THE EFFECTIVE INTEGRATION OF DIGITAL GAMES AND LEARNING CONTENT

Matthew Peter Jacob Habgood, BSc. (Hons)

Thesis submitted to the University of Nottingham
for the degree of Doctor of Philosophy

July 2007

ABSTRACT

This thesis is concerned with how the coveted user-engagement of digital games can be usefully harnessed for educational goals. Educational software has traditionally used gaming elements as a separate reward for completing learning content. The early 'edutainment' sector became synonymous with this cursory "chocolate-covered broccoli" approach (Bruckman, 1999): tagging games on to learning content in order to make it more palatable. However, such methods have often proved ineffective (Kerawalla & Crook, 2005; Trushell, Burrell, & Maitland, 2001) and have been criticised for combining the worst elements of both games and education (Papert, 1998) as well as for following extrinsically motivating design models (Lepper, 1985; Parker & Lepper, 1992).

This thesis provides a theoretical and empirical exploration of game designs that follow a more integrated approach. Five studies are described which detail the development and evaluation of a new theory for creating intrinsic integration based on integrating learning content with the game mechanics of a game. This includes the development of Zombie Division: a game that teaches mathematics to children through swordplay with skeletal opponents. Two experimental studies examine the motivational differences between integrated and non-integrated versions of Zombie Division by measuring time-on-task. Two more examine the educational effectiveness of integrated and non-integrated versions by measuring learning gains for a fixed amount of time-on-task. Statistically significant results are found which suggest that the integrated version is motivationally and educationally more effective than the extrinsic equivalent. Full results and implications are discussed.

PUBLICATIONS

Habgood, M. P. J. (2005). *Zombie Division: Intrinsic integration in digital learning games.* Paper presented at the 2005 workshop on Human Centred Technology Brighton, UK.

Habgood, M. P. J. (2007). *Zombie Division: Evaluating principles of intrinsic integration.* Paper presented at the 2007 Serious Games Summit: GDC, San Francisco, CA.

Habgood, M. P. J., Ainsworth, S., & Benford, S. (2005). *The educational content of digital games made by children.* Paper presented at the 2005 conference on Computer Aided Learning, Bristol, UK.

Habgood, M. P. J., Ainsworth, S., & Benford, S. (2005). Endogenous fantasy and learning in digital games. *Simulation and Gaming,* 36(4).

Habgood, M. P. J., Ainsworth, S., & Benford, S. (2005). *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.

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

ACKNOWLEDGEMENTS

To Jenny who inspired me to start and baby Elsie who inspired me to finish.

Many people deserve my thanks for their help in bringing this thesis into the world, but none more than Jenny for her endless emotional support and patience, as well as inspiring (and financially supporting) me with her gift for teaching. Shaaron Ainsworth also deserves considerable credit for all her valuable input: proving that you can be a great researcher and teacher at the same time. I'd also like to thank Steve Benford for teaching a lazy undergraduate the (object-oriented) language of video games all those years ago.

More thanks go to Jenny Habgood (again), Louise Ash, Elaine Cockburn, Lucy Button, Lizzie Evans, Sarah Peacock and all the staff at Reignhead Primary School, Gilt Hill Primary School and Sheffield E-Learning Centre West for their help and support with these studies. The creative talents of Ricki Martin, Matt Sandford, John Guerin and Stephen Lyons also deserve an honourable mention for helping to give Zombie Division a touch of class.

I'd also like to thank Judy Robertson for all her feedback and Katie Fraser for confusing skeletons with zombies! More thanks goes to Katie along with all the other students at the LSRI for being helpful and amiable colleagues! A final mention goes to Sumo-Digital for letting me do a full-time job part-time.

TABLE OF CONTENTS

CHAPTER 1

Introduction

"There is a widely held view that games software is capable of developing a degree of user engagement which could be usefully harnessed in an educational context" (McFarlane, Sparrowhawk, & Heald, 2002).

Interest in the field of games and learning has grown significantly in the four years since the work in this thesis began. Digital games are increasingly subject to serious study within the disciplines of psychology, education, computer-science and sociology – all in the name of learning. This thesis comes from the multidisciplinary perspective of the learning sciences, but incorporates one more perspective: that of the game development community itself. Developing digital games has been this author's profession for over a decade as well as a life-long interest and passion. This broad base has been applied to all the research in this thesis in the hope of providing a new contribution to the field.

This thesis is concerned with the way that learning content is integrated into the design of educational digital games. Historically, educational titles have often used gaming elements as an entirely separate reward for completing learning content. The early 'edutainment' sector became synonymous with this cursory "chocolate-covered broccoli" approach (Bruckman, 1999): tagging games on to learning content in order to make them more palatable. However, such methods have often proved ineffective (Kerawalla & Crook, 2005; Trushell et al., 2001) and have been criticised for combining the worst elements of both games and education (Papert, 1998) as well as for following extrinsically motivating design models (Lepper, 1985; Parker & Lepper, 1992).

This thesis provides a theoretical and empirical exploration of game designs that follow a more integrated approach, broadly known as *intrinsic integration* (Kafai, 2001). This is also often called *intrinsic* or *endogenous fantasy,* as it was fantasy elements that were considered key to creating this integration in the early research (Malone, 1981; Malone & Lepper, 1987). However, this thesis proposes an alternative viewpoint that identifies gameplay mechanisms or *game mechanics* as more critical to effective integration than fantasy. Both formative and summative evaluations are described which develop and test this new theory using a prototype mathematics game called Zombie Division.

1.1: THESIS STRUCTURE

Chapter two provides a review of the psychological literature that forms the theoretical basis for the engagement power of digital games. The empirical origins of intrinsic motivation are explored as part of the development of cognitive evaluation theory (Deci, 1975), flow theory (Csikszentmihalyi, 1988) and the cognitive curiosity associated with assimilation and accommodation (Piaget, 1950). Alongside this motivational research, the educational concepts of reflection (Schön, 1983) and transfer (Thorndike & Woodworth, 1901) are also examined as potential barriers to the effective use of games for learning.

Chapter two also focuses on the concept of intrinsic or endogenous fantasy (Malone, 1981) providing a critique of the theoretical and empirical foundations of this seminal work. Concerns are raised about the empirical basis of this work and a theoretical critique concludes that endogenous fantasy is a misnomer, in so far as the integral and continuing relationship of fantasy cannot be justified as a critical means of improving the effectiveness of educational digital games. An alternative perspective for the intrinsic integration of learning content is described, based on game mechanics. This proposal forms the main thesis of this work, and the basis of the novel contribution of this research.

Chapter three describes the first phase of design and development of Zombie Division: a mathematics game for seven and eight year olds. The game concept was conceived to test the value of our own proposed approach for creating intrinsic integration. The initial design concept for Zombie Division is described, including the learning content, core mechanics and fantasy context. The iterative design process is then detailed, including a series of informal trials with small groups of children from the target age group. This process resulted in the first prototypes of the game including intrinsic, extrinsic and control versions.

Chapter four describes the first empirical study undertaken to evaluate Zombie Division (study one). This set out to compare learning gains between the intrinsic and extrinsic versions of the game, in order to see which was the most educationally effective over a fixed amount of time-on-task. Unfortunately, the testing system proved too easy, and high pre-test scores across all conditions may have masked any significant differences between groups. Nonetheless the study provided a range of useful insights into the effectiveness of the game and the experimental methodology. These insights were used to inform the subsequent development of the game and the methodology used to evaluate it.

Chapter five describes the second phase in the design and development of Zombie Division. Changes were made to both the game design and methodology in order to address the failings of study one. Methodological changes included: a new testing instrument with longer, computer-based tests in order to help avoid future problems with high pre-test scores; a challenge-level that provides a game-based comparison between identical questions in the game and test; and more detailed process data to facilitate data mining. Changes to the game included: replacing the hundred-square with a multiplication grid that was more familiar to children; adding a help system to provide a framework for scaffolding the children's mathematical strategies; and improving the save-game system.

Chapter six describes two separate studies carried out to evaluate the motivational potential of Zombie Division. Study two set out to compare the time children spent playing intrinsic and extrinsic versions of the game when non-educational games were available as an alternative. Study three compared the total time spent playing the intrinsic and extrinsic versions of the game when children could choose to switch between them at will. Study two showed no significant difference between the times children spent playing different versions of the game when non-educational alternatives were available. Conversely, study three showed a significant preference for the intrinsic version when children could switch between versions. Based on this it is suggested that the intrinsic version is more motivating, but that the high production values and a competitive environment have a big impact on the motivational appeal of the extrinsic version.

Chapter seven describes the fourth and final study that evaluated learning outcomes in a fixed time-on-task situation similar to study one. However, this study incorporated the key changes from chapter five in order to address critical issues identified with the methodology and game functionality in the first study. This time a significant group difference was found for learning gains, suggesting that the intrinsic group achieved better learning than the other groups. This study also revealed a number of insights into transfer, reflection and deep learning as it relates to the design of digital learning games.

The results of all the studies are reviewed and discussed in detail in chapter eight. It is concluded that intrinsic games have the potential to create a) a higher level of motivational appeal and b) improved learning outcomes, over extrinsic equivalents. Concerns about the difficulty of transferring embedded learning content from intrinsic games have also been shown to be unfounded. Some concerns remain over the role of flow in inhibiting reflection-in-action in the intrinsic game, but no evidence was found that extrinsic equivalents are any better at promoting reflection. Future directions and implications are discussed.

1.2: MAJOR AIMS AND HYPOTHESES

The major aims of this thesis have been to develop and evaluate a new theory for the effective integration of digital games and learning content. This has challenged the established fantasy-based perspective on effective integration with one based on *game mechanics*. It suggests that learning content is better integrated with a game when it is embodied within the rule-systems that make the game fun, rather than the fantasy context.

The fantasy context of chess may be a good analogy for feudal society, but this is not the player's focus while playing the game. You could swap the feudal playing pieces for the latest Disney characters without changing the way that the game is played. Therefore the fantasy provides only a superficial way of integrating learning content within a game, when compared to the game's underlying rule systems or *game mechanics*.

Therefore this thesis suggests that educational games may be more effective if they have intrinsic learning content, which:

1. Delivers 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. Embodies learning material within the structure of the gaming world and the player's interactions with it, providing an external representation of learning content that is explored through the core mechanics of the gameplay.

From this standpoint this thesis sets out to test the following hypotheses about intrinsic games:

Hypothesis One: Deep Learning

The first hypothesis predicts that intrinsically integrated games are more effective then extrinsically integrated games as a result of creating a deeper connection with the learning content. This deeper connection will create greater engagement with the learning content in fixed time-on-task situations that should result in greater learning gains.

Hypothesis Two: Motivation

The second hypothesis predicts that intrinsically integrated games are more effective than extrinsically integrated games as a result of creating a superior level of motivational engagement with the game. This superior engagement will lead to a greater amount of playing time in free time-on-task situations that should eventually result in even greater learning gains.

Alternative Hypotheses: Transfer and Reflection

Nonetheless, a literature review in this area has suggested that the educational concepts of transfer and reflection may represent barriers to the effective application of intrinsic games. Intrinsically integrated games may actually be less effective as a result of the embedded intrinsic learning content transferring less effectively to contexts different from the game. Intrinsically integrated games may also be less effective as a result of essential reflective processes being impeded by integrating a game's learning content with its flow-experience.

The prototype mathematics game, Zombie Division, was created specifically for the purpose of generating empirical evidence to test these hypotheses. The four studies described in this thesis address the issues of motivation, deep learning, transfer and reflection as they relate to intrinsically integrated games. These studies have aspired to provide a body of empirical evidence, beyond that which has come before, to demonstrate the relevance of an integration theory based on game mechanics to the field of games and learning.

CHAPTER TWO

Engagement and Learning in Digital Games

This thesis does not seek to provide a complete review of the literature in games and learning, as there have been a deluge of reviews and reports in this area over the preceding years (de Freitas, 2006; Egenfeldt-Nielsen, 2005; H. Ellis, Heppell, Kirriemuir, Krotoski, & McFarlane, 2006; FAS, 2006; Hays, 2005; Kirriemuir & McFarlane, 2004; Mitchell & Savill-Smith, 2004; Sandford & Williamson, 2005) Instead, this chapter offers a grounding in the psychological and educational principles that are most pertinent to understanding this thesis. The chapter begins by providing a background to this work in the form of a number of relevant definitions and a short history of research into games and learning. The motivational interest in games is then explored through the literature on intrinsic motivation and its links to deep learning. The concept of intrinsic fantasy (Malone, 1981) emerges from this research as the literature's answer to the effective integration of digital games and learning content. Consequently this theory is reviewed in some detail, resulting in a new interpretation that raises educational concerns over the role of reflection and transfer. Therefore the chapter concludes with a review of these two concepts and their potential role in the creation of effective educational games.

2.1 BACKGROUND

This section provides a short history of research into games and learning in order to place this work within a wider historical context. However, it begins by providing rudimentary definitions for the terms 'digital game' and 'learning games'. These definitions are working assumptions, provided for reasons of practical clarity, and not intended as theoretical standpoints in their own right.

2.1.1 Definitions

Digital Games

The gaming literature provides an overwhelming number of different approaches to defining the essence of a game (Caillois, 1961; Crawford, 1982; Habgood & Overmars, 2006; Huizinga, 1950; Juul, 2005; Koster, 2005; Pearce, 2004; Rollings & Morris, 1999; Rouse, 2001; Salen & Zimmerman, 2004). One method for arriving at a working definition would be to draw out the commonalities between different definitions (e.g. Salen & Zimmerman, 2004, p.80). Yet these differences only serve to highlight Wittgenstein's (1953) observation that "you will not see something that is common to all, but similarities, relationships, and a whole series of them at that" (aphorism 66). His philosophy of language suggested that we cannot create a unifying definition because the term 'game' can only be expressed using examples from our personal experience (aphorism 75).

It is arguable then, that we should not attempt a definition, but in the interests of practicality we will take the simplest definition offered above which defines a game as an "interactive challenge" (Habgood & Overmars, 2006). This is an indirect definition that seeks to highlight the main differences between games and other forms of entertainment, rather than the similarities between games (p.87). This standpoint suggests that games contain an interactive element that distinguishes them from films, and prescribed challenges that distinguish them from toys. Nonetheless, the common role of entertainment is implicit in this distinction, and should be included in the definition. 'Digital' is used here to encompass all platforms which are digital in nature, including computers (and incorporating web platforms), home consoles, arcade consoles and other electronic platforms. This is primarily to prevent confusion between common variations in understanding of the terms 'computer game' and 'video game'. Put together this provides us with a working definition of a digital game as "an interactive challenge on a digital platform, which is undertaken for entertainment".

Learning and Educational Games

In the interdisciplinary field of the learning sciences, it is not uncommon to find different implicit assumptions about the words 'learning' and 'educational'. While it is beyond the scope of this thesis to engage in an analysis, it will help to declare the assumptions made by this document. Learning is therefore considered to be an all-inclusive term which includes both incidental and intentional learning (J. Anderson, 2005) of skills, knowledge and behaviours in all contexts. In contrast educational is considered to be a subset of learning bounded by a curriculum which has been determined by the wider society.

The empirical work of this thesis is based around an *educational* digital game as research in school contexts is always facilitated by conforming to the curriculum. Nonetheless, the theoretical and empirical arguments within this thesis equally apply to a learning content which was not bound by the curriculum.

2.1.2 Historical Context

The 1980s saw a meteoric rise in the popularity of videogame arcades, leading to the first significant wave of academic research into digital games. Much of this early research centred on attempts to demonstrate the dangers of this new media, in terms of its aggression (C. Anderson & Ford, 1986; Braun & Giroux, 1989), addictive traits (Selnow, 1984) and associations with "deviant behaviours" (D. Ellis, 1984). Nonetheless, there was already some study of the therapeutic use of video games in medicine (Kolko & Rickard-Figueroa, 1985) and the educational potential of the videogame phenomenon did not go unnoticed either. Although computer simulations, including games had been used in university teaching since the 1950s (Cullingford, Mawdesley, & Davies, 1979), it was the raw engagement power of videogames like Pac Man that excited a new generation of educationalists (Bowman, 1982). However, despite making a promising start (e.g. Lepper & Malone, 1987; Loftus & Loftus, 1983) the resulting generation of 'edutainment' products was widely recognized as having failed to effectively harness the engagement power of digital games (e.g. Papert, 1998). So while the mainstream games industry boomed throughout the 1990s, the educational sector was left behind in terms of technology, revenues and commercial interest.

The work in this thesis was completed during a second wave of academic research into digital games, which coincides with another revolution in the popularity of digital games. At the turn of the millennium companies like Sony, Nintendo and Microsoft have helped to turn digital games into a fashionable and indelible part of youth culture. Digital games are now discussed within the context of being more popular with children than television (Greenburg, 2004) and with revenues to rival the film industry (H. Ellis et al., 2006). Riding on the back of the popularity of video games, academic researchers have regrouped under the banner of 'serious games' and gained the patronage of large commercial (e.g. ELSPA: H. Ellis et al., 2006) and government institutions (e.g. BECTA: Dawes & Dumbleton, 2001). A number of popular texts extolling the potential of games for learning have helped to raise the profile and interest in serious games (e.g. Aldrich, 2004; Gee, 2003; Prensky, 2001; Shaffer, 2006) and collaborations have begun to take place between the industry and academia (e.g. Futurelab and EA: Sandford, Ulicsak, Facer, & Rudd, 2006). The industry itself has experienced commercial success with popular learning games such as 'Brain Training' and 'Big-Brain Academy'. At least three of the world's top one hundred game development studios now have their own serious gaming divisions to exploit the potential of games designed for more than just entertainment[1].

2.2 MOTIVATIONAL BENEFITS

The concept of intrinsic motivation lies at the heart of interest in the use of digital games for education. Children (and adults) will willingly spend hours of their time engaging with digital games in a way that many educationalists would like to transfer to an educational context. However, some researchers suggest that there is more to intrinsic motivation than increasing time-on-task and that intrinsically motivated learners also achieve deeper learning. Consequently this section will review both the concepts of intrinsic motivation and deep learning in order to provide an insight into the educational potential of digital games.

2.2.1 Intrinsic Motivation

Growing out of work by a number of different researchers (e.g. Csikszentmihalyi, 1975b; Deci, 1975; Lepper & Greene, 1975) it is commonly surmised that a person is "intrinsically motivated to perform an activity when he receives no apparent rewards except the activity itself" (Deci, 1971). The self-motivation of game players is certainly apparent, but it is perhaps the work of Csikszentmihalyi (1975a), and the concept of flow that provides the most striking example of the engagement power of digital games. Being 'in the zone' is a common experience for gamers, and one that could potentially provide a significant force for learning if it could be effectively applied to digital learning games.

[1] Square Eniq (£6,518,697 : 41), Sumo Digital (£5,194,784 : 52), Blitz Games (£3,809,307 : 70). 2006 revenues and world ranking from Develop (2007).

Lepper and Malone (1987, p.258) categorise theoretical models for intrinsic motivation into three different approaches. The first, which is characterised by the work of Csikszentmihalyi (1988), defines the optimal experience produced by intrinsic motivation in terms of challenge. Flow theory proposes that clear goals, achievable challenges and accurate feedback are all required to achieve a state of flow in an activity (p34). This 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" (p.30). There are clear parallels between this and the way that 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). Perhaps it is unsurprising then that the concept of flow or 'being in the zone' is so commonly associated with digital games. Feelings of total concentration, distorted sense of time, and extension of self are experiences that are as common to game players as Csikszentmihalyi's chess players, rock climbers, dancers and surgeons (Csikszentmihalyi, 1975a).

The second theoretical model for intrinsic motivation is characterised by curiosity, drawing upon research into novelty, complexity and incongruity (Berylne, Craw, Salapatek, & Lewis, 1963) alongside Piagetian ideas of cognitive conflict in the accommodation of new knowledge (D. Wood, 1998). It suggests that "we derive pleasure […] from activities that provide us with some […] level of surprise, incongruity or discrepancy" (Lepper & Malone, 1987, p.258). Contemporary game theorists have gone further to suggest that the pleasure derives from the wider learning process involved in solving gaming puzzles (Koster, 2005). Nonetheless there is a mapping between this model of learning and the experience of playing many digital games (see Gee's *probing* and *intertextual* principles). Most games begin by requiring the player to assimilate knowledge about the game into the player's existing understanding of the genre (e.g. in many platform games, enemies can be defeated by jumping on their heads); they then experience curiosity and surprise as a result of cognitive conflict (discovering that an end-of-level boss[2] can't be defeated in this way); before accommodating their knowledge to cope with a new discovery (that end-of-level bosses have periods of vulnerability during which they must be jumped on numerous times). In this way, game designers continually build and challenge the player's understanding of the gaming world to keep their curiosity engaged.

The third approach highlighted by Lepper and Malone (1987) focuses on the role of control and self-determination in intrinsic motivation. This has its foundations in the work of researchers like Deci (1971) who were interested in the effect of external rewards on intrinsic motivation. Citing deCharms (1968), Deci (1971) suggests that "when external rewards are given for an intrinsically motivating activity, the person perceives that the locus of control […] shifts to an external source" resulting in a decrease in intrinsic motivation (p.105). Together with Ryan he went on to develop this thesis into the field of self-determination theory (Ryan & Deci, 2002). This focuses on what they consider to be three areas of innate basic psychological needs, namely competence (similar to the concept of challenge above), relatedness (connection with and acceptance of others) and autonomy (being the perceived origin of one's own behaviour) (p.7). This concept of autonomy can be applied to the field of digital games in explaining the short-lived interest in the laserdisc gaming platform. At the time the new technology provided players with an unparalleled graphical experience, but gave limited autonomy in controlling the games. The format was eventually superseded by less visually impressive platforms which could provide greater autonomy to the player.

[2] Bosses are powerful computer opponents encountered at significant points in a game. They normally must be defeated in order for the player to progress further.

One, more recent refinement of self-determination theory attempts to decompose intrinsic motivation into three basic types (Vallerand & Ratelle, 2002). Intrinsic motivation to *know* implies engagement from "learning, exploring and understanding new things" (c.f. cognitive curiosity); intrinsic motivation to *accomplish* implies engagement to "surpass-oneself creating or accomplishing something" (c.f. flow) and intrinsic motivation to *experience* implies engagement from "stimulating sensations associated with it". The last of these is similar to Caillois's idea of *vertigo* as an aspect of many games that stimulate or confuse the senses for enjoyment (Caillois, 1961). Yet, although there is nothing new about some of these categories they do highlight the potential separation between the intrinsic motivations that may be differentially experienced by players. The cognitive curiosity behind the motivation to *know* sounds like it has potential for integrating educational content into games, but would it provide the same benefits for players that are motivated by *accomplishment* or *experience*?

These three perspectives on intrinsic motivation (challenge, curiosity and control) formed the core of Malone and Lepper's "Taxonomy of Intrinsic Motivations for Learning" (Malone & Lepper, 1987). This suggested that together with fantasy, competition and co-operation, these six factors could be used to explain the intrinsically motivating characteristics of digital games. It is interesting to note that the inter-personal motivations of competition and co-operation are included in the relatedness aspect of self-determination theory. The role of fantasy and specifically its part in the integration of games with their learning content will be discussed in much greater detail in the next section.

More recently, researchers examining digital games have suggested a plausible link between the flow phenomenon and addictive behaviours (Chou & Ting, 2003). Some attribute this to an interaction between personality type and the ability of the challenges of a game to create arousal (Griffiths & Dancaster, 1995). Researchers have also examined the role of flow more generally as part of a framework for studying the use of information technology (Bryce & Higgins, 2000; Ghani & Deshpande, 1994). This, and educational research in the same area, suggests that flow may be linked to more exploratory behaviours (Martens, Gulikers, & Bastiaens, 2004). However, the same study found no improvement in learning outcomes as a result of high levels of intrinsic motivation. In fact there has been research that suggests providing the learner with the ability to engage in exploratory behaviours may not be beneficial for learning (Laurillard, Stratfold, Luckin, Plowman, & Taylor, 2000; Trushell et al., 2001). Nonetheless other studies found encouraging results for incorporating flow into mathematics games (Sedighian, 2007), and suggest it may be particularly appropriate to this domain.

2.2.2 Deep Learning

Deep learning is characterised by an insight and understanding of underlying principles and rules within a particular domain rather than just factual knowledge (Sandberg & Barnard, 1997). Research by Biggs (1987) linked the contrast between 'deep' and 'surface' approaches to the concept of intrinsic motivation (Deci, 1975). This suggested that students that are extrinsically motivated engage in surface approaches to learning in order to meet the minimum requirements asked of them, focussing on reproductive or rote learning techniques (Biggs, 1987, p.12). In contrast, intrinsically motivated students engage in deep learning by reading widely and interrelating their learning with previous knowledge. This link between intrinsic motivation and deep learning provides some hope for educational game designers, in terms of the metacognitive benefits of the powerful engagement they create with their audience.

Some researchers (Gee, 2005; Shaffer, 2006) have suggested an alternative model for how digital games create deep learning experiences. This draws upon learning research around communities of practice (Lave & Wenger, 1991) and reflection-in-action (Schön, 1983) for the concept of 'epistemic games'. These are games that allow players to experience learning from within the epistemological practices of a real world role such as doctors, lawyers, architects, engineers, journalists etc. (Shaffer, 2004). They suggest that adopting identities in this way triggers an extended commitment and deep investment on the part of the player (Gee, 2005, p.7). Nonetheless this approach raises questions about how well the knowledge developed within a simulated community of practice would transfer to a different real-world community, or indeed what subversive role the meta-community of practice (gamers) might play in this relationship.

2.3 INTRINSIC INTEGRATION

The term 'intrinsic integration' has been used to refer to the effective integration of a game idea with its learning content (Kafai, 2001). Borne out of the goal of effectively harnessing the intrinsic motivation of digital games, this concept has its roots in the work of Lepper and Malone (1987) discussed above. This focussed on the role of fantasy in creating this intrinsic relationship, concluding that the educational effectiveness of a digital game depends on the way in which learning content is integrated into the fantasy context of the game. Therefore it was claimed that content which is intrinsically related to the fantasy will produce better learning than that which is merely extrinsically related (p. 240).

Clearly the concept of intrinsic fantasy and the educational claims made about it are highly relevant to this thesis. However, while many subsequent works, such as Reiber (1996), Dempsey (Dempsey, Lucassen, Gilley, & Rasmissen, 1993) and Asgari (Asgari & Kaufman, 2004) cite the concept of intrinsic fantasy without reanalysis, others including Kafai (1996), Fabricatore (2000), Prensky (2001) and Egenfeldt-Nielsen (2005) offer their own implicit reinterpretations. Many of these reinterpretations do not focus on the role of fantasy, yet none provide a full review and critique of the original concept. Therefore this section will explore the empirical origins behind intrinsic fantasy and attempt to rationalize the potential significance of fantasy in creating a more intrinsic relationship between educational games and their learning content.

2.3.1 A Fantasy-Based Perspective

The concept of intrinsic fantasy attributes educational benefits to the distinction between intrinsic and extrinsic fantasy (Malone, 1980), later relabelled endogenous and exogenous fantasy by Malone and Lepper (1987). An educational game is said to contain intrinsic fantasy when “the skill being learned and the fantasy depend on each other” and when “there is an integral and continuing relationship between the fantasy context and the instructional content being presented” (Malone & Lepper, 1987, p. 240). An educational game with extrinsic fantasy is defined as “one in which the fantasy depends on the skill being learned but not vice versa”. Furthermore, it is proposed that “endogenous fantasies are both more interesting and more educational than exogenous fantasies” (p. 240). This forms the basis for the fantasy-based perspective on intrinsic integration, which has become established within the literature.

2.3.2 Empirical Work on Intrinsic Fantasy

Although the taxonomy of gaming motivations (Malone & Lepper, 1987) was based around a number of empirical studies, only one of these (Malone, 1981) directly addressed the concept of intrinsic fantasy (for a review of the other studies see Habgood, Ainsworth, & Benford, 2005b). This study applied an ablation technique[3] to an educational computer game called Darts, based around learning fractions (see Figure 2.1). Players attempted to enter fractional numbers corresponding to the height of three balloons placed at random positions on a number line. Each guess would launch an arrow across the screen at the specified height. Arrows aimed on target would pop the corresponding balloon, and misses would remain on the number line with the incorrect guess written next to them (Malone, 1981, p. 349).

Intrinsic fantasy was identified as one of the potentially motivating features of this game, as the fantasy of ‘balloons and arrows’ was said to be intimately related to

[3] Techniques that attempt to determine the function of a system by observing the wider effects of its removal.

the skill of estimating fractions. An extrinsic fantasy version of the game was also created as a comparison. This attempted to break the intrinsic relationship by displaying the balloons as a separate scoring mechanism away from the number line and replacing the targets with plain rectangles (see figure 2.1).

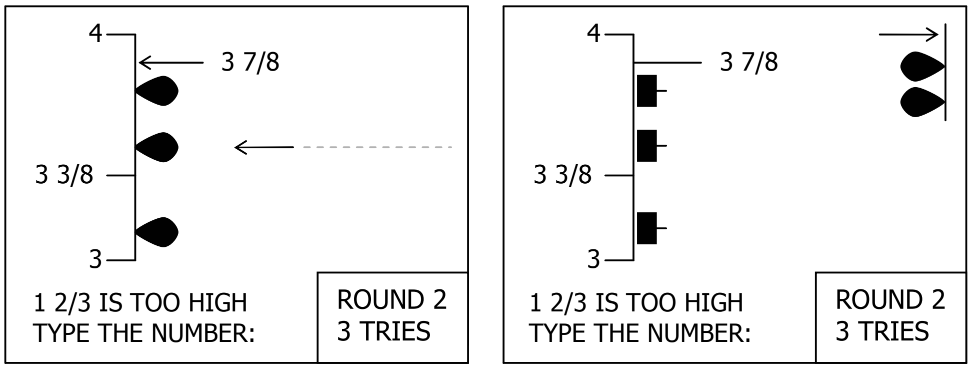


Figure 2.1: The screen layout of the intrinsic (left) and extrinsic (right) versions of the Darts game. Adapted from Malone (1981).

The results of this study showed no difference in the time boys spent playing the intrinsic and extrinsic fantasy versions of this game. However, girls played the intrinsic fantasy version for significantly less time than the extrinsic version. Malone (1981, p. 354) argues that this suggests the girls disliked the intrinsic version the most because this integrated the fantasy – which they disliked – most effectively. These results offer very limited (if any) support for the ability of intrinsic fantasy to produce more interesting educational games. Furthermore, this study was not designed to measure learning outcomes so it cannot provide any support for the claim that games with intrinsic fantasies are more educational.

2.3.3 Games and Fantasy

The studies behind Malone and Lepper's taxonomy (Malone, 1981) present a confused picture of the role of fantasy in their games. An earlier paper states that "Non-fantasy games involve only abstract symbols" (Malone, 1980, p. 164). Yet the rapid progress of technology has left the Darts game's arrows and balloons looking very abstract to the modern eye. A conceptual difference between a balloon-shaped blob and a rectangle is hard to justify when you compare them both to contemporary 3D computer graphics. Throughout this work there seems to be an underlying conceptual viewpoint of fantasy as something that permeates all aspects of digital games. The feedback provided by the visual representation of guesses in the Darts game is attributed to the "fantasy world of balloons on a number line" (Malone, 1981, p. 361). In another game (Breakout: see figure 2.2) a wall of bricks is described as a "visually compelling fantasy goal" (Malone, 1981, p. 348), crediting the fantasy with producing goals and a scoring mechanism as well as a visual effect.

Exchanging bricks or balloons for 'abstract symbols' such as crosses might decrease the motivational appeal of these games, but they would still function in the same way. The decomposition of fantasy is the subject of an unresolved debate in psychology, and specifically as to whether fantasy can occur without imagery (Klinger, 1971, p. 136). Nonetheless, the game development process routinely decomposes the fantasy context of a computer game into the elements that are required to implement it (e.g. Habgood & Overmars, 2006). This includes the contextual elements of the graphics, sound and music that set the game within a specific time and place; as well as the narrative and characterization that are created through the story. However, these fantasy elements are ultimately only a surface-level representation of the game, and can be completely changed without affecting the essence of the gameplay produced by that game. This viewpoint comes from a ludic perspective (Frasca, 2003), which sees the fantasy elements of games as quite separate from – and less significant to – the underlying rule systems of a game.

The Darts experiment also provides a concrete example of an intrinsic fantasy against which to compare the definition. This game is offered as a good example of an intrinsic fantasy, "[…] where the fantasy (the positions of the arrows and balloons on the number line) is intimately related to the skill being used (estimating fractions)" (Malone, 1981, p. 350). Darts is contrasted with the extrinsic fantasy of Hangman, which is said to be, "only weakly related to the skill being used (spelling and vocabulary)". The extrinsic nature of Hangman is further justified by suggesting that the learning content of the hangman fantasy could be swapped for something different. Yet in the same way 'abstract' rectangles and lines are swapped for the 'fantasy of balloons and arrows', the 'fantasy of balloons and arrows' could be swapped for the fantasy of 'elephants and currant buns'. If this was achieved by simply redrawing the graphics then there is no reason to believe that this new game would be educationally distinct than the old one. If this switching of fantasies is possible, then can there truly be "an integral and continuing relationship" between any of these fantasies and the learning content, or are there common structures to the implementation of both these fantasies that are the real intrinsic factors at work?

2.3.4 Games and Mechanics

The gameplay produced by digital games is not directly attributable to the fantasy context of a game but the "mechanism through which players make meaningful choices and arrive at a meaningful play experience" (Salen & Zimmerman, 2004, p. 317). These mechanisms are commonly referred to by game developers as the game mechanics (Habgood & Overmars, 2006). Game mechanics are the procedural mechanisms of a game that provide the essential interactions required to create a meaningful gaming activity. So the core game mechanic of Breakout is in controlling the horizontal position of one object in order to repeatedly intercept another moving object and keep it bouncing around a confined space (figure 2.2).

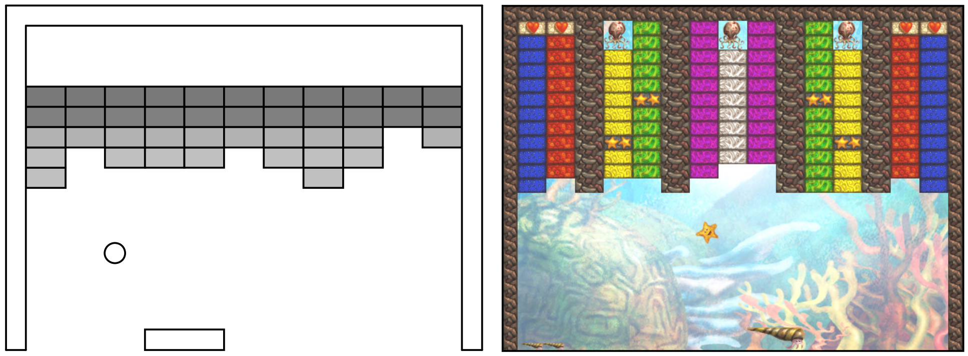


Figure 2.2: Different fantasy contexts for similar game mechanics. Breakout © 1976, Atari (left) and (right) Super-Rainbow Reef (Habgood & Overmars, 2006).

