam University*

Shaaron E. Ainsworth
*School of Psychology and Learning Sciences Research Institute*
*University of Nottingham*

The concept of *intrinsic motivation* lies at the heart of the user engagement created by digital games. Yet despite this, educational software has traditionally attempted to harness games as extrinsic motivation by using them as a sugar coating for learning content. This article tests the concept of *intrinsic integration* as a way of creating a more productive relationship between educational games and their learning content. Two studies assessed this approach by designing and evaluating an educational game called Zombie Division to teach mathematics to 7- to 11-year-olds. Study 1 examined the learning gains of 58 children who played either the intrinsic, extrinsic, or control variants of Zombie Division for 2 hr, supported by their classroom teacher. Study 2 compared time on task for the intrinsic and extrinsic variants of the game when 16 children had free choice of which game to play. The results showed that children learned more from the intrinsic version of the game under fixed time limits and spent 7 times longer playing it in free-time situations. Together, these studies offer evidence for the genuine value of an intrinsic approach for creating effective educational games. The theoretical and commercial implications of these findings are discussed.

Correspondence should be addressed to M. P. Jacob Habgood, Department of Computing, Sheffield Hallam University, Furnival Building, Sheffield, S1 2NU, UK. E-mail: J.Habgood@shu.ac.uk

The use of computer games and simulations in education dates back to the 1950s (Cullingford, Mawdesley, & Davies, 1979), when computing was still in its infancy and the commercial videogame industry had yet to emerge. Nonetheless, it was the raw engagement power of 1980s videogames like Pac-Man that inspired a new generation of educationalists to consider the learning potential of this exciting new medium (Bowman, 1982). These early protagonists were quick to identify the motivational power of videogames as their key asset (e.g., Lepper & Malone, 1987; Loftus & Loftus, 1983) and were able to apply a range of existing motivational (e.g., Csikszentmihalyi, 1975; Deci, 1975; Lepper & Greene, 1975) and behavioral (Ferster & Skinner, 1957) theories to their rationales. However, despite this promising start, the resulting generations of "edutainment" products have been widely recognized as failing to effectively harness the engagement power of digital games (e.g., Hogle, 1996; Kerawalla & Crook, 2005; Papert, 1998; Trushell, Burrell, & Maitland, 2001). So although the mainstream games industry boomed throughout the 1990s, the educational sector was left behind in terms of technology, revenue, and commercial interest. However, the turn of the millennium has seen a rejuvenation of interest in game-based learning, with a number of texts extolling the potential of games (e.g., Aldrich, 2004; Gee, 2003; Shaffer, Squire, Halverson, & Gee, 2005), paralleled by the commercial success of "self-improvement" titles such as Brain Training and Big-Brain Academy (Nintendo).

This article offers empirical evidence for the value of a design approach that may help to explain the failure of edutainment to fulfill its educational promise. This approach hinges upon the ability of learning games to effectively harness the *intrinsic motivation* (Deci, 1975) of a game for educational goals by creating an *intrinsic integration* (Kafai, 1996) between a game and its learning content. Furthermore, we suggest that such an integration is created through an intrinsic link between a game's *core mechanics* (Lundgren & Björk, 2003) and its learning content.

Zombie Division is a computer game specifically created to empirically examine the concept of intrinsic integration. The game integrates mathematics into the core mechanic of a 3D adventure through a combat system in which opponents are mathematically divided in order to be defeated. Three variations of this game were created for evaluation: an *intrinsic* version that integrated mathematics into combat, an *extrinsic* version that had non-mathematical combat and placed identical mathematical multiple-choice questions between levels instead, and a *control* version that contained no mathematics at all. The first study compared learning gains between all three versions as a measure of the relative educational effectiveness of the intrinsic approach. The second study compared time on task between the intrinsic and extrinsic versions of the game as a measure of the relative motivational appeal of the intrinsic approach.

## DEFINING INTRINSIC INTEGRATION

The concept of intrinsic integration in educational games is rooted in the more familiar concept of "intrinsic motivation." It is commonly surmised that a person is intrinsically motivated to perform an activity when he or she receives no apparent rewards except the activity itself (Deci, 1975). Although modern videogames *can* provide external rewards (such as those produced by farming virtual game resources; see Steinkuehler, 2006), they are largely autonomous pursuits that create their own internal motivations for continuing the activity. Game designers can create these internal motivations through the inclusion of aspects such as *challenge*, *control*, *fantasy*, and *curiosity*, whereas interpersonal motivations can be added through factors such as *competition*, *cooperation*, and *recognition* (Malone & Lepper, 1987). The inclusion of *challenge* in this taxonomy is derived from the work of Csikszentmihalyi (1988) into flow theory and optimal experience. This proposes that clear goals, achievable challenges, and accurate feedback are all required to achieve a state of flow in an activity that requires "a balance between the challenges perceived in a given situation and the skills a person brings to it," suggesting that "no activity can sustain it for long unless both the challenges and the skills become more complex" (Csikszentmihalyi, 1988, p. 30). There are clear parallels between this and the way in which game designers carefully structure the difficulty curves of their games to provide the optimal level of challenge as a player's skills develop (Habgood & Overmars, 2006, p. 158). It is perhaps unsurprising then that feelings of total concentration, a distorted sense of time, and extension of self are as common to game players as they are to Csikszentmihalyi's (1988) rock climbers and surgeons. There is also emerging evidence that, when measured correctly, flow is predictive of learning (e.g., Engeser & Rheinberg, 2008).

The gaming literature provides an overwhelming number of different approaches to defining the essence of a game (Caillois, 1961; Crawford, 1982; Huizinga, 1950; Juul, 2005; Koster, 2005; Salen & Zimmerman, 2004). Yet these differences only serve to highlight Wittgenstein's (1953) observation on games that "you will not see something that is common to all, but similarities, relationships, and a whole series of them at that" (Aphorism, 66 p. 27). Therefore, in the interests of practicality, we use a definition of game that seeks to highlight the main differences between games and other forms of entertainment rather than all the similarities between things we might refer to as a game. This pragmatic definition defines a *game* as simply an "interactive challenge," suggesting that games contain an interactive element that distinguishes them from films and prescribed challenges that distinguish them from toys (Habgood & Overmars, 2006, p. 87). We therefore see games as something that encompass a wide spectrum of digital and nondigital applications—including many simulations—and we hope this research could potentially have relevance to all of them. It should also be noted that our definition deliberately avoids assigning a motivational aspect to a game, as the experience of intrinsic motivation is subjective (does a game stop being a game if it stops being fun?). Nonetheless, the ability of games or simulations to create intrinsic motivation is clearly central to this argument, and uninspiring games are not a good model for creating motivating learning games either.

Although digital games may be capable of providing activities that are intrinsically motivating in their own right, it is critical to consider the effect of adding learning content to an intrinsically motivating game. Game designers have come to recognize the role of learning in good game design (e.g., Crawford, 1982; Gee, 2003; Habgood & Overmars, 2006; Koster, 2005). This is not about commercial games containing educational content but about how the enjoyment of games derives from the process of learning itself (i.e., "the fundamental motivation for all game-playing is to learn," Crawford, 1982, p. 17). Unfortunately, edutainment products have traditionally taken a "chocolate-covered broccoli" (Bruckman, 1999) approach when combining learning content with gameplay. This is where the gaming element of the product is used as a separate reward or sugar coating for completing the educational content.

It was Malone's (1980) and Malone and Lepper's (1987) seminal work on videogames that first considered the problem of creating a more integrated approach to designing educational games. This originally proposed the concept of an intrinsic fantasy as providing "an integral and continuing relationship between the fantasy context and the instructional content being presented" (Malone & Lepper, 1987, p. 240). This was contrasted with an extrinsic fantasy in which "the fantasy depends on the skill being learned, but not vice versa," and it was suggested that the learning content of extrinsic fantasies could be easily replaced with something different. Furthermore, it was suggested that "in general, intrinsic fantasies are both more interesting and more instructional than extrinsic fantasies" (Malone, 1980, p. 164). We can attempt to clarify this definition by considering contrasting examples. The classic Maths Blaster series included a game called Trash Zapper that required the player to provide the answer to a simple arithmetic sum ($9 + 6 = ?$) by shooting a moving item of rubbish that had the answer written on it. However, one could replace arithmetic sums with a spelling task (e.g., ELE?HANT) and attach letters to the rubbish without having a significant impact on the fantasy context of the game. Therefore, this game could be considered to be an extrinsic fantasy. Conversely, the same definition might consider football management games as providing intrinsic fantasies for mathematical learning content. This kind of game allows control over team budgets and team statistics, and clearly it would completely break the fantasy context of the game to try and balance a payroll, or a team's abilities, by spelling words correctly.

However, we have argued that this focus on the intrinsic nature of fantasy is misplaced, suggesting that such fantasy contexts are often purely arbitrary and could be swapped for another so long as the basic mechanics of the game are not altered (Habgood, Ainsworth, & Benford, 2005a). The football example may be intrinsic according to the definition above, but the fantasy context of football could be replaced by Smurf volleyball so long as the rule systems and player interactions with the budgets and player statistics remained the same. The game may not appeal to the same audience, but it would be equally as "instructional" as the original. It is therefore the core mechanics that embody the rule systems and player interactions that are intrinsic to the educational value of football management games and not the fantasy context at all.

Consequently, we consider the term *intrinsic integration* to be a more appropriate way of describing a situation in which "a designer integrates the subject matter with the game idea" (Kafai, 1996, p. 82). This is not to suggest that the theoretical argument for the role of core mechanics has been "won," and many researchers still prefer to describe the concept of effective integration with reference to fantasy (Asgari & Kaufman, 2004; Gunter, Kenny, & Vick, 2008). Others acknowledge the wider debate between researchers in attributing the intrinsic motivation produced by games to the role of narrative context (fantasy) or intrinsic goals and rewards (core mechanics; Dondlinger, 2007). Nonetheless, it is clear that when some researchers refer to the intrinsic role of fantasy they mean this to include the role of both fantasy and core mechanics in the intrinsic relationship (e.g., Paras, 2005).

The definition of intrinsic integration evaluated in this article was first developed in Habgood et al. (2005a) and has two central components:

1. Intrinsically integrated games deliver learning material through the parts of the game that are the most fun to play, riding on the back of the flow experience produced by the game and not interrupting or diminishing its impact.
2. Intrinsically integrated games embody the learning material within the structure of the gaming world and the player's interactions with it, providing an external representation of the learning content that is explored through the core mechanics of the gameplay.

We can now consider the two earlier examples of Maths Blaster and football management games with respect to this new definition. Trash Zapper certainly delivers learning material through the parts of the game that are the most fun to play and rides on the back of the flow experience. As a result, it does not diminish the impact of the game, and the additional cognitive demands of the arithmetic potentially add to the challenge of an otherwise trivial gaming exercise. However, the relationship with the core mechanics and structure of the gaming world is not embodied. Although the mathematical content is attached to the core mechanic, the two are not actually integrated at all—which is why the learning content could so easily be replaced with spelling. This thin integration means that mathematical representations are not part of the structure of the gaming world, so it does not provide scope for the kind of constructivist interactions associated with microworlds and simulations (see below). It is interesting that football management games arguably fail the first part of our definition for intrinsic games, as they do not deliver the learning content through the part of the game that is most fun to play! The model for this genre of game often involves completing the "numerical chores" between matches in order to reach the reward of watching one's side play based on one's team selection and substitutions. However, some players choose to skip the opportunity to watch their team play altogether and get straight back to the "chores," which suggests that administration between games provides a flow experience in its own right. Nonetheless, football management games do seem to embody some learning material within the structure of the gaming world and players' interactions in a way that leads players to engage cognitively with the learning content in the game.