Whether the game uses the fantasy context of a bat and ball or (as in figure 2.2) a mollusc and a starfish, it makes no difference to the fundamental gaming activity. The core mechanic of the Darts game is entering fractional values that make one object hit another object, based on its position along the length of a third object. As already suggested these could just as well be elephants and currant buns as balloons and arrows, and there would be no reason to expect that the game would have any less educational value.

This argument suggests that it is not the "integral and continuing relationship" of fantasy that creates an intrinsic relationship between a game and its learning content, but the game mechanics. The choice and quality of fantasy are extremely important in engaging players' emotional interests in a game, but theoretically any fantasy could be swapped for one of equal merit without changing the educational effectiveness of that game. Therefore, we suggest that intrinsic fantasy is a misnomer, which clouds our ability to distinguish the effect of fantasy contexts from the more precise distinctions between games. Intrinsic integration (Kafai,

1996) is therefore a more useful term and this section will continue to provide its own definition of what creates an intrinsically integrated educational game.

2.3.5 A Mechanics-Based Definition

Many of the intrinsically motivating components of digital games are realized through their core mechanics. Game mechanics are the fundamental building blocks for providing the clear goals, achievable challenges and accurate feedback that Csikszentmihalyi identifies as central to creating the flow experience (1988, p. 34). Furthermore, these seem to be the very kind of experiences that are missing in the majority of edutainment products, and could be a major factor in the distinction between extrinsic and intrinsic learning in digital games. Some edutainment products interrupt the flow of the gameplay with their learning content, and others keep the learning quite separate from it, but few manage to make the learning content intrinsic to this flow experience.

The relationship between fantasy and game mechanics can also be seen as a way of embodying a logical system of rules within a complex visual representation. Even an arbitrary fantasy can present a useful visual representation if the underlying mechanics embody the learning content. Visual representations of knowledge can support the construction of mental models (Schnotz & Bannert, 2003), enhance learners' metacognitive strategies, and encourage them to learn complex topics more completely (Ainsworth & Loizou, 2003). By employing visual representations in environments such as Microworlds and Simulations (de Jong & van Joolingen, 1998; Papert, 1980), learners have been encouraged to participate in the interactive exploration of learning content (Miller, Lehman, & Koedinger, 1999; Papert & Talcott, 1997). Therefore when the learning content of an educational game is intrinsically integrated with the game's core mechanics, then it could naturally provide a form of interactive representation for that content as well. While visual representations are often employed to aid understanding in edutainment software it is rarely possible for the learner to interact with them in an active way. Yet the interactive manipulation of visual environments allows learners to take active control of their own learning and in so doing construct a deeper understanding of the subject (Martin & Schwartz, 2005).

This line of argument suggests that educational games may be more effective if they have intrinsic learning content, which:

3. Delivers 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.

4. Embodies 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.

While this may seem to represent a clearer approach to intrinsic integration than intrinsic fantasy, this definition actually makes it easier to see how an integrated approach might produce *less* effective learning. It is possible that including learning content within the flow experience of a game may impede reflective processes that allow metacognition and the acquisition of declarative knowledge. Furthermore, embedding learning content within the structure of the gaming world may contextualise the learning, and make it more difficult for learners to transfer it to other contexts. The role of reflection and transfer in learning with digital games is therefore explored in more detail in the following section.

2.4 EDUCATIONAL CONCERNS

Our review of intrinsic integration has raised potential concerns with the role of reflection and transfer in the educational effectiveness of an intrinsic approach. Therefore this final section provides a review of these two important educational concepts and their significance to creating digital learning games.

2.4.1 Reflection

In an educational context the term 'reflection' brings to mind thoughtful or meditative processes associated with learning. A century ago, John Dewey (1910) defined the term as "Active, persistent, and careful consideration of any belief or supposed form of knowledge in light of the grounds that support it, and further conclusions to which it tends" (p.6) He used the term to describe a model of deductive scientific reasoning: reflecting upon an existing body of knowledge in order to apply it to a new situation. Dewey's viewpoint associates reflection with forward-looking thought processes, using knowledge from the past to build new understandings upon which future actions or conclusions can be based.

However, more recently, Donald Schön (1983) used the term in a different way to describe thoughtful processes focussed on ongoing activity in the present. He used 'reflection-in-action' to describe the "art by which practitioners sometimes deal well with situations of uncertainty, instability, uniqueness and value conflict" (p.50). This focussed on tacit knowledge which is somewhat at odds with Dewey's very logical ideas of deductive reflection. Schön used the term 'reflection-on-action' to describe the retrospective thought processes that we more usually associate with reflection today e.g. "reflecting upon your behaviour". Commonly discussed within the context of improving the practice of teaching (e.g. Matchett, 2005; McAlpine, Weston, Beauchamp, Wiseman, & Beauchamp, 1999) this kind of reflection is a form of debriefing: dissecting and learning from a recent experience while it is still fresh in the mind. Nonetheless Dewey and Schön's contrasting ideas about reflection show that there are several different ways of approaching this educational concept.

Reflection is often considered to play an essential role in developing metacognitive skills (self-regulation of the learning process) that are critical to effective learning (e.g. Bransford, Brown, & Cocking, 2000, p.97). Yet when reflection is discussed within the context of digital games there is an implicit suggestion that games do not naturally provide opportunities for reflective learning. Prensky (2001, p.50) suggests that educational designers need to "invent ways to include reflection and critical thinking […] with the learning [of a game]". Others have gone to great lengths to include structured reflective activities in order to get the most educational value out of a commercial game (K. D. Squire, 2004). Indeed Prensky's concept of a 'twitch-speed generation' certainly seems to conflict with Dewey's idea of reflection as a contemplative process of deductive reasoning.

Nonetheless, learning is considered a central aspect of commercial game design by a growing number of academics and game designers (e.g. Gee, 2003; Koster, 2005). This is not about commercial games containing educational content, but how the enjoyment of games derives from the process of learning how to play the game itself: i.e. "the fundamental motivation for all game-playing is to learn" (Crawford, 1982, p.17). Games stop being fun when the player stops learning, and game designers attend to the 'learning curves' of their games accordingly (Habgood & Overmars, 2006, p.156). Designers even have an implicit understanding of behavioural learning processes such as reinforcement, insofar as they apply to motivation in digital games (Loftus & Loftus, 1983).

So if the learning process is central to the success of non-educational digital games, and reflection is critical to learning, then can successful commercial games really be devoid of reflective opportunities? Indeed a number of researchers have examined the role that online communities play as repositories of gaming knowledge (Burn, 2006; Gee, 2003; Steinkuehler, 2004), where players engage in pursuits that could certainly be described as reflection-on-action. There is also some research which suggests that playing commercial games in social groups promotes behaviours indicative of reflection-in-action. Cooperative working in educational games has been shown to provide potential advantages to productivity and motivation (Inkpen, Booth, Klawe, & Upitis, 1995), but Schott and Kambouri (2006) also observed advantages for a group of children playing a commercial game designed for one player. They described the social interactions of the group as a form of 'situated learning' (Lave & Wenger, 1991) and Twidale, Wang and Hinn (2005) found a similar process of 'peripheral participation' in their studies of college students playing commercial multi-player games.

Schön's original studies of reflection-in-action were based on observing reflective discourse between professional practitioners and their students (Schön, 1983). Like the gaming examples above, it was this discourse that allowed the reflective process to be studied effectively. Nonetheless, it does not preclude individuals from engaging in a process of reflection-in-action on their own. Indeed it is easy to draw parallels between Schön's description of a reflective practitioner and an individual engaging in a 'reflective conversation' with a digital game:

> *The practitioner allows himself to experience surprise, puzzlement, or confusion in a situation which he finds uncertain or unique. He reflects on the phenomenon before him, and on the prior understandings which have been implicit in his behaviour. He carries out an experiment which serves to*

generate both a new understanding of the phenomenon and a change in the situation. (Schön, 1983, p.68)

He suggests that professionals enjoy and actively seek out uncertain situations: the same kind of situations that embody aspects of cognitive curiosity and challenge thought to be critical to the engagement power of games (Lepper & Malone, 1987). The application of prior understanding is very similar to the way in which gamers apply experience between games of similar genres (Gee, 2003). Yet even games of the same genre have unique elements and players are forced to adapt their understanding of the genre through an interactive process of experimentation and feedback with the game (see the example in section 2.2.1). Schön even talks of the role of the practitioner's mentally constructed 'virtual worlds' as "a constructed representation of the real world of practice" which are a "crucial component of his ability […] to experiment rigorously" (p.157).

So while digital games may naturally offer some opportunities for both reflection-in-action and reflection-on-action, reflection is most often associated in the literature with slow-paced and even group-based processes. It must therefore remain an area of concern for any educational game designers attempting to create 'twitch-speed' games.

2.4.2 Transfer

When Dewey spoke of reflection as considering knowledge in light of the "further conclusions to which it tends" (Dewey, 1910), he was alluding towards the role of reflection in supporting the transfer of knowledge into new contexts. A decade earlier, Edward Thorndike had begun to study "the influence of improvement in one mental function upon the efficiency of other functions" (Thorndike &

Woodworth, 1901). These early works recognised the difficulties of applying knowledge learnt within formal schooling, to contexts outside of the classroom. Thorndike undertook a range of experiments exploring the transfer of skills between different contexts and came to the conclusion that "the spread of practice occurs only where identical elements are concerned in the influencing and influenced function" (p.250). He rejected the established view that 'mental functions' such as attention, memory, observation or accuracy are faculties that can be developed as generic skills, suggesting instead that the mind makes "particular reactions to particular situations" (p.249).

More contemporary work agrees that "transfer between tasks is related to the degree to which they share common elements, although the concept of elements must be defined cognitively" (Bransford et al., 2000, p.78). This cognitive approach sees these 'elements' as "abstract knowledge structures that enjoy a wider range of transfer" (J. Anderson, 2005, p.306) rather than the simple stimulus-response mechanisms associated with Thorndike's earlier behavioural work. Nonetheless, there are those who reject the entire concept of transfer as an 'impoverished caricature' of learning (Carraher & Schliemann, 2002). They argue that transfer suggests a "carrying over of procedures from one situation to another" (p.21) which over-simplifies the role that prior knowledge and experience play in the process of assimilation and accommodation involved in learning new concepts (Piaget, 1950).

Transfer is often spoken of in terms of a number of different scales: it is considered to be *near* or *far* transfer depending on how closely related the original context is to the transfer context (e.g. Van Eck & Dempsey, 2002); and *high* or *low road* transfer depending on whether the transfer is achieved through the "automatic triggering of well-practiced routines" (low) or "deliberate mindful abstraction" (high) (Perkins & Salomon, 1988). The previous example of gamers applying experience between games of similar genres (Gee, 2003) might be considered low road, near transfer as it is based on extensive practice with games that have significant similarities. However, transferring knowledge learnt within a game into a classroom setting may conceivably require high road, far transfer in order to be successful. Research by Nunes, Schliemann & Carrher (1993) showed that Brazilian children selling items on the street demonstrated mathematical skills that did not transfer to a classroom setting. The children had developed successful oral-based strategies which they were unable to apply to the more abstract mathematical procedures used in the classroom (Nunes & Bryant, 1996, p.107). If educational games fail to create transfer then they have the potential for creating highly contextualised learning of this kind. Street mathematics may be a useful life-skill in its own right but football management mathematics (McFarlane et al., 2002) may not.

Some researchers suggest that the problem of transfer is accentuated by learning environments (such as traditional schooling) that create 'inert' knowledge by treating it as an end in itself (Bransford, Sherwood, Hasselbring, Kinzer, & Williams, 1990). They offer Anchored Instruction as a model for avoiding this by "creating an anchor or focus that generates interest and enables students to identify and define problems" (p.123). This is an intrinsically motivating and personally meaningful situation that allows learners to identify "critical features of problem situations" so that they might 'anchor' their learning within it. For example: they used the opening sequence of the film, "Raiders of the Lost Ark" as a basis of planning a trip to the same location, anticipating the problems they might encounter (p.128). Bransford and colleagues have conducted a number of studies that show improved results in transfer tasks as a result of taking this approach (e.g. A. Michael, Klee, Bransford, & Warren, 1993). There are obvious parallels for creating intrinsically motivating and personally meaningful anchors using digital games as well. This has started to be explored by a number of researchers, producing some complex interactions that suggest the role of competition (Van Eck & Dempsey, 2002) and interactivity (Moreno & Mayer, 2005) may have a negative effect on transfer.

It is the transfer of aggressive behaviours, beliefs and attitudes that forms the basis of concerns about violence in digital games (C. Anderson, 2002). Proponents of the theory believe that behavioural transfer can take place by priming aggressive scripts and perceptual schemata, increasing arousal and creating an aggressive affective state (p.104). Michael and Chen (2005) suggest that the US military attribute a number of transferable skills to military recruitment and training games such as America's Army. These include an improved ability to multitask, improved target differentiation, target prioritisation, teamwork with minimal communication, desensitisation to shooting human targets and a willingness to take aggressive action (p.59). While not all of this appears to be fully substantiated, there is evidence to suggest that perceptual skills can transfer from digital games, despite the normal lack of transfer between individual visual training tasks[4] (Green & Bavelier, 2006a). This has been demonstrated in terms of hand-eye co-ordination (Griffith, Voloschin, Gibb, & Bailey, 1983), reaction times (Orosy-Fildes & Allan, 1989) and peripheral attention (Green & Bavelier, 2006b).

Nonetheless, knowledge transfer represents another significant area of concern for educational game designers. The concept of transfer is a complex and often confounding one for all researchers in the field of education, and so it is certainly an issue that will impact on the design of effective educational games.

[4] c.f. Thorndike above.

2.5 CONCLUSION

This theoretical review has proposed a new perspective for the concept of intrinsic integration, focussing on game mechanics rather than fantasy. This suggests that greater integration may be achieved by integrating learning material within the core mechanics that underpin the flow experience and interactive representation provided by a learning game. In doing so it has also highlighted both motivational benefits and educational concerns that could potentially impact the effectiveness of this intrinsic approach. Chapter three describes the design of Zombie Division, a game concept which was conceived to embody these theoretical guidelines and provide a basis upon which to test the value of the intrinsic approach.

CHAPTER THREE

Zombie Division Game Design

This chapter describes the first phase of the design and development of Zombie Division. This game concept was conceived to embody the theoretical guidelines set out in chapter two for creating an intrinsic integration between gameplay and learning content. Zombie Division was created to be used in a series of empirical evaluations comparing an intrinsic variation of the game with an extrinsic equivalent. This would allow the underlying hypotheses of chapter two to be examined, and evidence sought for the value of the intrinsic approach. This chapter details the initial design concept for Zombie Division including the learning content, core mechanics and fantasy context. The ongoing iterative design process is also described, including a series of informal trials with small groups of children from the target age group. This process resulted in the first prototype version of the game in the form that it was used in study one (chapter four)[5].

3.1 AIMS

3.1.1 Theoretical Constraints

Our theoretical analysis of intrinsic games suggests that they may provide motivational benefits, which in turn may have positive and/or negative effects on learning. An empirical evaluation of these effects requires a fair comparison between intrinsic and extrinsic design approaches. Therefore two versions of the same game needed to be created that were identical in all respects except for the way in which they integrated their learning content (intrinsically and extrinsically).

[5] The reader may further familiarise themselves with the game by playing the final version of the game (used in study four) included on the DVD-ROM.

This was the primary goal for the game design, and most of the key design decisions were based around this fundamental design constraint.

Therefore the intrinsic version of the game needed to embody the theoretical guidelines for creating intrinsic integration set out in chapter two. These can be summarised as follows (Habgood et al., 2005b):

1. 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. 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 explored through the game's core mechanics.

Our analysis had suggested that this approach could provide motivational benefits, which create a greater level of engagement and a deeper connection with the learning content. However, the same analysis also suggested that such benefits might be outweighed if embodied learning content transfers less effectively to other contexts than unembodied content. Players might also find it more difficult to reflect upon learning content that is delivered as part of a game's flow-experience. These alternative hypotheses can therefore be used to define a contrasting set of 'extrinsic guidelines' that can be summarised as follows:

1. Keep the learning material separate from the parts of the game that are the most fun to play, avoiding the distraction of the flow experience produced by the game.

2. Separate the learning material from the structure of the gaming world, providing a direct mapping of the learning content that must be completed in order to proceed with the gameplay.

Consequently the game design required a flow-inducing core mechanic that embodied the learning content at the heart of the game in the intrinsic version, yet still allowed the learning content to be removed from the extrinsic version. Furthermore, its removal must not significantly change the motivational experience the gameplay creates for the player – otherwise the comparison would be unfair. Identical learning content also needed to be included in the extrinsic version of the game, in an 'unembodied' form, separate from the flow experience produced by the game.

3.1.2 Ethical Constraints

It was decided at an early stage to target a primary school (7-11) audience for the game and its associated studies. Primary schools in the UK can provide groups of children who are routinely in the same place and under the charge of a single teacher. This makes permissions and planning more straightforward, but also imposes a number of practical and ethical constraints on the design of the game and methodology. Although the game would be teaching some form of learning content, it would have been undesirable to take up too much of children's classroom time. It was decided that any disruption should be minimised to around half an hour of each child's daily lessons for a period of a week. This would provide a total of about 2½ hours of playing time within which to produce measurable learning outcomes using the game. Zombie Division therefore needed to provide a 'gaming episode' (rather than a complete game) that focussed on a specific area of learning content that could be delivered within this time frame.

Naturally, ethical considerations also meant that the content of the game needed to be age-appropriate as well. However, our experience from previous work making games with children (Habgood, Ainsworth, & Benford, 2005a) suggested that unnecessarily sanitised content may fail to engage a gaming audience. In order to gauge the content properly, the European ratings system (PEGI) was used to provide guidelines for producing game content that was appropriate for primary aged children. The design of the game took into account that the PEGI guidelines for children over the age of seven allow the following level of violent/scary content in games:

1. Occasional violence to non-realistic fantasy characters.
2. Pictures or sounds likely to be scary or frightening to young children

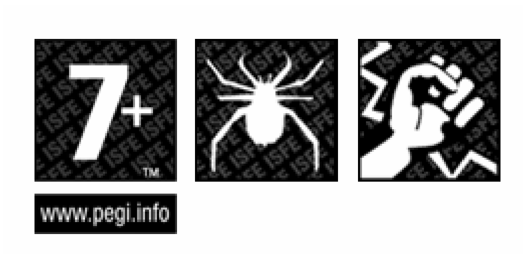


Figure 3.1: PEGI content rating symbols.

3.2 GAME CONCEPT

3.2.1 Learning Content

The game's learning content was allowed to evolve in parallel with the game design as there were already many constraints on the design process. Nonetheless, it was acknowledged from the outset that the content should be based around the National Curriculum in order to provide value for schools participating in any studies. Mathematics also seemed offer the most easily pliable learning content to work into an intrinsic game design. Eventually the concept settled around one of the UK's National Curriculum targets for Key Stage 2 (7-11 year olds) based on number patterns and sequences:

> *Recognise and describe number patterns, including two- and three-digit multiples of 2, 5 and 10, recognising their patterns and using them to make predictions; recognise prime numbers up to 20 and square numbers up to 10 x 10; find factor pairs and all the prime factors of any two digit integer.*

The strong history of research into children's mathematical learning was another reason for choosing this focus for the game's learning content. Whilst the current trend in the game-based learning literature is often for more 'progressive' learning content (e.g. healthy eating, economics) the aims of this thesis are to compare contrasting design approaches – not test the suitability of learning content for games. The more thoroughly a learning domain is understood, the easier it is to examine the specific questions that concern this thesis without becoming sidetracked by other issues.

3.2.2 Core Mechanics

The Intrinsic Game

The initial concept was an action-adventure game, based around a combat mechanic in which the player must use different attacks to mathematically divide numbered opponents in hand-to-hand combat. The core mechanic could be described as "defeating enemies in combat by attacking each enemy with a divisor that divides their dividend into whole parts". Each of the player's attacks would have a different animation which embodies that divisor and reinforces the association between the divisor and attack. Archaic combat weapons have been chosen to illustrate these relationships here, but the same effect could have been achieved with futuristic or imaginary weapons:

- Divide by 2 – a single swipe of a sword.
- Divide by 3 – a barge with a triangular shield.
- Divide by 4 – two swipes of a sword.
- Divide by 5 – a punch with a (five-fingered) gauntlet.
- Divide by 6 – a single swipe of a sword and a barge with a triangular shield.
- Divide by 7 – a seven bladed discus which is thrown at the enemy.
- Divide by 8 – three swipes of a sword.
- Divide by 9 – two barges with a triangular shield.
- Divide by 10 – a single swipe of a sword and a punch with a gauntlet.

The structure of these attacks embodies additional mathematical relationships in the way that weapons combine: dividing by 8 is the same as dividing by 2, three times and therefore it is represented by three swipes of the sword. Dividing by 6 is the same as dividing by 2 and then by 3, so it is represented by a swipe of the sword and a barge with the shield. This added further depth to the core mechanic and helped the design to comply with the second part of the intrinsic design guidelines (that of embodiment).

Only three attacks would be available to the player at once, ensuring that there are always some enemies that cannot be defeated with the available attacks. Each game level would contain about twenty enemies, and the player would be told how many enemies are left on the level to 'divide'. Using an inappropriate attack against an enemy would result in the player losing health and when their health reaches zero they must 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.

This concept naturally lends itself to the use of number patterns associated with multiples of 2, 5 and 10, as these can help the player to make fast predictions about which opponents should be engaged in combat. This encompasses the National Curriculum objectives above, as well as providing scope for continuing these patterns for numbers that divide by 3, 4, 6, 7, 8 and 9.

An Extrinsic Equivalent

The extrinsic version of the game required a comparable core mechanic that replaced the mathematical element of the combat system with an equivalent one unrelated to division. In the intrinsic version, the dividend displayed on each skeleton's chest would provide the player with a way of determining its vulnerability to different attacks. The same result would be achieved in the extrinsic version by replacing the dividend with a symbolic representation of which attacks can divide that skeleton. These would be exactly the same weapons (divisors) as in the intrinsic version, but the mathematical relationship is hidden because the dividend is no longer displayed.

Using this system, a direct mapping was created between each dividend in the intrinsic version and one or more symbols in the extrinsic version. Table 3.1 shows how this works for a number of examples. Note that nearly all dividends can be represented in three or fewer symbols because many attacks are subsets of others. 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 four (two swords) and a divisor of two (one sword) as well. This has 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 the intrinsic version. It also means that for dividends within the range of 1-99 divided by divisors in the range of 1-10, only the numbers 60 and 90 need to be represented by more than three symbols.

Table 3.1. Examples of the symbolic equivalents of dividends.

Dividend	Potential Divisors	Symbols
22	2	sword
21	3	shield
19	none	none
6	2 and 3	sword + shield
16	2, 4 and 8	sword x 3
12	2, 3 and 4	sword x 2 + shield
30	2, 3, 5, 6 and 10	sword + shield + gauntlet
60	2, 3, 4, 5, 6 and 10	sword x 2 + shield + gauntlet
90	2, 3, 5, 6, 9 and 10	sword x 2 + shield x 2 + gauntlet

By using these symbols, the mathematical content would be removed from the core-mechanic of the extrinsic game. The same content would then be reintroduced at the end of each level in the form of a multiple-choice quiz. This quiz would require 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. The extrinsic version would thereby provide identical learning content delivered away from the flow-inducing gameplay, and presented as 'unembodied' abstract mathematical questions (as required by the guidelines).

Additional Mechanics

On its own the combat mechanic described this far might be dismissed as a drill and practice exercise. Nonetheless, this still would not affect the validity of using the game to explore the central question of motivation (and its effect on learning) that concerns this thesis. There is also significant research that suggests procedural competencies in mathematics need to develop alongside the conceptual (Rittle-Johnson, Siegler, & Alibali, 2001). However, Hibert and Lefevre (1986) suggest that rote learning in mathematics does not embody the meaningful relationships between knowledge that are essential for its successful understanding and application. Therefore a number of additional sub-mechanics were also added to include further potential for exploring relationships between mathematical knowledge contained in the game.

The interrelated structure of attacks in the game already embodies some additional 'meaningful relationships' within the mathematical mechanics of the game (dividing by 8 is the same as dividing by 2 three times, etc). However, Greer (1992) suggests that successful mathematics teaching needs to be rooted in an understanding of the real world situations that they serve to model. The real world 'anchors' of anchored instruction (Bransford et al., 1990) are also based on a similar perspective of mathematics teaching. Therefore this was included into the initial design by representing the physical division of opponents as the result of a successful attack. In practice this would mean that dividing an opponent by the divisor 2, would produce two opponents, half the size each with quotients half the size of the original dividend. Likewise, dividing an opponent by the divisor 8 would produce eight opponents, one-eighth of the physical size and quotient of the original opponent and its dividend. These smaller opponents would not usually present a new threat to the player and would disperse on their own. This feature would help to reinforce the link between the symbolic mathematical operations involved in division and the real world act of dividing physical objects into parts.

Squire and Bryant (2003b) identify the inverse relation between divisor and quotient as being critical to a better understanding of the division operation in mathematics. The design attempted to build the inverse relationship into the game mechanics in a meaningful way using giant-sized opponents. These would be physically larger than normal opponents, but (initially) defeated in the usual way. However, if the quotients of the resultant parts were greater than 10 then the new opponents would continue to attack the player with increased ferocity. In such situations the player would be told that they need to divide the giant into smaller pieces in order to avoid this happening. This would then set up the challenge to explore and discover the inverse relationship between the size of weapon divisors and the size of the resultant mini-opponent quotients.

One more mechanic was included in the initial game concept in order to try and build further relationships between mathematical content explored in the game. This equipped some opponents with weapons as a defence against the player's attacks. These were the same weapons that are available to the player, and each weapon would defend against any of the player's attacks that incorporated that weapon. So an opponent with a sword could parry attacks that divide them by 2, 4, 6, 8 and 10, one with a shield could parry attacks that divide them by 3, 6 and 9 and one with a gauntlet could parry attacks that divide them by 5 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 them. This feature would also help to highlight the numerical relationships between the combinations of weapons that are used in each attack.

3.2.3 Fantasy Context

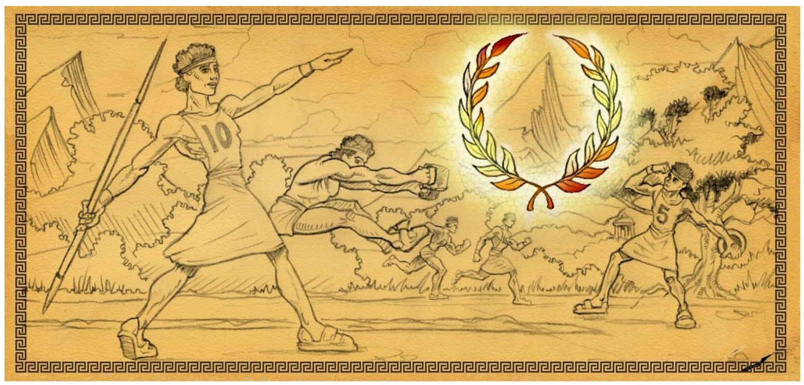
So far the game design has been described independently of the fantasy context in which it would be set. This helps to clarify the explanation of the game mechanics and also illustrates that core mechanics are independent of the fantasy context of the game (Habgood et al., 2005b). In reality the game's fantasy context was developed in parallel to the learning content and game mechanics. Nonetheless, it is important to note that the fantasy was tailored to fit the constraints of the game mechanics – not the other way around.

The fantasy context would be described to the player in an introductory sequence at the beginning of the game, accompanied by appropriate illustrations and narration. The final narration sequence is included here as an effective way of explaining the storyline to the reader:

This is the story of our adventure…

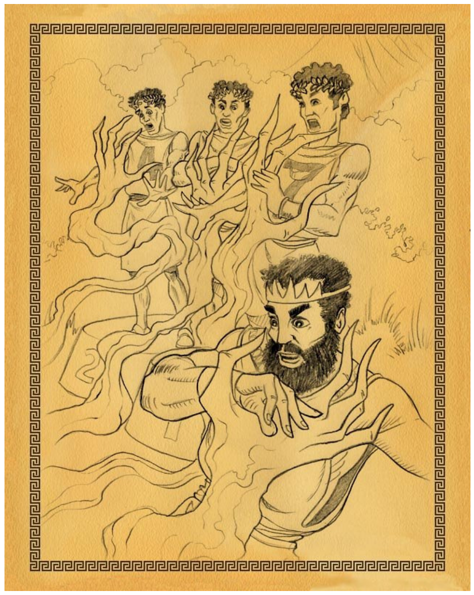
It is from a time when history was told through myths and legends, so I cannot say how much of it is true – but I hope you will find that there is still much to be learned, from a tale of heroes and dark creatures.

In ancient times, the kingdoms of Greece lived peacefully alongside each another and their peoples enjoyed happy and prosperous lives.

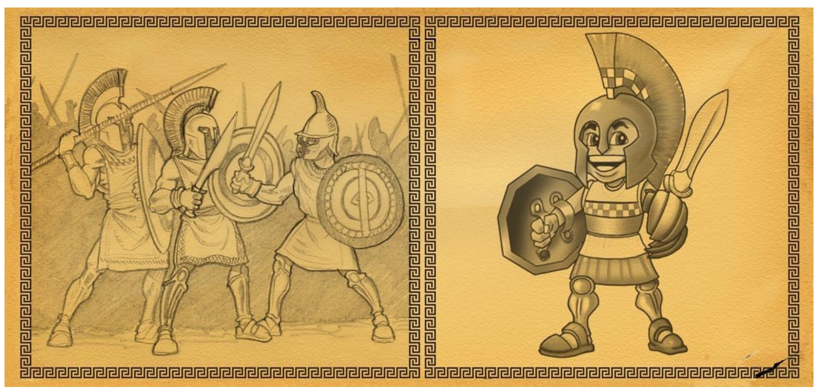


Every four years, each kingdom sent their finest athletes to compete in the Olympic games. The heroes of each event received a magical athlon, which gave them the courage and strength to maintain the peace.

However, there was one bad King who wanted all the Athlons for himself, so one year, he decided to cheat and use magic potions to help his athletes win every single event!



But instead of blessing his athletes with courage and strength, the Athlons cursed them all to a terrible fate, and they were never seen again.



Unfortunately, the Athlons disappeared as well, and without them the kingdoms fell into war. Yet the people live in hope that one day a hero will come, who will find the lost athlons and bring peace to the land once more…

The fantasy context was designed to engineer a scenario in which the player can face an indefinite series of numbered opponents in hand-to-hand combat. The cursed athletes referred to in the story would appear in the game as skeletal athletes, still wearing their competitor numbers on their rib cages (albeit in Arabic numerals). These numbers would also be the source of the Olympic curse, which is only dispelled when a skeleton's number is divided with magical attacks. The player would control a character called Matrices who must recover the lost Olympic athlons from the cursed King's spooky dungeons. In order to complete each dungeon all of the dividable skeletons on that level must be defeated using the weapons described in the mechanics section (sword, shield, gauntlet and discus). Matrices would also be given a magical hourglass, which stores the ashes of defeated skeletons and shows how many are left to divide on each level. The gods would enchant Matrices' weapons in different ways so that different attacks were available on each level. Matrices would be accompanied by Gargle – a clockwork gargoyle who provides useful game playing advice when it is needed.

Figure 3.2: from left to right: Gargle, Matrices and a reincarnated skeletal athlete.

3.3 GAME FEATURES

This section will detail the main aspects of the game design, highlighting the key differences between the intrinsic and extrinsic versions of the game. Where no differences are described, it can be assumed that the feature will remain identical between versions.

3.3.1 Weapons and Attacks

The weapons and attacks were central to the core mechanic of the game concept and would remain identical between the two versions. The player would have up to three different attacks available to him or her for defeating all the skeletons on each level. For ease of use, these attacks might have been assigned to three different keys in close proximity on the keyboard. However research by Sedighian and Westrom (1997) suggests that learning can be more effective when the user's interactions are directly linked to the concepts being taught. Therefore it was decided that attacks should be assigned to keys corresponding to the divisor that they represent. As the numerical keys do not include the number ten, the function keys were used for this purpose instead.

Iconic representations of these attacks (and the combinations of weapons from which they are comprised) would be displayed on a panel at the top of the screen as part of the in-game interface (see figure 3.3). The number of the function key that activates the attack would be displayed next to each icon (in the intrinsic version this is also the weapon's divisor). Attacks that were unavailable on the current level would be greyed out on the weapons panel. Each attack would have its own unique animation for Matrices showing him using the correct weapons in sequence to form the attack.

Figure 3.3: The weapons panel from the finished prototype.

From an early stage of development, the discus weapon (divisor 7) was considered a lower priority than the others. This was a relatively complicated game feature to implement (as the only ranged attack) and it was also considered one of the most difficult divisors for the children to learn (occurring at a later stage in the curriculum). Therefore as the game only needed to provide 2½ hours of gameplay, this feature was never actually implemented for the studies described in this thesis.

3.3.2 Health and Combat

In both versions of the game the player would be given three 'health points' that are represented by three hearts on the in-game interface. If the player used an incorrect attack against a skeleton then the skeleton would retaliate and the player loses one of these health points. If a player's health reaches zero then they would return to the start of the level and begin it again. This would encourage the player to think carefully about their attacks and avoid employing random strategies to divide skeletons. Random strategies could potentially undermine the learning content in the intrinsic version of the game and give children an easy way of 'gaming the system' (Baker et al., 2006). Nonetheless, this would still be a natural gaming mechanism that would not adversely affect the extrinsic version or seem out of place in either version of the game.

3.3.3 Skeletons

The chest panels on each skeleton would be the most significant (and obvious) difference between the two versions. Nonetheless skeletons in both versions would have both front and back chest panels in a large and bold style that can be clearly seen from a distance. Skeletons would also naturally turn to face towards the player as they approached, so that their panels would rarely be obscured from the players view. In the intrinsic version these panels would display the skeleton's dividend and the extrinsic version would display the attack symbols described previously (see figure 3.4).

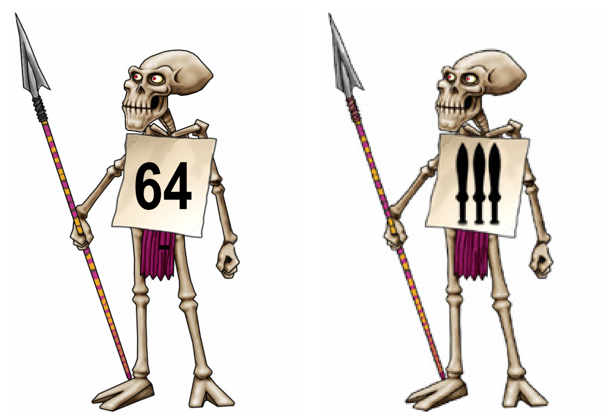


Figure 3.4: The same skeleton with its vulnerability to attacks represented numerically (left) and symbolically (right).

The behaviour of skeletons would be identical between versions, but there would also be four different kinds of skeleton behaviours encountered in the game:

1. ***Guards*** – these skeletons stand in one position and would only attack the player in retaliation to a failed attack. This provides the player with an indefinite period to consider their attacks before making them, and would be the most common type of skeleton encountered at the start of the game.

2. ***Patrols*** – these would follow a set path between different points in order to give the impression of patrolling the dungeon. They would also only attack in retaliation, but movement adds a greater sense of urgency to the situation.
3. ***Blockers*** – actively block a player from crossing an invisible line on the level. In addition to retaliating, blockers would also attack the player if they try and force their way past, knocking the player back beyond the invisible line.
4. ***Pursuers*** – these skeletons would actively pursue a player that is in the same room as them. In addition to retaliating, pursuers would automatically attack the player once they are close enough. These skeletons would only appear later in the game, initially as individuals, but later in small groups as well.

In order to maintain a fair comparison between the intrinsic and extrinsic versions, the number of skeletons on each level would also need to be balanced between versions. The extrinsic version would include a quiz in addition to the game levels, so each level would take longer to complete. Over time, this would mean that the extrinsic group would be exposed to less learning content than the intrinsic group. In order to keep this in balance, skeletons would be removed from the extrinsic version to keep the average time per level the same.

3.3.4 End of Level Quiz

The extrinsic version of the game would deliver its identical learning content in the form of a quiz at the end of each level. This would be a multiple-choice drill and practice exercise designed to deliver the content in a way that is directly comparable to the intrinsic version. The questions for each level would offer the player the choice of three different divisors with which to divide a given dividend. The quiz divisors would be identical to those of the weapons used on the same level in the intrinsic version, while the dividends would be identical to the dividends of the skeletons. A 'none of these' option would also be offered, comparable with choosing not to attack a skeleton at all in the intrinsic version (required when the number is indivisible). Each question would have to be answered correctly in order to proceed and the player would have to return to the start of the quiz after three incorrect answers. These 'chances' would be recorded on the screen in a similar way to the health scoring mechanism in the main game. In this way the players would be exposed to the same learning content in a form that is appropriate to the design guidelines for their condition.

3.3.5 The Hundred Square

Both the intrinsic and extrinsic versions would provide a hundred-square as a tool to support players with division. In the intrinsic version this would be displayed in game, while in the extrinsic version this would appear during the quiz (i.e. at the point at which it is relevant in each version). The hundred-square would highlight all the numbers between 1 and 100 that can be divided by each divisor. It would slowly cycle through all the available divisors on each level in order to reinforce the patterns and relationships that exist between different multiples (e.g. all numbers that can be divided by 2 end in 0, 2, 4, 6 or 8). It would also provide the player with a reliable means of looking up the answer when they were stuck. To integrate the number square into the game mechanics in the intrinsic version it would also be used to show the player which skeletons remain to be divided on each level. The extrinsic version would show the remaining skeletons using a two dimensional map showing their relative position to the player's current position (see figure 3.5). Divisible skeletons would be shown in blue and indivisible skeletons would be shown in red.

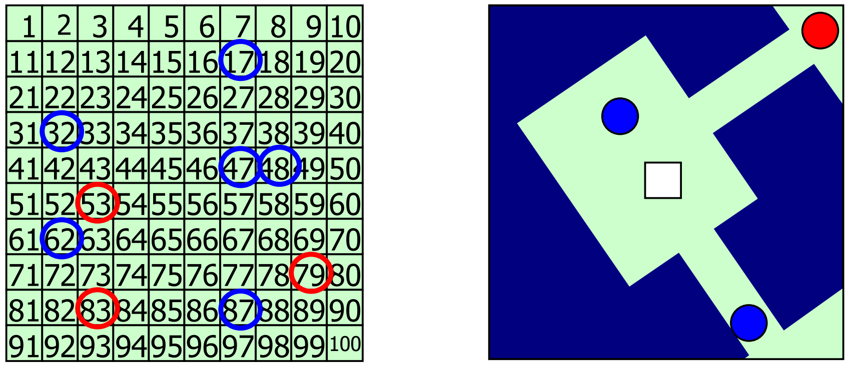


Figure 3.5: Different ways of showing the remaining skeletons on a level. The intrinsic hundred-square (left) and the extrinsic map (right).

3.3.6 Levels and Navigation

The levels and the positioning of skeletons within them would be identical between versions. The levels would be fairly simple in design, representing dungeons as a series of square rooms interconnected by corridors. There would be three different visual themes for the levels, getting gradually gloomier as the game advances in order to give the player a sense of progression. Players would navigate through the levels from a third person perspective using the arrow keys on the keyboard and the camera would follow them in a simple fashion.

3.3.7 Help and Tutoring

It was decided at this stage not to include tutoring mechanisms for the learning content, as this would have to be provided in both intrinsic and extrinsic versions (to be fair) and so wouldn't help to distinguish between versions. However, help and tutoring would be provided for the *gaming* content through the sidekick character, Gargle. An initial training level would be constructed where the player could fight passive 'clockwork' skeletons before meeting them for the first time in the game. Gargle would also provide oral instructions as to how to play the game, including keys and task direction.

3.3.8 Feature Comparison

To compare the similarities and differences of the three versions in detail, all of these features were documented in the tables shown below. These tables were then used as reference to construct the initial usability prototypes discussed in the next section:

Table 3.2. Comparison of environment features in each version of the game.

Ref	Type	Intrinsic	Extrinsic	Control
E1	Player	Greek warrior, Matrices	Identical	Identical
E2	Weapons	Sword, shield, gauntlet and discus	Identical	Identical
E3	Levels	Themed dungeon rooms (e.g. classical, slimy, spooky)	Identical	Identical
E4	Opponents	Reincarnated skeletal athletes with numbers on rib cages	Weapon symbols instead of numbers	Weapon symbols instead of numbers

Table 3.3. Comparison of screen features in each version of the game.

Ref	Type	Intrinsic	Extrinsic	Control
S1	Health	Shows remaining health points	Identical	Identical
S2	Target Indication	Remaining skeletons' dividends are highlighted on the 100 square.	2D map showing player's position relative to nearby skeletons.	2D map showing player's position relative to nearby skeletons.
S3	Attack buttons	Number of divisor/function key written next to weapon icons	Number of function key written next to weapon icons	Number of function key written next to weapon icons

Table 3.4. Comparison of gameplay features in each version of the game.

Ref	Type	Intrinsic	Extrinsic	Control
G1	Movement	Controlled using arrow keys	Identical	Identical
G2	Attack	Controlled using function keys	Identical	Identical
G3	Indication of appropriate attack	Number written on skeleton's chest	Symbols on chest	Symbols on chest
G4	Result of unsuccessful attack	Player knocked back and loses health. Sent back to last restart position when health reaches zero	Identical	Identical
G5	Result of destroying skeletons	Splits into a proportional number of smaller skeletons according to quotient and divisor. Usually run off but will stay and fight if they are large enough in size	Identical	Identical

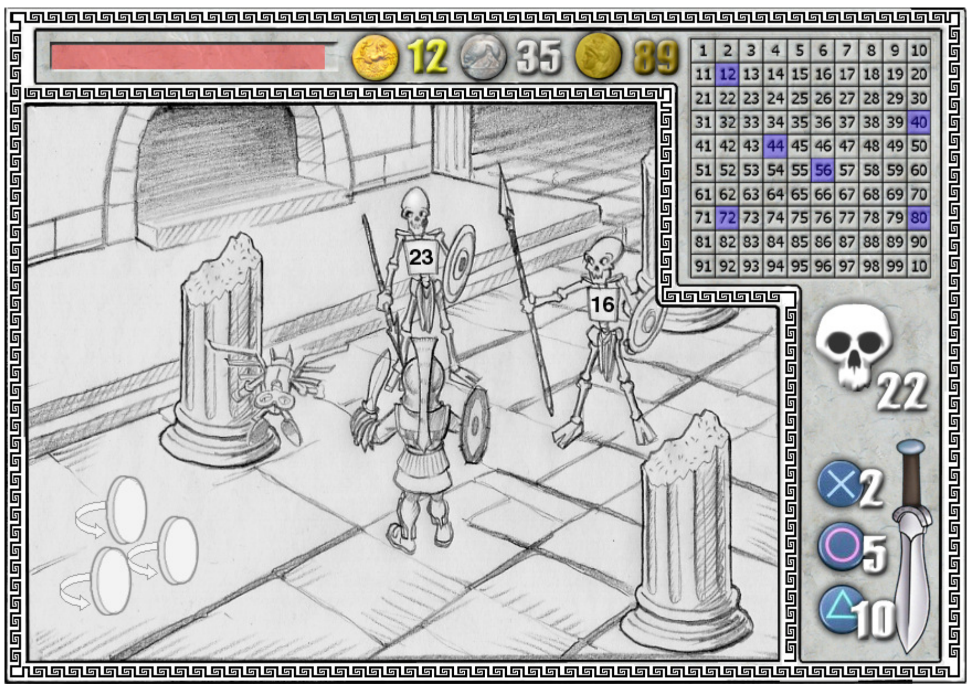
Table 3.5. Comparison of learning features in each version of the game.

Ref	Type	Intrinsic	Extrinsic	Control
L1	Division Context	Integrated with G3	Multiple-choice questions appear in place of 3D view. Identical dividends on each level	None
L2	Division	Integrated with G2	Controlled using mouse	None
L3	Hundred square representation	Integrated with S2. Deliberate number patterns are chosen for each level	Appears in place of map when questions are shown	None
L4	Incorrect choice	Integrated with G4	Told answer is incorrect and asked to choose again	None
L5	Correct choice	Integrated with G5	Told answer is correct	None
L6	Reinforcement of operation	Integrated with G5	Dividend, divisor and quotient are displayed	None

Table 3.6. Comparison of gameplay tutoring in each version of the game.

Ref	Type	Intrinsic	Extrinsic	Control
T1	Game play and controls	Tutored in moving player and attacking passive clockwork skeletons	Identical	Identical
T2	New weapon tutorial	Introduced to new weapons and the numbers that they divide	Introduced to new weapons and the symbols that they divide	Introduced to new weapons and the symbols that they divide

Figure 3.6: An early concept sketch of the game layout.



3.4 ITERATIVE DEVELOPMENT

The initial concept described above fed into an iterative design process that incorporated three stages of focus group testing with seven and eight year olds. The first of these used a low-tech paper prototype of the game, while the subsequent two used early playable versions of the computer game. This section details the outcomes of these focus groups and the effect they had on the final game design.

3.4.1 Participants

All the focus groups were held at the same primary school where the game-making study had been run in the previous school year. However, none of the children in this year group had been involved in the previous study. This was a large primary school on the outskirts of a large city in the north of England. The school is situated in a low-income area with a very small percentage of students having origins outside of the UK. The school has an average number of students with special educational needs, and the attainment of their intake is below the national average. The percentage of final year students achieving expected levels in mathematics was below national averages for the preceding year.

3.4.2 Paper Prototype

The paper prototype incorporated the game's core combat mechanic into a board game. This took the form of a deck of skeleton cards (complete with numbers on their chests) dispersed around a map of a dungeon (see figure 3.7). Players were given a combination of three attack cards and told that they had to defeat skeletons in order to win keys and progress to the next level. Furthermore, in order to defeat a skeleton they needed to be able to divide its number by one of the numbers on their attack cards. In order to indicate that they wished to attack a particular skeleton, players were asked to touch the skeleton with one of their attack cards. If the attack was successful then the researcher would turn over the skeleton card and give the player a key. If the attack failed then the researcher would remove one of the player's five life points (represented by heart tokens), and when they had no life points left they had to start the level again.

On each level there were six keys to collect from a total of ten skeletons, but four skeletons on each level could not be divided with the available attacks. Both giant skeletons (that divided into smaller skeletons) and parrying skeletons (armed with weapons) were also included in the levels in various combinations. For each level the player was given a number square with the dividable skeletons circled on it. They were also given acetate overlays for each attack that could be placed over the number squares to highlight the numbers each attack could divide. The game concluded after six levels with an end of game boss in the form of a giant skeletal spider. Each of the spider's legs had its own number that had to be divided to finally defeat the creature and finish the game.

Figure 3.7: The paper prototype

The prototype was individually played with ten children across a spectrum of ability over two days of sessions. The reaction of the children to the game was extremely positive, regardless of their preference and ability in mathematics. All the children were clearly motivated by the task and only one child didn't reach the end of the prepared levels. All the children seemed to demonstrate an understanding of the core combat mechanic and were able to relate this to their mathematical knowledge in at least the simplest cases (dividing by 2 and 10). The understanding of the other mechanics was more difficult to interpret. Parrying skeletons (skeletons that parry attacks where their weapon is contained within that attack) seemed to be understood by most of the boys, but created some confusion amongst some of the girls. Many of the boys in this class regularly played fantasy card games such as Yu-Gi-Oh, and so were probably more used to picking up card game rules of this kind. The results for giant skeletons (skeletons that divide into smaller skeletons if their quotient after an attack is more than 10) were more ambiguous. Although most children appeared to approach them in the correct way (by dividing them by the highest divisor), this was also usually the easiest divisor (i.e. 10) making it hard to gauge their true understanding. The success of the number squares and acetate overlays was mixed, but nonetheless many of the children did refer to them at some point while playing the game.

The paper prototype fed into the final stages of creating the concept design described in the previous section, and as such was responsible for creating rather than modifying this design. Nonetheless, the overall success of this initial prototype was encouraging enough to begin work on a software prototype in order to trial these features in more detail.

3.4.3 Software Trial 1

The first version of the software prototype included only the intrinsic version of the game, with three complete levels and attacks for the divisors 2, 5, and 10. During early development dividing skeletons into smaller skeletons was found to put too large a graphical load on the system. Therefore the small skeletons (made up of many polygons) were replaced by small ghosts instead (made up of just one polygon). These ghosts wore their division quotients in the same way as the mini skeletons but slowly faded away. However when a giant was divided and the quotient was larger than 10, the ghosts would drift down to the floor and rise again as new skeletons. The levels contained some giant skeletons, but the skeletons with parrying weapons had not yet been implemented. There was also no training mode at this stage and the trial was used as an opportunity to develop effective ways of explaining the game to children.

Participants in this trial were initially asked to take a draft paper pre-test designed to test their mathematical understanding of problems related to the learning content in the game. This test contained twenty multiple-choice questions that were read aloud to each child before playing the game. Twelve participants across the range of abilities took part in the trials, which were conducted on laptops during the course of a normal school day. Only two of the participants had already played the paper prototype, but all of them finished the three levels and completed the multiple-choice test.

Qualitative observations were made about the software based on closely observing each child as they played the game. The main findings were related to the basic playability of the game. Children found it quite difficult to navigate the dungeons and kept accidentally returning to places they had already visited. The game's simple environment meant that all the rooms tended to look very similar once the skeletons had been cleared. This made it easy to get lost in areas of the dungeon that the player had visited before. The children also had difficulty avoiding combat with skeletons that could not be divided. This was because children would walk up to skeletons before working out that they could not be defeated. Unfortunately, standing this close to a skeleton caused the player's character to enter "combat mode" where the character automatically faces towards his opponent, and so tends to orbit around it in a circular motion. The player could only disengage by walking backwards away from the skeleton, something that was not obvious to most players who usually ended up attacking the skeleton out of desperation instead.

Both of these observations fed directly back into the development of the game. The navigation problem was addressed by including locked doors that require a key to open, but lock shut again behind the player once they walk through. These doors were then used to divide the dungeons up into manageable chunks with the key being released once the last skeleton in each area was defeated. The combat problem was addressed by reducing the distance at which the player's character would enter combat mode and by adding a timeout that automatically deactivated combat mode after a couple of seconds if the player hadn't used any attacks.

A design change was also made in the number square as a result of the observations in this trial. The children had been told that dividable skeletons were circled in blue and undividable skeletons were circled in red, so they were using the number grid to work out which skeletons should be avoided. However, they were performing this as an automated look-up – disregarding the possibility of trying to work it out for themselves, and certainly oblivious to the mathematical relationships afforded by the number square. This meant that they were not only ignoring the benefits of the number square, but that they had one of the most complicated mathematical tasks in the game automatically performed for them. Realising that 42 divides by two is just a case of identifying it as an even number, but realising that 41 does not divide 2, 5 or 10 requires relating a number of different kinds of mathematical information. Therefore the circles were removed at this stage as they only served to allow children to subvert the mathematical goals of the game. This also meant that it was no longer necessary to implement the equivalent map-based system for the extrinsic version (see figure 3.5).

The most significant insights into the learning content of the game actually came from the draft pre-test used in this trial. Several different ways of phrasing the multiple-choice problems were trialled in order to see which the children most easily understood. For example:

- Which of these numbers is in the 2 times table?
- Which times table is 8 a member of?
- Which of these numbers can be shared equally between 5 groups?
- How many groups can 25 be shared equally between?
- Which of these numbers can be divided by 10?
- Which of these numbers can 80 be divided by?

All of these problems can be quickly solved with a basic knowledge of the 2, 5 and 10 times tables, but they seemed to produce very different levels of success with many children. It quickly became apparent that there were some fundamental differences in the underlying assumptions behind these questions, and thus the children's ability to solve them. In many cases, even though children knew the relevant times tables facts, they were unable to apply this knowledge to help them solve the problems phrased in terms of division or grouping. To apply their knowledge of times tables to division requires children to appreciate that multiplication is the opposite of division and some had not yet come to this realisation. The 'easiest' form of phrasing seemed to be the one that directly referred to the relevance of times tables in solving the answer. However, even here some of the poorest students were unable to relate their understanding of "counting in 2's, 5's and 10's" (which they could all do when prompted) to solving problems with times tables.

Research by Greer (1992) suggests that these problems also embody conceptual differences between partiative and quotiative division (sharing into a known number of groups to find the size of each group, and sharing into groups of known size to find the number of groups). This further adds to the number of conceptual insights required to solve some of the pre-test problems. It was decided that partiative division questions should be used in the final test as this seemed to represent the middle level of difficulty and mapped most closely to the act of dividing skeletons. It was also decided that emphasis should be given to the inverse relationship between multiplication and division by referring to the number square as the "magic book of times tables" which helps us to "divide evil creatures".

Significant changes were also made to the way in which skeletons divided into smaller (quotient) ghosts. In the initial version of the prototype, ghosts just appeared as skeletons died. However, there wasn't a clear link between the division of a skeleton's number and the appearance of ghosts bearing the quotient of the division operation. An animation sequence was therefore added to reinforce this link. This showed the 'spirit' of the defeated skeleton separating from its body as a whole unit. This whole spirit would then split into equal sized portions, (depending on the divisor) and each one would grow into a small ghost bearing the quotient. It was hoped that by making the physical operation of division more explicit it would help players to link the symbolic mathematical operations and the real world act of dividing physical objects into parts.

3.4.4 Software Trial 2

The second prototype included both intrinsic and extrinsic versions of the game and had a full training level and five finished dungeon levels. It incorporated all of the attacks except for 7, 8 and 9 and contained skeletons that could parry with a sword. This version also included the additional locked doors and a refined combat mode added as a result of the first prototype. Twelve children took part in this trial, playing alternating versions of the game (intrinsic and extrinsic) on a pair of laptops over two days. Seven of these children had not played the game before and played all the way from the training level through to the end of level five. One boy was unable to get past level three after failing to listen to any instructions provided by the researcher or the game, and was reluctantly (on their part) moved on to free up the laptop for another child. The four participants who had played the game before were timed completing the same levels in both the intrinsic and extrinsic versions (in a random order), in order to obtain an estimate of the amount of content children were likely to get through in each version.

The training level seemed to be very effective at teaching the children the basics of how to play the game, and with the one exception, all the children managed to complete this level with almost no additional help. The subsequent three levels introduced the first three attacks (2, 5 and 10) by giving the player one more attack on each level. As each new attack was introduced, Gargle would explain that the player could now divide skeletons by the new divisor (or in the extrinsic version, with a new symbol) using the appropriate function key. An image of this key would also be displayed on the screen at the same time. However, some children became stuck after thinking that the new attack was a replacement for the old one and not an additional attack. As a result, the help speech was changed in order to reinforce that these were additional weapons, and images of all the currently active weapon keys were displayed on the screen when new weapons were added.

In this trial session it also became apparent that the giant skeletons were not having the intended effect. At the start of the game children paid some attention to the ghosts in the game, but they were gradually ignored as it became apparent that they did not (usually) play a significant role in the game. Therefore when they encountered giants later on, most children did not realise that they were being attacked by new skeletons as a result of the size of the ghosts they had created. Consequently in the final version of the game the camera was always made to cut away and focus on the skeletons, whilst Gargle explained why the new skeletons were being created.

The multiple-choice question and answer session in the extrinsic version worked as planned, although some children needed prompting to switch to a mouse-based control method at this point. As a result, a spoken prompt from Gargle was added to remind them of this in the final version of the game.

3.4.5 Pre/Post-Test Trial

In addition to trialling the game prototype, the paper and pencil pre/post-test was also tested on the same class of children. The first full-scale study was planned to take place in a different school, using a class of seven year olds *and* a class of eight year olds. This class only included seven year olds so only the top set of this class took part in the trial to account for the extra ability of the older children. Twenty-five children were given twenty minutes to answer twenty multiple-choice questions and were allowed to ask for help reading any of the questions. All of the children attempted all of the questions within the allotted time and there were no significant misunderstandings raised. The children's mean score was 11.64 and the distribution was negatively skewed as a result of high scores. Consequently any question that over 70% of the children got correct was either changed or replaced with harder questions in the final version of the test.

3.5 SUMMARY

Using the theoretical guidelines from chapter two, a mathematics game was designed that could support both intrinsic and extrinsic approaches using identical learning content. This would provide the practical basis for a series of interventions that would empirically examine our alternative hypotheses about the value of the intrinsic approach. Designing the game was an iterative process that was heavily informed by trials with children from an early paper prototype, right through to the software implementation of the game. This iterative process would not stop here, and both the software and methodology used in the interventions continued to be refined throughout the interventions described in the remainder of this thesis.

CHAPTER FOUR

Study 1: Evaluating Learning

This chapter describes the first empirical study undertaken to evaluate Zombie Division. This study was designed to examine if our theoretical hypotheses about the intrinsic game would result in a more or less effective learning experience when compared to the extrinsic equivalent. Therefore study one compared learning outcomes for the intrinsic, extrinsic and control versions of the game over a fixed amount of time-on-task.

4.1 AIMS

The educational interest in digital games is founded in the apparent engagement power that they possess. The manifestation of this engagement often shares many similarities with the concept of flow (Csikszentmihalyi, 1988) in terms of total concentration, distorted sense of time, and extension of self. It also epitomizes the definition of intrinsic motivation as "an activity with no apparent rewards except the activity itself" (Deci, 1971). Furthermore, it has been proposed that intrinsic motivation leads to deeper learning (Biggs, 1987), which may suggest that the motivational benefits of games could create a greater level of engagement and a deeper connection with learning content.

The concept of intrinsic integration (Habgood et al., 2005b; Kafai, 2001; Malone, 1981; Reiber, 1996) suggests that the engagement power of digital games may be most usefully harnessed for educational aims by more closely integrating a game with its learning content. Edutainment titles are criticised for including games as separate extrinsic motivation or reward for completing learning content (Lepper & Malone, 1987). Habgood et al (2005b) suggest that this integration should be between a game's core mechanics and its learning content, in order to "embody the learning material within the structure of the gaming world and the player's interactions with it" (p. 494). However they also acknowledge that creating an intrinsic integration in this way will not necessarily create more effective learning.

One way in which intrinsic integration could produce less effective learning than an extrinsic approach, is through the role of transfer. "Transfer between tasks is related to the degree to which they share common elements" (Bransford et al., 2000, p.78). For this reason research has shown that children find it difficult to transfer mathematical skills between different contexts (Nunes et al., 1993). The embedded nature of intrinsic learning content may therefore create a very specific context, which makes it harder to transfer learning to the different context of a classroom.

A second reason relates to the potential conflict between flow and reflective learning processes. Reflection is often considered to play an essential role in developing metacognitive skills that are critical to effective learning (e.g. Bransford et al., 2000, p.97). However, many games provide intensely interactive 'twitch-speed' (Prensky, 2001) experiences that seem unlikely to promote contemplative reflective process such as these.

The primary aim of this study was therefore to compare the relative learning gains for intrinsic and extrinsic versions of Zombie Division in order to provide some evidence for the potential benefits or drawbacks of the intrinsic approach. Although the central benefit of motivation may be to produce greater time-on-task, this first study aimed to provide a like-for-like evaluation of learning potential. Therefore, learning gains were compared over a fixed amount of time-on-task, against a control-group version of the game without the educational content. These learning gains were used as a measure of the relative educational effectiveness of intrinsic and extrinsic approaches for Zombie Division. Detailed process data was also collected to help support the main findings and provide insights into the effectiveness of other aspects of the game's design.

4.2 METHOD

4.2.1 Design

This study used a two-factor mixed design. The first factor, 'group' was between-subjects, giving each child one of the three different versions of the game to play (intrinsic, extrinsic or control). The second factor, 'test' was within-subjects, providing two repeated measures of learning outcome for each child in paper-based pre and post-tests. Sixty-six children were assigned to one of the three conditions based on matched pre-test scores. Children were sorted according to their scores and allocated alternately to each condition using a randomised block design to ensure an equal balance of gender and year group within each condition. Two children were away during the study week, so twenty-two children were ultimately assigned to the intrinsic condition, twenty-one to the extrinsic and twenty-one to the control.

4.2.2 Participants

All children attended a primary school on the outskirts of a medium-sized city in central England. The school is situated in a low-income area with a very small percentage of students having origins outside of the UK. There are an above average number of students with special educational needs, and the attainment of their intake is around the national average. The percentage of final year students achieving expected levels in mathematics was just above national averages for the preceding year.

Sixty-four primary children between the ages of 7 years 6 months and 9 years 7 months took part in the study week (mean 8 years 8 months). This was made up of one complete class of thirty mixed-ability year 3 children and another complete class of thirty-four mixed ability year 4 children. There were thirty-two girls and thirty-two boys in total, all of whom had some prior experience of using the computers in the school's ICT suite. The study was carried out over the last three weeks of the final term in the school year.

4.2.3 Materials

Facilities

The tests and interventions were both carried out within the school's ICT suite. The suite contained twenty, relatively new PCs running Windows 2000 with accelerated 3D graphics support, and audio output through stereo headphones. Two additional laptops were also provided to make up the numbers for the large size of the groups. Each machine ran the study's software via its own CD-ROM disk, and saved process data onto a floppy disk.

Test Materials

The paper and pencil test consisted of 20 multiple choice questions with four options in each case (generally 1 correct + 3 distractors). This portion of the test was identical for both pre and post-tests. These were all division questions mostly comprised of two recurring formats, either asking the child to select the divisor that divides a given dividend (dividend-based):

e.g. Circle one number that 45 can be divided by:

4, 6, 9 or none of these

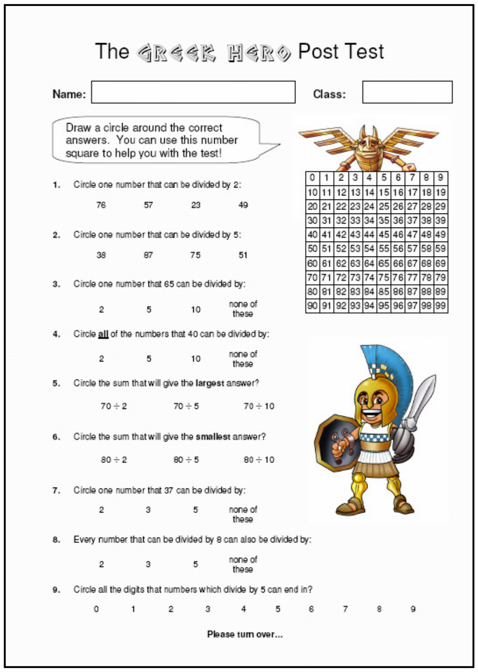
Or to select the dividend that can be divided by a given divisor (divisor-based):

e.g. Circle one number that can be divided by 5:

35, 13, 29 or 41.

About a third of the questions were divisor-based, a third were dividend-based and a third were conceptual questions. Conceptual questions were included to test for knowledge of heuristic patterns (e.g. all numbers that divide by 5 end in 0 or 5) as well as understanding the relationships between divisors, or for applying rules outside of normal limits (i.e. dividends greater than 10 times the divisor). The test began with a few of the easiest questions and then continued in a random difficulty order. Children were given up to twenty minutes to answer all the questions under test conditions appropriate to this age group. A hundred square was printed alongside the questions – similar to the one found in the intrinsic and extrinsic versions of the game (see figure 4.1)

The post-test paper also contained three multiple-choice questions about each child's use of technology. These were included to allow correlations between technology use and various other competencies measured during the intervention to be explored. These asked children to rate the frequency with which they used different technologies at home as either "never", "sometimes", "often" or "always". The three types of technology were games consoles, edutainment and general computer use.



The GREEK HERO Post Test

Name: | **Class:**

Draw a circle around the correct answers. You can use this number square to help you with the test!

0	1	2	3	4	5	6	7	8	9
10	11	12	13	14	15	16	17	18	19
20	21	22	23	24	25	26	27	28	29
30	31	32	33	34	35	36	37	38	39
40	41	42	43	44	45	46	47	48	49
50	51	52	53	54	55	56	57	58	59
60	61	62	63	64	65	66	67	68	69
70	71	72	73	74	75	76	77	78	79
80	81	82	83	84	85	86	87	88	89
90	91	92	93	94	95	96	97	98	99

1. Circle one number that can be divided by 2:
   76 57 23 49
2. Circle one number that can be divided by 5:
   38 87 75 51
3. Circle one number that 65 can be divided by:
   2 5 10 none of these
4. Circle **all** of the numbers that 40 can be divided by:
   2 5 10 none of these
5. Circle the sum that will give the **largest** answer?
   $70 \div 2$ $70 \div 5$ $70 \div 10$
6. Circle the sum that will give the **smallest** answer?
   $80 \div 2$ $80 \div 5$ $80 \div 10$
7. Circle one number that 37 can be divided by:
   2 3 5 none of these
8. Every number that can be divided by 8 can also be divided by:
   2 3 5 none of these
9. Circle all the digits that numbers which divide by 5 can end in?
   0 1 2 3 4 5 6 7 8 9

Please turn over...

Figure 4.1: The pencil and paper test.

e.g. How often do you play PlayStation, Gamecube or X-Box games at home?

Never, Sometimes , Often, Always

4.2.4 Procedure

Pre-tests

The pre-tests were carried out on the Friday of the week before the study. The children completed the pre-test in their class groups at approximately the same time. The task was explained to children by their class teacher, emphasising the presence of the hundred-square to help them with the test, without explaining how it was used. They were given a 20-minute time limit, which allowed all but one child to attempt every question. Children were also told that they could ask the teacher for help with reading any of the questions in the test, but not with the mathematics.

Game Intervention (100 Minutes)

The children first played Zombie Division on the Monday morning of the study week, three days after the pre-test. Throughout the study, each playing session lasted for approximately twenty minutes, 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 9:30am and the last group finishing at 11:00. This meant that the last group on each day missed a portion of their morning break, but the children consented to this arrangement. Each child's position in the game was saved at the end of each playing session and the game restarted them from the beginning of their current level in the next playing session. The children played the game for the whole week until they had accumulated a total of one hundred minutes playing time. At this point the software automatically stopped the game and the child was sent back to their class. A number of catch-up sessions were run for absentees to ensure that all children had played for their allotted time before the end of the week.

Post Tests

The post-tests were performed on the Monday morning of the week following the intervention. These were carried out in the same way to the pre-tests, with the additional questions on computer use. Children were given an additional 10 minutes (30 minutes in total) to answer these extra questions.

Measures

Each test provided a score out of twenty. The post-test included the same score with an additional three questions about technology use. In addition, the games also logged a large body of process data for each subject, providing a time stamped 'commentary' over the entire length of the intervention.

4.3 RESULTS

The results section for this study is divided into five sub-sections based around the different data sources available. The first section uses the pre and post-test results to examine the impact of different versions of the game on overall learning outcomes. The second section reports the results obtained from the game process data including outcomes of learning tasks and a breakdown of scores for each divisor. The results from the outcome and process data are compared in the third section as an indication of transfer between the game and tests. The fourth section examines the children's profile data for correlations that may help to explain the effect that differences in technology use have on the game. The final section provides an analysis of profile data and learning outcomes with respect to gender, to examine its role in this relationship.

4.3.1 Learning Outcomes

A total of three children were removed from the analysis of learning outcomes. One extrinsic subject was removed from the study completely as a result of a technical mix-up that switched them to playing the intrinsic version of the game half way through the intervention. Another two children (one extrinsic and one control) were removed from the learning outcome data, after failing to turn over their post-test paper to answer the questions on the other side. However, the process data for these two children was not affected by this oversight and so they are included in later analyses. There were also two absentees for the study week, both of which were in the control group. The groups were therefore rearranged on the day in order to rebalance the numbers in each group. Unfortunately these changes adversely affected the mean group scores, so that the control group's average pre-test scores were slightly (but not significantly) lower than the others.

Numerical scores were examined using a two-way repeated measures ANOVA with two levels of the within-subjects factor 'test' (pre and post) and three levels of the between-subjects factor 'group' (intrinsic, extrinsic and control). The analysis of numerical scores (table 4.1) revealed a significant effect of test ($F(1,58) = 5.87$, $MSE = 24.41$, $p < 0.05$, $\eta^2 = 0.09$) but no main effect of group, or group by test interaction. Overall numerical scores at post-test were therefore significantly higher than numerical scores at pre-test.