## POTENTIAL ADVANTAGES AND DISADVANTAGES OF INTRINSIC INTEGRATION

Having described our approach to intrinsic integration we can now turn to considering what advantages or indeed disadvantages it might bring for learning. The central claimed benefit of educational games and intrinsic integration lies in the potential to more effectively motivate and engage the player in the learning content of a game (e.g., Garris, Ahlers, & Driskell, 2002; Rieber, 1991). As such, flow is often considered to be critical in creating and maintaining this motivational appeal. Integrating learning content into the very parts of the gameplay that give rise to the flow experience should ensure that the benefits of the flow are directed toward educational goals. Conversely, edutainment or extrinsic games that provide gameplay as a reward for learning content are more likely to disrupt flow if players are asked to regularly switch to another non-flow-inducing activity. Moreover, the flow state in extrinsic games is therefore experienced in the service of game but not educational goals. However, although intrinsic integration in educational games may increase motivation and flow, it is not completely clear how this translates into increased learning (Pintrich, 2003). Mechanisms that have been postulated include persistence, more focused attention, increased arousal, increased affect, and alternative strategies (Garris et al., 2002; Martens, Gulikers, & Bastiaens, 2004; Parkin, Lewinsohn, & Folkard, 1982; Pintrich, 2003; Vollmeyer & Rheinberg, 2000).

Another central benefit of intrinsic integration comes from embodying the learning content (the tasks learners must address, the actions they perform to do so, and the feedback they receive as a consequence) within the core representational structure of a gaming world. There is a vast literature on the importance of external representations in learning, with much evidence that using appropriate representations (or combinations of representations) can enhance learning outcomes (Ainsworth, 2006; Scaife & Rogers, 1996; and Winn, 1987, all provide reviews). The value of interactive representations has long been recognized in educational simulations (de Jong & van Joolingen, 1998) and microworlds (Papert, 1980), where structured learning environments attempt to embody a particular learning domain by providing interactive representations within a self-contained, rule-governed world (Edwards, 1998) and the synergies between microworlds and digital games are evident (Rieber, 1996). However, although microworlds provide a carefully structured learning environment, they do not generally attempt to structure the motivational environment and manage the flow experience in the same way digital games do. So we suggest that learning should be enhanced if the representational structure embodies the core gameplay mechanisms that give rise to the central flow experience of the game.

However, it is also possible that intrinsic integration may be disadvantageous. It is well known that children find it difficult to apply mathematical knowledge acquired in one context to a different one, even if the mathematical principles are the same (e.g., Nunes, Schliemann, & Caraher, 1993). So one concern is that intrinsic integration could create knowledge that is highly specialized to the specific condition of application so that learning in the game "stays in the game" rather than transferring to school mathematics. Children may also apply their learning from extrinsic games more effectively, as the format is often much closer to the abstract form of a school context, and situations that are similar tend to promote both enhanced recall (e.g., Tulving & Thompson, 1973) and transfer (e.g., Gentner, 1989).

Our approach to intrinsic integration raises these concerns about transfer for two main reasons. First, intrinsic integration typically involves concrete representations rather than abstract ones. Some (although by no means all) existing research suggests that children (in particular) can find it difficult to transfer their understanding from concrete representations to alternative representations (e.g., DeLoache, 1991; Goldstone & Son, 2005; Kaminski, Sloutsky, & Heckler, 2008). Interacting and playing with concrete representations can make this situation worse (e.g., Uttal, O'Doherty, Newland, Hand, & DeLoache, 2009). Second, intrinsically integrating learning content within frantic action-based games could make it harder to learn the educational content, as the learner must cope with two forms of competing demands simultaneously (the educational and gameplay elements). This is likely to be more true of action-led games such as Zombie Division than simulations or epistemic games (Shaffer, 2004). There may be a concern that it may also inhibit contemplative reflective activity and so hinder the development of appropriate strategies; as a consequence, there may be less transfer from the game (e.g., Berry, 1983).

Given that equally compelling arguments could be made for the advantages and disadvantages of intrinsic integration, we decided to test these arguments by developing and then evaluating an intrinsic and extrinsic version of the same educational game.

## ZOMBIE DIVISION: THE EDUCATIONAL GAME

The learning content of the game is based upon the United Kingdom's National Curriculum targets for Key Stage 2 (7- to 11-year-olds) focusing on number patterns and sequences:

> Recognise and describe number patterns, including two- and three-digit multiples of 2, 5 and 10, recognizing their patterns and using them to make predictions; recognise prime numbers up to 20 and square numbers up to 10 × 10; find factor pairs and all the prime factors of any two digit integer. (Qualifications and Curriculum Authority, 1999, p. 22)

The game itself is a 3D adventure game based around sword fighting in which the player (acting as the hero Matrices) must use different attacks to mathematically divide opponents according to the numbers displayed on their chests (see Figure 1). The core mechanic could be described as defeating enemies in combat by attacking each enemy with a divisor that divides its dividend into whole parts. Each of the player's attacks has a different animation that embodies that divisor and reinforces the association between the divisor and the attack. Archaic combat weapons illustrate these relationships (e.g., Divide by 2—a single swipe of a sword, Divide by 3—a barge with a triangular shield, Divide by 5—a punch with a (five-fingered) gauntlet, Divide by 10—a single swipe of a sword and a punch with a gauntlet). Thus, the structure of these attacks embodies additional mathematical relationships in the way that weapons combine. In this way, the learning content is integrated within the core mechanic of the gameplay. The game also includes secondary game mechanics that revolve around exploration and collection (exploring a nonlinear 3D dungeon and collecting keys), but these are not integrated with the mathematical content of the game. An arbitrary fantasy context for the game is provided (linking the numbers on skeletons' chests to cursed Olympic athletes), but we argue that this fantasy is extrinsic and could very easily be replaced with an entirely different fantasy context (e.g., evil robot enemies with monitors on their chests) without changing the intrinsic relationship of the learning content with the game's core mechanic. One could also replace the secondary game mechanics that revolve around exploration and collection with others (e.g., tower defense:

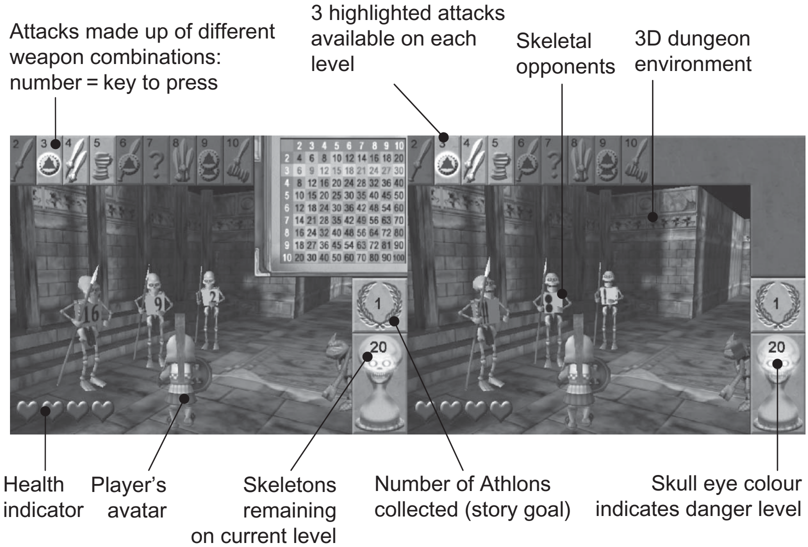


FIGURE 1 Screenshots of an identical gameplay situation in the intrinsic (left) and extrinsic (right) versions of Zombie Division. All labeled features are common to both versions.

defending a static object from attack) provided the combat mechanic remained the same.

Each game level contains about 20 enemies (zombie skeletons). When divided with an appropriate attack, the "spirit" of a defeated skeleton rises and then splits into equal-size portions (depending on the divisor) that grow into small ghosts bearing the quotient (these normally disperse, but see below). Using an inappropriate attack against an enemy results in the skeleton fighting back and the player losing health, which, when exhausted, forces the player to start the level again. In this way, the player is not just asked to choose between three divisors for each opponent but must consider whether opponents are dividable at all using their current attacks.

Two additional mechanics help children develop their mathematical competency. As players progress through a game, skeletons gain weapons with which to parry attacks. For example, an opponent with a sword could parry attacks that divide them by 2, 4, 6, 8, and 10. In these cases, the player would be forced to consider other (potentially less obvious) divisible factors of the opponent's dividend in order to defeat the opponent. This mechanism, for example, stops children from dividing all even numbers by 2. The other mechanic is the inclusion of giant skeletons that can only be defeated outright by dividing by a large number such that the quotients of the resultant parts are less than 10. Quotients greater than 10 rise again as new skeletons in their own right and continue to attack the player with increased ferocity. Again, this is included to encourage children to use a fuller repertoire of divisors.

Only three different attacks (divisors) are available to the player at any point in the game in order to ensure that there are always some enemies that cannot be defeated with the available choices. Players begin with the divisors 2, 5, and 10 and gradually gain access to different attacks as the game progresses. The skeletons themselves gain more ferocity, which encourages faster responses from children (in the initial levels, skeletons stand around rather passively, but as levels increase they move around, block areas of the dungeon, or pursue Matrices). More docile skeletons are generally found in levels with new divisors, and the hostility of the skeletons on each level is indicated through the color of the skeleton's eyes on the player's user interface (ranging from green to glowing red).

Children receive two main sources of in-game help: Gargle (an animated pedagogical agent) and a magical book of times tables (the multiplication grid). Gargle accompanies Matrices throughout the game providing just-in-time instructions and help. He is first seen in an initial training level in which the players learn to navigate and fight passive "clockwork" skeletons before meeting them for the first time in the game. He also helps children understand how the multiplication grid can be used to help decide whether a number is divisible by a particular divisor (which is provided in a small tutoring component of the game based around sharing bones). As the game progresses, Gargle can provide oral instructions as to how to play the game, including task direction, but he is typically silent unless children are experiencing problems.

The game was created specifically for this study (and different versions were created to test our approach to intrinsic integration; see below). The design and development of Zombie Division was undertaken as an iterative process with regular input from both girls and boys from the target audience. The instigation of the project followed an intensive period working with children of the same age, teaching them how to make their own educational games (Habgood, Ainsworth, & Benford, 2005b). The initial design for Zombie Division was created as a cardboard prototype and piloted with boys and girls with a range of different mathematical abilities. Despite our own initial reservations about the gender neutrality of the fantasy context, it appeared to have cross-gender appeal for the target age group when presented within a school setting. It is also worth noting that the game has a number of similar game mechanics to the "Zelda" series of games that have become notable as a franchise that appeals to both male and female audiences. Regular piloting also allowed us to address gameplay issues (such as navigation and combat systems) that may otherwise have favored success by the more game-literate children (usually boys).