Table 4.1. Mean total numerical scores by group and test

	Intrinsic (n=22)	Extrinsic (n=19)	Control (n=20)
Pre-test (max of 20)	10.55 (4.73)	10.58 (4.34)	8.90 (3.49)
Post-test (max of 20)	11.50 (4.69)	12.26 (4.69)	8.95 (3.61)

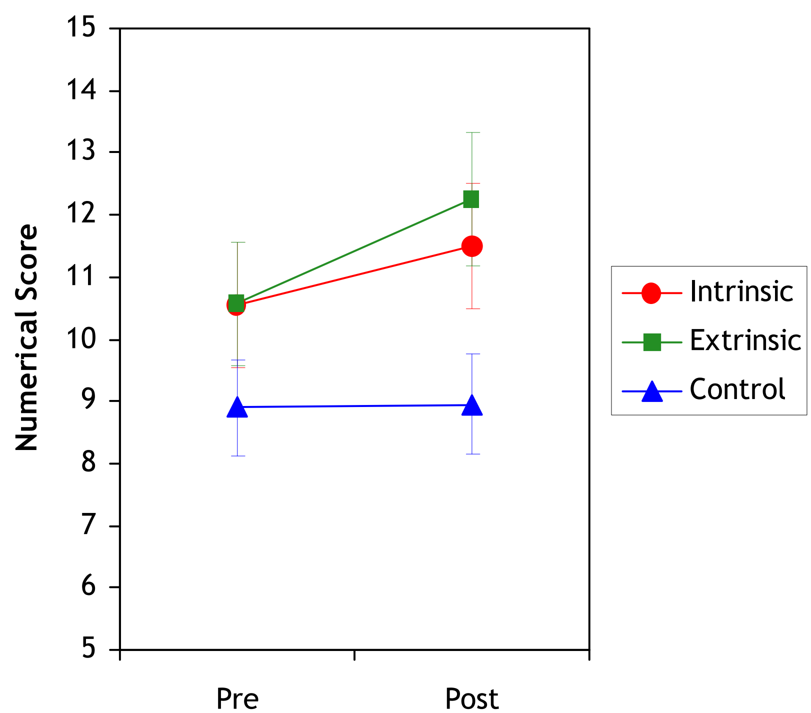


Figure 4.2. Mean total numerical scores by group and test

The numerical scores were further subdivided into dividend-based, divisor-based and conceptual questions in order to examine learning within these three sub-categories. These three measures were examined using a two-way repeated measures MANOVA with two levels of the within-subjects factor 'test' (pre and post) and three levels of the between-subjects factor 'group' (intrinsic, extrinsic and control). Multivariate tests showed a significant main effect of 'test' ($F(3,56) = 5.21$, $p< 0.05$, $\eta^2 = 0.22$), but no main effect of 'group' or test by group interaction. Univariate tests revealed this effect was only significant for conceptual questions ($F(1,58) = 1.05$, $MSE = 10.14$, $p< 0.05$, $\eta^2 = 0.15$). This suggests that taken as a whole, the children in this study only improved significantly in their performance in the conceptual questions.

Table 4.2. Mean categorised scores by group and test

Category		Intrinsic (n=22)	Extrinsic (n=19)	Control (n=20)
Dividend-based score out of 7	Pre-test	2.82 (1.87)	3.11 (2.00)	2.50 (1.47)
	Post-test	3.32 (1.76)	3.84 (1.74)	2.50 (1.10)
Divisor-based score out of 6	Pre-test	4.23 (1.71)	3.89 (1.49)	3.10 (1.62)
	Post-test	4.09 (1.44)	4.00 (1.80)	2.85 (1.79)
Conceptual score out of 7	Pre-test	3.50 (1.71)	3.58 (1.64)	3.30 (1.81)
	Post-test	4.09 (2.07)	4.42 (1.84)	3.60 (1.73)

Correlations with Learning

Bivariate correlations were also performed between measures of each subject's mathematical ability (pre-test score), total learning gains (post-test score - pre-test score) and the final game level they reached during the study. Only pre-test scores were significantly correlated with level reached ($r(61) = 0.41$, $p < 0.01$). This suggests that prior mathematical ability had an impact on children's progress in the game, but the absence of any correlation with learning gains suggests that the children's progress in the game did not have an impact on their learning.

4.3.2 Learning Process

The process logs produced by the game provide a versatile source of data in the form of a time-stamped commentary on the game as it is being played. For this study, over one thousand log files were mined for the purposes of post-hoc analysis. Our hypotheses suggested that children might find it more difficult to transfer their learning from the intrinsic game to the test than from the extrinsic. The mining therefore focussed on obtaining broadly comparable data for 'skeletal divisions' and 'quiz divisions' from the intrinsic and extrinsic versions of the game. A comparison between in-game and test accuracy scores should help to identify any potential near-transfer problems in the different versions of the game. This should provide a first indication of whether the experimental methodology should be developed to allow a more detailed analysis of transfer.

Game Task Learning Outcomes

It is relatively easy to interpret the extrinsic 'quiz' data in terms of accuracy scores, but the free-roaming nature of skeletal encounters in the intrinsic game makes it much more complex and open to interpretation. In particular, determining the beginning and end of an encounter with a skeleton requires assumptions about the user's engagement with division problems that may not always be true. For the purposes of this analysis an encounter was said to begin when the player enters the same room as a skeleton, and said to end if the player attacks the skeleton or leaves the room again without dividing it. The different outcomes of encounters can be therefore be defined as follows:

We will define a *maths task* as a single encounter with a divisor-based problem in the game. A maths 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 maths task is a dividend-based question with a choice of up to three divisors to divide a given dividend. The choice of divisors remains the same throughout all maths tasks on the same game level. 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 the divisors provided if none of them would divide the dividend. In the intrinsic version this involves manoeuvring to avoid combat with the skeleton, while in the extrinsic version a player selects an alternative answer marked 'none of these'. We will refer to dividends that are dividable by one of the available divisors as *target dividends,* and those that cannot as *distractor dividends*. A maths task is not concluded until either it is answered correctly, or the level/quiz is restarted.

Several attempts can be made during one encounter, and subsequent encounters with the same maths task will occur when a) a level or quiz is restarted after a player loses all their health or b) when the player returns to the start of the level after loading a saved game. Nonetheless, the outcome of a maths task is assessed here in terms of the first attempt made upon the dividend in the player's first encounter with each maths task in the game. This is because the pre/post-tests do not include lives and always accept the first answer given. Therefore, to allow meaningful comparisons to be made between game and test process data, only the attempts in the first encounters with each maths task are generally included in the analyses. This means there can be one of six outcomes depending on whether the dividend is a target or distractor:

Correct Outcomes:

1. Target Attacked Correctly (TAC) – the player correctly divides the dividend by one of the available divisors on the first attempt.
2. Distractor Left Correctly (DLC) – the player correctly rejects all of the available dividends for an indivisible dividend on the first attempt.

Incorrect Outcomes:

3. Target Attacked Incorrectly (TAIN) – the player incorrectly attempts to divide the dividend by one of the available divisors on the first attempt.
4. Target Left Incorrectly (TLIN) – the player rejects all of the available divisors for a dividable dividend on the first attempt.
5. Distractor Attacked Incorrectly (DAIN) – incorrectly attempts to divide an indivisible dividend by one of the available divisors on the first attempt.

Undefined Outcomes:

6. Undefined (U) – A restarted level or an error such as a crash or a glitch in the log causes an encounter to be unresolved.

Table 4.3a shows the overall mean number of maths tasks for *all encounters*, as well as the mean number of first encounters with maths tasks, and the mean accuracy of first encounters with maths tasks. Table 4.3b shows a breakdown of the mean number of first encounters with maths tasks according to their outcomes. The figure for first encounter accuracy from table 4.3a is the sum of both correct outcomes over the total number of first encounters.

A one-way MANOVA was performed on all the measures in tables 4.3a and 4.3b with two levels of the between-subjects factor, 'group' (intrinsic and extrinsic). This revealed a significant overall effect of group at the multivariate level ($F(1,33) = 7.41$, $p < 0.001$, $\eta^2 = 0.64$). The univariate analyses then showed a significant effect of group on the following measures: targets attacked incorrectly ($F(1,40) = 10.82$, $MSE = 98.62$, $p < 0.005$, $\eta^2 = 0.21$) and targets left incorrectly ($F(1,40) = 17.81$, $MSE = 791.29$, $p < 0.001$, $\eta^2 = 0.31$). Targets attacked incorrectly and targets left incorrectly were both significantly higher in the intrinsic group. These measures suggest the intrinsic group was less successful at in-game maths tasks than the extrinsic group.

Table 4.3a. Mean number and percentage accuracy of maths tasks by group.

	Total encounters with maths tasks	1^{st} encounters with maths tasks	Accuracy of 1^{st} encounters
Intrinsic (n=22)	363.68 (108.89)	180.72 (56.01)	81.10 (6.73)
Extrinsic (n=20)	298.55 (104.36)	177.15 (85.71)	83.93 (12.30)
Combined (n=42)	332.67 (110.47)	179.02 (70.81)	82.45 (9.77)

Table 4.3b. Mean no. of first encounters with maths tasks by group and outcome.

	Targets Attacked Correctly	Distractors Left Correctly	Targets Attacked Incorrectly	Targets Left Incorrectly	Distractors Attacked Incorrectly	Undefined
Intrinsic (n=22)	127.31 (42.97)	20.05 (7.02)	7.32 (3.73)	14.59 (8.40)	6.55 (3.58)	4.91 (6.62)
Extrinsic (n=20)	133.35 (69.60)	22.70 (12.94)	4.25 (1.94)	5.90 (3.93)	7.30 (3.39)	3.65 (3.38)
Combined (n=42)	130.19 (56.57)	21.31 (10.23)	5.86 (3.36)	10.45 (7.91)	6.90 (3.47)	4.31 (4.88)

<u>Game Task Analysis of Divisors</u>

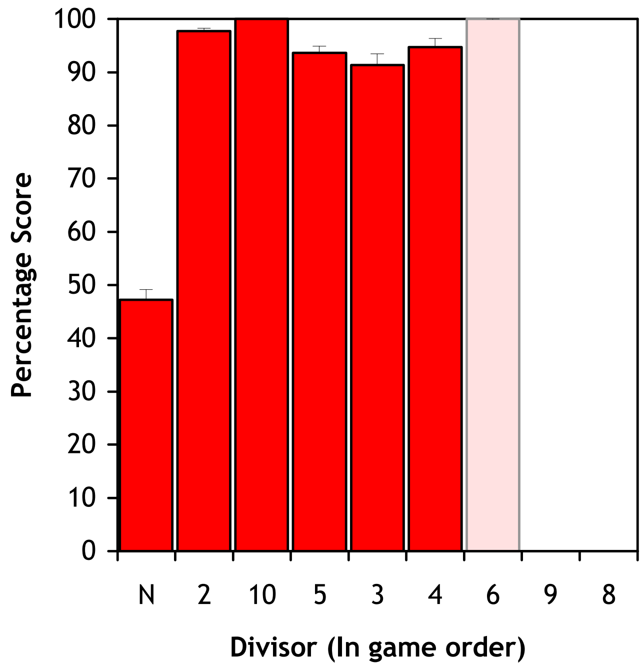
The previous analysis provides an overview of the children's mathematical success while playing the game, but the same process data can be used to provide a more detailed analysis broken down by divisor. This allows us to find out the proportion of new mathematical content encountered during the interventions and the relative success of the groups for new and familiar content. Table 4.4 shows the combined group breakdown of divisors used in first encounters with maths tasks. Just over 58% of all maths tasks were solved using the divisors 2, 5 and 10. The corresponding multiplication tables are part of the national curriculum requirements for children of this age, so only about 42% of the overall learning content in the games could be considered as unfamiliar to this audience.

Table 4.4. Breakdown of divisors used in first encounters with maths tasks.

	None (n=42)	2 (n=42)	10 (n=40)	5 (n=41)	3 (n=36)	4 (n=24)	6 (n=11)	9 (n=1)	Total
Number of first Encounters	1526	2467	539	1349	766	554	118	19	7338*
Proportion of total 1st encounters	20.30	32.81	7.17	17.94	10.19	7.37	1.57	0.25	97.6*

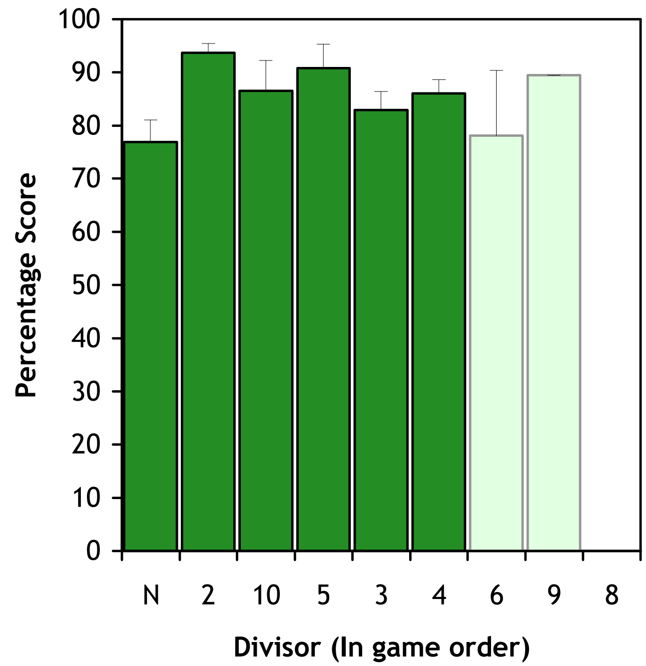
Divisors listed in the order they appear in the games (none = indivisible).
* Remaining encounters are undefined

Figures 4.3 and 4.4 show a breakdown of the percentage accuracy of in-game maths tasks by divisor for the intrinsic and extrinsic groups. The number of samples in a multivariate analysis is limited to the minimum number of samples available for every divisor included in the test. Therefore only the divisors with 50% or more participants contributing from each group (none, two, ten, five, three and four) were included in the analysis. A one-way MANOVA was performed on these measures. This had two levels of the between-subjects factor, 'group' (intrinsic and extrinsic) which was significant at the multivariate level ($F(1,16) = 26.35$, $p < 0.001$, $\eta^2 = 0.91$). Univariate effects were present for the divisors none ($F(1,21) = 142.33$, $MSE = 6666.49$, $p < 0.001$, $\eta^2 = 0.87$), five ($F(1,21) = 6.85$, $MSE = 94.41$, $p < 0.05$, $\eta^2 = 0.25$), three ($F(1,21) = 8.70$, $p < 0.01$, $\eta^2 = 0.29$) and four ($F(1,21) = 7.36$, $p < 0.05$, $\eta^2 = 0.26$). The extrinsic group had significantly higher percentage scores for the divisors 'none' and five, and the intrinsic group have significantly higher scores for the divisors three and four. This suggests that the intrinsic group were more successful at the less familiar learning content and the extrinsic group were more successful at the familiar learning content.



N = No divisor; Faded for < 50% of participants

Figure 4.3. Mean in-game percentage scores by divisor for the intrinsic group (skeletons).



N = No divisor; Faded for < 50% of participants

Figure 4.4. Mean in-game percentage scores by divisor for the extrinsic group (quiz).

4.3.3 Comparing Outcome and Process Data

Comparing the learning outcome scores and learning process scores provides a means of examining the issue of transfer. The in-game accuracy scores and dividend-based questions in the test were based on the same type of mathematical problems in different contexts. If children achieved significantly higher scores in the game than in the post-test, this could indicate that the learning is not transferring between these contexts effectively. A two-way repeated measures ANOVA was performed between the accuracy of 1st encounters in-game and percentage values for dividend-based questions derived from the post-test score data (table 4.5). This had two levels of the within-subjects factor 'method' (game and assessment) and two levels of the between-subjects factor, 'group' (intrinsic and extrinsic). This revealed a significant overall effect of method ($F(1,39) = 73.48$, $MSE = 20006.45$, $p < 0.001$, $\eta^2 = 0.65$) but no significant interaction with group. Therefore both groups may have experienced problems transferring their learning from the contexts of the games to the post-test.

Table 4.5. Test and game accuracy for dividend-based questions by group

	Percentage score for dividend-based questions at post-test	Accuracy of 1st encounters
Intrinsic (n=22)	47.40 (25.09)	81.10 (6.73)
Extrinsic (n=19)	54.89 (24.86)	83.84 (12.63)
Combined (n=42)	50.87 (24.95)	82.37 (9.88)

4.3.4 Profile Data

The profile data provided background information on each subject relating to their use of technology in the home. Children rated themselves on a score of 0-3 according to the frequency with which they used game consoles, edutainment software and home PCs (never, sometimes, often, always). We were particularly interested in how children's prior experience with technology might impact their performance and learning with the games, so these measures were analysed for correlations with measures of each subject's mathematical ability (pre-test score) total learning gains (post-test score - pre-test score) and the final game level they reached during the study.

Bivariate correlations were performed on these six measures. Edutainment frequency was found to be significantly correlated with PC frequency ($r(63) = 0.33$, $p< 0.01$), console frequency was significantly correlated with level reached ($r(63) = 0.26$, $p< 0.05$), and pre-test scores were significantly correlated with level reached ($r(61) = 0.41$, $p< 0.01$). This suggests that both prior console experience and mathematical ability had an impact on children's progress in the game, but the absence of correlations with learning gains suggests that none of these measures had an impact on their learning.

4.3.5 Gender Differences

Playing video games has traditionally been a male dominated activity, so the role of gender generates a significant amount of interest with respect to research based on games. Therefore, although it is not the central focus of this thesis it seemed pertinent to examine parts of the data for gender differences. The received wisdom would suggest that boys might perform better in learning activities related to games, because they have more experience with them in other contexts. We can test this by re-examining the profile data with respect to gender. Table 4.6 shows the mean and standard deviation of each of these values by gender. These scores were examined using a one-way MANOVA with two levels of the between-subjects factor 'gender'. A significant multivariate effect of gender was found ($F(1,53) = 4.10$, $p < 0.005$, $\eta^2 = 0.35$) with significant univariate effects for the measures: console frequency ($F(1,59) = 5.73$, $MSE = 4.93$, $p < 0.05$, $\eta^2 = 0.09$), edutainment frequency ($F(1,59) = 4.08$, $MSE = 2.86$, $p < 0.05$, $\eta^2 = 0.07$), computer frequency ($F(1,59) = 6.84$, $MSE = 4.55$, $p < 0.01$, $\eta^2 = 0.10$), and level reached ($F(1,59) = 13.81$, $MSE = 146.11$, $p < 0.001$, $\eta^2 = 0.19$). Console frequency and level reached were significantly higher for boys while computer and edutainment frequency were significantly higher for girls.

Table 4.6. Mean scores by gender.

	Male (n=31)	Female (n=30)
Console freq	1.94 (0.96)	1.37 (0.89)
Edutainment freq	1.00 (0.82)	1.43 (0.86)
Computer freq	1.39 (0.84)	1.93 (0.79)
Pre-test score	10.68 (4.29)	9.33 (4.15)
Learning gain	1.23 (2.54)	0.53 (3.26)
Level reached	11.13 (3.56)	8.03 (2.91)

To see whether these differences had an impact on the children's learning with the game, we can re-examine overall learning gains, for gender differences moderated by condition. Table 4.7 shows the numerical scores for each test by group and gender. These were examined using three-way repeated measures

ANOVA with three levels of the within-subjects factor 'test' (pre, post and delayed), three levels of the between-subjects factor 'group' (intrinsic, extrinsic and control) and two levels of the between-subjects factor, 'gender'. No significant interaction effect of gender was found for numerical scores.

Table 4.7. Mean and S.D. of numerical scores by group, gender and test

	Intrinsic		Extrinsic		Control	
	Male (n=11)	Female (n=11)	Male (n=8)	Female (n=11)	Male (n=11)	Female (n=9)
Pre-test	11.09 (4.21)	10.00 (5.35)	11.73 (4.67)	9.00 (3.51)	8.89 (3.76)	8.91 (3.45)
Post-test	12.45 (5.59)	10.55 (3.59)	12.82 (5.33)	11.50 (3.85)	11.90 (4.78)	9.87 (4.04)

4.4 DISCUSSION

The central aim of this study was to evaluate the educational effectiveness of intrinsic and extrinsic approaches for a fixed amount of time-on-task. It was designed to provide evidence to help reject one or more of our conflicting hypotheses about the value of intrinsic games. One hypothesis predicts greater learning gains in the intrinsic version as a result of a motivational engagement and a deeper connection with the learning content. However, the other two hypotheses predict smaller learning gains: the first as a result of embedded intrinsic learning content transferring less effectively than the extrinsic; and the second because integrating learning content within a game's flow-experience may impede reflection-in-action.

4.4.1 Learning Outcomes

Children in the study as a whole demonstrated learning between pre and post-test, but no one group learned significantly more than any other. As expected, the control group demonstrated a negligible average increase in score (0.05 points), but the relatively large standard deviations in all groups (4.27 points) meant that the increases in the intrinsic (0.95 points) and extrinsic (1.68 points) groups were not large enough to distinguish themselves from each other or the control. Analysis of learning outcomes in different categories of questions (dividend-based, divisor-based and conceptual) confirmed the overall learning effect, but did not reveal any differences between groups either. Interestingly this analysis also showed that the learning between pre and post-test was only significant for conceptual questions.

These results appear to indicate that the high pre-test scores were affecting the scope for detecting improvement. The questions in the testing system had already been made more difficult as the result of a trial run with a class of a similar age group. However, the mean score in this study was still 10.9 out of a maximum of 20 (55%), with over a quarter of participants achieving a score of 14 or above (70%). This creates poor conditions for observing learning gains, with some children having little scope for improvement. Although an overall improvement was observed, this improvement was not significant for the easiest, dividend and divisor-based questions, but it was significant for the hardest conceptual questions. This is not surprising when you consider that over 20% of children achieved full marks in the divisor-based category at pre-test. This further suggests that the scope for improvement in children's scores may have masked any significant differences between groups.

The data was also analysed for correlations between learning outcomes, prior mathematical ability (pre-test score) and level progress. No correlation was found between learning gains and level progress either, but this is not necessarily surprising given the apparent overall difficulty in detecting improvement caused by the high pre-test scores.

Unfortunately this makes it difficult to come to any conclusions about our hypotheses using the data for learning outcomes. The intrinsic version did not produce the lower learning gains predicted by the adverse transfer or reflection theories, or the higher learning gains predicted by the deep learning theory. However, it is very unlikely that the absence of a significant difference between the intrinsic and extrinsic groups offers genuine evidence against any of these theories. After all, no significant difference was detected between the experimental and control groups either.

Ultimately, it is the very small learning gains that are most critical to the failure of this study to meet its primary aims. The game, methodology and instruments have simply not demonstrated a large enough improvement in learning in order to reach any conclusions based on the data. It is therefore important that these shortcomings are addressed in future studies before useful data can be obtained from playing the games.

4.4.2 Learning Process

The process data showed significant differences in the outcome of maths task encounters between the intrinsic and extrinsic groups. The intrinsic group made more mistakes dividing skeletons, both in terms of using incorrect divisors (targets attacked incorrectly) and leaving skeletons that can be divided (targets left incorrectly). However, the latter may be explained by the relatively uncertain nature of identifying 'left skeletons' when compared to the absolute certainty of equivalent quiz questions answered as 'none of these'. A lack of action on the part of the intrinsic player does not necessarily mean they have considered or even noticed the skeleton, but the extrinsic player is forced to consider and answer each quiz question in turn. Therefore the data will tend to overestimate the number of 'targets left incorrectly' in the intrinsic version.

The more detailed analysis of in-game accuracy by divisor revealed that less than half (42%) of the learning content experienced in the study could be considered new to children of this age. Learning models based on assimilation and accommodation (Piaget, 1950) suggest that cognitive conflict is necessary to make children move from familiar strategies (e.g. reciting times tables) to new ones (applying the hundred-square). New divisors may represent the ideal situation for creating this conflict because they cannot be assimilated into the children's existing schemas (most don't know their 3 times tables). Therefore the balance of learning content may benefit from a lower frequency of familiar divisors (2, 5 and 10) in future studies. Interestingly the analysis of divisors also showed that the intrinsic group were more successful at the less familiar learning content (divisors 3 and 4), while the extrinsic were more successful at some of the more familiar content (the nominal divisor 'none of these' and the divisor 5). Therefore this could suggest that the intrinsic game is actually more successful at creating the necessary cognitive conflict to alter the children's strategies.

4.4.3 Comparing Outcomes and Process

Mean accuracy scores in-game were over 30 percentage points higher than scores for comparable dividend-based content in the post-test. This could certainly be taken as an indication of a significant problem transferring learning from both versions of the game to the test situation. However, the tests were conducted in the children's normal class environment, whereas the games were played in the relatively unusual and appealing ICT suite. Therefore motivational differences could explain some or all of this contrast in scores, and future studies would benefit from a methodology that addresses this. However, it is notable that the results of this study *do not* indicate that the intrinsic game creates greater transfer issues than the extrinsic. Furthermore, the high in-game scores also offer some vindication of the value of games (in general) as motivating contexts for learning – or at the very least for demonstrating existing knowledge.

4.4.4 Profile Data

Reported prior use of gaming consoles did predict children's progress in the games (as measured by level reached), but did not predict their learning outcomes. Pre-test scores were also found to predict level reached suggesting that overall performance in the game was a result of both gaming and mathematical skills. However, the overall difficulty detecting improvement in this study means it would be wrong to read too much into the learning outcome data.

4.4.5 Gender Differences

The profile results seem to support the idea that there are significant gender differences in technology use. The analysis found that boys reported that they used games consoles at home significantly more often than girls, while girls reported that they used computers and edutainment products significantly more than boys. This supports our preconceptions about the male-dominance of video gaming as a leisure activity. Boys also reached significantly higher levels in the game than the girls, which concurs with the correlation observed between console frequency and level reached. However, these gender differences do not appear to have affected the learning outcomes of the study with no significant difference between genders in the learning gains achieved during the study. Although overall learning outcomes were small, this may also suggest that the Zombie Division game does not present girls with a significant disadvantage for learning.

4.4.6 Implications for Reflection and Deep Learning

The process data provide limited scope for interpreting children's reflective behaviour during the course of the game. Nonetheless, observations at the time seemed to suggest that many children managed to remain oblivious to the intended focus of the game's learning content (applying the hundred-square to solving division problems). The most common discussions were about comparing level status, and about how to find out what level they had reached (as this was somewhat hidden). Children were regularly observed helping each other, but usually this involved telling another child the answers directly, rather than reflecting upon the rules and mathematical systems of the games. Many of the children also seemed to ignore the explanations and advice offered by the game's guide character. Very few (if any) children were observed using the hundred-square in the way that the game advocates. All this leaves the impression that the 'flow' of the game had indeed distracted them from the aims of the learning content. However, there was no obvious difference between the intrinsic and extrinsic groups in their (lack of) reflective behaviours, with just as little attention being paid to the character and hundred-square in both. So although empirical evidence is not available, this study certainly left the impression that both versions of the game failed to promote any kind of useful reflective behaviour on the mathematical content.

The study does not provide any obvious way of examining the depth of learning created by different versions of the game either. A delayed post-test might provide some indication of the strength of the children's connection with the learning content taught by the game. However, a repeat of the test was not feasible due to the proximity of the study to the end of the school year.

Nonetheless, our hypothesis attributes any deeper learning to greater motivation and this study was able to highlight a key flaw in the design of the study relating to motivation. All versions of the game used a save-game feature that only recorded the level reached by a child at the end of each playing session – not their exact position in the game. In the extrinsic version this meant children returned to the start of the gaming content at the beginning of each playing session: ensuring they always had a complete level of motivating game play before being exposed to a test. Later on in the study this meant that some children were still being rewarded with the motivation of playing the game despite failing to progress with the learning content. Conversely, this also meant that the extrinsic group was (theoretically) at a learning disadvantage because none of these repeated gaming levels contained any learning content (although in practice one half-hour playing session was long enough for most children in the extrinsic group to complete at least one level, including the quiz[6]). Nonetheless, the save feature does make it difficult to defend the design of this study as a fair comparison between the two approaches. Future studies should therefore include a save game feature that saves the player's exact position and state within the game, in order to redress this.

[6] After all 5 playing sessions the mean level in the extrinsic group was 8.9.

4.5 CONCLUSIONS

This study has provided some evidence to support the idea that children may experience problems transferring learning between the gaming and testing contexts. However, there were no apparent differences between the intrinsic and extrinsic groups and it is still unclear what other motivational or contextual differences may be influencing the high in-game scores. Unfortunately small learning gains and limited scope for improvement from pre-test scores have prevented the study from providing useful learning outcome data. Therefore this study has not been able to provide conclusive evidence to support or reject any of the three initial hypotheses about intrinsic games.

Nonetheless, the high accuracy of in-game scores demonstrated the value of Zombie Division as a potentially motivating context for learning. The study has also provided valuable insights into the testing instrument, learning content, save-game mechanism and intervention environment that can be used to inform the future development of the software and methodology. Incorporating these changes will help to ensure that future studies will be able to detect any significant differences in the learning outcomes of the two design approaches.

CHAPTER FIVE

Changes to Game and Methodology

This chapter describes the second phase in the design and development of Zombie Division. Each study instigated developments in the game design and experimental methodology, but the changes made between study one and study two were much more than just iterative adjustments. A number of actual and potential failings were identified as part of the experience of the first experimental trial of the game. This chapter provides a review of the practical outcomes of study one in terms of how they affected the continued direction of the project. It details all the changes made to the game design and methodology as well as the ongoing theoretical arguments for making these changes.

5.1 TESTING

The main failing of study one was the testing instrument used to measure children's learning gains from the intervention. This was developed and refined using a group of children who turned out to be unrepresentative of the participants in the final study – despite being of a similar age group. The paper and pencil test was also relatively short (just 20 questions) and so only provided a limited buffer for any variance between groups. The scores recorded in study one were very high, limiting the scope for improvement and potentially concealing any interesting differences between groups. It was also apparent from the children's workings that many were using paper and pencil division strategies that contrasted with the methods they were using in the game. As a result, the ability of these tests to provide a representative measure of the children's learning from study one was highly questionable. Therefore the most important change that needed to be addressed following study one was in the way in which the tests were administered.

5.1.1 Computer Based Test

It was decided that the best way to address the shortcomings of the paper and pencil testing system was to create a computer-based equivalent. Principally this was because a computer-based system could administer a larger potential number of questions for the same time period. Trials of the initial paper-based system (and teacher feedback) had suggested that a test of more than twenty questions might prove too long for some children – particularly the second time around (i.e. at post-test). Consequently, the test length was kept short in case children with less mathematical knowledge refused to engage with a test that they thought they could not complete. However, a computer-based test would allow a larger total number of questions to be available without revealing this to the children. From the children's perspective it would be a test of how many questions they could answer correctly in a set time limit, rather than a test with a fixed length. This would then provide significant extra scope for the best performing students to improve on their scores without de-motivating the others.

Although increasing the test length was the principle reason for creating the computer-based testing system, it could also provide a number of beneficial side effects. The computer-based format would help to discourage the children from employing pencil and paper mathematical methods in the test that were different from those they were using in the game. Automating the testing system also meant that marking (and eventually grouping) could be performed automatically by the software. Another useful side effect was that it was possible to record and analyse the time it took each child to answer every question in each test.

Comparisons could then be made between group averages and successive tests in order to gain a deeper insight into the development of children's learning from the intervention.

<u>The Software Implementation</u>

Figure 5.1 shows the final testing system that was implemented and used in the remaining studies of this thesis. Unlike other aspects of the game and methodology, this did not change at all over subsequent studies. The test was comprised of sixty-three multiple-choice questions (three times as many as the paper-based test) to be completed within a 15-minute time limit. Each question had four options with just one correct answer and three distractors. The test began with three practice questions, which were not assessed or timed, in order to ease students into the test:

e.g. Select the number of legs that a dog has:

4, 5, 6 or 7

Of the remaining 60 questions, 40 were division questions equally comprised of two recurring formats, either asking the child to select the divisor that divides a given dividend (dividend-based):

e.g. Select one number that 45 can be divided by:

4, 6, 9 or none of these

Or to select the dividend that can be divided by a given divisor (divisor-based):

e.g. Select one number that can be divided by 5:

35, 13, 29 or 41

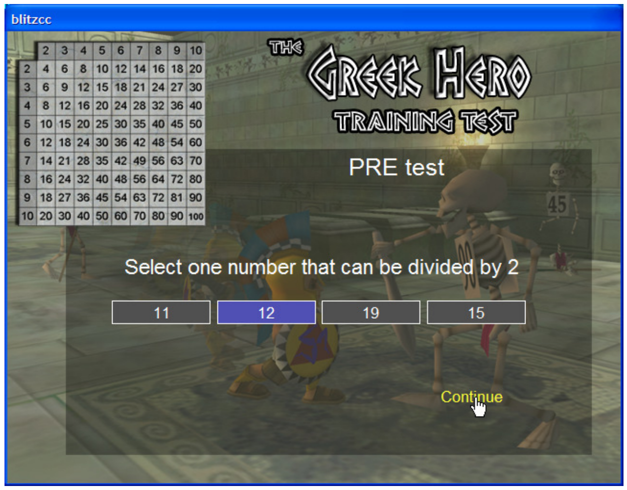


Figure 5.1: The computer-based test.

Amongst these 40 questions there were 5 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 addition to these 40 questions another 5 dividend-based questions were added where the answer was 'none of these'.

The remaining 15 questions were conceptual: 3 tested for knowledge of the heuristic patterns associated with numbers that divide by 2, 5 and 10; and 12 more tested for an understanding of relationships between divisors, or for applying rules outside of normal limits (i.e. dividends greater than 100).

The order of the questions was initially randomised, but remained consistent between-subjects and between tests. The software timed 15 minutes from the start of question 4 and automatically stopped the test at the end of this period. However, the time was not displayed on the screen and no feedback was provided on the choices made. Providing feedback could introduce repeat effects in the data as children learned from the testing system – rather than from the intervention. A grid was also displayed, in the corner of the screen, similar to the one found in the intrinsic and extrinsic versions of the game (see figure 5.1)

5.2 LEARNING CONTENT

Our observations from study one strongly suggested that children did not use the number-grid to help them to identify mathematical patterns and solve division problems. The original game design had tried to integrate the grid as part of the gameplay, by using it as a way of showing the remaining dividable and undividable skeletons on each level. However, this was quickly removed after it became apparent that this merely provided a way of avoiding undividable skeletons without considering the mathematics. Nonetheless, this did make the number grid a redundant feature to the gameplay, with little or no real incentive for children to refer to it during play. Furthermore, questions and comments made by some children suggested they did not see the relevance of the grid to solving division problems. Class teachers involved in the study also commented that the grid may represent one too many levels of abstraction from the children's classroom experience of multiplication and division. They recommended using a multiplication grid as an alternative, as their greater familiarity with it might help them to see its relevance to the game's learning content.