Zombie Division was designed to be consistent with a 7+ rating under the Pan European Game Information (PEGI) age-rating system (e.g., includes occasional violence to nonrealistic fantasy characters). Its theatrical title seemed to provide instant appeal to its target audience, but the game itself did not contain any of the gruesome or gory content that the word *zombie* might evoke. The skeletons were rather comical, and children were more frightened of the doors (which slammed behind them in a spooky way) than the skeletons. Nonetheless, even the youngest children enjoyed the creepy atmosphere, and it seemed to add to their engagement with the game rather than putting them off.

## STUDY 1: THE IMPACT OF INTRINSIC INTEGRATION ON LEARNING

This study was designed to assess learning outcomes to test the effectiveness of intrinsic integration in educational games as used within traditional classroom settings. We created three versions of the game that differed only in how the math content was delivered: intrinsic, via in-game action; extrinsic, end-of-level quiz; or control, none at all. The classroom setting imposed a number of additional constraints on the study. The overall time children spent learning was fixed in terms of the time per session (15–20 min) and strictly controlled in terms of the total playing time (2¼ hr). This amount of time was chosen to be realistic within the framework of the UK Numeracy Strategy, which would typically devote this amount of time to teaching multiplication and division every day for a week and then revisit the topic frequently under a cyclic curriculum. Finally, we also wanted teachers to play an active role in supporting children's learning with the game.

Post-play debriefing is considered by many to be critical to the effective application of educational simulations and games (e.g., Garris et al., 2002; Lederman, 1992; Squire, 2004). Sandford and Williamson (2005) went further to suggest that "the outcomes of any lesson-based computer activity will depend on the introduction of the task, the interventions made during the activity and the way that the activity is set in the context of students' wider educational experience" (p. 11). Even outside of games, the role of the teacher is often seen as an important facet of computer-based learning (e.g., Tabak & Reiser, 1997). Therefore, the way in which Zombie Division is framed and supported within the children's wider educational context is likely to have a large effect on the learning outcomes of the game. The potential learning gains in both intrinsic and extrinsic conditions would have almost certainly been maximized by the addition of supporting activities pre, post, and in parallel with the gaming interventions. However, the purpose of this study was not to maximize overall learning gains but to compare learning gains between intrinsic and extrinsic approaches over a limited amount of time on task. We did not want to bias the learning gains produced by just over 2 hr of gaming with another 2 hr of teaching.

We therefore decided to include a controlled teacher-led reflection session delivered after the children had had a chance to become familiar with the game but before they had gotten too far. This would aim to get children to reflect on the mathematical context of the game (in the intrinsic and extrinsic versions) and scaffold the conceptual process of making the link between solving division problems using multiplication facts. The reflection session would be structured to include identical mathematical content in all three conditions but made relevant to the context of each version of the game (except for the control group, for which the learning content has no relevance to their version of the game and so would be taught in the abstract case). We believed this would provide the right balance between the requirement for some form of debriefing to make effective use of the game and the need not to make this a study about teaching mathematics around the instructional anchor (Bransford, Sherwood, Hasselbring, Kinzer, & Williams, 1990) of a videogame.

## Method

*Participants.* All children attended a primary school in a low-income area on the outskirts of a large city in the north of England. National tests show that the attainment of the school's incoming pupils was below average, and the percentage of final-year students achieving expected levels in mathematics was below national averages for the preceding year. The 30 girls and 28 boys were between 7 years, 1 month, and 8 years, 10 months, in age ($M = 8$ years, 0 months). All participants had prior experience using the computers in the school's Information and Communications Technology (ICT) suite.

*Design.* This study used a two-factor ($3 \times 3$) mixed design. The first factor, game, was a between-groups factor with three levels (intrinsic, extrinsic, or control) that determined which version of the game children played. The second factor, time, was a within-groups factor and also had three levels (pre-, post-, and delayed test). The 58 children were randomly assigned to one of the three conditions such that pretest scores, gender, and age did not differ within each condition. Consequently, 20 children were assigned to the intrinsic, 20 to the extrinsic, and 18 to the control condition.

*Materials.*

Facilities. Children took tests and played with Zombie Division within the school's ICT suite, which contained 20 personal computers (PCs) running Windows 2000 with accelerated 3D graphics support and audio output through stereo headphones. The teacher-led reflection sessions were carried out in a standard classroom using an interactive whiteboard running PowerPoint to present the teaching material.

Game versions. Three versions of the game were created for this study. The base version was the intrinsic version (as described above), and the extrinsic and control versions were based upon it. The key practical and theoretical difference in the extrinsic and control versions was that dividing skeletons no longer involved any mathematical content. In the extrinsic version the mathematical content was provided as an end-of-level quiz, and in the control version it was excluded altogether. In the intrinsic version, the dividend displayed on each skeleton's chest provided the player with a way of determining the skeleton's vulnerability to different attacks. The same result was achieved in the extrinsic and control versions by replacing the dividend with a symbolic representation of which attacks could divide that skeleton (e.g., displayed by combinations of sword, gauntlets, and shield). Thus, exactly the same dividend and divisors were present in the extrinsic version, but the mathematical relationship was hidden because the numbers were no longer displayed. For example, the number 16 can be divided by 8, 4, and 2, but the symbol for a divisor of 8 (three swords) naturally includes symbols for a divisor of 4 (two swords) and a divisor of 2 (one sword) as well. This had the additional bonus of making the symbols require a level of logical interpretation, keeping the challenge of defeating skeletons at a more comparable level to that in the intrinsic version. It also meant that for dividends within the range of 1 to 99 divided by divisors in the range of 1 to 10, only the numbers 60 and 90 needed to be represented by more than three symbols.

In the extrinsic version, the mathematical content was reintroduced at the end of each level in the form of a multiple-choice quiz. This quiz required the player to divide the same dividends as found on the skeletons in the intrinsic game using exactly the same choice of divisors (weapons) that were available to defeat those skeletons (including "none of these" to be equivalent to leaving a skeleton). The extrinsic version therefore provided identical learning content delivered away from the flow-inducing gameplay and presented as abstract mathematical questions (see Figure 2). The control version simply omitted the end-of-level mathematics quiz. Comparisons of the different versions of the game are presented in Table 1 (stressing the gameplay) and Table 2 (highlighting the mathematical content).

### *Test Materials.*

Outcome test. The time-limited computer-based test consisted of 63 multiple-choice questions with four options in each case (one correct + three distractors). The first three questions were interface practice questions (e.g., "Select

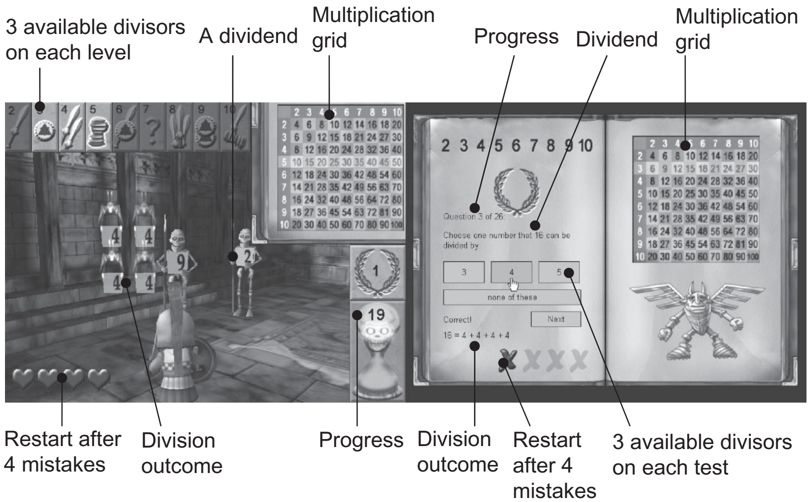


FIGURE 2 Explanations of mathematical learning content in the intrinsic (left) and extrinsic (right) versions of Zombie Division. Comparable features are labeled.

TABLE 1
Comparison of the Key Gameplay Differences Between the Intrinsic, Extrinsic, and Control Versions of Zombie Division

| *Feature* | *Intrinsic* | *Extrinsic* | *Control* |
|---|---|---|---|
| Skeleton | Numbers on chests | Weapon symbols on chests | As in the extrinsic |
| Attacks | Number relates to divisor and function key | Number relates to function key | As in the extrinsic |
| Successful attack | Skeleton splits into a proportional number of ghosts bearing the quotient | No ghosts | As in the extrinsic |
| New weapon tutorial | Introduced to the numbers that they divide | Introduced to the symbols that they divide | As in the extrinsic |

the number of legs that a dog has: 4, 5, 6, or 7”). Of the remaining 60 questions, 40 were division questions equally composed of two formats: (a) dividend based, in which the child was asked to select the divisor that divides a given dividend (e.g., “Select one number that 45 can be divided by: 4, 6, 9, or none of these”), and (b) divisor based, in which the task was to select the dividend that can be divided by a given divisor (e.g., “Select one number that can be divided by 5: 35, 13, 29, or 41”). There were five questions on each divisor from 2 to 10 (excluding the divisor 7, as it was not included as part of the games’ learning content). In

TABLE 2
Comparison of the Key Mathematical Features of the Intrinsic, Extrinsic, and Control Versions of Zombie Division

| *Feature* | *Intrinsic* | *Extrinsic* | *Control* |
|---|---|---|---|
| Division context | Dividing skeletons | Multiple-choice questions at end of level | None |
| Division control | Function keys | Controlled using mouse | None |
| Multiplication grid | Appears during gameplay | Appears with questions | None |
| Feedback on incorrect choice | Player is knocked back and loses health. Is sent to the restart position if health reaches zero. | Player is told the answer is incorrect and is asked to choose again. Is sent back to the start of the test if chances run out. | None |
| Feedback on correct choice | Splits into a proportional number of ghosts bearing the quotient | Players is told the answer is correct and the dividend, divisor, and quotient are displayed | None |

addition, five dividend-based questions were included for which the answer was "none of these." The remaining 15 questions were more conceptual in nature: 3 tested knowledge of the heuristic patterns associated with numbers that divide by 2, 5, and 10; and 12 tested an understanding of relationships between divisors (e.g., that all numbers that divide by 9 can also be divided by 3) or of applying rules outside of normal limits (i.e., dividends greater than 100).

The order of the questions was initially randomized but remained consistent between subjects and between tests. The software timed 15 min from the start of Question 4 (the end of the practice questions) and automatically stopped the test at the end of this time period. However, the time was not displayed on the screen, and no feedback was provided on the choices made. A multiplication grid similar to the one found in the game was provided in the corner of the screen.

Challenge level. Given the concerns raised about the potential for intrinsic integration to reduce transfer, we created "challenge" levels to allow us to compare questions in the abstract pre-/posttests and the same questions contextualized within the challenge levels of the game. Thus, the challenge level acted as a game-based test and consisted of two specially constructed levels of the game that directly replicated a portion of the outcome test's division problems within the game environment. All three groups played the gaming elements of these challenge levels with the learning content embedded (or omitted) as appropriate for the group's condition.

In the extrinsic version, these questions were asked in the normal way at the end of each level (e.g., "Select one number that can be divided by 5: 35, 13, 29, or 41").

In the intrinsic version, each question was posed within the context of a separate room within the challenge levels. The weapons (and therefore divisors) available to the player changed to match each question as the player entered its associated room. Divisor-based questions were posed in terms of offering the player a choice of three weapons with which to divide a single skeleton. An exit to the room was also included to provide an option equivalent to "none of these." A correct answer was recorded only if the skeleton was defeated with the correct attack on the player's first attempt. Dividend-based questions were posed in terms of a choice of four skeletons to divide with a single weapon, and, again, a correct answer was recorded only if the right skeleton was chosen on the player's first attempt. In addition, the gameplay demands were reduced within the challenge levels, as the dungeons were linear, the skeletons were immobile, and keys were provided when required.

**Teacher-led reflection.** Reflection sessions were included to help children reflect on the mathematical content of the game and to scaffold the conceptual process of making the link between solving division problems using multiplication facts. Children were taught in three separate groups according to their experimental condition. All of the teaching materials were created by the researchers and tailored to the context of each group's game, but they contained identical learning content (including the numerical examples) and followed the same structure. Each session lasted for half an hour and consisted of 15 min of direct instruction followed by 10 min of collaborative exercises. The instruction addressed three issues:

- *Division as sharing.* Children were shown a number of objects (bones or balls) and asked how they would work out whether they could be divided into two equal-size sets. A volunteer was then asked to come and draw circles around the sets. The class then confirmed this by counting the number of objects in each set.
- *Tables and rules.* The class was asked to suggest other techniques they could use to work out whether a number of objects could be divided equally into a number of sets. This continued until the class offered "using times tables" as a solution or the teacher eventually intervened with this suggestion. The class was also reminded of the numeric patterns for the 2, 5, and 10 times tables, if they had not already been discussed.
- *The multiplication grid.* The class was asked how they could solve division problems for times tables they did not know without counting objects. They were presented with the multiplication grid and shown how it can be used to answer division problems. They then worked through four example questions, checking whether a specific dividend could be divided by a specific divisor.

This was followed by 10 min of exercises carried out in pairs or groups of three. The exercises consisted of 12 divisor-based division problems with an option of three divisors to divide a given dividend. A multiplication grid was available. Worksheets also contained three blank questions for the children to create their own questions for their partners at the end. During this period the teacher provided individual support to any child who needed it.

*Procedure.* Figure 3 shows the schematic for the study. Children spent a total of 4 hr in the study, which was spread over 34 days.

Stage 1: Pretest. The pretests were carried out in three half-hour sessions 10 days before the main body of the study. Groups of up to 20 children were selected at random to complete the 15-min timed pretest in the ICT suite. The task was explained to children with the aid of a demo that emphasized the presence of the multiplication grid to help them with the test. They were informed of the 15-min time limit but told that they were not expected to finish all of the questions and encouraged not to treat it as a race. Each child was then allocated a PC and allowed to begin the test in his or her own time. Children who finished before the end of their time limit were asked to sit quietly until the entire group had finished.

Stage 2: The game. The children first played Zombie Division 10 days after the pretest. Each group (intrinsic, extrinsic, and control) played its version of the software without children from the other group present. Each playing session lasted for approximately 20 min, with a half hour turnaround on successive groups. The order of groups was rotated on each day of the study, with the first group beginning at 10:00 and the last group finishing at 11:30. Each child's position in the game was saved at the end of each playing session, and the game resumed from precisely the same point at the start of the next one. Each child played the game twice in this stage.

Stage 3: Teacher-led reflection. All of the reflective sessions were delivered by the same practicing teacher who taught older children within the school. None of the children had been formally taught by this teacher before, although some level of familiarity through everyday school life can be assumed. These reflections took place immediately preceding the children's third game session when all groups had played the game for an average of 40 min.

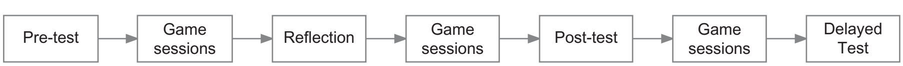


FIGURE 3 Schematic of the study.

All sessions were observed by the researcher, and there was no diversion from the teaching materials provided.

Stage 4: Further game sessions. The children played the game on two more days until they had accumulated a total of 100 min playing time. At this point, the software automatically stopped the game and the child was sent back to his or her class. A number of catch-up sessions were run for absentees to ensure that all children had played for their allotted time before taking the posttest.

Stage 5: Posttest. The posttests were carried out on the day after the children had completed 100 min of playing time with the game. Children were divided into three new groups containing an equal number of children from each condition. These groups were tested in three consecutive sessions in an identical way as for the pretest but with the addition of the challenge level. In order to prevent any distraction, children were not allowed to begin the game-based test until the entire group had finished their outcome tests.

Stage 6: Final game session. Two weeks after the posttest, children had a final opportunity to play their version of the game. This brought their total playing time up to exactly 135 min.

Stage 7: Delayed test. The delayed tests were carried out on the same day as the final playing session. The children were divided into mixed-condition groups and tested in three consecutive sessions as before. The challenge levels were taken in the same way, 2 days later.

## Results

Of the 54 children who completed all stages of the study, 3 were excluded from the analysis. One child was identified as having special educational needs in mathematics, and two demonstrated significantly better mathematical knowledge than the other children before the study started, as they scored 81.5% and 71.8% at pretest (3.6 and 2.9 *SD* above the mean, respectively). Analyses were conducted on the data from the remaining 51 children to explore the impact of the game condition on both the process and outcomes of learning.

*Learning outcomes.* Learning was measured by examining the percentage of correct answers that children gave on the tests (correct answers / total answers $\times$ 100). As the tests were timed, this measure was chosen to make sure that strategies that may have taken longer to perform but were more accurate as a result (such as using the multiplication grid) were not penalized.

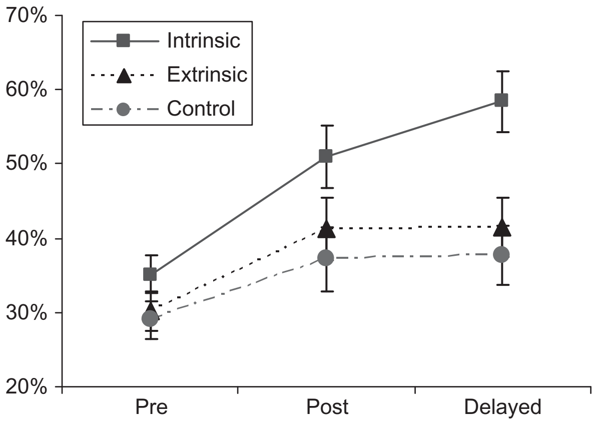


FIGURE 4 Mean percentage score by time and game condition.

A two-way mixed measures analysis of variance (ANOVA) with three levels of the within-subjects factor time (pre-, post-, and delayed test) and three levels of the between-subjects factor game (intrinsic, extrinsic, and control) was performed (see Figure 4). This revealed a significant main effect of time, $F(2, 96) = 34.86$, $MSE = 4006.41$, $p < .001$, ${}_{p}\eta^2 = 0.42$. Post hoc comparisons using the Bonferroni adjustment for multiple comparisons showed that students' posttest scores were higher than their pretest scores (11.69, $p < .001$), as were delayed test scores (14.46, $p < .001$), but there was no significant difference between posttest and delayed test scores.

Analysis also revealed a Time × Game interaction, $F(4, 96) = 5.86$, $MSE = 1025.794$, $p < .006$, ${}_{p}\eta^2 = 0.11$. Simple main effects analysis showed that all groups improved over the tests: intrinsic, $F(2, 47) = 24.89$, $p < .001$, ${}_{p}\eta^2 = 0.51$; extrinsic, $F(2, 47) = 6.78$, $p < .003$, ${}_{p}\eta^2 = 0.22$; and control, $F(2, 47) = 3.97$, $p < .025$, ${}_{p}\eta^2 = 0.15$. Multiple comparisons (Bonferroni corrected) showed that children who had played the intrinsic game scored significantly higher at posttest than they had at pretest ($M = 15.91$, $p < .001$) and gained still more from posttest to delayed test ($M = 7.04$, $p < .04$). The extrinsic group improved from pre- to posttest ($M = 10.97$, $p < .006$) but made no further improvement, whereas for the control group the delayed test score was significantly higher than their pretest score ($M = 8.78$, $p < .03$). The only test that showed any differences between the groups was the delayed test, $F(2, 48) = 7.49$, $p < .001$, ${}_{p}\eta^2 = 0.24$. Post hoc comparisons using the Bonferroni adjustment for multiple comparisons showed that children who played the intrinsic game scored significantly higher than children in either the extrinsic group ($M = 16.94$, $p < .14$) or the control group ($M = 20.66$, $p < .002$). In summary, children in all conditions learned from the experience (game plus teacher-led reflection), but children in the intrinsic condition learned the most.

This analysis was repeated with the additional between-group factor of gender. There were no significant differences between boys and girls, $F(1, 45) = 1.16$, and no interaction between gender and any other factor, including game condition.

**Time to answer questions.** As there were differences between conditions in accuracy of responses to tests, it is also informative to consider whether there were differences in the time taken to answer questions (see Table 3). This measure was examined using a 3 × 3 mixed ANOVA with three levels of the within-subjects factor time and three levels of the factor group (intrinsic, extrinsic, and control). This revealed a significant main effect of time, $F(2, 96) = 5.09$, $MSE = 145.3$, $p < .008$, ${}_{p}\eta^2 = 0.10$. Pairwise comparisons (Bonferroni corrected) showed that children took longer to answer the questions at the delayed test than at the pretest ($M = 2.95$, $p < .02$). A main effect of group, $F(2, 48) = 3.67$, $MSE = 423.6$, $p < .04$, ${}_{p}\eta^2 = 0.13$, found that children in the control condition took longer to answer questions than children in either of the other two conditions (intrinsic $M = 1.19$, $p < .06$; extrinsic $M = 2.95$, $p < .02$). However, overall children who took more time to answer questions were more accurate: pretest ($r = .35$, $p < .02$), posttest ($r = .5$, $p < .001$), and delayed test ($r = .41$, $p < .003$).

**Challenge levels.** The challenge levels provide a direct comparison between questions in the abstract pre-/posttests and the same questions contextualized within the context of the game as a means of exploring transfer. Twenty of the questions from the assessment were repeated (in appropriate format) in the challenge levels of the intrinsic and extrinsic games (there was no math content in the control condition). Children played these levels on two separate occasions: once following the posttest and once following the delayed test. Table 4 shows the mean percentage scores for the challenge levels and for the equivalent 20 items in the outcome tests (note that one child in the intrinsic condition missed the test).