5.2.1 More Paper Prototyping

Research by Nunes and Bryant (1996) stresses the importance that framing mathematical problems has on children's choice of appropriate procedures. It is unclear exactly what mathematical procedures children were using to solve division problems in study one, but it seems unlikely that these procedures were structured or reinforced by the number grid. While a multiplication grid may prove more relevant, it was important to gain some insight into the kind of procedures that children actually use before committing to this fundamental change. This led us to return to the classroom with the paper prototype. The number grid was replaced by a multiplication grid showing all multiples within the range of 2x2 to

10x10. The dividends displayed on skeletons' chests were also limited to within the range of 10 times each divisor so that the answers could always be read from the multiplication table. In order to frame the division tasks within a concrete context, children were given the following explanation of combat:

Each skeleton has a number on its chest which shows the number of magic bones it has. To defeat these creatures you must use your weapons to divide these magic bones into groups of equal size. This restores their magical balance so that their souls can finally rest in peace. However, if you try to divide them into unequal groups they attack you with increased strength!

To reinforce this context they were also provided with a pile of 'magic bones' which they could use to help them solve division problems in the game. The game was played with a number of children, specifically asking them to explain or show their working as they decided which skeletons to attack with which weapons. Very few of the children naturally used the multiplication table to help them solve the division problems – despite being reminded that it was there to do just that.

Children's approaches to dividing skeletons were highly variable. Figure 5.2 shows some of the different approaches the children took to dividing skeletons and the mathematical concepts that relate to those approaches. Some children used quantitative division strategies, employing repeated subtraction to keep taking the divisor from the pile of bones until they could do so no more. Many children employed partiative division to try and share the bones equally into a number of piles corresponding to the divisor. Some applied heuristic rules, incorporating number patterns and sequences for well known divisors to help them. However, none of these strategies directly relates to the mathematical concept of an inverse relationship between multiplication and division. Children that understood this relationship were able to use their knowledge of multiplication-tables to help them solve division problems. Many did this from memory but most were generally able to make the transition to using the multiplication grid to help them with divisors, when they did not know their multiplication tables for that divisor.

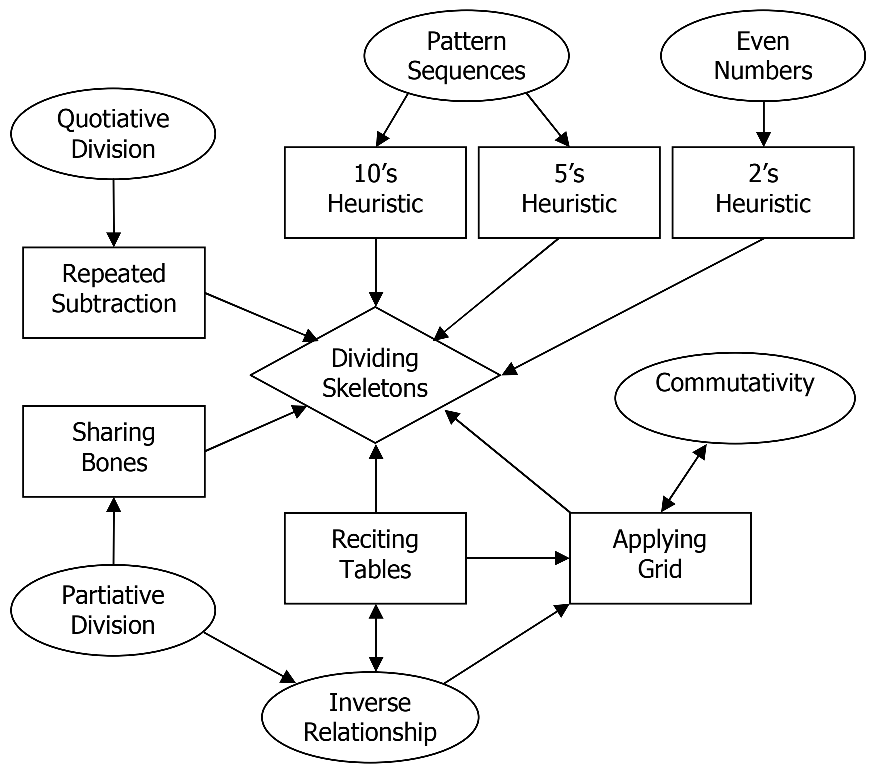


Figure 5.2: A diagrammatic representation of the range of concepts (ovals) and procedures (rectangles) applied to the mathematics of dividing skeletons by children in the trials. This illustrates the wide range of prior knowledge that different children applied to Zombie Division, and how this framed their initial mathematical approaches to the game.

It was apparent that even children with very similar school experiences could approach the same learning content from very different perspectives. Indeed, Bransford, Brown & Cocking (2000) suggest that people come to formal education with "a range of prior knowledge, skills, beliefs and concepts that significantly influence what they notice about the environment and how they organise and interpret it". Research by Siegler and Jenkins (1989) also shows that children continue to switch between different mathematical strategies long after they have more efficient strategies at their disposal. Nonetheless, some of the procedures chosen by the children could potentially obstruct their ability to learn from the game, for example: quotiative division strategies using repeated subtraction, rather than partition (sharing). The difference may seem subtle from an adult perspective, but research by Squire and Bryant (2003a) highlights the importance of using partiative procedures in children's understanding of division. Crucially for us, repeated subtraction is conceptually different from multiplication and division and so less likely to naturally steer players towards applying the inverse relationship between the two.

Other procedures were sometimes problematic too. Many children attempted to apply heuristic procedures for the divisors 2, 5 and 10, but mistakes were common. These included wrongly identifying even numbers, thinking that only numbers ending in 5 divide by 5 or believing that numbers that began with a 1 would divide by ten. Nonetheless, even children who could apply these heuristics correctly were often left without a working procedure when they advanced to levels which included divisors for which they didn't have heuristics. Overall, most children seemed either unaware of, or unable to apply, the inverse relation between multiplication and division, and so they did not naturally see the relevance of the multiplication grid.

Putting these observations alongside the results from the first study, we decided to adopt a more structured approach to the learning content in the game. It was clear that within the context of Zombie Division, simply providing children with appropriate mathematical tools and opportunity to learn from them was not enough. Unless the game was framed within an appropriate mathematical context, it was inevitable that many children would fail to make measurable learning gains within our limited time constraints.

We returned to the classroom with the paper prototype once more in order to develop this new approach. This time skeletal division was framed in terms of magic bones as before, but with additional support to scaffold the children's adoption of the multiplication grid procedure. So once a few skeletons had been successfully divided by counting bones, each child was shown how they could use their knowledge of multiplication tables as a shortcut to the counting process. After applying their tables from memory a few times they were shown how to use the multiplication grid to work out how to divide any skeleton. In this way the framing of the problem was broken up into logical steps which were introduced at the point at which their relevance was most obvious to the child. This process was developed and refined from one child to the next in order to find a scaffolding system that could be implemented in the game. The level of support varied greatly between children, but it was not difficult to get most children applying the multiplication grid without further prompting in a short period of time (15-20 minutes). This was then used as the basis of the framing system that formed the main addition to the leaning content added to the game.

5.2.2 A Framing System

The in-game framing system was designed to mirror the improvised scaffolding process that had proved successful alongside the paper prototype. This can be viewed as a very basic form of contingent tutoring system (H. Wood & Wood, 1999). The system would be contingent insofar as it would have a minimal presence unless unsuccessful attempts were made to divide skeletons. Some mandatory framing content would be provided at the start of the game and when new divisors were introduced, but otherwise the system would only intervene when its support was required. It would also be possible for the player to voluntarily choose to activate the system at any time during play if desired. When the system was activated by choice it would be possible to quit out of the system without completing the process, but otherwise it would have to be successfully followed through before continuing with the game.

The system was set within the context of a thought-bubble that appeared over the top of the main playing area. Separating the system from the main game in this way was not a deliberate choice, but development constraints meant that a fully-integrated 3D system was not feasible at this stage. The process was split into a number of steps which provided the player with two separate paths, or levels of framing (see figure 5.3). The first path (stages marked with the letter A) involved sharing bones into piles, and was the only path available at the start of the game. This concluded by showing the player the division expression that resulted from the process of sharing they had just performed manually. This reinforced the concrete context of the division task and provided players with a very slow, but reliable way of testing whether a skeleton could be defeated using a particular weapon.

The second path (stages marked with the letter B) showed the player how to find a skeleton's dividend on the multiplication grid and read off the multiplication expression that produces that dividend. It then went on to show how this expression could be rearranged into an equivalent division expression that quickly shows that the skeleton's bones could be divided into equal piles. This potentially provided a much quicker method for deciding if a weapon would divide a given skeleton. To create a progression between these two stages, the 'magic book of times tables' was moved forward to the fourth level of the game, and the second path (B) was made available to the player at the same time as they found the book. In this way the player was allowed to experience the counting method for a few levels before being offered a more efficient alternative.

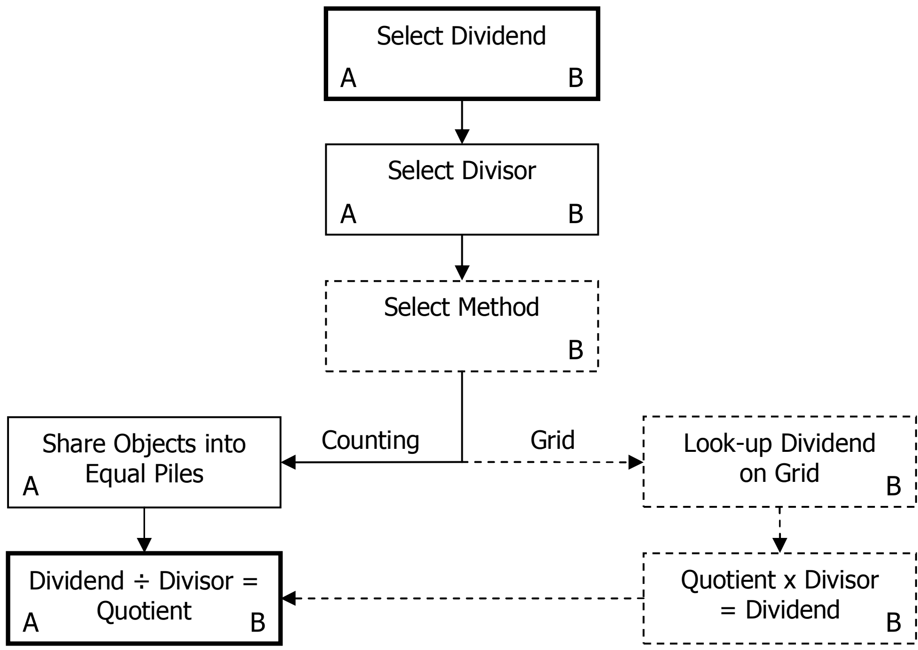


Figure 5.3: The Flow of the Framing System.

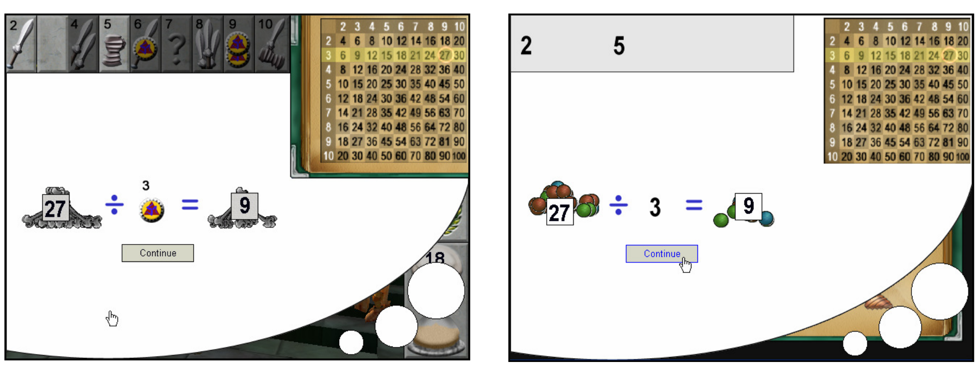


Figure 5.4 Screenshots from intrinsic (left) and extrinsic (right) framing systems.

The framing system was implemented for both the intrinsic version of the game (using skeletons, weapons and bones) and the extrinsic version (using numbers and marbles). The extrinsic framing took place within the end of level quiz, but it used the same underlying code as the intrinsic version, ensuring that the system was offering equal levels of support to both. Figure 5.4 shows screenshots of the two final versions running alongside each other.

5.2.3 Divisors and Dividends

The switch from using a hundred-square to a multiplication grid also had a knock-on effect on the divisors and dividends used in the game. One of the original reasons for using a hundred-square was because a multiplication grid can only display dividends up to ten times each divisor. Many dividends used in the levels were greater than ten times the divisor that was provided to divide them (e.g. the dividend 98 and the divisor 2). These dividends were therefore modified to keep them within the new limits. However, because this restricted the range of dividends these changes actually accentuated the problem of the game containing too much familiar content (dividends that could be divided by 2, 5 or 10). Consequently the balance of dividends and divisors would be adjusted again before study four in order to address this.

5.3 RESEARCH GOALS

The new testing system and learning content were by far the most significant changes made to the game at this stage. However, a number of smaller changes and tweaks were made to the game in order to try and address the project's research goals more effectively. Our hypotheses about intrinsic games suggest that they may provide motivational benefits that create a greater level of engagement, and a deeper connection with the learning content. They also suggest that intrinsic learning content may transfer less effectively to test situations and that the game's flow-experience may impede reflection. Study one was only designed to measure overall learning outcomes with limited potential for examining the basis of these hypotheses: motivation, deep learning, transfer and reflection. Changes were therefore made to the software and methodology to try and provide the potential for examining these areas in more depth.

5.3.1 Motivation

Our central tenet is that intrinsic games are more motivating than extrinsic equivalents, yet study one was not designed to include any direct measures of motivation. Task-performance is often used as a dependant variable in studies of motivation (e.g. Callahan, Brownlee, Brtek, & Tosi, 2003), but the potential interaction of transfer and reflection in games means that lower learning gains may not necessarily indicate less motivation. Consequently, a more direct measure of motivation would help to provide a deeper insight into the real relationship between learning and motivation in the game. Motivation is commonly measured in two other ways: via self-reported measures (usually questionnaires) and as a function of free time-on-task. Tubbs and Trusty (2001) suggest that the validity of self-reported measures of motivation are variable for adults, and the distorting effect of the social desirability response is generally considered to be worse for younger children (Dadds, Perrin, & Yule, 1998). There is a long history of measuring motivation as a function of time-on-task, including some of the early seminal research on intrinsic motivation (Daniel & Esser, 1980; Deci, 1971). Therefore time-on-task was chosen as our primary dependant variable with which to compare the relative motivational pull of the two versions of the game.

It was not necessary to make any changes to the game in order to record time-on-task, as the total playing time was already recorded for each player. This could therefore provide the primary dependant variable within the context of a study designed to allow free time-on-task. However, some other changes needed to be made to the game in order to ensure that a time-on-task comparison would be fair between the two versions. The existing extrinsic version of the game removed skeletons from levels in order to keep an equal balance of learning content between the two versions. This was arguably the fairest thing to do for a comparison of learning outcomes, but as skeletons are part of the motivational content of the game, it would be less fair for a comparison of motivational appeal. Consequently the system that removed skeletons in the extrinsic version would be disabled in all studies that were primarily intended to examine motivation rather than learning outcomes (studies 2 and 3).

Another necessary change was brought to light in the save-game system used in study one. The existing version did not save the player's exact position on a level, and sent them back to the start of their current level each time play was resumed from a saved game. In the case of the extrinsic version, this meant that children mid-way through the quiz would get to play the gameplay part of the level again. The limited length of sessions meant that children who struggled to complete quizzes on later levels, still spent most of their time in each session replaying the motivational content of levels. Not only is this unfair for a comparison of learning content it also seemed to undermine any motivational comparison as well. Consequently a more sophisticated save-game system was implemented at this stage to record and restore their exact position on a level and their question number within a quiz.

5.3.2 Deep Learning

Deep learning has been linked to intrinsic motivation (e.g. Biggs, 1987; Chin & Brown, 2000) and consequently formed part of our hypothesis for the benefit of intrinsic games. It would be extremely useful for our studies to include some measure of deep learning, but this is traditionally examined using detailed questionnaire data. Questionnaires such as The Learning Process Questionnaire (Biggs, 1987) provide general data on attitudes that provide valuable correlates or ways of measuring long-term global changes in learning. However they seem less appropriate for relatively short interventions within specific domains. They were also not originally designed for use with children as young as our target group. Completing detailed questionnaires would add additional time requirements on the testing-phase of the studies, with the strong possibility of fatigue and loss of concentration on the part of the young children.

Deep learning has also been linked to conceptual learning (e.g. Chin & Brown, 2000), so the conceptual questions included in the pre and post-tests could be seen as a potential measure of deep learning as well. Nonetheless, the majority of the test questions were not conceptual in nature and short-term improvements in overall scores could easily result from shallow learning approaches. However, shallow approaches are less likely to result in long-term improvements, so an additional round of delayed-tests could provide a further indication of the depth of learning that has taken place (Biswas et al., 2004). As a result, both conceptual improvements and delayed test results would be used as measures of deep learning in the final study (study four).

5.3.3 Transfer

One of the potential drawbacks of intrinsic games was perceived to be the difficulty children could experience transferring their learning to mathematical contexts outside of the game. The pre and post-tests provide a range of near-transfer tasks in an abstract mathematical context, including both dividend and divisor-based questions. However, the game only includes dividend-based questions, so there it was not possible to make a direct comparison between the game and test as a measure of the children's ability to transfer their learning. Therefore an additional phase of testing was designed to replicate a portion of the test's division problems within the game environment. The game-based test would consist of two specially constructed levels of the game, played after taking the post or delayed-test. All groups would play the gaming elements of these 'challenge levels', but each would have the learning content embedded (or omitted) appropriately for their condition.

In the extrinsic version these questions would be asked in the normal way at the end of each level, with provision for the alternative phrasing of division problems posed in the test (divisor-based questions):

e.g. Select one number that can be divided by 5:

35, 13, 29 or 41

In the intrinsic version, each question would be posed within the context of a separate room within the challenge levels. The weapons (and therefore divisors) available to the player would change to match each question as they entered its associated room. A divisor-based question would be posed in terms of offering the player a choice of three weapons with which to divide a single skeleton. An exit to the room would provide an option similar to 'none of these' (see figure 5.5).

Figure 5.5: A divisor-based question in the challenge level. The player must either choose to divide the skeleton using one of the three available weapons or leave via the unlocked door.

However, if the player left via the door, but the skeleton was dividable, then a second locked door would be placed immediately behind the exit. The game recorded an incorrect answer if the skeleton was attacked with the wrong weapon, or if the player left the room when the skeleton could be defeated. A correct answer would only be recorded if the skeleton was defeated with the correct attack on the player's first attempt. This would destroy the skeleton in the normal way and leave behind the key to the next room. Using the wrong weapon would cause the skeleton to attack the player before vanishing and leaving the key to the next room.

A dividend-based question would be posed in terms of a choice of four skeletons to divide with a single weapon. The exit would be locked and the key would not be provided until the player attacked one of the skeletons. (see figure 5.6).

Figure 5.6: A dividend-based question in the challenge level. The player must choose one of the four skeletons to divide with a single weapon, before they can proceed through the locked door.

An incorrect choice would cause the skeleton to attack the player before vanishing, and a correct choice would destroy the skeleton in the usual way. Attacking the correct skeleton would also cause all the remaining skeletons in the room to vanish. The health scoring-mechanism used in the main game would also be removed to try and replicate the lack of a score in the tests (effectively making the player invulnerable). Through these mechanisms the format of the questions in the test was replicated in gaming-terms that could provide a reasonably fair comparison as a measure of the children's ability to transfer their learning.

5.3.4 Reflection

Another potential drawback of integrating learning content within the flow experience of a digital game could be a negative effect on reflection-in-action (Schön, 1983). Nonetheless, it is arguable that providing a motivating fantasy context for learning content could also have a positive effect on reflection-on-action. However, while the concept of reflection has very established roots (Dewey, 1910) it is not a concept that is easily measured, and intensive qualitative data collection and analysis are usually required (e.g. Mansvelder-Longayroux, Beijaard, & Verloop, 2007). Such approaches are beyond the scope of these studies, but it is possible that analysis of the detailed process data produced by the game may be able to provide some indication of reflective behaviours. Therefore additional changes were made to the system to ensure that the data allowed each skeleton and its location to be tracked, so that the player's progress through the game could be analysed more precisely.

Changes were also made to the game design in order to help foster a more reflective environment at certain stages of the game. This attempted to create a clearly flagged separation between parts of the game that were intended for slower reflective learning and parts of the game that were for more intensive proceduralisation of that learning. This would be implemented through the use of 'danger levels' expressed through the colour of the skeleton's eyes on the in-game interface. The eyes would pulse with a green light on levels containing new learning content (new divisors). This would indicate that skeletons were currently docile, and their behaviour would be moderated so that they would not move around and only attack the player in retaliation to an incorrect attack. This would introduce the new learning content during a more relaxed period of gameplay free of the usual sense of urgency that may discourage reflection-in-action.

Figure 5.7: The red danger level. A continuous stream of skeletons appears from the magical area in the centre of the room to attack the player.

As the levels progress the 'danger level' would advance to amber, where the skeleton's eyes would pulse with a larger and more obvious yellow light. In this state skeletons would be livelier, moving along set paths and attacking the player if they got in the way. Each set of learning content (six levels) would conclude with a red 'danger level' in which the player competed to win one of the lost athlons they were searching for as part of the storyline. In this level the skeleton's eyes would pulse with a red light and the music would change to an intense climactic theme. None of the learning content on these final levels would be new, but a constant stream of skeletons would chase and attack the player until they had all been defeated (see figure 5.7). This would be the part of the game that encouraged proceduralisation, before the game returned to the green danger level for the start of new learning content on the next level.

5.4 SUMMARY

This chapter has described a number of developments in the game design and experimental methodology that will be used to evaluate Zombie Division. Improvements to the game included: replacing the hundred square with a multiplication grid in order to present less of a conceptual leap for children; including a framing system to provide a coherent mathematical strategy for children with a range of prior experience; and an improved save-game system to allow children to continue from exactly where they left off. Methodological changes have included: a computer-based testing system to reduce the possibility of high pre-test scores; a game-based transfer test to provide a direct comparison between questions in the test and game conditions; and more detailed process data output to allow data to be mined for reflective behaviours. Together it is hoped that these changes will facilitate the evaluation of Zombie Division in future studies so that conclusions can be reached about the relative value of intrinsic and extrinsic approaches.

CHAPTER SIX

Studies 2 and 3: Evaluating Motivation

This chapter describes two separate studies carried out to evaluate the motivational potential of Zombie Division using task persistence measures. The first set out to compare the time children spent playing intrinsic and extrinsic versions of the game when non-educational games were available as an alternative. The second compared the total time spent playing each version of the game when children could choose to switch between versions at will. Both studies used the new builds of the games, although most of the changes related to learning rather than motivation (see chapter five). However, this allowed the learning outcomes from study two to be used as a way of trialling these changes before attempting another study focussed on learning gains (study four).

6.1 AIMS

The educational interest in digital games is founded in the apparent engagement power that they possess. The manifestation of this engagement often shares many similarities with the concept of flow (Csikszentmihalyi, 1988) in terms of total concentration, distorted sense of time, and extension of self. It also epitomizes the definition of intrinsic motivation as “an activity with no apparent rewards except the activity itself” (Deci, 1971).

The concept of intrinsic integration (Habgood et al., 2005b; Kafai, 2001; Malone, 1981; Reiber, 1996) suggests that the engagement power of digital games may be most usefully harnessed for educational aims by more closely integrating a game with its learning content. Edutainment titles are criticised for including games as separate extrinsic motivation or reward for completing learning content

(Lepper & Malone, 1987). Habgood et al (2005b) suggest that this integration should be between a game's core mechanics and its learning content, in order to "embody the learning material within the structure of the gaming world and the player's interactions with it" (p. 494).

Task persistence was a central element of early research into intrinsic motivation (e.g. Deci, 1971) and Csikszentmihalyi's (1975a) concept of flow suggests that people will willingly engage with tasks for large periods of time if they contain clear goals, achievable challenges and accurate feedback. Malone's (1981) taxonomy of computer game motivations also suggested that good digital games contain these same elements. The central benefit of intrinsic integration may therefore be to produce greater time-on-task, which may then lead to better learning outcomes.

Study 2 : Time-On-Task

With Non-Educational Alternatives

The primary aim of this study was to compare the time children chose to spend playing the game when they were assigned to different versions, but provided with non-educational games as an alternative. This time-on-task was used as a measure of the relative motivational appeal of intrinsic and extrinsic approaches for Zombie Division. Learning outcome data and process data was also collected as a way of trialling the new changes made to the software and methodology.

6.2 METHOD

6.2.1 Design

Study two used a single between-subjects factor of 'group', giving each child one of two different versions of the game to play (intrinsic or extrinsic) and recording a single measure for time-on-task. Forty-four children were assigned to one of the two conditions based on matched pre-test scores. Children were sorted according to their scores and allocated alternately to each condition using a randomised block design to ensure an equal balance of gender in each condition. The data for two children were lost due to damaged data storage leaving twenty-three children in the intrinsic group and twenty-one in the extrinsic.

<u>System Changes</u>

All the changes made to the game and learning content following study one are summarised below. Please refer to chapter five for more detail on each aspect and the full reasoning behind it:

- Switch to multiplication grid

  Less of a conceptual leap for children. Had a knock-on effect of reducing the range of usable dividends to ten times each divisor.

- A mathematical 'framing' system

  To provide a relevant and coherent mathematical strategy for children entering with a range of prior experience.

- Improved save-game system

  To allow children to continue from exactly where they left off, addressing any motivational and educational differences between groups.

- Proceduralisation levels and reflection-flagging

  To provide gaming cues for reflective opportunities and structured progress towards proceduralising mathematical skills.

<u>Methodological Changes</u>

All the changes made to the study methodology and tools following study one are summarised below. Please refer to chapter five for more detail on each aspect and the full reasoning behind it:

- Computer-based testing system

  Longer, automated test to reduce the possibility of high pre-test scores. Also collects timing data and allows percentage scores to be calculated.

- Game-based transfer test

  To provide a direct comparison between identical dividend and divisor-based questions in test and game conditions.

- More detailed process data

  To allow data to be mined for reflective behaviours.

Measures

Each child had a maximum of one hundred and thirty-five minutes of playing time to spend on the game (2¼ hours). The game automatically logged the number of seconds spent playing the game as the primary measure of time-on-task for this study. In addition, direct pre and post-test measures of learning were taken using the new computer-based testing system. Each test provided two direct measures of learning outcomes: the total number of questions answered, and the number of correct answers made. These values were then used to calculate a percentage score (correct / total x 100) as an alternative measure of learning outcome. This additional measure can take into account strategies that may take longer to perform but are more accurate as a result (such as using the multiplication grid).

6.2.2 Participants

Children were taken from several primary schools in the catchment area of a single City Learning Centre[7] based in a large city in the north of England. The CLC is situated in an average-income area with about ten percent of students having origins outside of the UK. There are an above average number of students with special educational needs, and the attainment of their intake is below the national average. Nonetheless, the local schools achieve a high level of value-added with an above average percentage of students reaching expected levels in mathematics by their final year at primary school.

Forty-six mixed-ability children between the ages of 7 years 10 months and 8 years 8 months took part in the study (mean 8 years 2 months). The study was repeated on two successive days with twenty-four children attending day one and

[7] City Learning Centres are centres of technology and teaching expertise that work with schools to improve the use and practice of ICT in the school curriculum.

twenty-two different children on day two. There were eighteen girls and twenty-eight boys in total. The study was carried out during a half-term holiday towards the end of the school year.

6.2.3 Materials

Facilities

The tests and interventions were both carried out within one of the City Learning Centre's ICT suites. The suite contained thirty-two, brand new PCs running Windows XP with accelerated 3D graphics support, and audio output through stereo headphones. Each PC was provided with a CD-ROM disk containing the study's software, and a USB memory stick to save the process data.

Test Materials

The computer-based test consisted of 63 multiple choice questions with four options in each case (1 correct + 3 distractors). The first 3 questions were non-assessed practice questions designed to ease students into the test:

e.g. Select the number of legs that a dog has:

4, 5, 6 or 7

Of the remaining 60 questions, 40 were division questions equally comprised of two recurring formats, either asking the child to select the divisor that divides a given dividend (dividend-based):

e.g. Select one number that 45 can be divided by:

4, 6, 9 or none of these

Or to select the dividend that can be divided by a given divisor (divisor-based):

e.g. Select one number that can be divided by 5:

35, 13, 29 or 41

Amongst these 40 questions there were 5 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 addition to these 40 questions another 5 dividend-based questions were added where the answer was 'none of these'.

The remaining 15 questions were conceptual: 3 included to test for knowledge of the heuristic patterns associated with numbers that divide by 2, 5 and 10; and 12 more tested for an understanding of relationships between divisors, or for applying rules outside of normal limits (i.e. dividends greater than 100).

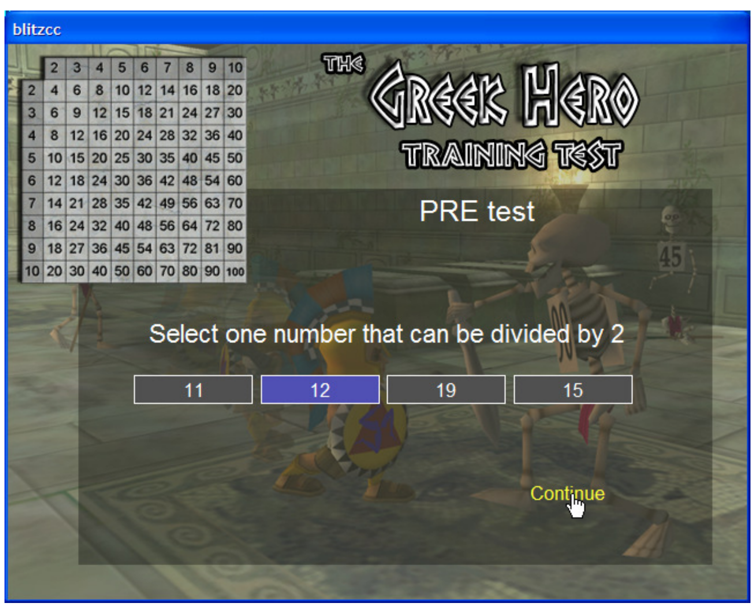


Figure 6.1: The computer-based test.

The order of the questions was randomised, but remained consistent between-subjects and between tests. The software timed 15 minutes from the start of question 4 and automatically stopped the test at the end of this period. However, the time was not displayed on the screen and no feedback was provided on the choices made. A multiplication grid was provided in the corner of the screen similar to the one found in the intrinsic and extrinsic versions of the game (see figure 6.1)

Alternative Gaming Materials

Unique to this study, the software also included a game menu that appeared when the player exited Zombie Division. This allowed the children to select which game they wanted to play next from a range of different games – including returning to Zombie division (see figure 6.2).

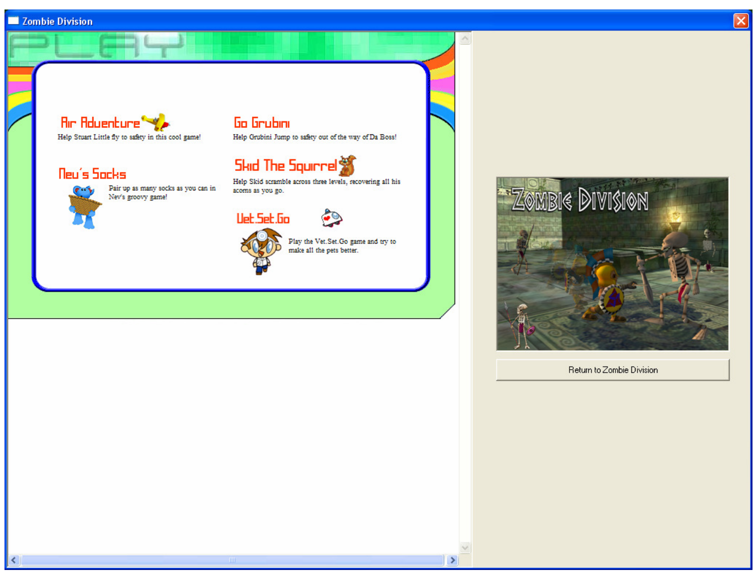


Figure 6.2: The game menu.

A number of flash-based games from the BBC website were offered as alternatives. None of these included explicit educational content, and their selection was limited to five web games to avoid too much time being spent experimenting with the different options. These provided motivating alternatives to playing Zombie Division in order to help to distinguish the relative motivational appeal of the intrinsic and extrinsic versions. The main game recorded the time each player spent playing Zombie Division so that the average group times could then be compared.

Game-Based Test Comparison

The game-based test consisted of two specially constructed levels of the game, designed to replicate a portion of the test's division problems within the game environment. This was given to the children immediately after the post-test to provide direct comparisons between the game and test that could indicate any transfer issues. However, the challenge levels were merely piloted in this study, and were included at this stage more as a motivational incentive for completing the post-test than for studying transfer. Both groups played the gaming elements of these 'challenge levels'. However, while the learning content was embedded appropriately in the intrinsic version of the game-based test, it was omitted from the extrinsic condition altogether as the system was still under development.

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 they entered its associated room. A divisor-based question was 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 similar to 'none of these' (see figure 6.3).

Figure 6.3: A divisor-based question in the challenge level. The player must either choose to divide the skeleton using one of the three available weapons or leave via the unlocked door.

However, if the player left via the door, but the skeleton was dividable, then a second locked door was placed immediately behind the exit. The game recorded an incorrect answer if the skeleton was attacked with the wrong weapon, or if the player left the room when the skeleton could be defeated. A correct answer was only recorded if the skeleton was defeated with the correct attack on the player's first attempt. This would destroy the skeleton in the normal way and leave behind the key to the next room. Using the wrong weapon would cause the skeleton to attack the player before vanishing and leaving the key to the next room.

A dividend-based question was posed in terms of a choice of four skeletons to divide with a single weapon. The exit would be locked and the key would not be provided until the player attacked one of the skeletons. (see figure 6.4).

Figure 6.4: A dividend-based question in the challenge level. The player must choose one of the four skeletons to divide with a single weapon, before they can proceed through the locked door.

An incorrect choice would cause the skeleton to attack the player before vanishing, and a correct choice would destroy the skeleton in the usual way. Attacking the correct skeleton would also cause all the remaining skeletons in the room to vanish. The health scoring-mechanism used in the main game was also removed to try and replicate the lack of a score in the tests (effectively making the player invulnerable). Through these mechanisms the format of the questions in the test was replicated in gaming-terms that could provide a reasonably fair comparison between the same questions in different contexts.