Analysis by 2 (posttest, delayed test) × 2 (challenge, outcome) × 2 (intrinsic, extrinsic) ANOVA showed a main effect of time, $F(1, 31) = 32.3$, $MSE = 2598.5$,

TABLE 3
Time per Question in Seconds by Condition and Time

| *Time* | *Intrinsic (n = 17)* | *Extrinsic (n = 17)* | *Control (n = 17)* |
|---|---|---|---|
| Pretest | 16.8 (7.2) | 12.7 (5.3) | 14.1 (4.6) |
| Posttest | 18.0 (9.8) | 12.5 (5.6) | 13.0 (5.7) |
| Delayed test | 21.6 (10.8) | 15.6 (10.8) | 15.1 (4.7) |

TABLE 4
Outcomes Scores by Environment, Time, and Game Condition

| | *Intrinsic (n = 16)* | | *Extrinsic (n = 17)* | |
|---|---|---|---|---|
| *Time* | *Challenge* | *Outcome* | *Challenge* | *Outcome* |
| Posttest | 59.1% (23.0) | 62.1% (18.8) | 45.3% (16.2) | 47.9% (16.4) |
| Delayed test | 75.6% (16.2) | 57.9% (20.0) | 57.9% (20.8) | 49.0% (10.8) |

$p < .001$, ${}_{\mathrm{p}}\eta^2 = 0.51$, with children scoring higher on the delayed test; and a main effect of condition, $F(1, 31) = 6.41$, $MSE = 8400.9$, $p < .02$, ${}_{\mathrm{p}}\eta^2 = 0.17$, with children in the intrinsic condition scoring higher than children in the extrinsic condition. Overall, children's performance on the different types of test did not differ, but there was a Test × Time interaction, $F(1, 31) = 7.92$, $MSE = 1081.4$, $p < .008$, ${}_{\mathrm{p}}\eta^2 = 0.20$. There were no differences between the challenge and outcome test at the posttest, $F(1, 31) = 0.39$; but children scored higher on the challenge level at the delayed test, $F(1, 31) = 9.15$, $p < .005$. Children also significantly increased their scores from posttest to delayed test but only on the challenge level, $F(1, 31) = 32.3$, $p < .001$. There was no Test Type × Condition interaction, $F(1, 31) = 0.1$, so children in the extrinsic condition whose tests had similar formats to the quiz section of their game did not perform better than children in the intrinsic condition whose tests were dissimilar.

Game performance. The process logs produced by the game provide a valuable source of data in the form of a time-stamped commentary on the game as it is being played. For this study, more than 2,500 log files from the extrinsic and intrinsic condition were mined for the purposes of post hoc analysis.

Before performance on the intrinsic and extrinsic versions can be compared, it is necessary to operationalize measures of performance that are as equivalent as possible. Some variables are the same in both games (e.g., "level" refers to the amount of the game explored), but others vary depending upon the game version. A math task is always presented in the form of a skeleton in the intrinsic version and a quiz question in the extrinsic version. In both cases, a math task is a dividend-based question with a choice of up to three divisors to divide a given dividend. These divisors are provided in the form of different weapons in the intrinsic version and multiple-choice answers in the extrinsic version. The player also has the option of rejecting all of the divisors provided if none of them will divide the dividend. In the intrinsic version this involves maneuvering to avoid combat with the skeleton, whereas in the extrinsic version a player selects an alternative answer marked "none of these." Furthermore, for the purposes of this analysis a math task is assessed in terms of the first attempt made upon the dividend. This is said to begin when the player enters the same room as the skeleton and end if the player attacks the skeleton or leaves the room again without dividing it. This means that there can be one of five outcomes depending on whether the dividend is a target or distracter:

- Correct Outcomes
  1. Target Answered Correctly—the player correctly divides the dividend by one of the available divisors on the first attempt (extrinsic chooses the right number for the math quiz item, intrinsic chooses the correct weapon for the skeleton).
  2. Distracter Left Correctly—the player correctly rejects all of the available dividends for an indivisible dividend on the first attempt (extrinsic chooses "none of these," intrinsic does not fight the skeleton).
- Incorrect Outcomes
  3. Target Answered Incorrectly—the player incorrectly attempts to divide the dividend by one of the available divisors on the first attempt (extrinsic chooses the incorrect number for the math quiz item, intrinsic chooses the incorrect weapon for the skeleton).
  4. Target Left Incorrectly—the player rejects all of the available divisors for a dividable dividend on the first attempt (extrinsic chooses "none of these" when a correct response exists, intrinsic chooses not to fight a skeleton that could be defeated).
  5. Distracter Answered Incorrectly—the player incorrectly attempts to divide an indivisible dividend by one of the available divisors on the first attempt (extrinsic chooses a number for the math quiz item when the correct response is "none of these," intrinsic chooses to fight a skeleton that is not defeatable).

A total of 2.46% of the data were not able to be analyzed because of errors such as the software crashing, an error in data logging, or restarting the level.

One-way multivariate analysis of variance (MANOVA) on the variables shown in Table 5 revealed a multivariate effect of condition, $F(4, 29) = 7.12$, $p < .001$, ${}_p\eta^2 = 0.50$. Univariate analyses showed that children reached the same level regardless of condition, $F(1, 32) = 0.36$; and performed the same number of unique math tasks (quiz questions or skeleton dividing), $F(1, 32) = 1.00$. However, children in the extrinsic condition performed significantly fewer math tasks overall, $F(1, 32) = 15.07$, $MSE = 23796$, $p < .001$, ${}_p\eta^2 = 0.32$; and were more accurate, $F(1, 32) = 7.86$, $MSE = 683.65$, $p < .01$, ${}_p\eta^2 = 0.20$.

It might be expected that children's performance in the game would relate to both their initial mathematical understanding and what they learned from their experiences. Although children's pretest scores were significantly related to all

TABLE 5
Game Measures by Condition

| *Variable* | *Intrinsic (n = 17)* | *Extrinsic (n = 17)* |
|---|---|---|
| Level reached | 7.59 (3.59) | 6.88 (3.24) |
| Unique math tasks | 156.8 (84.5) | 120.9 (77.0) |
| Attempts at all math tasks | 382.2 (180.5) | 198.2 (75.0) |
| Accuracy | 70.1% (7.9) | 79.1% (10.5) |

TABLE 6
Correlations Between Indices of Game Performance, Pretest, and Learning Outcome

| *Variable* | *1* | *2* | *3* | *4* | *5* | *6* |
|---|---|---|---|---|---|---|
| 1. Level reached | — | .99*** | .49** | .41* | .43* | .18 |
| 2. Unique math tasks | | — | .55** | .38* | .44* | .20 |
| 3. All math tasks | | | — | −.23 | .37* | .25 |
| 4. Accuracy | | | | — | .06 | −.14 |
| 5. Pretest | | | | | — | −.01 |
| 6. Gain | | | | | | — |

* $p < .05$. ** $p < .01$. *** $p < .001$ (two-tailed).

game measures except accuracy, there was no relation between game performance and what children learned (see Table 6)

Analysis of the result of a math task (percentage of a particular result given the total number of unique math tasks) is displayed in Table 7. This analysis helps reveal which particular aspects of mathematics and/or gameplay children found difficult in the different conditions.

MANOVA on the variables shown in Table 7 showed a multivariate effect of condition, $F(5, 28) = 22.83$, $p < .001$, ${}_{\mathrm{p}}\eta^2 = 0.80$. Univariate analysis showed that the children in the extrinsic condition were significantly better at answering a

TABLE 7
Results of Math Task by Condition

| | *Intrinsic (n = 17)* | | *Extrinsic (n = 17)* | |
|---|---|---|---|---|
| *Result* | *M* | *SD* | *M* | *SD* |
| Target answered correctly | 56.85% | 7.13 | 66.65% | 10.98 |
| Distracter left correctly | 13.27% | 3.00 | 12.86% | 3.39 |
| Target answered incorrectly | 4.75% | 2.81 | 3.07% | 2.19 |
| Target left incorrectly | 20.56% | 6.40 | 7.78% | 5.92 |
| Distracter answered incorrectly | 4.24% | 2.48 | 5.70% | 2.95 |

target divisor question that had a valid dividend, $F(1, 32) = 9.53$, $MSE = 0.82$, $p < .004$, ${}_{p}\eta^2 = 0.24$; but not at correctly leaving a distracter question that did not have a valid dividend, $F(1, 32) = 0.57$. In terms of incorrect responses, children in the extrinsic condition were much less likely to leave a target incorrectly, $F(1, 32) = 36.57$, $MSE = 0.14$, $p < .001$, ${}_{p}\eta^2 = 0.53$.

## STUDY 2: THE IMPACT OF INTRINSIC INTEGRATION ON CHOICE OF GAME

Study 1 showed that the intrinsic version of Zombie Division produced greater learning gains than either an extrinsic version or a control version without learning content. This study was conducted under a strict in vivo experimental regime whereby time on task was completely controlled. However, although educational games are used in schools, they are also increasingly available as leisure activities. Thus, it is important not just that intrinsic games are effective but that children choose to use them. It could be that the central benefit of intrinsic integration is to produce greater time on task. Consequently, Study 2 compared the time children chose to spend playing the different versions of the game when they were provided with a free choice of which game to play.

There is a long history of measuring motivation as a function of time on task, including in some of the early seminal research on intrinsic motivation (Daniel & Esser, 1980; Deci, 1971). Furthermore, task persistence is considered one of the key mechanisms by which motivation can influence learning (e.g., Dweck, 1986; Vollmeyer & Rheinberg, 2000). For example, Vollmeyer and Rheinberg (2000) found that learners with high motivation showed increased task persistence and that when this was combined with initially lower levels of knowledge, it resulted in increased learning. Therefore, time on task was chosen as the primary dependant variable with which to compare the relative motivational appeal of intrinsic and extrinsic approaches for Zombie Division.

### Method

*Participants.* All children attended the same primary school as the students in Study 1 but were taken from a different (older) year group. The 5 girls and 11 boys were between the ages of 9 years, 10 months, and 11 years, 2 months ($M =$ 10 years, 4 months). All participants had prior experience using the computers in the school's ICT suite and were members of an after-school computer club.

*Design.* This study used a single repeated measure design, game, with two levels (intrinsic or extrinsic) that reflected the amount of time children played each game.

#### *Materials.*

**Facilities.** The intervention was carried out within the ICT suite at the school using the normal facilities used for the after-school club. The suite contained 20 relatively new PCs running Windows 2000 with accelerated 3D graphics support and audio output through stereo headphones. Children could switch between different versions of the game using a menu that appeared each time the game was launched. This allowed children to choose between the intrinsic and extrinsic versions of the game and provided them with a visual reminder of the differences between the two versions. The order that the options appeared in the menu was randomized each time so that either version would appear on the left or right with an equal probability. Quitting the game would return the player to this menu, where he or she could switch versions again. When switching versions the player's exact position was resumed with intrinsic skeletons becoming extrinsic or vice versa. In this way, it was ensured that switching between versions provided neither a gameplay penalty nor an advantage, so children could not use it as a way to "game the system" (e.g., Baker, Corbett, Koedinger, & Wagner, 2004).

#### *Procedure.*

**Introduction.** Children were introduced to Zombie Division as a group through a demonstration of the two different versions running side by side on two separate PCs. Both games were saved at identical positions within the same game level so that the differences between the intrinsic and extrinsic versions were apparent. Children were shown how combat worked in both versions, with an emphasis on the mathematical content of the intrinsic version, alongside the quiz that appeared at the end of each level in the extrinsic version. They were introduced to the "intrinsic" and "extrinsic" terminology and shown how the version switching menu worked. Children were told that their game position would not be lost by their switching versions and that they were expected to try both versions. This introduction took 10 min.

**Game intervention.** All of the children played Zombie Division on their own PCs for the remainder of the first club session. In subsequent sessions each child could choose to continue playing the game or return to his or her normal club pursuits (and freely switch between the two). Each club session lasted for approximately 1 hr, with around 45 to 50 min of playing time. This continued for two more club sessions after the first, providing a maximum of around 135 min ($2\frac{1}{4}$ hr) playing time for each child. The children's positions in the game were saved at the end of each playing session, and the game resumed from precisely the same point at the start of the next session. The children were reminded several times over the course of the sessions that they were expected to try playing both versions of the game.

Group interview. In the fourth and final club session the children took part in a group interview with the two different versions running side by side. They were asked to summarize the differences between the two versions and state which they preferred. Each child was given the opportunity to explain why he or she preferred the option he or she did, and the group was encouraged to discuss which version was the most fun to play and which was the most educational.

### Results

The mean number of minutes children spent playing the two different versions of the game was analyzed using a paired sample $t$ test. Participants spent more than 7 times longer playing the intrinsic version at 75.7 min ($SD = 35.5$) compared to the extrinsic version at 10.28 min ($SD = 10.28$), $t(15) = 7.38$, $p < .001$, $r = .89$. Analysis of the influence of gender showed that overall girls ($Mdn = 114$) chose to spend longer playing Zombie Division than boys ($Mdn = 73$; $U = 10$, $p < .052$, $r = .5$), however they did not differ in the proportion of time they spent on the intrinsic version ($U = 23$). Table 8 shows the results of the group interview concerning children's responses to different versions of the game. Children not only had more positive perceptions of the intrinsic version of the game but they also had quite a sophisticated appreciation of the mechanisms of intrinsic and extrinsic integration. These data suggest that children's decision to spend more time playing the intrinsic version was clearly deliberate and resulted from their positive perceptions of the intrinsic game mechanics.

## DISCUSSION

### Learning Outcomes

The results of these studies provide a strong argument in favor of the intrinsic integration of a game with its learning content (Kafai, 1996; Malone, 1980; Rieber, 1996) in contrast to extrinsic environments, which provide a separate extrinsic motivation or reward for completing learning content. At the beginning of Study 1, children had limited familiarity with the mathematics addressed in Zombie Division, as they scored an average of 31% at pretest. Over the course of the study children in all conditions improved their understanding to end with an average delayed test score of 46%. However, although all children learned, children in the intrinsic condition improved the most. At the delayed test, these children (with a mean score of 58%) significantly outperformed children in both the control (38%)

TABLE 8
Children's Responses to the Different Game Versions

| *Response* | *Intrinsic* | *Extrinsic* |
|---|---|---|
| Likes | "It's not as hard—it's quick and easy." | "It can help you like learning your times tables and doing your SATs." |
| | "It's easier to learn division [. . .] instead of having to figure out what the symbols are you just have to figure out what to divide by." | |
| | "It's easier [. . .] because you get to learn division." | |
| | "It's better to learn doing it by intrinsic, because it's quicker." | |
| | "It's easier to learn your times tables." | |
| | "It's fun." | |
| | "You don't have to do a test at the end." | |
| | "More fun because it's like subliminal advertising with maths." | |
| | "It's like mixing paint [. . .] the maths in the game with the fun [. . .] you don't really think you're doing that much." | |
| Dislikes | "It's not faster because on the beginning of every level there's [. . .] a help thing." | "Because you need to do all the maths at the end and that's what you might lose interest in." |
| | [Teachers would think it's] "too much fun—and hasn't got a test." | "The version with the math test at the end wasted time [. . .] you don't get as far as you do in the other one." |
| | | "It just tells you what to use" [fighting skeletons]. |
| | | "It's not a challenge" [fighting skeletons]. |
| | | "The maths test at the end was just got boring." |
| | | "Very slow and boring." |
| | | "You think: oh I've had the fun part, now I have to do a test—I'm just going to turn it off and not bother." |

and extrinsic (41%) conditions (see Figure 4). This improvement was the same for both boys (12.84%) and girls (16.25%).

The intrinsic result might also initially seem surprising given that the children's performance during the game had not favored the intrinsic condition. There was no difference between conditions in the level reached or the number of unique math tasks, but children in the extrinsic condition were more accurate (the measure most closely related to the tests). We must acknowledge that this may be partly an artifact of the way that accuracy was operationalized. In this extrinsic condition, a child had to make a conscious decision to choose "none of these" when a correct response existed, but it was impossible to know whether a child in the intrinsic condition chose not to fight a skeleton or just exited the room without seeing it. Thus, our data mining used a “best guess” heuristic that was likely to overestimate the number of conscious rejections of skeletons. According to this heuristic, around 21% of the responses children made in intrinsic condition were of this form compared to only 8% in the extrinsic condition.

However, this overestimation of rejections is unlikely to be the sole explanation for the difference between the intrinsic and extrinsic conditions. When performing math tasks in the extrinsic condition, children only need to manage the demands of the division problems. In contrast, during the intrinsic game they must navigate dungeons, seek out keys, and respond to increasingly assertive skeletal attacks as the levels advance. It is therefore unsurprising that children in the extrinsic condition (67%) were more accurate at answering a target question than children in the intrinsic condition (57%). Arguably, children in the intrinsic condition may even have had better mathematical skills within the game as well but that these were being masked by the difficulties of the gaming aspects of Zombie Division. Evidence for this proposal can be seen from the challenge level data. In these levels, the mathematics remains the same but the gameplay demands are significantly reduced in the intrinsic condition (skeletons do not attack, there is less need to navigate, etc.). Children in the intrinsic condition scored higher than those in the extrinsic condition on these levels at both posttest and delayed test (scoring an impressive 76% in the delayed challenge levels).

One of the reservations about the use of intrinsic games is that they encourage the development of overspecialized knowledge that does not transfer to everyday school mathematics, but we found no evidence to support this concern. The learning outcomes tests presented the math problems in an abstract quiz form that was much closer to the form of the extrinsic condition, and yet results still favored the intrinsic condition. We also developed challenge levels within the game to directly test this concern, as they contained a subset of the outcome tests with 20 questions presented in appropriate game format (skeleton or quiz). Consequently, a significant decrease in performance between the challenge level and the test would have indicated that children were failing to transfer their mathematical skills from the computer game to math tasks more generally. Overall, children performed equally well on the test and the challenge but at delayed test they scored significantly better on the challenge level than on the test itself. Thus, there is some evidence that children were more engaged with the math content in the game than they were when the same content was presented as a test. However, this difference between challenge level and test performance was the same in both conditions, so there is no evidence to suggest that learning content transfers less effectively from the intrinsic version of the game (in which practice and challenge tasks were on skeletons and test tasks were on quiz items) than the extrinsic (in which all tasks were in quiz format).

Another concern might be that intrinsic games, with their increased time pressure, affect, and arousal, might encourage children to respond too rapidly to questions, thereby promoting speed at the expense of accuracy. This had been a concern in the design of the game and had led to the introduction of "slow" levels with passive skeletons as well as penalties for guessing (health reduction) to encourage mathematical thought and reflection rather than stabbing at the keyboard! We are happy to report that nothing in these data supports the worry that children had learned to guess rather than work out the answers. Children took longer to answers questions on the delayed test than they did at the pretest, and children who took longer to respond were more accurate. Again, there were no differences between the intrinsic and extrinsic conditions on this measure, suggesting that intrinsic players who had been under more pressure to respond faster during learning had not come to rely on a guessing strategy.

*Why did intrinsic integration foster learning?.* We argued that intrinsic integration could support learning by ensuring that the activities that give rise to the central flow experience of the game involve appropriate external representations that players interact with using core gameplay mechanisms. Thus, our definition integrates motivational and cognitive processes to explain how intrinsic games can help learning.

Researchers have proposed a number of mechanisms by which motivation and flow could enhance learning, including persistence, more focused attention, increased arousal, increased affect, and alternative strategies (Dweck, 1986; Garris et al., 2002; Martens et al., 2004; Parkin et al., 1982; Pintrich, 2003; Vollmeyer & Rheinberg, 2000). Study 2 was a test of the assertion that increased motivational appeal should lead to increased task persistence, and the results strongly support this claim. However, task persistence cannot provide an explanation for the increased learning outcomes of Study 1, as time on task was strictly controlled in this study. Consequently, other explanations are required to explain increased learning.

Zombie Division was created with the explicit intention of increasing learners' attention, arousal, and affect—all components of the flow experience. However, most of the research on flow does not describe how flow enhances learning (beyond increasing persistence). Therefore, all we can do is postulate some possible mechanisms by which this may have increased learning outcomes for the intrinsic game. It is well established that direct attention during encoding of material enhances its recall (e.g., Baddeley, Lewis, Eldridge, & Thomson, 1984; Murdock, 1965). Arousal has a more complicated relationship to performance and learning, albeit one that has been understood for a considerable time. Yerkes and Dodson (1908) famously described a U-shaped curve whereby arousal increases performance up to a level before decreasing again. One can therefore speculate that optimizing tasks so that they are challenging but achievable (as in the way that flow theory predicts) should aim to keep learners in an optimal state of arousal. The intrinsic gameplay naturally provides players with some level of control over their own state of arousal during the mathematics, as they can decide how quickly to seek out and engage with skeletons. Finally, compared to learning mathematics in the extrinsic condition, learning in the intrinsic condition is more emotionally charged (although presumably learners' emotions would change as they win or lose to skeletons). Research on the relationship between affect and learning is still relatively young (e.g., Pintrich, 2003), but there is reason to believe, for example, that increased affect during encoding (e.g., gameplay) can enhance retrieval at a delay (test; e.g., Parkin et al., 1982).

The intrinsic version of Zombie Division may also have encouraged children to use better strategies to learn the mathematics. Partly this was designed into the structure of the game, as new game mechanics were introduced as the levels advanced to encourage exploration of different strategies (e.g., parrying and "giant" skeletons). However, few children in the studies we conducted actually progressed to the levels inhabited by giants. Nevertheless, it could be that children in the intrinsic condition were spontaneously using more effective strategies to combat the skeletons (and hence solve the math problems). Evidence for this comes indirectly from other research that has shown that games that include more fantasy can encourage children to use more complex mathematical operations (Cordova & Lepper, 1996) and can encourage the use of more systematic strategies (Vollmeyer & Rheinberg, 2006) and more exploratory behavior (Martens et al., 2004). Unfortunately, the design of the present studies did not allow us to explore this directly, but we can report our observation of children pausing the intrinsic game to work out their approaches to skeletons lurking in the next room. This also serves to highlight the fact that although fantasy may not be the correct focus for intrinsic integration, it should not detract from its value within the overall game concept.