6.2.4 Procedure

Pre-tests

The pre-tests were carried out at the very start of each day of the study. Children were allocated a computer as they arrived and helped to enter their own name, gender and birthday. Afterwards, the test was explained to the whole group emphasising the presence of the multiplication grid to help them with the test – without explaining how it was used. They were informed of the 15-minute time limit, but told that they were not expected to finish all the questions and encouraged not to treat it as a race. They were also told that they could ask an adult for help reading any of the questions in the test, but not with the mathematics. Children who finished before the end of their time limit were asked to sit quietly until the entire group had finished.

Once the test was complete, the software broadcast all the results across the network to be collated by a server program. This automatically sorted children according to their scores and allocated them alternately to each condition using a randomised block design. It then broadcast the correct groups back to each child's machine and started the appropriate version of the game for them. While the learning outcomes of this study were secondary to the motivational ones, a child's motivation and performance on a particular set of learning content are often correlated, so grouping by ability still seemed like a reasonable approach.

Compulsory Intervention (45 Minutes)

Children were initially given a period of 45 minutes to play their version of the game. During this time no mention of the alternative games or the menu was made. At the end of this session children were asked to quit and save their game position – with the assurance that they would be able to continue from where they left off in the next playing session. The children then took a fifteen-minute playground break to rest their eyes and hands.

Optional Intervention (90 Minutes)

After returning from their break the children were presented with the game menu, and told that they now had a choice of which game to play. Children then played their choice of game for another 45 minute session followed by another short break and a final 45 minutes. During this time, children were prevented from surfing the web or touching each other's machines, but they were not discouraged from chatting or watching other children play.

Post Tests

After the third playing session the children took a 30-minute lunch break, followed by another 15 minute playground break. Upon returning to the classroom the children re-took the computer-based test, with the advertised incentive of a chance to play the challenge levels afterwards.

Challenge Levels

In order to prevent any distraction, children were not allowed to begin the challenge levels until the entire group had finished their post-tests. They were then given another 20 minutes to complete the challenge, after which any remaining children were told to stop and save their progress.

6.3 RESULTS

The results section for this study is divided into four sub-sections based around the different data sources available. The first section uses time-on-task data to examine group and gender affects on motivation in order to address the primary aims of this study. The second section reports the results obtained from the game process data to obtain a breakdown of the proportion of maths tasks undertaken for each divisor. This provides an insight into the proportion of learning content encountered by the children that could be considered new to them. The third section reports the results of the learning outcome data, and the success of the challenge levels as indications of overall learning and transfer. The final section analyses the test process data in more detail to explore timing as an indication of changes in mathematical strategies.

6.3.1 Motivation

The data for two children was lost due to damaged memory sticks leaving twenty-three children in the intrinsic group and twenty-one in the extrinsic. The primary measure for this study was the amount of time children spent playing their version of the game. This provides an indication of the relative motivational appeal of the different versions of the game. Table 6.1 shows the mean and standard deviation of time-on-task by group and gender. Gender differences are not the primary focus of this thesis, but study one had indicated that there were differences in reported gaming experience between boys and girls and it is conceivable that these differences could impact on motivation for the games.

These measures were examined using a two-way univariate ANOVA with two levels of the between-subjects factor 'group' (intrinsic and extrinsic) and two levels of the between-subjects factor 'gender'. A significant effect of gender was found ($F(1,40) = 8.41$, $p< 0.01$, $\eta^2 = 0.17$), but there was no significant effect of group or a group by gender interaction. Boys therefore spent significantly longer playing their versions of Zombie Division than the girls, but there was no difference between versions.

Table 6.1. Mean and S.D. of time-on-task by group and gender

	Intrinsic		Extrinsic	
	Male (n=13)	Female (n=10)	Male (n=14)	Female (n=7)
Time-on-task (s)	5819.31 (996.41)	4957.70 (1489.01)	6532.71 (983.70)	5234.14 (1431.38)

6.3.2 Learning Process

The process logs produced by the game provide a versatile source of data in the form of a time-stamped commentary on the game as it is being played. For this study, over one thousand log files were mined for the purposes of post-hoc analysis. The mining in this study focussed on determining the proportion of maths tasks undertaken for each divisor in the games. Study one had suggested that less than half (42%) of the content encountered by children in the intervention could be considered new to them. This may be unhelpful for promoting the conflict necessary to adopt the new mathematical strategies promoted by the game. Much of this content was changed after study one as a result of the switch to using a multiplication grid that could only display dividends up to ten times each divisor. This analysis was therefore useful to examine the effect that these changes made to the amount of familiar learning content encountered during the intervention.

We will define a *maths task* as a single encounter with a divisor-based problem in the game. A maths task is always presented in the form of a skeleton in the intrinsic version, and a quiz question in the extrinsic version. All maths task data in this study refers to a player's first encounter with each maths task in the game (as well as their first attempted answer at that task). Table 6.2 shows the combined group breakdown of divisors used in first encounters with maths tasks. Just over 72% of all maths tasks were solved using the divisors 2, 5 and 10. The corresponding multiplication tables are part of the national curriculum requirements for children of this age, so only about 28% of the overall learning content in the games could be considered as unfamiliar to this audience.

Table 6.2. Breakdown of divisors used in first encounters with maths tasks.

	None (n=44)	2 (n=44)	10 (n=42)	5 (n=41)	3 (n=18)	4 (n=6)	6 (n=1)	9 (n=1)	Total
Number of first Encounters	1016	2028	738	1049	263	82	32	29	5237*
Proportion of total 1st encounters	19.40	38.72	14.09	20.03	5.02	1.57	0.61	0.55	98.7*

Divisors listed in the order they appear in the games.
* Remaining encounters are undefined

6.3.3 Learning Outcomes

As gender had been shown to be a significant factor in time-on-task it was also included as a factor in the analysis of learning gains. Numerical scores were examined using a three-way repeated measures ANOVA with two levels of the within-subjects factor 'test' (pre and post), two levels of the between-subjects factor 'group' (intrinsic and extrinsic) and two levels of the between-subjects factor 'gender'. The analysis of numerical scores (table 6.3) revealed a significant effect of test ($F(1,40) = 11.94$, $MSE = 198.41$, $p< 0.005$, $\eta^2 = 0.23$) and a test by gender interaction ($F(1,40) = 4.24$, $MSE = 70.49$, $p< 0.05$, $\eta^2 = 0.10$), suggesting that girls and boys performed significantly differently from each other. Paired samples t-test comparisons[8] of the percentage score for each gender revealed that the difference between pre and post-test numerical scores was only significant for the girls ($t(9) = 3.67$, $p < 0.005$). Overall numerical scores were therefore significantly higher in the post-test than in the pre-test but numerical learning gains were only significant for the girls.

Table 6.3. Mean numerical scores by group and test

	Intrinsic		Extrinsic	
	Male (n=13)	Female (n=10)	Male (n=14)	Female (n=7)
Pre-test	20.69 (6.02)	21.20 (7.67)	21.50 (9.57)	19.00 (3.70)
Post-test	20.92 (7.74)	26.00 (7.78)	23.79 (8.03)	24.14 (2.80)

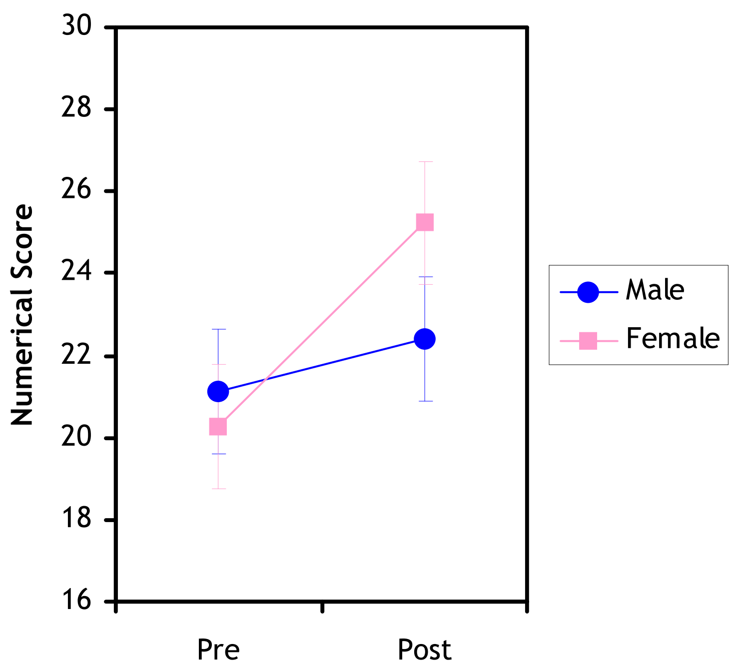


Figure 6.5. Mean numerical scores by gender and test.

[8] Not corrected for multiple comparisons.

The computer-based test records the total number of questions attempted as well as the number that were correct. Therefore this allows a percentage score to be calculated that can take into account strategies that may take longer to perform but are more accurate as a result (such as using the multiplication grid).

Table 6.4. Mean percentage scores by group and test

	Intrinsic		Extrinsic	
	Male (n=13)	Female (n=10)	Male (n=14)	Female (n=7)
Pre-test	36.34 (7.38)	38.73 (15.05)	39.77 (14.12)	39.99 (8.29)
Post-test	34.87 (12.90)	47.12 (16.99)	43.15 (20.19)	40.86 (4.98)

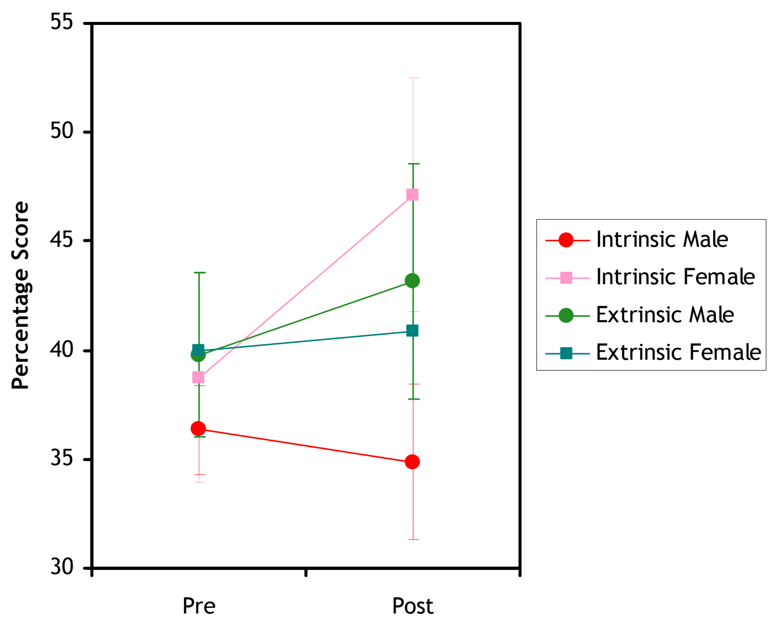


Figure 6.6. Mean percentage scores by group, gender and test

Percentage scores were examined using a three-way repeated measures ANOVA with two levels of the within-subjects factor 'test' (pre and post), two levels of the between-subjects factor 'group' (intrinsic and extrinsic) and two levels of the between-subjects factor 'gender'. The analysis of percentage scores (table 6.4) revealed no significant effect of test and no test by group or test by gender interactions. However, it did reveal a significant test by group by gender interaction ($F(1,40) = 4.87$, $MSE = 195.40$, $p < 0.05$, $\eta^2 = 0.11$), suggesting that one of the genders in one of the groups performed significantly differently from the others. Paired samples t-test comparisons[9] of the percentage score for each of the four gender/groups revealed that the difference between pre and post-test percentage scores was only significant for the girls in the intrinsic condition ($t(9) = 3.93$, $p < 0.005$). This suggests that only the girls in the intrinsic condition made significant percentage learning gains in this study.

Correlations with Time-on-Task

Implicit in the concept of the motivational value of intrinsic games is the idea that more time-on-task will create greater learning gains. Therefore bivariate correlations were also performed comparing time-on-task with level progress and learning gains (numerical and percentage). Time-on-task was significantly correlated with level progress ($r(44) = 0.65$, $p < 0.01$) but not learning gains.

Challenge Levels

The challenge levels were included to provide a direct comparison between students' ability to solve the same division problems in the assessment test and in the game. Twenty of the questions from the assessment were replicated in the challenge levels of the intrinsic version. Unfortunately there was not enough time

[9] Not corrected for multiple comparisons.

to add similar functionality to the extrinsic version for this study, but the system was only included at this stage to test its operation. Table 6.5 shows the mean percentage scores for the challenge levels alongside their equivalent percentage scores derived from the assessments. One-way repeated measures ANOVAs were performed on percentage scores with just one within-subjects factor of 'method' (assessment or challenge). No significant effect of method was found in this case suggesting that there was no significant difference in the scores of the intrinsic group between game and test situations.

	Intrinsic (n=23)
Assessment Percentage	44.78 (12.75)
Challenge Percentage	39.77 (17.96)

Table 6.5. Mean and S.D. of percentage scores by test-type.

6.3.4 Testing Process

The process data for the pre, post and delayed tests was used to obtain each child's mean time for answering questions in each test (see table 6.6). This measure was then examined using a three-way repeated measures ANOVA with two levels of the within-subjects factor 'test', two levels of the between-subjects factor, 'group' (intrinsic and extrinsic) and two levels of the between-subjects factor 'gender'. This revealed a significant main effect of time ($F(1,40) = 7.02$, $MSE = 268.64$, $p< 0.05$, $\eta^2 = 0.15$), but no significant effect of gender and no interactions. The average time per question in the post-test was therefore significantly shorter than in the pre-test.

Table 6.6. Mean number of seconds per question by group and test

	Intrinsic		Extrinsic	
	Male (n=13)	Female (n=10)	Male (n=14)	Female (n=7)
Pre-test	13.44 (9.68)	13.58 (4.81)	14.19 (12.14)	18.56 (6.30)
Post-test	9.11 (3.44)	11.65 (6.12)	10.94 (5.54)	13.58 (2.07)

Bivariate correlations were also performed between percentage scores and average question timings in each test as an indication of a speed/accuracy trade off. Such a trade off may indicate the use of the multiplication-grid as a slower, but more accurate strategy. Pre-test question timings were not significantly correlated with percentage scores at pre-test, but post-test question timings were significantly correlated with percentage scores at post-test ($r(44) = 0.72$, $p < 0.01$). This may be indicative of some children adopting the multiplication grid method in the post-test as a result of the intervention.

6.4 DISCUSSION

The aim of this study was to compare time-on-task for different versions of the games as a measure of their relative motivational appeal. Our theoretical analysis of intrinsic games suggests that they may provide motivational benefits, which create a greater level of engagement and a deeper connection with the learning content. However, the central benefit of intrinsic motivation may actually be to produce greater time-on-task. This study compared the time children chose to spend playing the game when they were assigned to different versions, but provided with non-educational games as an alternative.

6.4.1 Motivation

The children in the intrinsic and extrinsic groups did not demonstrate a clear difference in their motivation as measured by the total amount of time they chose to spend playing their version of Zombie Division. Overall children chose to play Zombie Division for an average of 65% of their 'optional' playing time ((5757-2500)/5000[10]) and boys played the game for significantly longer than girls (74% as opposed to 51%). The boys' percentage is particularly high when you consider that they were choosing an educational game over non-educational ones. Nonetheless, this data would seem to suggest that there was no real difference between the motivational appeal of the intrinsic and extrinsic versions of the game. However, observations at the time suggested that there might have been motivational differences between the versions that were not revealed by measuring total time-on-task. Several children in the extrinsic groups (mostly boys) expressed a strong dislike of the quiz questions, and frequently asked how to skip past or avoid them. One boy was even reduced to tears of frustration by

[10] Based on an average of 2500 seconds of possible playing time in each of the three 45 minute sessions (allowing for loading/settling times), and only the second two sessions being 'optional'.

the quiz questions, yet he refused to play the BBC games, asserting that he wanted to get back to the ‘fun’ part of Zombie Division. So it appeared that for some children in the extrinsic group the positive motivation of the ‘chocolate’ gameplay was more than enough to outweigh any negative motivations of the ‘broccoli’ quiz.

Competition between children in the two groups also played a strong motivational role during the intervention. In this study, both versions of Zombie Division included exactly the same number of skeletons. However, because the extrinsic group also had to complete a quiz at the end of each level it took them longer to progress through the levels in the game. This was a deliberate choice made in this study to ensure that the motivational content of the game (fighting skeletons) was identical between versions[11]. However, it meant that children in the intrinsic group soon found themselves ahead on levels, and so were perhaps more inclined to try the BBC games while their extrinsic classmates struggled to catch up.

6.4.2 Learning Process

The learning process data was not examined in great detail in this study, as this was not its main focus. Nonetheless, the process data did reveal that 72% of the learning content encountered by the children was for the familiar divisors 2, 5 and 10. This was much larger than study one where it was only 58%, but the switch from a hundred-square to a multiplication grid had made it necessary to change many of the dividends (as multiplication grids do not include as many dividends as

[11] The studies focussed on learning outcomes (1 and 4) reduce the number of skeletons in the extrinsic version so that children are exposed to roughly the same amount of learning content per unit time. The studies focussed on motivational outcomes (2 and 3) have the same number of skeletons between versions so that the motivational content is consistent.

a hundred-square). Clearly these changes had an adverse effect, so all the dividends and divisors in the game would be reviewed following this study.

6.4.3 Learning Outcomes

Study two was used as a way of piloting some of the new features and changes made to the game before undertaking a new study focussing on learning outcomes (study four). Although learning outcomes were not the focus of this study, they do not suggest a great improvement over the versions of the game used in study one: an overall gain was observed in numerical scores, but the combined mean improvement was less than 3 marks, with the girls only making a slightly better improvement of 5 marks. Furthermore, no difference was found between numerical gains in the intrinsic and extrinsic conditions. Interaction effects should be interpreted with caution as no corrections were made for multiple comparisons (see results section). Nonetheless, while no combined mean gain was observed in percentage scores at, girls in the intrinsic group did significantly distinguish themselves from the rest with an improvement of around 5 percentage points.

One factor that may have influenced low observed learning gains was the long duration of the study over a single morning. Children had to perform the test twice in a single day and it was clear that their motivation in the post-test was much lower than it had been in the pre-test. Another factor could be the high percentage (72%) of learning content that was already familiar to the children. Familiar content is less likely to change the children's approach to division problems as they can rely on existing strategies and recalled knowledge in order to solve the problems. Nonetheless, the modest improvement in percentage scores amongst girls in the intrinsic group offers some support for the value of the intrinsic version. The difference between girls and boys scores was apparent in the attitudes they demonstrated towards the game during the study. Although boys did not seem to object to the intrinsic learning content (the skeletons) many did seem very reluctant to listen to the instructional content presented by the framing system. Some discarded their headphones and others used it as an opportunity to compare their progress with their neighbour. Overall the level of engagement many children demonstrated for the framing system clearly did not match their engagement with the rest of the game.

Part of the problem with the framing system was that it can quickly appear very repetitive. If the player gives it their full attention from the start of the game then its presence can actually be very minimal. However, if they ignore it and begin applying a 'trial and error' approach it will regularly 'nag' them with instructional content. Unfortunately in many cases it seems that the more the content appears the less likely many children are to listen. This may be particularly true when the child is competing to progress in the game and sees the framing system as something that slows down their progress. Overall this study gave the impression that the framing system was simply not an appropriate way of introducing boys to the learning content in the game.

Relationships Between Playing Time and Learning

Implicit in the concept of the motivational value of intrinsic games is the idea that more time-on-task will create greater learning gains. However, no correlation was found between time-on-task and learning gains in this study – despite a correlation between time-on-task and level reached. This could suggest that some children may not have understood the framing system, yet still persisted with the game – and/or understood the framing system quickly, but chose not to continue playing.

Challenge Levels

The new challenge level feature was designed to provide an insight into transfer issues in the children's learning. It does this by providing a subset of the pre/post test questions in a gaming form appropriate to each group's condition. In this study the (incomplete) challenge feature only supported comparative test content for the intrinsic version of the game. No statistical difference was observed between challenge scores and comparative test scores, suggesting that transfer problems may not be occurring. However, this result carries little weight without an extrinsic comparison. Nonetheless the completed challenge feature should be able to provide this data in future studies.

6.4.4 Testing Process

The timing data showed that children's mean time per question increased between pre and post-test with a significant correlation between post-test times and post-test percentage scores. This offers some evidence that the game was encouraging more children to use the multiplication grid (and so make the speed/accuracy trade off) – even if this only translated into a significant percentage improvement in learning gains for girls in the intrinsic group.

6.5 SUMMARY

This study demonstrated no significant difference between the time children chose to spend on the intrinsic and extrinsic versions of the game when non-educational alternatives were available. However, children were not directly choosing between the intrinsic and extrinsic approaches. This study also demonstrated significantly higher percentage gains for girls in the intrinsic group. This may offer the first suggestion that the intrinsic approach may be more effective than the extrinsic in some situations.

Study 3: Time-On-Task With Free Switching

Our theoretical analysis of intrinsic games suggests that they may provide motivational benefits, which create a greater level of engagement and a deeper connection with the learning content. However, the central benefit of intrinsic motivation may actually be to produce greater time-on-task. Consequently, this study compared the time children chose to spend playing the different versions of the game when they were provided with a free choice between them. This time was used as a direct measure of the relative motivational appeal of intrinsic and extrinsic approaches for Zombie Division.

6.6 METHOD

6.6.1 Design

Study three used a single within-subjects factor, 'version' with measures of time for playing both the intrinsic and extrinsic versions of the game. Sixteen children took part in the study during an existing after-school computer club at their school. Children were exposed to identical conditions which allowed them to freely switch between intrinsic and extrinsic versions of the game.

Measures

Each child had a practical limit of one hundred and thirty-five minutes of playing time to spend on the game (2¼ hours) over the course of three weekly club sessions. The game automatically logged the number of seconds spent playing each version of the game as the primary measures of time-on-task for this study. No measures of learning were taken in this study as it would be difficult to attribute gains to either version of the game, and it would also have affected attendance in an optional after-school club. In a fourth club session directly after the others, children spent the entire session providing feedback and comments on the relative merits of the two different versions of the game. This session was recorded and transcribed for analysis.

6.6.2 Participants

All children attended a large primary school on the outskirts of a large city in the north of England. The school is situated in a low-income area with a very small percentage of students having origins outside of the UK. The school has an average number of students with special educational needs, and the attainment of their intake is below the national average. The percentage of final year students achieving expected levels in mathematics was below national averages for the preceding year.

Sixteen primary children between the ages of 9 years 10 months and 11 years 2 months took part in the study (mean 10 years 4 months). This was made up of thirteen mixed-ability children from year 5 and three children from year 6, with five girls and eleven boys in total. All of the children had extensive experience of using the computers in the school's ICT suite as part of an after-school club. The study was carried out at this club, over the last four weeks of the final term in the school year.

6.6.3 Materials

<u>Facilities</u>

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 twenty, relatively new PCs running Windows 2000 with accelerated 3D graphics support, and audio output through stereo headphones. Each PC was provided with a CD-

ROM disk containing the study's software, and a USB memory stick to save the process data.

<u>Version Switching</u>

Version switching was provided through an additional menu that appeared each time the game was launched (see figure 6.8). This allowed children to choose between the intrinsic and extrinsic versions of the game, and provided them with a visual reminder of what the differences between the two versions were. The order that the options appeared in the menu was randomised each time so that either version would appear on the left or right with an equal probability.

Figure 6.8: The version switching interface.

Quitting the game would return the player to this menu, where they could switch versions again. Crucially, players did not have to return to the beginning of the adventure (or even the level) when switching versions: their exact position was resumed with intrinsic skeletons becoming extrinsic or visa versa. In this way it was ensured that there was no gameplay penalty for switching between versions.

6.6.4 Procedure

Introduction (10 Minutes)

Children were introduced to Zombie Division as a group by demonstrating 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 intrinsic and extrinsic versions were apparent. Children were shown how combat worked in both versions, emphasising 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. Emphasis was given to the fact that their game position would not be lost by switching versions and that they were expected to try playing both versions.

Game Intervention (135 Minutes)

All 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 returning to their normal club pursuits (and freely switch between the two). Each club session lasted for approximately one hour, with around 45-50 minutes of playing time. This continued for two more club sessions after the first, providing a maximum of around 135 minutes (2¼ hours) 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 group 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 children were given a group interview in front of the two different versions running side by side. They were asked to summarise the differences between the two versions and which versions they preferred. Each child was given the opportunity to explain why they preferred the option they did, and the group was encouraged to discuss which version was the most fun to play and which was the most educational.

6.7 RESULTS

The results section for this study is divided into two sub-sections based around the different data sources available. The first section uses the timing data to examine the group's preferences for either version. The second section uses the interview data to corroborate and add additional depth to these findings.

6.7.1 Motivation

Table 6.5 shows the mean number of seconds children spent playing the two different versions of the game. These were analysed using a two-way repeated measures ANOVA with two levels of the within-subjects factor 'version' and two levels of the between-subject factor 'gender'. This showed a main effect of version ($F(1,14) = 84.42$, $MSE = 133689150.5$, $p < 0.001$, $\eta^2 = 0.86$) and a group by gender interaction ($F(1,14) = 7.35$, $MSE = 11641511.14$, $p < 0.05$, $\eta^2 = 0.34$). Paired samples t-test comparisons[12] between versions for each gender showed a significant difference in the time between versions for both boys ($t(10) = 5.97$, $p < 0.001$) and girls ($t(4) = 6.71$, $p < 0.005$). Independent samples t-tests between genders for each version showed a significant difference between genders for the intrinsic version ($t(14) = 2.59$, $p < 0.05$). Both girls and boys therefore spent significantly longer playing the intrinsic version than the extrinsic, but girls played the intrinsic version for significantly longer than the boys.

[12] Not corrected for multiple comparisons.

Table 6.5. Mean number of seconds played by gender and version.

	Male (n=11)	Female (n=5)	Combined (n=16)
Intrinsic	3747.82 (1824.50)	6278.60 (1777.49)	4538.69 (2128.27)
Extrinsic	639.36 (589.22)	567.60 (574.19)	616.94 (566.17)

6.7.2 Interview Data

The children were interviewed for an extended period of time (nearly an hour) about the relative merits of the two versions. They expressed a strong overall preference for the intrinsic version in line with the results for time-on-task. However, the reasons given for and against may provide a deeper insight into their reasoning behind this preference.

Table 6.6. Reasons given for liking or disliking the intrinsic version.

Intrinsic Version	
Likes	"it's not as hard – it's quick and easy" "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" [teachers would think it's] "too much fun – and hasn't got a test"

Table 6.6 shows the comments children made about liking or disliking the intrinsic version of the game. Reasons for liking the intrinsic version seemed to focus around it being easier, quicker to progress and more fun than the extrinsic version. Reasons for disliking it focussed on it not being quicker to progress and perceived adult disapproval.

Table 6.7 shows the comments children made about liking or disliking the extrinsic version of the game. The only reason given for liking the extrinsic version focussed around its similarity to standard educational testing. Reasons for disliking the extrinsic version revolved around it being less fun, slower to progress and less challenging.

Table 6.7. Reasons given for liking or disliking the extrinsic version.

Extrinsic Version	
Likes	"it can help you like learning your times tables and doing your SATs"
Dislikes	"because you need to do all the maths at the end and that's what you might lose interest in" "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 […] too easy." "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"

Table 6.8 shows the negative comments children made about the in-game help system in both versions of the game. These reasons revolved around the Gargle character being annoying and slow. There were no positive comments made.

Table 6.8. Reasons given for disliking the help system (both versions).

Instructional Help System	
Dislikes	"it did my nut in" "it slows me down" "how do you skip the help?" "it was annoying because of that gargoyle's odd voice" "the annoying clockwork gargoyle talks too slow"

6.8 DISCUSSION

The aim of this study was to compare time-on-task for the two versions of the game as a measure of their relative motivational appeal. Our theoretical analysis of intrinsic games suggests that they may provide motivational benefits, which create a greater level of engagement and a deeper connection with the learning content. However, the central benefit of intrinsic motivation may actually be to produce greater time-on-task. Therefore this study compared the time children chose to spend playing the game when they were allowed to switch freely between the intrinsic and extrinsic versions.

6.8.1 Motivation

The children in this study demonstrated a clear difference in their motivation for the intrinsic and extrinsic versions of Zombie Division as measured by the different amount of time they chose to spend playing them. Children spent an average of over seven times longer playing the intrinsic version of the game than the extrinsic. This result seems to suggest that in a direct comparison between the two approaches the intrinsic version has more motivational appeal. One child even discovered a strategy of playing a level in extrinsic mode and switching to the intrinsic at the end in order to avoid any learning content! However, he soon dropped this mixed method in favour of the intrinsic version, after becoming bored with the time it took to repeatedly load and save the game.

Interaction effects should be interpreted with some caution as no corrections were made for multiple comparisons (see results section). Nonetheless, in contrast to study two, girls played the game for significantly longer than boys. This appeared to be the result of boys trying not to lose face by withdrawing from the game when it became clear they were being beaten by the girls. Only the three most successful girls were still playing the game at the end of the study, but they regularly attracted a crowd of boys as the girls reached new levels and encountered different enemies and weapons. The difference between studies two and three in terms of the girls' level progression may be explained by the older age of children in this study. Study one suggested that girls generally had less gaming experience than boys, but it seems reasonable to expect that fewer girls have no experience with games as they get older. Nonetheless this does demonstrate that the game can appeal to both genders in the right contexts.

6.8.2 Interview Data

The interview data from study three supports the quantitative findings that suggested the children preferred the intrinsic version over the extrinsic. There were four recurring themes amongst the reasons children gave for preferring one version over the other, namely: speed, difficulty, fun and learning.

Speed was considered an important factor because competition was such a strong motivation for the children's interest in the game. Some children saw the extrinsic version as slower to complete (containing a game and a quiz) and therefore an impediment to their progress. However, other children saw the intrinsic version as slower because it forced learning content on the player when skeletons were attacked incorrectly (even though this also happened in the extrinsic quiz).

A number of children saw the intrinsic version as easier for learning division, some identifying the symbolic skeletons in the extrinsic version as a distraction to their learning. Others saw both the extrinsic quiz and fighting symbolic skeletons as too easy – not presenting the same challenge as the combination of maths and skeletons in the intrinsic version. It should be noted again that the children in this study were a year or two older than previous studies, which may account for their complaints about the learning content not being challenging enough. Nonetheless the children's class teacher advised that many of the children were not as good at maths as they were claiming in front of their peers.

The fun contained within the intrinsic version and the absence of it in the extrinsic was another recurring theme in the children's discussion. Astoundingly, some children were even able to make the connection between combining the fun and the learning content to make the intrinsic version a better game. This included the observation that the intrinsic version was, "like subliminal advertising with maths" as well as a paint-based analogy "[mixing] the maths in the game with the fun". Both children agreed that this made the intrinsic version more fun as a result. Others thought that the intrinsic version could be a problem for schools and teachers because it was, "too much fun – and hasn't got a test". While many children seemed to cite learning as a reason why the intrinsic version was better then the extrinsic, there was some acknowledgement that the extrinsic version conformed better to school teaching models, of tests and SATs.

The interviews suggest that some of the children's preference for the intrinsic version was a result of their perception of it as a faster way of getting through the levels. However, it is also clear that they were motivated by deeper differences between the versions as well. They understood that games should be challenging and felt that the intrinsic version provided the most enjoyable challenge. Some of (the older) children were even able to verbalise the integration between maths and the 'fun' in the intrinsic version – as well as the problem with separating the two. Overall the interview data proved extremely insightful, providing convincing support for the quantitative findings, and wildly exceeding our own expectations for this study. It is clear that the children were more motivated by the intrinsic version of Zombie Division, providing more support for the hypothesis that intrinsic games are more motivating than extrinsic ones.

6.9 OVERVIEW

These two studies were designed to evaluate the relative motivational appeal of the intrinsic and extrinsic versions of the game. Study two compared the time children spent playing the game when they were assigned a version, but had the alternative of playing non-educational games from the BBC website. Study three compared the time children spent playing the intrinsic and extrinsic versions of the game when they could freely switch between them. Motivation was primarily measured as a direct function of time-on-task, but study three also provided qualitative interview data to support the quantitative findings. Nonetheless, at a surface level the results of the two studies appear to conflict each other. Rationalising these contrasting results has forced us to recognise the highly significant role that two specific motivational factors have played in all our evaluations of Zombie Division:

Production Values

Zombie Division has clearly motivated children in all of the studies carried out for this thesis and study three demonstrated children's preference for the intrinsic version. Nonetheless, study two suggests that the motivation produced by Zombie Division's production values may be greater than any potential motivational distinction between intrinsic and extrinsic approaches. Children were choosing between 2D games from the BBC website without explicit learning content, and the real time 3D-rendered world of Zombie Division which included mathematics. The key distinction for most children seemed to be one of production values, and the behaviour of some children (mainly boys) suggested that this was the only factor important to them. The child that refused to stop despite being reduced to tears by the extrinsic quiz, shows that the motivation of high production values can be enough to outweigh any distain for learning content. Naturally, this does not create ideal affective conditions for learning, but it does challenge our hypothesis that the central benefit of the motivation produced by intrinsic games would be to produce greater time-on-task. Extrinsic motivation (particularly alongside competitive peer pressure) can clearly be powerful enough to make children participate in a learning activity – even if that participation is at a superficial level.

It was this observation that inspired the change of approach applied in study three. Rather than giving children a choice between the production values of 2D and 3D games, this study gave them direct control over the version of Zombie Division that they were playing. This meant that the only factor affecting children's time-on-task was the difference between an integrated and non-integrated approach. This time the results were far more conclusive and children spent an average of over seven times longer playing the intrinsic version of the game. Nonetheless, it is clear that high production values have a large part to play in creating the motivational appeal of Zombie Division and the results of studies two and three may have been completely different without them.

Competition

The second motivational factor that appears to have had a huge impact on all of the studies so far is that of group competition. Every time Zombie Division is introduced into a classroom context, children find a way of competing against each other. Even in study one, when children were not directly provided with a means of comparing themselves against each other, they went to great lengths to keep track of their progress (by monitoring the debug output on the loading screen). As this competition seemed impossible to suppress, level progress was deliberately displayed in subsequent studies. The competitive spirit seems strongest at the start of each study, when the children's progress is fairly equal and winning seems like an 'achievable challenge' (Csikszentmihalyi, 1988) for the majority of children. The competition appears to tail off as the distribution of the children's progress spreads out and fewer children feel they are within reach of being the best.