In contrast to the motivational aspects of intrinsic integration, which had predicted almost uniformly that it should lead to better learning, the review of the more cognitively oriented literature concerning the role of interacting with concrete fast-paced external representations revealed a more mixed picture. For example, there were concerns that making the mathematical symbols more concrete in the intrinsic game would lead to decreased learning and transfer (e.g., DeLoache, 1991; Goldstone & Son, 2005; Uttal et al., 2009). Moreover, intrinsic integration breaks may of the "rules" for using concrete representations (e.g., Brown, McNeil, & Glenberg, 2009). It encourages learners to play and interact with representations (Uttal et al., 2009), potentially making it harder for learners to see them as representations rather than objects of interest in their own right (e.g., DeLoache, 1991) and certainly giving significance to features of the environment that are not relevant for learning (zombie skeletons!). However, the representations in Zombie Division are different from the concrete representations that have typically been researched. First, children do not have extensive experience with zombie skeletons in everyday life, which might encourage them to ground their understanding in inappropriate ways. Second, the skeletons are a concrete context for presenting abstract symbols rather than an alternative representational system such Cuisenaire or Dienes blocks. Moreover, because children must engage mathematically with the skeletons in order to solve the problems there is no sense in which the environment is "doing the work" for them (e.g., Martin, 2009).

*Why did children prefer intrinsically integrated games?.* The children in Study 2 demonstrated an overwhelming difference in preference for the intrinsic and extrinsic versions of Zombie Division. They spent on average more than 7 times longer playing the intrinsic version of the game than the extrinsic. This provides clear support for the hypothesis that intrinsic integration increases motivation. There were only a small number of girls in this study (5 girls compared to 11 boys), probably because of its setting within a computer club. However, analysis of any gender differences suggested that although the girls played Zombie Division more than the boys, there were no gender differences in preference for the intrinsic or extrinsic version.

The interview data (see Table 8) reveal why the children preferred playing Zombie Division. They tended to see the intrinsic game as easier and quicker. Not surprisingly, it was also seen as more enjoyable. Only two explanations were provided for disliking the intrinsic version—one concerning the enforced in-game tutorial and one because the participant saw the game as not fitting into the school context, speculating that teachers would not approve of it. The children's perceptions of school requirements were echoed in the only justification provided for liking the extrinsic version when a child commented it would help them on tests. Children were correspondingly able to explain their dislike of the extrinsic version in many ways, seeing it as slower and less fun. But perhaps surprisingly they also saw it as too easy in terms of both the math and the game content. Thus, ease was seen as a positive attribute in the intrinsic version and a negative attribute in the extrinsic version, providing insight into the subtle and important nature of perceived challenge in educational games (e.g., Malone, 1981). It is intriguing that two children articulated principles of intrinsic integration, with one stating that it is "more fun because it's like subliminal advertising with maths" and another stating that "it's like mixing paint [. . .] the maths in the game with the fun [. . .] you don't really think you're doing that much."

*Is game-based learning for everyone?.* One concern that might arise when considering the use of intrinsically integrated games in classrooms is whether in using such games a particular subset of the population, for example non-gamers or girls, may be disadvantaged. The results of these studies do not find that Zombie Division disadvantaged these groups. In Study 1, girls improved their scores by an average of 16% compared to 13% for the boys, a nonsignificant difference. Just like the boys, they learned more from the intrinsic game than the extrinsic game (26% in intrinsic and 14% in extrinsic for girls compared to 20% and 9%, respectively, for boys). There were also no differences in in-game performance: On average girls progressed to Level 7 and boys to Level 8, and their in-game accuracy was 74% and 75%, respectively. Study 2 included only a few girls and was conducted in an after-school computer club, which are not ideal conditions for exploring gender differences. However, the girls spent considerably longer playing the intrinsic game than the boys did (as the boys spent less time playing Zombie Division and more time on other club activities). Consequently, there is no evidence that the central game mechanic—attacking skeletons—caused the girls in these studies any anxiety. However, it should be remembered that Zombie Division had been iteratively developed with both girls and boys, and some issues that had been observed to have potential concerns for girls were addressed (e.g., boys were found to more easily understand the parrying mechanic, and so more explanation was provided in the final version). Contrary to media concerns but in line with academic discussion and research (e.g., Kafai, 2008), it seems that there is no simple relationship between gender and games and that an ideal for developing games should involve the early participation of both boys and girls as developers to create intrinsic games with a wide range of core mechanics.

We also have no evidence that children's gaming skills influenced what they could learn from Zombie Division. Unfortunately, we do not have demographic data on children's use of digital technologies and games outside the school that would have allowed us to test this relationship explicitly. However, we can look at whether their performance in the game influenced what they learned. First, we find that prior mathematical knowledge (pretest scores) predicted game performance (e.g., accuracy in encounters, level reached), suggesting that math knowledge is important in progressing through the game. Second, no measure of game performance (or pretest score) predicted learning gains, suggesting that children with all levels of mathematical and gameplay skills can learn successfully from Zombie Division. Again, one reason for this successful result may be the iterative development of the prototype, whereby lack of gameplay skills that had been observed to cause children problems (such as navigation issues and problems withdrawing from skeletons) were ameliorated by changes to the design.

## Classroom Implications

Although Study 1 shows that children who played the intrinsic version of the game learned the most, children in the control group also made significant improvements during the course of the study (and indeed were not reliably different from children in the extrinsic condition). This illustrates the power of debriefing in combination with the motivational appeal of games. We had expected some improvement in the control group as a result of the teacher-led reflection session but are surprised by the degree of improvement, as this lesson had no relevance to their version of the game and so was not reinforced in any way before the posttests the following week. However, the teacher running these sessions reported that all three groups were unusually enthusiastic and attentive to her lesson—and she attributed this to the children's excitement about their involvement with the game. So it appears that the children's motivation for the game may have transferred to their learning in the classroom context as well.

Therefore, this study leaves the strong impression that that there is also significant potential for using games like Zombie Division as motivational anchors (Bransford et al., 1990) for classroom learning. There is certainly potential for creating a whole range of supporting materials based around the content and characters in Zombie Division. Furthermore, we believe that the intrinsic nature of the game naturally lends itself to the creation of intrinsic supporting materials that go beyond simply including visual images from the game. It is easy to conceive other characters that would add, subtract, or multiply the values of skeletons as well as a whole range of different mathematically based foes—all of which could be cheaply and easily included in paper-based classroom resources.

So in line with other research into game-based learning (e.g., Squire, 2004), our experiences with Zombie Division seem to support the idea that teachers have a critical role to play in maximizing the educational potential of intrinsic games. Although this is not something that our research has explicitly shown (as we controlled the teacher's contribution so as not to bias the results), we do not believe that games should—or could—replace traditional methods of education but should simply form another part of the toolkit available to teachers in creating engaging and effective learning experiences for their pupils.

## Design Implications

Given that the outcomes of intrinsic integration are desirable, both our research and practical experience would suggest that there is a logical hierarchy to designing an intrinsically integrated game. This prioritizes the learning content, followed by the game mechanics, and then finally the fantasy context (in line with our theorizing). Fertile learning content for creating intrinsically integrated games includes concepts that can exist and interact within a common world rather than separate unrelated content. These links can then be used to create layers of game mechanics that interact and create emergent gameplay strategies that reinforce the learning goals. Subordinate to this, the fantasy can then be worked around the game mechanics to bring them together into a coherent and motivating context. This approach is not an attempt to detract from the considerable motivational relevance of fantasy contexts in game design (e.g., Cordova & Lepper, 1996) but an acknowledgement of the primary role of core mechanics in creating an intrinsic relationship between games and their learning content. Furthermore, we would suggest that designers of educational games give equal consideration to the offline resources available to parents and teachers in order to support learning with their game, as our own findings add to the growing research that games on their own may be unable to offer a complete learning experience.

However, there are also practical and economic factors that present commercial barriers to the application of this research. Intrinsic games may be both more motivating and more effective than their extrinsic equivalents, but they are also more difficult and more expensive to develop, which makes it harder to justify a business case. The very nature of extrinsic games means that they are more separate from their learning content and so can be reapplied more cost effectively to new educational purposes. Intrinsic games, in contrast, are far more difficult to apply to new learning content and must be largely redeveloped from scratch in order to address different learning goals. Unfortunately, this is an issue that designers will have to wrestle with until the market can demonstrate a financial advantage to creating intrinsic games in addition to any motivational and learning benefits.

## CONCLUSION

Our research acknowledges the motivational significance of fantasy in games but argues that it is core mechanics—rather than fantasy—that is critical to creating an intrinsic relationship with the learning content of a game. We have explored the value that this definition of intrinsic integration can bring to educational games and have found benefits in terms of both motivation and learning outcomes. There is clearly much more work that could be done to tease apart the components of this intrinsic relationship or to explore the best way of using intrinsic games within a classroom context. Nonetheless, we believe that this work goes some way toward establishing the value of and relevance of this issue as worthy of future investigation within the field of game-based learning.

## ACKNOWLEDGMENTS

We wish to acknowledge and thank Jenny Habgood, Louise Ash, Lucy Button, and Lizzy Evans (Reignhead Primary School); Elaine Cockburn (Gilt Hill Primary School); Sarah Peacock (Sheffield E-Learning Centre West); Vincent Aleven (Carnegie Mellon University); and Steve Benford (University of Nottingham) for their help and support in running these studies. Our thanks are also extended to the many children who collaborated with us on the design and evaluation of Zombie

Division. This research was partially supported by a PhD studentship awarded to M. P. Jacob Habgood from the Learning Sciences Research Institute.

## REFERENCES

Ainsworth, S. (2006). DeFT: A conceptual framework for considering learning with multiple representations. *Learning and Instruction, 16*(3), 183–198.

Aldrich, C. (2004). *Simulations and the future of learning*. San Francisco, CA: Wiley.

Asgari, M., & Kaufman, D. (2004, July). *Relationships among computer games, fantasy, and learning*. Paper presented at the International Conference on Imagination and Education, Vancouver, British Columbia, Canada.

Baddeley, A. D., Lewis, V., Eldridge, M., & Thomson, N. (1984). Attention and retrieval from long-term memory. *Journal of Experimental Psychology: General, 113*, 518–540.

Baker, R. S., Corbett, A. T., Koedinger, K. R., & Wagner, A. Z. (2004). Off-task behavior in the cognitive tutor classroom: When students "game the system." *Proceedings of ACM CHI 2004: Computer-Human Interaction*, 383–390.

Berry, D. C. (1983). Metacognitive experience and transfer of logical reasoning. *Quarterly Journal of Experimental Psychology Section A, 35*(1), 39–49.

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

Bransford, J. D., Sherwood, R. D., Hasselbring, T. S., Kinzer, C. K., & Williams, S. M. (1990). Anchored instruction: Why we need it and how technology can help. In D. Nix & R. Spiro (Eds.), *Cognition, education and multimedia* (pp. 115–141). Hillsdale, NJ: Erlbaum.

Brown, M. C., McNeil, N. M., & Glenberg, A. M. (2009). Using concreteness in education: Real problems, potential solutions. *Child Development Perspectives, 3*(3), 160–164.