The power of this effect potentially represents a significant problem for all the group-based interventions in this thesis. Children's motivation in study two may have been cross-contaminated between the intrinsic and extrinsic conditions. Unfortunately it was not practical to separate children in different conditions, which meant that a child in the intrinsic group could be sitting adjacent to – and be directly competing with – a child from the extrinsic group. This could have resulted in an averaging out of the motivational appeal of both versions: the extrinsic child enduring his version of the game for longer in order to catch up with his more motivated friend in the intrinsic group. It is even possible that the same effect may take place (to a lesser extent) between groups in the other studies. The conditions may be separated during the interventions, but the children are all from the same class or school, and may well take their competition into the playground. The methodology used in study two does not allow us to determine how much of an effect (if any) this actually had on children's motivation. Nonetheless given the visibly significant role that competition appears to play in the children's motivation, it is worthy of consideration in any conclusions drawn from this thesis.

6.9.1 Conclusions

Study three provides convincing evidence that the children found the intrinsic version of Zombie Division more motivating than the extrinsic. However, the contrasting result from study two's alternative methodology has highlighted the importance of production values and competition in the children's motivations for using the game. It suggests that the motivational difference between the two contrasting theoretical approaches may not be as significant as these other factors. Of course this does not suggest that all motivations are equal when learning is involved, but it does challenge the hypothesis that the central benefit of the intrinsic approach over the extrinsic is to produce greater time-on-task.

Study two also provided the first tentative evidence that the motivation produced by the intrinsic version is more usefully applied to learning goals (for girls). While this finding certainly requires verification, the study provided more clues as to how the software and methodology could be improved in order to better achieve this. The game needs to provide a greater focus on unfamiliar learning content, and there was some suggestion that flow may not be compatible with boys' uptake of instructional information. This became the main argument for adjusting the methodology in study four to include a period of teacher-led reflection away from the distraction of the game in order to try and achieve a larger improvement in learning gains for both boys and girls.

The most rewarding experience of these studies was talking to the children from study three about Zombie Division. The ability of two different children to express the theoretical differences between intrinsic and extrinsic approaches was wholly unexpected. Their understanding and approval of the intrinsic approach seemed to confirm it as a theoretical concept that is relevant to its intended audience. One child was even able to suggest his own intrinsic idea for expanding the game to include end of level bosses that required the player to use multiple divisors to defeat it. In doing so he demonstrated the very combination of deep, conceptual and meta learning skills that all educationalists strive for their students to achieve.

CHAPTER SEVEN

Study 4: Re-Evaluating Learning

This chapter describes the fourth and final study undertaken for this thesis. Like study one, it was designed to examine our hypotheses about intrinsic integration by comparing learning outcomes between versions for a fixed amount of time-on-task. The results for learning outcomes in study one were inconclusive, so this study incorporated a number of key changes in order to address critical issues identified with the testing instrument and methodology in that study (see chapter five). Study four also included a second game intervention period and an additional test, two weeks after the post-test to examine of the persistence of learning outcomes (as an indication of deep learning). Potential enhancements were also made to the instructional content of the game by including a help system that provided a consistent mathematical framework for structuring the children's division strategies within the game. Nonetheless, piloting this system in study two had suggested that the boys' engagement with gameplay made them highly resistant to the interruptions of the help system – despite its relevance to their performance. Consequently all groups in this study were also given a reflection session that covered the same content as the help system, but away from the distraction of the game.

7.1 AIMS

The educational interest in digital games is founded in the apparent engagement power that they possess. The manifestation of this engagement often shares many similarities with the concept of flow (Csikszentmihalyi, 1988) in terms of total concentration, distorted sense of time, and extension of self. It also epitomizes the definition of intrinsic motivation as "an activity with no apparent rewards except the activity itself" (Deci, 1971). Furthermore, it has been proposed that intrinsic motivation leads to deeper learning (Biggs, 1987), which may suggest that the motivational benefits of games could create a greater level of engagement and a deeper connection with learning content.

The concept of intrinsic integration (Habgood et al., 2005b; Kafai, 2001; Malone, 1981; Reiber, 1996) suggests that the engagement power of digital games may be most usefully harnessed for educational aims by more closely integrating a game with its learning content. Edutainment titles are criticised for including games as separate extrinsic motivation or reward for completing learning content (Lepper & Malone, 1987). Habgood et al (2005b) suggest that this integration should be between a game's core mechanics and its learning content, in order to "embody the learning material within the structure of the gaming world and the player's interactions with it" (p. 494).

The results of study three suggest that the intrinsic version of Zombie Division does indeed have a greater motivational appeal than the extrinsic version. Nonetheless, study two also highlighted the potential power of the extrinsic version to make children participate in a learning activity as well. Study one also failed to provide convincing evidence for the superior educational effectiveness of either approach, so the primary aim of study four was to re-evaluate the relative learning gains for intrinsic and extrinsic versions of Zombie Division.

Learning Outcomes

In order to provide a fair evaluation of learning potential, learning gains were compared for a fixed amount of time on task, against a control-group version of the game. Measures of overall learning gains were used to test the relative educational effectiveness of intrinsic and extrinsic approaches for Zombie Division.

In this study, measures of learning were taken at pre-test, post-test, and delayed-test, two weeks after the main intervention. As a motivational incentive, the delayed-test was immediately preceded by another chance for the children to play the game. A traditional delayed-test design does not usually include an intervention between post and delayed-test, but we will continue to use the term for communicative clarity. This study was therefore able to provide two potential measures of deep learning, in the form of the delayed-test results and learning gains for the conceptual questions over the span of the intervention (refer to section 6.3.2 for justification).

Transfer

One way in which intrinsic integration may potentially produce less effective learning than an extrinsic approach, is through the role of transfer. "Transfer between tasks is related to the degree to which they share common elements" (Bransford et al., 2000, p.78). For this reason research has shown that children find it difficult to transfer mathematical skills between different contexts (Nunes et al., 1993). The embedded nature of intrinsic learning content may therefore create a very specific context, which makes it harder to transfer learning to the different context of a classroom.

Study one produced average in-game mathematical accuracy scores of over 80% while post-test accuracy scores for comparable dividend-based questions were closer to 50%. This statistically significant difference suggested that learning from both versions of the game may not have transferred effectively to the tests. However, this finding required verification, as the affective contexts of the tests and game were quite different: the tests were performed on paper in the children's usual classrooms and the games were played on computers in the school's ICT-suite. Therefore this study also included a challenge-level that replicated a portion of the test questions in a level of the game in order to more accurately examine transfer between the two contexts. Both the computer-based test and the games were carried out on computers in the same location (the school's ICT-suite) so that the affective comparison would be fair. This data would be used to check the validity of this finding so that more definitive conclusions can be drawn about the role of transfer in the educational effectiveness of the intrinsic version of the game.

<u>Reflection</u>

Another reason why the intrinsic integration may produce less effective learning relates to the potential conflict between flow and reflective learning processes. Reflection is often considered to play an essential role in developing metacognitive skills that are critical to effective learning (e.g. Bransford et al., 2000, p.97). However, many games provide intensely interactive 'twitch-speed' (Prensky, 2001) experiences that seem unlikely to promote contemplative reflective processes such as these.

More detailed process data was produced by the game in this study, in order to collect evidence relating to reflective behaviours taking place during gameplay. The experience of study two suggested that there was a significant problem with the ability of many boys to reflect upon the mathematical content of the game. The combination of an engaging interactive experience (fighting skeletons) in a competitive environment (competing to reach the highest levels) kept them focussed on the game's goals, but prevented them from appreciating the efforts of the help system to assist them in achieving those goals. Studies using commercial off-the-shelf games (COTS), have used guided reflection as a way of structuring their educational goals around the game (e.g. K. D. Squire, 2004). Our experience so far, suggests that reflection away from the game could prove more effective than a help system that attempts to impose reflection during play. Furthermore, the same competition that prevents reflection in the game could actually become a source of motivation and focus for a guided reflection session. Consequently this study included a teacher-led session away from the game to review the content covered by the help system in the game. This was delivered to the children in their experimental groups so that the mathematical content could be directed towards their experience of it in the game.

7.1.1 Key Changes

System Changes

All the changes made to the game and learning content following study one are summarised below. Please refer to chapter five for more detail on each aspect and the full reasoning behind it:

- Switch to multiplication grid

  Less of a conceptual leap for children. Had a knock-on effect of reducing the range of usable dividends to ten times each divisor.

- A mathematical 'framing' system

  To provide a relevant and coherent mathematical strategy for children entering with a range of prior experience.

- Improved save-game system

  To allow children to continue from exactly where they left off, addressing any motivational and educational differences between groups.

- Proceduralisation levels and reflection-flagging

  To provide gaming cues for reflective opportunities and structured progress towards proceduralising mathematical skills.

In addition to these, the balance of learning content and divisors was also redressed after study two as a result of the finding that 72% of the learning content encountered during the intervention was for the divisors 2, 5 and 10. This involved changing the dividends for the majority of skeletons in the game and reducing the number of levels that provided these divisors to the player.

Methodological Changes

All the changes made to the study methodology and tools following study one are summarised below. Please refer to chapter five for more detail on each aspect and the full reasoning behind it:

- Computer-based testing system

    Longer, automated test to reduce the possibility of high pre-test scores. Also collects timing data and allows percentage scores to be calculated.

- Game-based transfer test

    To provide a direct comparison between identical dividend and divisor-based questions in test and game conditions.

- More detailed process data

    To allow data to be mined for reflective behaviours.

This study also incorporated a reflection session away from the computer, in order to reinforce the game's instructional content, and promote reflection on its value in the game. The inclusion of this session makes it impossible to accurately separate the learning benefits of the games from those of the teaching in the session. However, this thesis does not seek to provide a comparison of games with traditional teaching and the comparison of intrinsic and extrinsic approaches remains valid.

7.2 METHOD

7.2.1 Design

This study used a two-factor mixed design similar to study one. The first factor, 'group' was between-subjects, giving each child one of the three different versions of the game to play (intrinsic, extrinsic or control). The second factor, 'test' was within-subjects, providing repeated measures for each child in the pre, post and delayed-tests. Fifty-eight children were assigned to one of the three conditions based on matched pre-test scores. The assignment of children was automatically performed by a computer program that sorted children according to their scores and allocated them alternately to each condition. It then used a number of heuristics to try and balance the means and standard deviations of each group. At the same time it also employed a randomised block design to ensure an equal balance of gender and year group within each condition. Twenty children were assigned to the intrinsic condition, twenty more to the extrinsic and eighteen to the control. Each resulting group had almost identical mean pre-test scores with similar standard deviations from the mean.

<u>Measures</u>

Each test provided two direct measures of learning outcomes: the total number of questions answered, and the number of correct answers made. These values were then used to calculate a percentage score (correct / total x 100), as an alternative measure of learning outcome. This additional measure can take into account strategies that may take longer to perform but are more accurate as a result (such as using the multiplication grid).

Both the games and tests also logged a large body of process data for each subject. The test process data provides timing information for each question, allowing measures for average times to be calculated. It also facilitates a breakdown and analysis of scores by divisor for each of the tests. Process data from the game is more versatile, providing a time stamped 'commentary' for the entire length of the intervention (see section 7.3.4). This can be used to ask many different post-hoc questions of the data, including comparable timings for in-game divisions as well as breakdowns of accuracy by divisor. They can also be used to provide an overview of how these and other measures change over time during the intervention.

7.2.2 Participants

All children attended a large primary school on the outskirts of a large city in the north of England. The school is situated in a low-income area with a very small percentage of students having origins outside of the UK. The school has an average number of students with special educational needs, and the attainment of their intake is below the national average. The percentage of final year students achieving expected levels in mathematics was below national averages for the preceding year.

Fifty-eight primary children between the ages of 7 years 1 months and 8 years 10 months took part in the study (mean 8 years 0 months). This was made up of forty-seven mixed-ability children from a single year group (2 complete year 3 classes) with eleven children from the top maths set of the year below. There were thirty girls and twenty-eight boys in total, all of whom had some prior experience of using the computers in the school's ICT suite. The study was carried out over the last four weeks of the final term in the school year.

7.2.3 Materials

Facilities

The tests and interventions were both carried out within the school's ICT suite. The suite contained twenty, relatively new PCs running Windows 2000 with accelerated 3D graphics support, and audio output through stereo headphones. Each PC ran the study's software via their own CD-ROM disk, and saved process data onto a USB memory stick. The teacher-led reflection sessions were carried out in the older year group's usual classroom, using an interactive whiteboard running PowerPoint to present the teaching material.

Test Materials

The computer-based test consisted of 63 multiple choice questions with four options in each case (1 correct + 3 distractors). The first three questions were non-assessed practice questions designed to ease students into the test:

e.g. Select the number of legs that a dog has:

4, 5, 6 or 7

Of the remaining 60 questions, 40 were division questions equally comprised of two recurring formats, either asking the child to select the divisor that divides a given dividend (dividend-based):

e.g. Select one number that 45 can be divided by:

4, 6, 9 or none of these

Or to select the dividend that can be divided by a given divisor (divisor-based):

e.g. Select one number that can be divided by 5:

35, 13, 29 or 41

Amongst these 40 questions there were 5 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 addition to these 40 questions another 5 dividend-based questions were added where the answer was 'none of these'.

The remaining 15 questions were conceptual: 3 tested for knowledge of the heuristic patterns associated with numbers that divide by 2, 5 and 10; and 12 more tested for an understanding of relationships between divisors, or for applying rules outside of normal limits (i.e. dividends greater than 100).

The order of the questions was randomised, but remained consistent between-subjects and between tests. The software timed 15 minutes 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 was provided in corner of the screen similar to the one found in the intrinsic and extrinsic versions of the game (see figure 7.1)

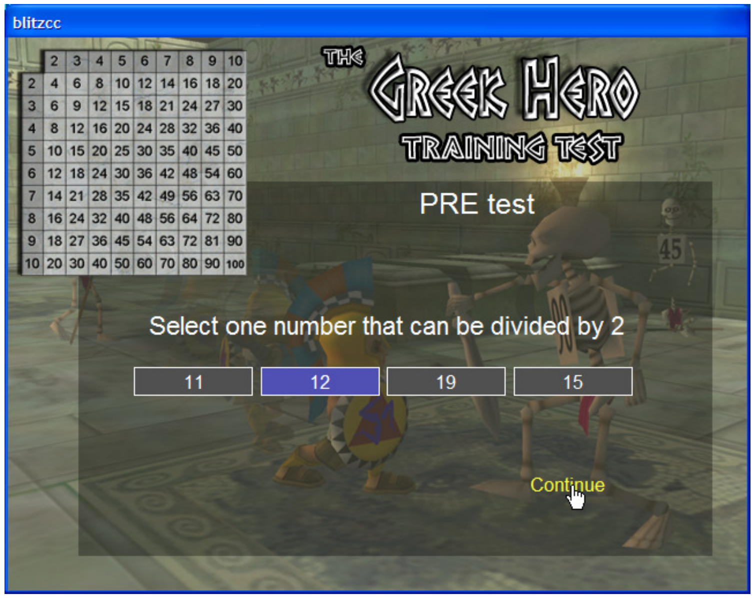


Figure 7.1: The computer-based test.

Game-Based Test Comparison

The game-based test consisted of two specially constructed levels of the game, designed to replicate a portion of the test's division problems within the game environment. This was included to provide a direct comparison with the test so that issues of transfer could be explored in more detail. All three groups played the gaming elements of these 'challenge levels' with the learning content embedded (or omitted) appropriately for their group's condition.

In the extrinsic version these questions were asked in the normal way at the end of each level, with provision for the alternative phrasing of division problems posed in the test (divisor-based questions):

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 they entered its associated room. A divisor-based question was 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' (see figure 7.2).

Figure 7.2: A divisor-based question in the challenge level. The player must either choose to divide the skeleton using one of the three available weapons or leave via the unlocked door.

However, if the player left via the exit and the skeleton was dividable, then they would find a second locked door immediately behind the exit. The game recorded an incorrect answer if the player left the first room when the skeleton could be defeated, or if the player attacked the skeleton with the wrong weapon. A correct answer was only recorded if the skeleton was defeated with the correct attack on the player's first attempt. This would destroy the skeleton in the normal way and leave behind the key to the next room. An incorrect attack would result in the skeleton attacking the player back, before vanishing into thin air and leaving behind the key to the next room. This keeps the challenge levels comparable with the tests by preventing the player from making a second attempt at a maths task.

A dividend-based question was posed in terms of a choice of four skeletons to divide with a single weapon. The exit would be locked and the key would not be provided until the player attacked one of the skeletons. (see figure 7.3).

Figure 7.3: A dividend-based question in the challenge level. The player must choose one of the four skeletons to divide with a single weapon, before they can proceed through the locked door.

Attacking a skeleton would cause all the remaining skeletons in the room to vanish. An incorrect choice would cause the skeleton to attack the player before vanishing, and a correct choice would destroy the skeleton in the usual way. Through these mechanisms the format of the questions in the test was replicated in gaming-terms that could provide a reasonably fair comparison. In study two the challenge levels also had their health scoring-mechanism removed to replicate the lack of a score in the tests. However, the apparent invulnerability this creates was obvious to the children and seemed to produce uncharacteristic changes in their behaviour. Therefore, in this study, the health-based scoring mechanism was re-introduced for the post-test challenge levels and tweaked to provide a points-based scoring mechanism for the delayed-test challenge levels.

7.2.4 Procedure

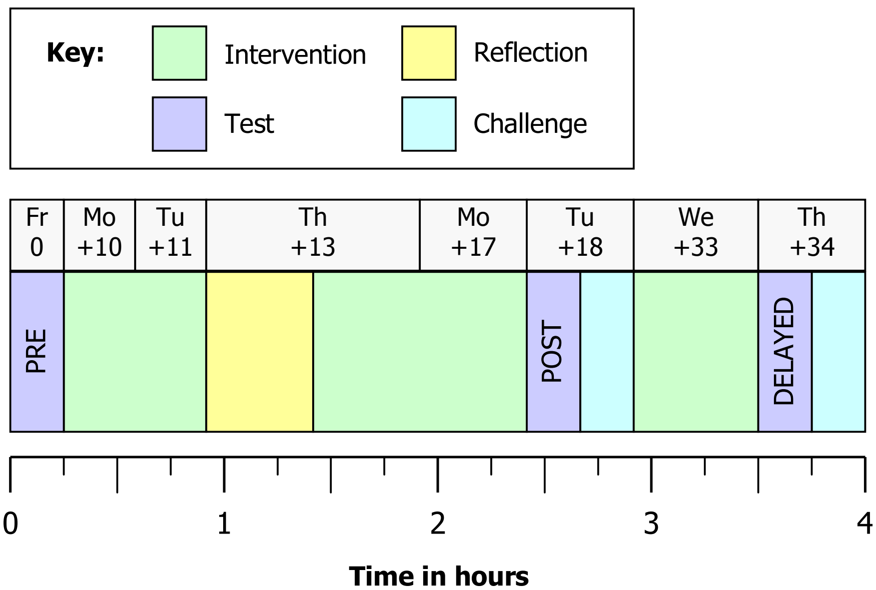


Figure 7.4: A breakdown of the total time spent on the study. The study took a total of four hours over a span of 34 days from the pre-test.

Pre-tests

The pre-tests were carried out in three successive half-hour sessions ten days before the main body of the study. Groups of up to twenty children were selected at random to complete the pre-test in the ICT suite. The task was explained to the children in front of a demonstration machine, emphasising the presence of the multiplication grid to help them with the test. However, the children were not told how to use the grid at this stage. They were informed of the 15-minute time limit, but told that they were not expected to finish all of the questions and encouraged not to treat it as a race. They were then allocated a PC of their own and allowed to begin the test in their own time. This helped to prevent copying by ensuring that most children were on different questions at any single point. Children that finished before the end of their time limit were asked to sit quietly until the entire group had finished.

Game Intervention (40 Minutes)

The children first played Zombie Division ten days after the pre-test. Each playing session lasted for approximately twenty minutes, 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 10am and the last group finishing at 11:30. This meant that the last group on each day missed a portion of their morning break, but the children consented to this arrangement. 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.

Reflection Session (30 Minutes)

The reflective sessions were delivered by a practicing teacher from the children's school who was not their normal class teacher. These took place immediately preceding the children's third playing session with the game, when all groups had played the game for an average of forty minutes. Children were taught in three separate groups according to their experimental condition. The teaching materials were tailored to the context of each group's game, but contained identical learning content (including the numerical examples) and followed the same structure. Each session lasted for half an hour beginning with 15 minutes of whole-group teaching in front of an interactive whiteboard. This included three parts, repeated for each group:

1. *Division as sharing*: the group was shown a number of objects and asked how they would work out if they could be divided into two equal-sized sets. A volunteer was then asked to come and draw circles around the sets. The group confirmed this by counting the number of objects in each set.

2. *Tables and rules*: the group was asked to suggest other techniques they could use to work out whether a number of objects can 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 group were also reminded of the numeric patterns for the 2, 5 and 10 times table, if they had not already been discussed.

3. *The multiplication grid*: the class were asked how they could solve division problems for times tables they didn't 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 if a specific dividend could be divided by a specific divisor.

The direct teaching was followed by 10 minutes of worksheet exercises carried out in pairs or groups of three. This contained twelve divisor-based division problems with an option of three divisors to divide a given dividend. A multiplication grid was included on the sheet to help them do this. The 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 that needed it.

The teaching session ended with a 5-minute plenary with the whole group back in front of the interactive whiteboard. Children were then presented with screenshots of six division problems from their own version of the game and volunteers were chosen to explain how to answer them in front of the group.

Throughout these sessions, the division problems were adapted to the context of each group's condition. Hence the intrinsic group were taught to use the grid to divide skeletons where dividends were presented as skeletons, divisors as weapons and objects as bones. The extrinsic group were taught to use the grid to answer quiz questions and the control group were taught to use the grid to solve division problems 'in general'. In the case of the two experimental conditions this mirrored the instructional content within the games, but delivered and supported by a practicing teacher.

Game Intervention (60 Minutes)

The children played the game on two more days until they had accumulated a total of one hundred minutes playing time. At this point the software automatically stopped the game and the child was sent back to their 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 post-test.

Post Tests

The post-tests were carried out on the day after the children completed their one hundred minutes playing time with the game. It was not possible to test all children at the same time, so they were divided into three new groups containing an equal number of children from each condition in each group. This was to ensure that any differences between the test sessions affected all conditions equally. The groups were then tested in three consecutive sessions in an identical way to the pre-tests, but with the addition of the 'challenge level' test. In order to prevent any distraction, children were not allowed to begin the challenge levels until the entire group had finished their post-tests.

Game Intervention (35 Minutes)

Two weeks after the post-test, the children were given another opportunity to play the game in their condition-based groups. This was done over a single forty-five minute playing session, until their playing time totalled 135 minutes.

Delayed Tests

The delayed-tests[13] were carried out on the same day as the previous playing session, after all groups had played for their allotted time. The children were divided back into their mixed condition groups and tested in three consecutive sessions identically as before. The challenge levels were taken in the same way, two days later. The children appeared to treat the post-test challenge levels as a race to the finish, so this time a points-based scoring mechanism was used to encourage accuracy over speed.

[13] A traditional delayed-test design does not usually include an intervention between post and delayed-tests, but the term is used here for communicative clarity.

7.3 RESULTS

The results for this study are divided into six sub-sections broadly based around the different data sources available. The first section uses test results to examine the effect of group and gender interactions on overall learning outcomes, as well as outcomes by question-type. This addresses the primary aims of the study as well as providing specific information on learning gains for conceptual questions as an indication of deep learning. The second section uses the challenge level data to investigate the difference between challenge and test scores as an indication of transfer. The third section analyses the test process data to explore trends in the amount of time taken to answer questions between the three tests, as an indication of changes in mathematical strategies. The fourth section reports the results obtained from the game process data including outcomes of learning tasks, a breakdown of scores for each divisor, and a comparison between game and test accuracy. This provides another indication of transfer between the game and test contexts as well as an analysis of the learning content encountered by children in the game. The fifth section reports the results of data-mining the game process data to provide an analysis of behavioural changes over time for the intervention. This allows behaviours to be explored that relate to strategies for creating reflective opportunities in the game. The final section includes an analysis of primary learning outcomes with respect to gender differences.

7.3.1 Learning Outcomes

One subject was discounted from the study after being identified as having special educational needs in mathematics. Four children were also absent during the week of the delayed-test, which was also the last week of the school year.

Numerical Scores

The primary measures of numerical and percentage scores were examined using a two-way repeated measures ANOVA with three levels of the within-subjects factor 'test' (pre, post and delayed) and three levels of the between-subjects factor 'group' (intrinsic, extrinsic and control). The analysis of numerical scores (table 7.1) revealed a main effect of test ($F(2,100) = 26.52$, $MSE = 687.22$, $p < 0.001$, $\eta^2 = 0.35$) but no test by group interaction.

Table 7.1. Mean numerical scores by group and test

	Intrinsic (n=17)	Extrinsic (n=18)	Control (n=18)
Pre-test	17.53 (6.38)	17.61 (6.63)	17.33 (6.84)
Post-test	24.71 (10.03)	24.50 (11.21)	22.50 (11.50)
Delayed-test	24.35 (10.58)	24.17 (10.68)	22.11 (10.58)

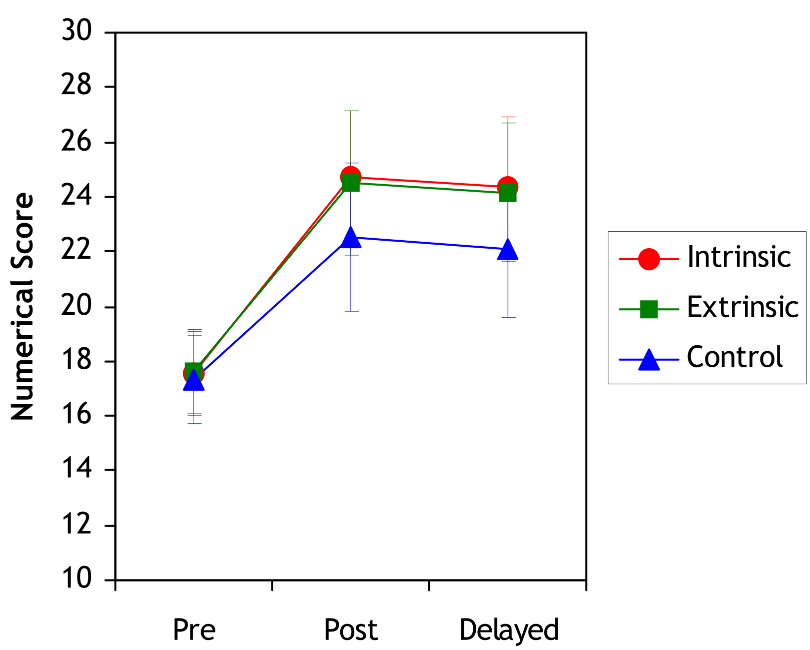


Figure 7.5. Mean numerical scores by group and test

The questions in each test could be divided into three types: dividend-based, divisor-based and conceptual. A significant improvement of scores in the conceptual category could be seen as indicative of deep learning. Table 7.1b shows the breakdown of numerical scores for each question type (dividend-based, divisor-based and conceptual). This was examined using a two-way repeated measures MANOVA with three levels of the within-subjects factor 'test' (pre, post and delayed) and three levels of the between-subjects factor 'group' (intrinsic, extrinsic and control). The multivariate analysis revealed a significant main effect of test ($F(6,45) = 8.73$, $p < 0.001$, $\eta^2 = 0.54$) but no test by group interaction. The univariate analyses only showed a significant main effect of test for dividend-based questions ($F(2,100) = 29.91$, $MSE = 221.53$, $p < 0.001$, $\eta^2 = 0.37$) and divisor-based questions ($F(2,100) = 10.82$, $MSE = 73.01$, $p < 0.001$, $\eta^2 = 0.18$). This suggests that overall children in this study made an improvement in their numerical scores for dividend-based and divisor-based questions, but not the conceptual.

Table 7.1b. Mean numerical scores by group, question type and test

	Intrinsic (n=17)			Extrinsic (n=18)			Control (n=18)		
	Dividend (max of 25)	Divisor (max of 20)	Conceptual (max of 15)	Dividend (max of 25)	Divisor (max of 20)	Conceptual (max of 15)	Dividend (max of 25)	Divisor (max of 20)	Conceptual (max of 15)
Pre-test	7.29 (2.50)	6.76 (2.61)	3.47 (1.94)	7.22 (2.63)	6.22 (2.94)	4.17 (2.46)	6.83 (3.84)	6.89 (2. 68)	3.61 (1.91)
Post-test	12.18 (4.64)	9.24 (4.34)	3.29 (1.93)	11.33 (4.95)	8.17 (5.01)	5.00 (2.52)	9.17 (5.03)	8.50 (4.73)	4.83 (3.20)
Delayed-test	11.82 (5.16)	8.88 (3.92)	3.65 (2.37)	10.89 (4.36)	8.78 (3.87)	4.50 (3.49)	8.39 (3.91)	8.39 (5.15)	5.33 (3.53)

Percentage Scores

Numerical scores do not incorporate the total number of questions completed by each child, and so cannot account for their use of strategies that may take longer to perform but are more accurate as a result (such as using the multiplication grid). Percentage scores (table 7.2) were examined using a two-way repeated measures ANOVA with three levels of the within-subjects factor 'test' (pre, post and delayed) and three levels of the between-subjects factor 'group' (intrinsic, extrinsic and control). This revealed both a significant main effect of test ($F(2,100) = 38.41$, $MSE = 3213.30$, $p < 0.001$, $\eta^2 = 0.43$) and a test by group interaction ($F(4,100) = 3.11$, $MSE = 260.34$, $p < 0.05$, $\eta^2 = 0.11$). Post-hoc pair wise comparisons of the main effect of percentage scores revealed no significant differences between the groups. Paired samples t-test comparisons[14] for each group between each test showed significant differences in percentage score for: the intrinsic group between pre and post-tests ($t(18) = 4.68$, $p < 0.001$) and pre and delayed-tests ($t(16) = 5.68$, $p < 0.001$); the extrinsic group between pre and post-test ($t(19) = 3.08$, $p < 0.01$) and pre and delayed-tests ($t(17) = 4.00$, $p < 0.005$); and the control group between the pre and post-tests ($t(17) = 3.72$, $p < 0.005$) and pre and delayed-tests ($t(17) = 3.80$, $p < 0.005$). However, independent samples t-tests between pairs of groups for each test only showed significant differences in percentage score in the delayed tests, and only between the intrinsic and extrinsic group ($t(33) = 2.13$, $p < 0.05$) and the intrinsic and control group ($t(33) = 2.59$, $p < 0.05$). This suggests that the intrinsic group were significantly better than the other two groups in their delayed-test results.

[14] Not corrected for multiple comparisons.

Table 7.2. Mean percentage scores by group and test

	Intrinsic (n=17)	Extrinsic (n=18)	Control (n=18)
Pre-test	35.10 (11.71)	32.49 (14.17)	31.88 (15.19)
Post-test	51.01 (19.96)	43.97 (19.67)	40.38 (20.65)
Delayed-test	58.42 (22.60)	44.36 (16.11)	40.57 (18.10)

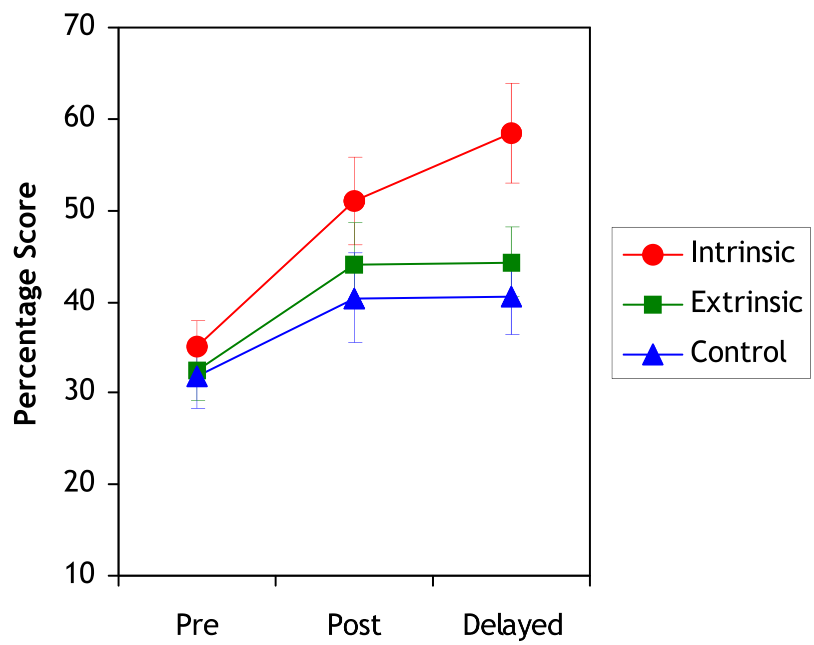


Figure 7.6. Mean percentage scores by group and test

Table 7.2b shows the breakdown of percentage scores for each question type (dividend-based, divisor-based and conceptual). This was examined using a two-way repeated measures MANOVA with three levels of the within-subjects factor 'test' (pre, post and delayed) and three levels of the between-subjects factor 'group' (intrinsic, extrinsic and control). The multivariate analysis revealed a significant main effect of test ($F(6,45) = 11.33$, $p < 0.001$, $\eta^2 = 0.60$) and a test by group interaction ($F(12,92) = 2.03$, $p < 0.05$, $\eta^2 = 0.21$). The univariate analysis showed a significant main effect of test for dividend-based questions ($F(2,100) = 35.03$, $MSE = 5108.25$, $p< 0.001$, $\eta^2 = 0.41$), divisor-based questions ($F(2,100) = 17.14$, $MSE = 3951.56$, $p< 0.001$, $\eta^2 = 0.26$) and conceptual questions ($F(2,100) = 3.59$, $MSE = 541.90$, $p< 0.05$, $\eta^2 = 0.07$). This suggests that overall children in this study made an improvement in their percentage scores for all three types of question.

The univariate analysis also showed a significant test by group interaction for dividend-based questions ($F(4,100) = 3.72$, $MSE = 543.40$, $p< 0.01$, $\eta^2 = 0.13$). Post-hoc pair-wise comparisons of the main effect of percentage scores revealed a significant difference between the intrinsic group and control ($p < 0.005$)[15]. This suggests that only the intrinsic group performed better than the control group in their percentage scores for dividend-based questions (see figure 7.6b).