Bruckman, A. (1999, March). *Can educational be fun?* Paper presented at the Game Developers Conference '99, San Jose, CA.

Caillois, R. (1961). *Man, play and games* (B. Mayer, Trans.). New York, NY: Free Press.

Cordova, D. I., & Lepper, M. R. (1996). Intrinsic motivation and the process of learning: Beneficial effects of contextualization, personalization, and choice. *Journal of Educational Psychology, 88*, 715–730.

Crawford, C. (1982). *The art of computer game design*. Retrieved from http://www.erasmatazz.com/free/AoCGD.pdf

Csikszentmihalyi, M. (1975). Play and intrinsic rewards. *Journal of Humanistic Psychology, 15*(3), 41–63.

Csikszentmihalyi, M. (1988). The flow experience and human psychology. In M. Csikszentmihalyi & I. S. Csikszentmihalyi (Eds.), *Optimal experience* (pp. 15–35). Cambridge, England: Cambridge University Press.

Cullingford, G., Mawdesley, M. J., & Davies, P. (1979). Some experiences with computer based games in civil engineering teaching. *Computers and Education, 3*, 159–164.

Daniel, T. L., & Esser, J. K. (1980). Intrinsic motivation as influenced by rewards, task interest, and task structure. *Journal of Applied Psychology, 65*, 566–573.

de Jong, T., & van Joolingen, W. R. (1998). Scientific discovery learning with computer simulations of conceptual domains. *Review of Educational Research, 68*(2), 179–201.

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

Deci, E. L. (1975). *Intrinsic motivation*. New York, NY: Plenum Press.

DeLoache, J. S. (1991). Symbolic functioning in very young children: Understanding of pictures and models. *Child Development, 62*, 736–752.

Dondlinger, M. J. (2007). Educational video game design: A review of the literature. *Journal of Applied Educational Technology, 4*(1), 21–31.

Dweck, C. S. (1986). Motivational processes affecting learning. *American Psychologist, 41*, 1040–1048.

Edwards, L. D. (1998). Embodying mathematics and science: Microworlds as representations. *Journal of Mathematical Behaviour, 17*(1), 53–78.

Engeser, S., & Rheinberg, F. (2008). Flow, performance and moderators of challenge-skill balance. *Motivation and Emotion, 32*(3), 158–172.

Ferster, C. B., & Skinner, B. F. (1957). Schedules of reinforcement. *Psychological Reports, 3*, 695.

Garris, R., Ahlers, R., & Driskell, J. E. (2002). Games, motivation, and learning: A research and practice model. *Simulation & Gaming, 33*, 441–467.

Gee, J. P. (2003). *What video games have to teach us about learning and literacy*. New York, NY: Palgrave Macmillan.

Gentner, D. (1989). The mechanisms of analogical learning. In S. Vosniadou & A. Ortony (Eds.), *Similarity, analogy, and thought* (pp. 199–241). New York, NY: Cambridge University Press.

Goldstone, R. L., & Son, J. Y. (2005). The transfer of scientific principles using concrete and idealized simulations. *Journal of the Learning Sciences, 14*, 69–110.

Gunter, G. A., Kenny, R. F., & Vick, E. H. (2008). Taking educational games seriously: Using the RETAIN model to design endogenous fantasy into standalone educational games. *Education Tech Research Development, 56*, 511–537.

Habgood, M. P. J., Ainsworth, S., & Benford, S. (2005a). Endogenous fantasy and learning in digital games. *Simulation & Gaming, 36*, 483–498.

Habgood, M. P. J., Ainsworth, S., & Benford, S. (2005b, July). *Intrinsic fantasy: Motivation and affect in educational games made by children*. Paper presented at the 2005 AIED workshop on Motivation and Affect in Educational Software, Amsterdam, The Netherlands.

Habgood, M. P. J., & Overmars, M. (2006). *The game maker's apprentice: Game development for beginners*. Berkeley, CA: APress.

Hogle, J. G. (1996). *Considering games as cognitive tools: In search of effective "edutainment."* Retrieved from http://www.twinpinefarm.com/pdfs/games.pdf

Huizinga, J. (1950). *Homo Ludens: A study of the play element in culture* (Unknown Trans.). Boston, MA: Beacon Press.

Juul, J. (2005). *Half real: Video games between real rules and fictional worlds*. Cambridge, MA: MIT Press.

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 & M. Resnick (Eds.), *Constructionism in practice: Designing, thinking and learning in a digital world* (pp. 71–96). Mahwah, NJ: Erlbaum.

Kafai, Y. B. (2008, June). *Considering gender in digital games: Implications for serious game designs in the learning sciences*. Paper presented at ICLS 2008, Utrecht, The Netherlands.

Kaminski, J. A., Sloutsky, V. M., & Heckler, A. F. (2008, April 25). The advantage of abstract examples in learning math. *Science, 320*, 454–455.

Kerawalla, L., & Crook, C. (2005). From promises to practices: The fate of educational software in the home. *Technology, Pedagogy and Education, 14*(1), 107–125.

Koster, R. (2005). *A theory of fun for game design*. Scottsdale, AZ: Paraglyph Press.

Lederman, L. C. (1992). Debriefing: Towards a systematic assessment of theory and practice. *Simulation & Gaming, 23*(3), 145–160.

Lepper, M. R., & Greene, D. (1975). Turning play into work: Effects of adult surveillance and extrinsic rewards on children's intrinsic motivation. *Personality and Social Psychology, 31*, 479–486.

Lepper, M. R., & Malone, T. W. (1987). Intrinsic motivation and instructional effectiveness in computer-based education. In R. E. Snow & M. J. Farr (Eds.), *Aptitude, learning and instruction: III. Conative and affective process analyses* (pp. 255–286). Hillsdale, NJ: Erlbaum.

Loftus, G. R., & Loftus, E. F. (1983). *Mind at play: The psychology of video games*. New York, NY: Basic Books.

Lundgren, S., & Björk, S. (2003, March). *Game mechanics: Describing computer-augmented games in terms of interaction*. Paper presented at the 2003 Technologies for Interactive Digital Storytelling and Entertainment conference, Darmstadt, Germany.

Malone, T. W. (1980, September). *What makes things fun to learn? Heuristics for designing instructional computer games*. Paper presented at the Association for Computing Machinery Symposium on Small and Personal Computer Systems, Palo Alto, CA.

Malone, T. W. (1981). Toward a theory of intrinsically motivating instruction. *Cognitive Science*, *5*, 333–369.

Malone, T. W., & Lepper, M. R. (1987). Making learning fun: A taxonomy of intrinsic motivations for learning. In R. E. Snow & M. J. Farr (Eds.), *Aptitude, learning and instruction: III. Conative and affective process analyses* (pp. 223–253). Hillsdale, NJ: Erlbaum.

Martens, R. L., Gulikers, J., & Bastiaens, T. (2004). The impact of intrinsic motivation on e-learning in authentic computer tasks. *Journal of Computer Assisted Learning, 20*, 368–376.

Martin, T. (2009). A theory of physically distributed learning: How external environments and internal states interact in mathematics learning. *Child Development Perspectives, 3*(3), 140–144.

Murdock, B. B. (1965). Effects of a subsidiary task on short-term memory. *British Journal of Psychology, 56*, 413–419.

Nunes, T., Schliemann, A.-L., & Caraher, D. (1993). *Street mathematics and school mathematics*. New York, NY: Cambridge University Press.

Papert, S. (1980). *Mindstorms: Children, computers and powerful Ideas*. New York, NY: Basic Books.

Papert, S. (1998, June). Does easy do it? Children, games and learning. *Game Developer*, pp. 87–88.

Paras, B. (2005, June). *Game, motivation, and effective learning: An integrated model for educational game design*. Paper presented at the Proceedings of DiGRA 2005 Conference: Changing Views—Worlds in Play, Vancouver, British Columbia, Canada.

Parkin, A. J., Lewinsohn, J., & Folkard, S. (1982). The influence of emotion on immediate and delayed retention: Levinger & Clark reconsidered. *British Journal of Psychology, 73*, 389–393.

Pintrich, P. R. (2003). A motivational science perspective on the role of student motivation in learning and teaching contexts. *Journal of Educational Psychology, 95*, 667–686.

Qualifications and Curriculum Authority. (1999). *The national curriculum for England: Mathematics, key stages 1–4*. Retrieved from http://curriculum.qcda.gov.uk/uploads/Mathematics%201999%20programme%20of%20study_tcm8-12059.pdf

Rieber, L. P. (1991). Animation, incidental learning and continuing motivation. *Journal of Educational Psychology, 83*, 318–328.

Rieber, L. P. (1996). Seriously considering play: Designing interactive learning environments based on the blending of microworlds, simulations, and games. *Educational Technology Research and Development, 44*(2), 43–58.

Salen, K., & Zimmerman, E. (2004). *Rules of play: Game design fundamentals*. Cambridge, MA: MIT Press.

Sandford, R., & Williamson, B. (2005). *Handbook on games and learning*. Bristol, England: Futurelab.

Scaife, M., & Rogers, Y. (1996). External cognition: How do graphical representations work? *International Journal of Human-Computer Studies, 45*(2), 185–213.

Shaffer, D. W. (2004). Pedagogical praxis: The professions as models for postindustrial education. *Teachers College Record, 106*, 1401–1421.

Shaffer, D. W., Squire, K. D., Halverson, R., & Gee, J. P. (2005). Video games and the future of learning. *Phi Delta Kappan, 87*(2), 104–111.

Squire, K. D. (2004). *Replaying history: Learning world history through playing Civilization III*. Unpublished doctoral dissertation, Indiana University, Bloomington.

Steinkuehler, C. A. (2006). The mangle of play. *Games and Culture, 1*(3), 199–213.

Tabak, I., & Reiser, B. J. (1997, December). *Complementary roles of software-based scaffolding and teacher-student interactions in inquiry learning*. Paper presented at the conference on Computer Support for Collaborative Learning.

Trushell, J., Burrell, C., & Maitland, A. (2001). Year 5 pupils reading an "interactive storybook" on CD-ROM: Losing the plot? *British Journal of Educational Technology, 32*, 389–401.

Tulving, E., & Thompson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review, 80*, 352–373.

Uttal, D. H., O'Doherty, K., Newland, R., Hand, L. L., & DeLoache, J. (2009). Dual representation and the linking of concrete and symbolic representations. *Child Development Perspectives, 3*(3), 156–159.

Vollmeyer, R., & Rheinberg, F. (2000). Does motivation affect performance via persistence? *Learning and Instruction, 10*(4), 293–309.

Vollmeyer, R., & Rheinberg, F. (2006). Motivational effects on self-regulated learning with different tasks. *Educational Psychology Review, 18*(3), 239–253.

Winn, B. (1987). Charts, graphs and diagrams in educational materials. In D. M. Willows & H. A. Houghton (Eds.), *The psychology of illustration: I. Basic research* (pp. 152–198). New York, NY: Springer.

Wittgenstein, L. (1953). *Philosophical investigations* (3rd ed.; G. E. M. Anscombe, Trans.). Malden, MA: Blackwell.

Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit formation. *Journal of Comparative Neurology and Psychology, 18*, 459–482.