Table 7.2b. Mean percentage scores by group, question type and test

	Intrinsic (n=17)			Extrinsic (n=18)			Control (n=18)		
	Dividend	Divisor	Conceptual	Dividend	Divisor	Conceptual	Dividend	Divisor	Conceptual
Pre-test	32.86 (10.23)	43.19 (16.87)	28.44 (15.31)	31.06 (12.01)	35.17 (19.79)	31.29 (21.63)	28.66 (16.72)	39.39 (19.69)	27.59 (16.59)
Post-test	57.01 (22.45)	60.24 (26.95)	27.48 (15.06)	47.59 (20.86)	45.05 (27.05)	36.32 (18.01)	38.31 (21.18)	46.55 (25.48)	35.38 (23.57)
Delayed-test	61.73 (25.61)	70.24 (26.90)	35.77 (22.27)	46.80 (16.90)	50.80 (20.24)	31.02 (23.69)	35.76 (16.87)	47.57 (26.86)	39.53 (24.13)

[15] Not corrected for multiple comparisons.

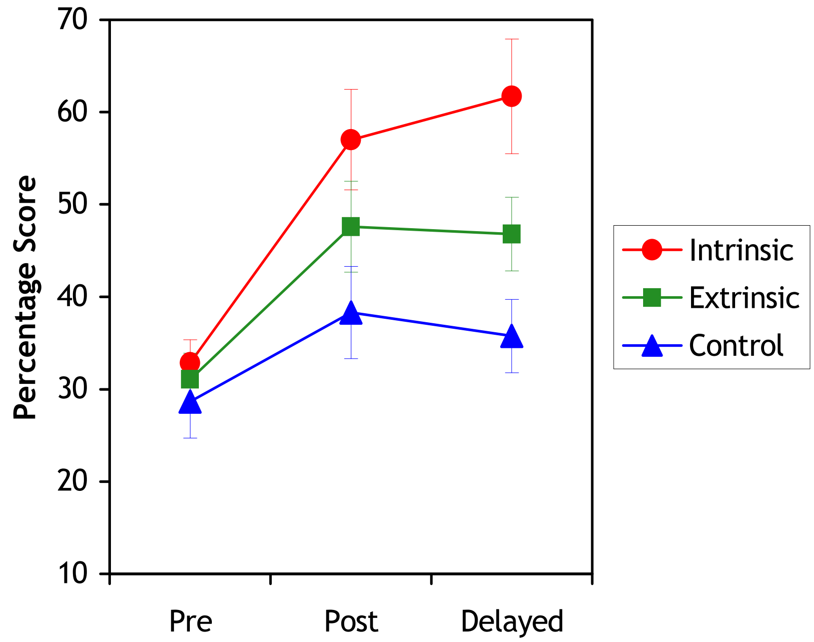


Figure 7.6b. Mean percentage scores for dividend questions by group and test

Learning Outcomes by Divisor

As well as examining different types of questions, the children's results can also be examined for each divisor. Significant differences between groups in learning gains for familiar and unfamiliar divisors could indicate different strategies taking place. Figures 7.7-7.9 show the mean percentage scores in each test by divisor, including questions where the answer was 'none of these' (N). These nine measures were analysed using a multivariate repeated measures analysis with three levels of the within-subjects factor 'test' and one between-subject factor of 'group' (intrinsic, extrinsic and control). The multivariate analysis showed a significant main effect of test ($F(18,33)=4.95$, $p< 0.001$, $\eta^2 = 0.73$), but no test by group interaction. Significant univariate results were found for the divisors two ($F(1.68,100) = 10.87$, $MSE = 5035.24$, $p< 0.001$, $\eta^2 = 0.18$)[2], three ($F(2,100) = 6.50$, $MSE = 4538.57$, $p< 0.005$, $\eta^2 = 0.12$), four ($F(2,100) = 13.74$, $MSE = 7411.68$, $p<0.001$, $\eta^2 = 0.22$), five ($F(2,100) = 10.71$, $MSE = 5677.64$, $p< 0.001$, $\eta^2 = 0.18$), six ($F(2,100) = 18.32$, $MSE = 8076.33$, $p<0.001$, $\eta^2 = 0.27$), eight

($F(2,100) = 9.82$, $MSE = 4093.41$, $p < 0.001$, $\eta^2 = 0.16$) and ten ($F(1.67,100) = 24.58$, $MSE = 14738.29$, $p < 0.001$, $\eta^2 = 0.33$)[16]. These results therefore show a significant overall improvement in test scores for questions involving all divisors except the divisor nine and questions where the answer was 'none of these'.

[16] The divisors two and ten failed Mauchly's test of sphericity, so the results for Greenhouse-Geisser are reported for these cases.

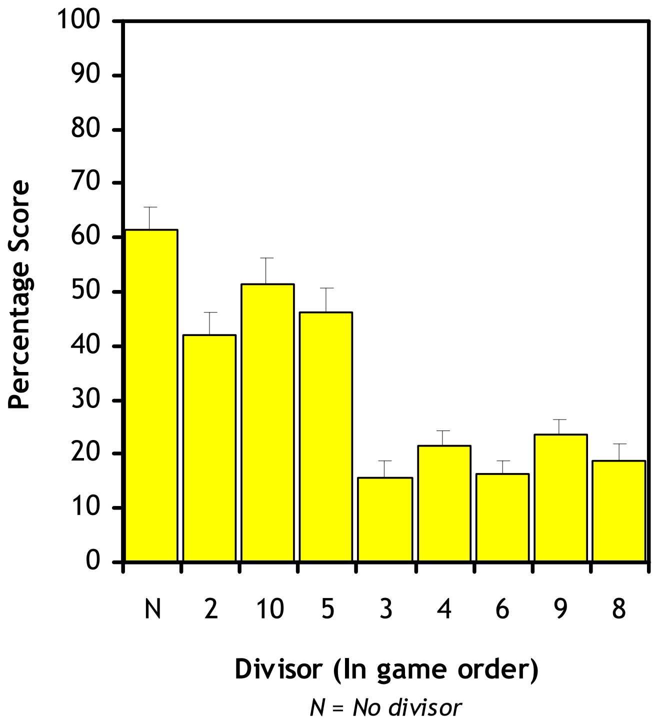


Figure 7.7. Mean percentage scores by divisor at pre-test

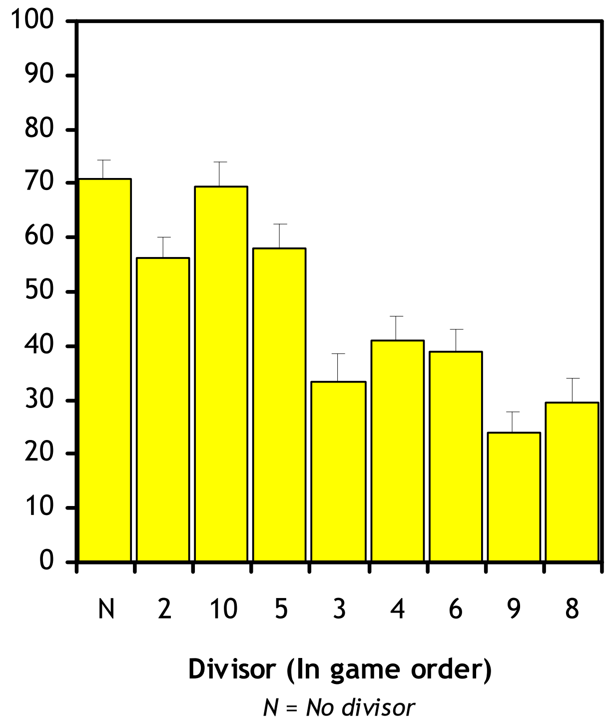


Figure 7.8. Mean percentage scores by divisor at post-test

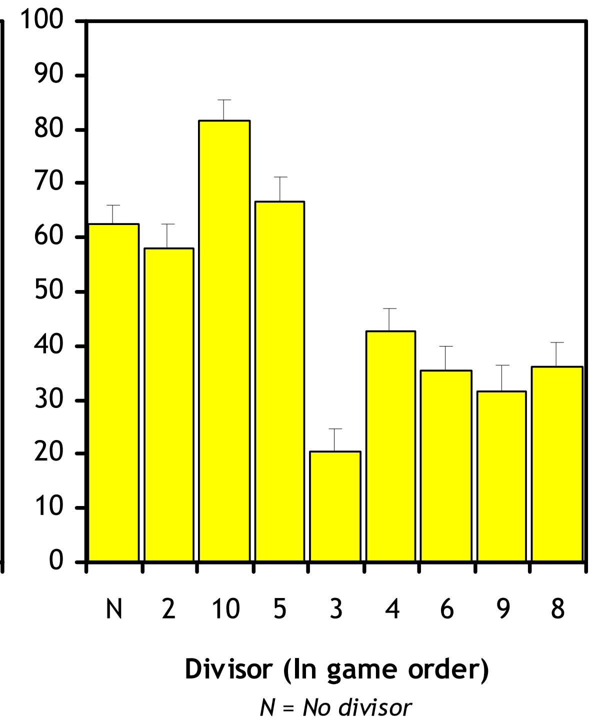


Figure 7.9. Mean percentage scores by divisor at delayed-test

7.3.2 Challenge Levels

Transfer

The challenge levels provide a direct comparison between students' ability to solve the same division problems in the assessment test and in the game. Twenty of the questions from the assessment are replicated in the challenge levels of the intrinsic and extrinsic versions. In this study, these levels were played on two separate occasions: once following the post-tests and once following the delayed tests. Table 7.3 shows the mean percentage scores for the challenge levels alongside their equivalent percentage scores derived from the assessments. A three-way repeated measures ANOVA was performed on percentage scores with two within-subjects factors, 'method' (assessment or challenge) and 'test' (post or delayed) and one between-subjects factor, 'group' (intrinsic or extrinsic).

Table 7.3. Mean and S.D. of percentage scores by test-type and group.

	Assessment		Challenge	
	Intrinsic (n=16)	Extrinsic (n=18)	Intrinsic (n=16)	Extrinsic (n=18)
Post-Test	59.95 (18.40)	50.00 (19.19)	57.63 (22.13)	47.00 (19.29)
Delayed-Test	67.29 (23.12)	51.88 (16.00)	75.63 (16.21)	60.28 (22.52)

This revealed a significant main effect of test ($F(1,32) = 25.07$, $MSE = 2597.79$, $p< 0.001$, $\eta^2 = 0.44$), and a test by method interaction ($F(1,32) = 8.12$, $MSE = 1076.32$, $p< 0.01$, $\eta^2 = 0.20$). Paired samples t-tests[17] of each method between the post and delayed tests only revealed a significant difference for challenge scores ($t(33)=5.73$, $p<0.001$). Paired samples t-tests of each test between

[17] Not corrected for multiple comparisons.

assessment and challenge methods only revealed a significant difference for the delayed-test ($t(33)=3.07$, $p<0.005$). Percentage scores achieved in the delayed-test challenge were therefore significantly higher than those in the post-test challenge. Percentage scores achieved in the delayed-test challenge were also significantly higher than those in the delayed-test assessment.

Racing to Finish

Our experience during the post-test challenge level led us to believe that children were treating the challenge level as a race, favouring strategies to complete the levels as fast as possible. Consequently, the health-based scoring mechanism was replaced with a point-based scoring mechanism in the delayed-test challenge. This change was made to favour accuracy over speed, so we were interested to see if there was a difference in timings between the two attempts at the challenge levels. Table 7.4 shows the mean overall time taken to complete each challenge. A two-way repeated measures ANOVA was performed with two levels of the within-subjects factor 'test' (post and delayed) and two levels of the between-subjects factor, 'group' (intrinsic and extrinsic). This revealed no main effects.

Table 7.4. Mean and S.D of challenge durations by test.

Challenge	Duration in seconds	
	Intrinsic (n=16)	Extrinsic (n=18)
Post-test	780.44 (191.76)	924.67 (270.74)
Delayed-test	741.00 (250.57)	847.67 (205.34)

Similarly, a correlation between challenge scores and the time taken to complete them would indicate whether the children's speed affected their performance. Bivariate correlations were performed between the challenge percentage scores, their test equivalents and the duration of the challenge levels. Percentage scores for the challenge were significantly correlated with their test equivalents at post-test ($r(37) = 0.62$, $p < 0.01$) and at delayed-test ($r(32) = 0.72$, $p < 0.01$). However there was no correlation between duration and scores for either challenge.

7.3.3 Testing Process

Speed Verses Accuracy

Percentage scores can take into account strategies that may take longer to perform but are more accurate as a result (such as using the multiplication grid). Therefore the presence of group differences in percentage scores may suggest that different strategies are also being used in different groups. Analysing the changes in the average time taken to answer questions between pre, post and delayed-test may help to confirm these potential differences in strategy.

The process data for the pre, post and delayed tests was used to obtain each subject's mean time for answering questions in each test (see table 7.5). This measure was then examined using a three-way repeated measures ANOVA with three levels of the within-subjects factor 'test', three levels of the factor 'group' (intrinsic, extrinsic and control) and two levels of the factor 'gender' (male and female). This revealed a significant main effect of time ($F(2,94) = 4.54$, $MSE = 133.16$, $p < 0.05$, $\eta^2 = 0.09$), but no interactions (see figure 7.10). Paired samples t-tests[18] of combined group average times showed a significant difference

[18] Not corrected for multiple comparisons.

between pre and delayed-tests ($t(52) = -2.04$, $p < 0.05$) and post and delayed-tests ($t(52) = -2.87$, $p < 0.01$). The average time per question in the delayed-test was therefore significantly longer than in either the pre or post-tests. This suggests that an overall increase in the amount of time taken to answer questions did take place, but there were no differences between groups.

Table 7.5. Mean number of seconds per question by group and test

	Intrinsic (n=17)	Extrinsic (n=18)	Control (n=18)
Pre-test	16.87 (7.19)	13.27 (5.68)	14.38 (4.58)
Post-test	18.01 (9.83)	12.64 (5.46)	12.85 (5.59)
Delayed-test	21.58 (10.85)	15.41 (10.53)	14.92 (4.65)

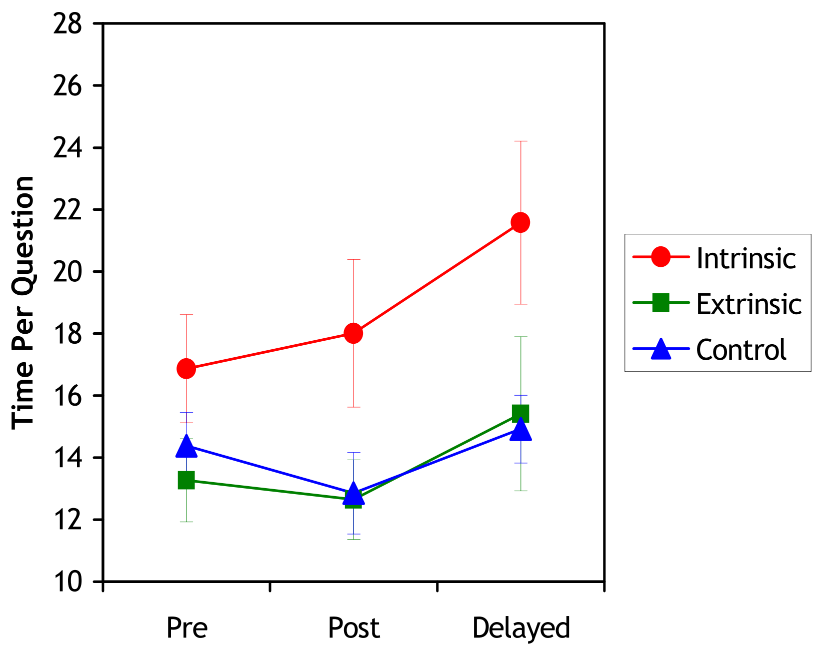


Figure 7.10. Mean number of seconds per question by group and test

Bivariate correlations were also performed between percentage scores (table 7.2) and average question timings in each test (table 7.5) as an indication of a speed /

accuracy trade off. Times were significantly correlated with percentage scores at pre-test ($r(57) = 0.35$, $p < 0.01$), post-test ($r(57) = 0.44$, $p < 0.005$) and delayed-test ($r(53) = 0.32$, $p < 0.05$). Children who took longer to answer questions did generally achieve higher scores, but there was no indication that this relationship changed over the course of the study.

7.3.4 Learning Process

The process logs produced by the game provide a versatile source of data in the form of a time-stamped commentary on the game as it is being played. For this study, over two and a half thousand log files were mined for the purposes of post-hoc analysis. The mining focussed on obtaining broadly comparable data for 'skeletal divisions' and 'quiz divisions' from the intrinsic and extrinsic versions of the game. Previous studies have revealed notably high mathematical accuracy scores in both the intrinsic and extrinsic versions of the game. A comparison between in-game and test accuracy scores can be used as an additional indication of any potential near-transfer problems in the different versions.

Game Task Outcomes

It is relatively easy to interpret the extrinsic 'quiz' data in terms of accuracy scores, but the free-roaming nature of skeletal encounters in the intrinsic game makes it much more complex and open to interpretation. In particular, determining the beginning and end of an encounter with a skeleton requires assumptions about the user's engagement with division problems that may not always be true. For the purposes of this analysis an encounter was said to begin when the player enters the same room as a skeleton, and said to end if the player attacks the skeleton or leaves the room again without dividing it. The different outcomes of encounters can therefore be defined as follows:

We will define a *maths task* as a single encounter with a divisor-based problem in the game. A maths 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 maths task is a dividend-based question with a choice of up to three divisors to divide a given dividend. The choice of divisors remains the same throughout all maths tasks on the same game level. 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 the divisors provided if none of them would divide the dividend. In the intrinsic version this involves manoeuvring to avoid combat with the skeleton, while in the extrinsic version a player selects an alternative answer marked 'none of these'. We will refer to dividends that are dividable by one of the available divisors as *target dividends,* and those that are not as *distractor dividends*. A maths task is not concluded until either it is answered correctly, or the level/quiz is restarted. However, the outcome of a maths task is assessed in terms of the first attempt made upon the dividend. This means there can be one of six outcomes depending on whether the dividend is a target or distractor:

Correct Outcomes:

1. Target Attacked Correctly (TAC) – the player correctly divides the dividend by one of the available divisors on the first attempt.
2. Distractor Left Correctly (DLC) – the player correctly rejects all of the available dividends for an indivisible dividend on the first attempt.

Incorrect Outcomes:

3. Target Attacked Incorrectly (TAIN) – the player incorrectly attempts to divide the dividend by one of the available divisors on the first attempt.
4. Target Left Incorrectly (TLIN) – the player rejects all of the available divisors for a dividable dividend on the first attempt.
5. Distractor Attacked Incorrectly (DAIN) – incorrectly attempts to divide an indivisible dividend by one of the available divisors on the first attempt.

Undefined Outcomes:

6. Undefined (U) – A restarted level or an error such as a crash or a glitch in the log causes an encounter to be unresolved.

Unless otherwise stated, all data in this study refers to a player's first encounter with each maths task in the game (as well as their first attempted answer at that task). Subsequent encounters with the same maths task will occur when a level or quiz is restarted after a player loses all their health. However, the pre/post/delayed-tests do not include lives and always accept the first answer given. Therefore, to allow meaningful comparisons to be made between game and test process data, only the first encounters with each maths task are generally included in the analyses. Nonetheless, table 7.6a shows the overall mean number of maths tasks for *all encounters*, as well as the mean number of first encounters with maths tasks, and the mean accuracy of first encounters with maths tasks. Table 7.6b shows a breakdown of the mean number of first encounters with maths tasks according to their outcomes. The figure for first encounter accuracy from table 7.6a is the sum of both correct outcomes over the total number of first encounters.

Table 7.6a. Mean number and percentage accuracy of maths tasks by group.

	Total encounters with maths tasks	1st encounters with maths tasks	% accuracy of 1st encounters
Intrinsic (n=19)	353.84 (190.59)	148.68 (83.40)	70.88 (7.93)
Extrinsic (n=20)	190.60 (76.60)	128.30 (81.70)	79.58 (12.12)
Combined (n=39)	270.13 (164.23)	138.23 (82.09)	75.34 (11.07)

Table 7.6b. Mean first encounters with maths tasks by group and outcome.

	Targets Attacked Correctly	Distractors Left Correctly	Targets Attacked Incorrectly	Targets Left Incorrectly	Distractors Attacked Incorrectly	Undefined
Intrinsic (n=19)	86.53 (55.02)	19.47 (10.39)	6.79 (4.51)	29.68 (18.83)	5.47 (3.23)	0.74 (1.52)
Extrinsic (n=20)	92.90 (67.70)	16.25 (10.60)	2.75 (1.83)	7.05 (4.75)	5.20 (2.63)	4.15 (1.31)
Combined (n=39)	89.79 (61.12)	17.82 (10.49)	4.72 (3.94)	18.08 (17.62)	5.33 (2.91)	2.49 (2.22)

These measures provide some indication of the children's relative success for maths tasks undertaken during the game. A one-way MANOVA was performed on all the measures in tables 7.6a and 7.6b with two levels of the between-subjects factor, 'group' (intrinsic and extrinsic). This revealed a significant main effect of group at the multivariate level ($F(1,30) = 17.76$, $p< 0.001$, $\eta^2 = 0.83$). The univariate analysis then showed a significant main effect of group on the following measures: total encounters ($F(1,30) = 12.55$, $MSE = 259647.03$, $p< 0.005$, $\eta^2 = 0.25$), first encounter accuracy ($F(1,30) = 6.95$, $MSE = 736.76$, $p< 0.05$, $\eta^2 = 0.16$), targets attacked incorrectly ($F(1,30) = 13.65$, $MSE = 158.99$, $p< 0.005$, $\eta^2 = 0.27$), targets left incorrectly ($F(1,30) = 27.12$, $MSE = 4991.71$, $p< 0.001$, $\eta^2 = 0.42$) and undefined encounters ($F(1,30) = 56.58$, $MSE = 113.51$, $p< 0.001$, $\eta^2 = 0.61$). Undefined encounters and first encounter accuracy were both significantly lower in the intrinsic group, while total encounters, targets attacked incorrectly and targets left incorrectly were all significantly higher in the intrinsic group. These results suggest that the intrinsic group was less successful at in-game maths tasks than the extrinsic group, although they completed more tasks in total.

Figures 7.11 and 7.12 show a breakdown of the percentage accuracy of in-game maths tasks by divisor for the intrinsic and extrinsic groups. This percentage is calculated as the number of 'targets attacked correctly' (TAC) using a particular divisor, over the total number of first encounters where that divisor was used (TAC + 'targets attacked incorrectly' + 'distractors attacked incorrectly'). Therefore this provides some indication of how successful children were at applying that divisor in the game. The number of samples in a multivariate analysis is limited to the minimum number of samples available for every divisor included in the test. Therefore only the divisors with 50% or more children contributing from each group (none, two, ten, five and three) were included in the analysis.

A one-way MANOVA was performed on the measures for the divisors two, ten, five and three. This had two levels of the between-subjects factor, 'group' (intrinsic and extrinsic) which was significant at the multivariate level ($F(1,29) = 11.38$, $p< 0.001$, $\eta^2 = 0.66$). Univariate main effects of group were present for the divisors two ($F(1,33) = 8.47$, $MSE = 677.95$, $p< 0.01$, $\eta^2 = 0.20$), three ($F(1,33) = 5.23$, $MSE = 2399.43$, $p< 0.05$, $\eta^2 = 0.14$) and five ($F(1,33) = 51.14$, $p< 0.001$, $\eta^2 = 0.61$). In all three cases the percentage accuracy for the extrinsic group was significantly higher than the intrinsic group.

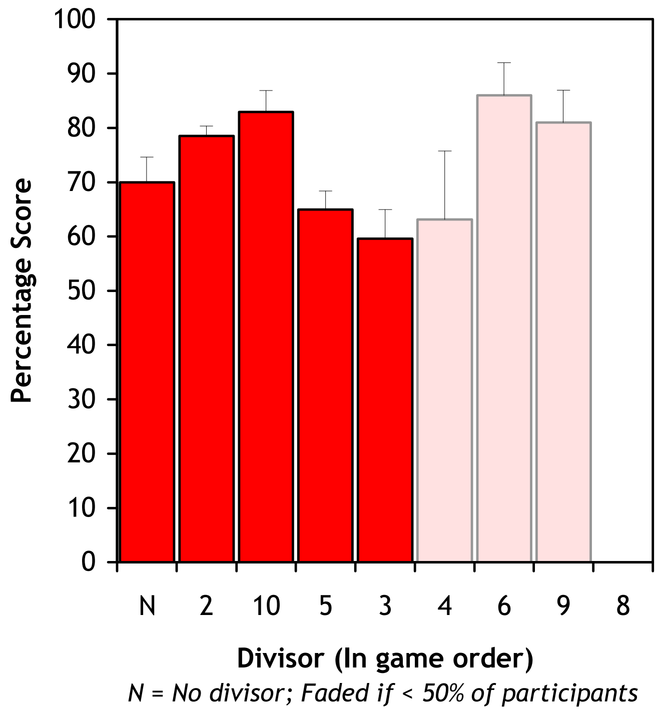


Figure 7.11. Mean in-game percentage scores by divisor for the intrinsic group.

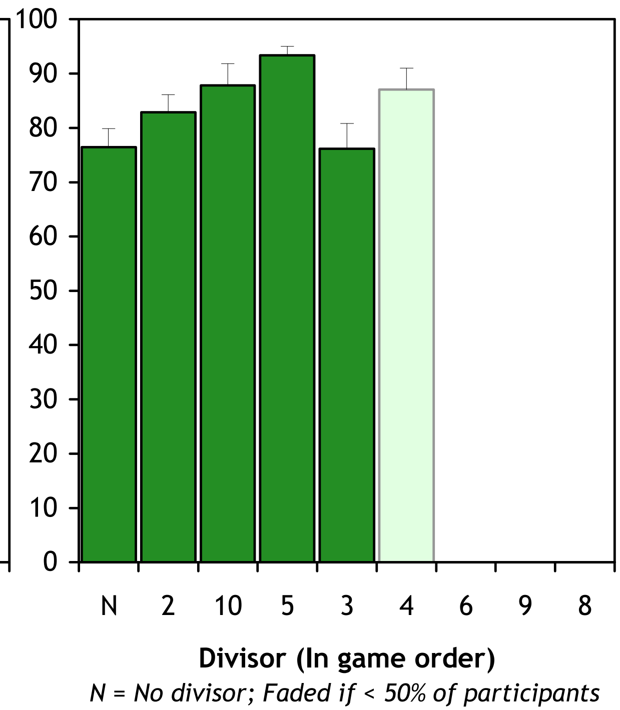


Figure 7.12. Mean in-game percentage scores by divisor for the extrinsic group.

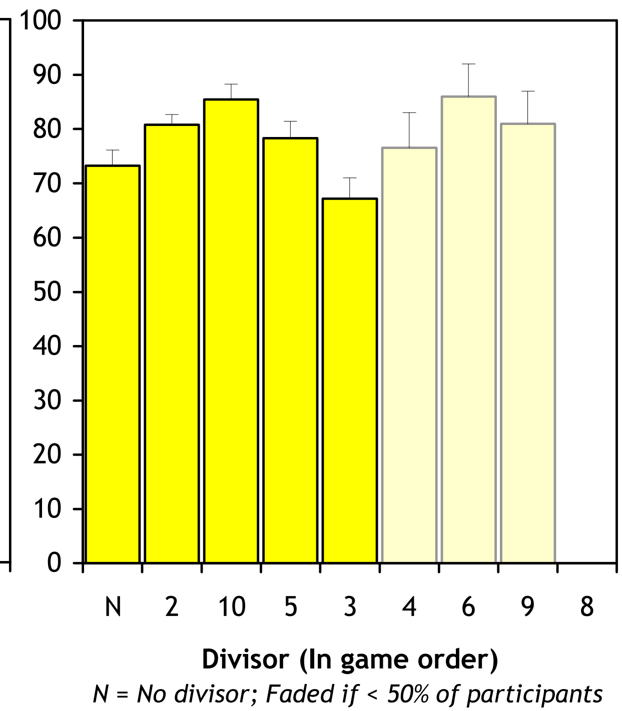


Figure 7.13. Mean in-game percentage scores by divisor for the combined groups.

Analysis of Game Task Divisors

Study two had suggested that less than a third (28%) of the content encountered by children in the intervention could be considered new to them. This may be unhelpful for promoting the conflict necessary to adopt the new mathematical strategies promoted by the game. All of the divisors and dividends were reviewed after study two and the game process logs from this study were used to examine the effect that these changes made to the amount of familiar learning content encountered during the intervention. Table 7.7 shows the combined group breakdown of divisors (including 'none of these') used in first encounters with maths tasks. Just over 55% of all maths tasks were solved using the divisors 2, 5 and 10. The corresponding multiplication tables are part of the national curriculum requirements for children of this age, so about 45% of the overall learning content in the games could be considered as unfamiliar to this audience.

Table 7.7. Breakdown of divisors used in first encounters with maths tasks.

	None (n=39)	2 (n=39)	10 (n=39)	5 (n=36)	3 (n=35)	4 (n=16)	6 (n=2)	9 (n=2)	Total
Number of first Encounters	942	1536	357	1090	1034	278	30	27	5294*
Proportion of total 1st encounters	17.47	28.49	6.62	20.22	19.18	5.16	0.56	0.50	98.2*

Divisors listed in the order they appear in the games.
* Remaining encounters are undefined

Game Task vs. Test Comparison

The accuracy in-game (figure 7.13) was also compared to the accuracy in the delayed-test assessment (figure 7.9) as another indication of transfer. A two-way repeated measures MANOVA was performed on divisors with 50% or more participants contributing in-game data (none, two, ten, five and three). This had two levels of the within-subjects factor 'method' (game and assessment) and two levels of the between-subjects factor, ‘group’ (intrinsic and extrinsic). Multivariate tests revealed a significant main effect of method ($F(1,29) = 16.52$, $p < 0.001$, $\eta^2 = 0.74$) as well as a significant group by method interaction ($F(1,29) = 4.73$, $p<0.005$, $\eta^2 = 0.45$). Univariate tests showed significant main effects of method for the divisors two ($F(1,33) = 4.06$, $MSE = 3756.72$, $p < 0.005$, $\eta^2 = 0.26$) and three ($F(1,33) = 49.00$, $MSE = 37881.5$, $p < 0.001$, $\eta^2 = 0.60$). Significant method by group interactions were found for the divisors ‘none of these’ ($F(1,33) = 4.49$, $MSE = 2364.29$, $p < 0.05$, $\eta^2 = 0.12$) and five ($F(1,33) = 13.14$, $MSE = 7148.66$, $p < 0.005$, $\eta^2 = 0.29$). Pair wise t-test comparisons[19] for each group showed that mean scores in-game were only significantly higher than scores at delayed test in the extrinsic group for the divisors ‘none of these’ ($t(19)=3.19$, $p<0.01$) and five ($t(16)=3.51$, $p<0.005$). Therefore, scores for ‘none of these’ and the divisor five fell significantly between game and delayed-test assessment for the extrinsic group and scores for the divisors two and three fell significantly for both groups.

7.3.5 Behavioural Process

All the analyses performed on the process data so far have been average measures over the entire intervention. However, the time-stamped data also allows changes in measures to be examined over time at well. This allows us to perform post-hoc analyses on changes in various measures that might be indicative of behavioural changes or differences over time. Nonetheless, these analyses are both time-consuming (each requiring custom software to process the data) and potentially increase the risk of false positive errors as a result of performing so many analyses on the data. Consequently just four measures were chosen to analyse over time for this study.

[19] Not corrected for multiple comparisons.

Overall accuracy was the primary measure of interest, observing changes over time as an indication of learning. Previous analyses have used percentage scores in first encounters as their measure of accuracy because this provides a direct comparison with the test data. The tests do not provide a second chance at questions; so comparing anything other than first encounters would not have been fair for observing transfer between the two contexts. However, the accuracy figure for the intrinsic version may be distorted, as leaving a room occupied by skeleton(s) is always interpreted as a conscious rejection of the available divisors. This could artificially increase the number of recorded TLIN outcomes (Target Left Incorrectly) for the intrinsic group and lower their overall percentage scores. Furthermore, subjective observations during the study suggested that there was a change in room exiting behaviour over time i.e. children ran away from aggressive skeletons. This could also affect TLIN outcomes and potentially obscure other interesting changes in accuracy. Consequently an alternative measure of accuracy was used for this analysis, which recorded the number of division attempts made per encounter, over all encounters (not just the first encounters). By including all encounters, this measure should average out the effect of 'unconscious rejections' in the intrinsic accuracy scores.

The second measure chosen was the number of 'targets left incorrectly' over time. The chance of 'unconscious rejections' is likely to increase as the player encounters larger numbers of more mobile skeletons. Examining changes in this measure for the intrinsic group may help to support our hypothesis about the distortion of this measure for the intrinsic group and its relationship to the difficulty of the gameplay rather than the mathematical content.

The other two measures were chosen to explore other observations about behaviours in the intrinsic group. Children were observed pausing the game to provide them with time to plan their attack. Pausing would help children to consider the mathematical and gameplay challenges presented by more aggressive skeletons. Therefore the number of pauses and the amount of time spent paused were also analysed over time for this study.

<u>Changes Over Time</u>

Mean values for each of the four measures were calculated four times, once for each quartile period of the children's total playing time. Figures 7.14 – 7.17 show plots of how these four measures change over time for each group. All four measures were examined using a two-way repeated measures MANOVA with four levels of the within-subjects factor 'progress', and two levels of the between-subjects factor 'group' (intrinsic and extrinsic). Multivariate tests showed no significant main effect of 'progress', but revealed a significant progress by group interaction ($F(12,294) = 3.28$, $p < 0.001$, $\eta^2 = 0.12$). Univariate tests revealed this interaction was significant for division attempts ($F(3, 99) = 3.84$, $MSE = 0.01$, $p < 0.05$, $\eta^2 = 0.10$), TLIN ($F(1.81,59.81) = 6.64$, $MSE = 353.67$, $p < 0.005$, $\eta^2 = 0.17$) and pause time ($F(1.68,55.55) = 3.67$, $MSE = 9578.87$, $p < 0.05$, $\eta^2 = 0.10$)[20]. Therefore division attempts, TLIN, and pause time all showed changes over time that were significantly different between the intrinsic and extrinsic groups.

Independent samples t-tests[21] between groups for the first and final quartiles showed that division attempts were significantly lower in the intrinsic group in the

[20] The measures for TLIN and pause time failed Mauchly's test of sphericity, so the results for Greenhouse-Geisser are reported for those cases.

[21] Not corrected for multiple comparisons.

first quartile ($t(33) = 4.48$, $p < 0.001$), but there was no significant difference in the final quartile; there was no significant difference in TLIN in the first quartile, but TLIN was significantly higher in the intrinsic group in the final quartile ($t(33) = 3.40$, $p < 0.005$); and there was no significant difference in pause time in the first quartile, but pause time was significantly higher in the intrinsic group in the final quartile ($t(33) = 2.29$, $p < 0.05$). Paired samples t-test comparisons for each group between the first and final quartiles showed significant differences for: TLIN in the intrinsic group ($t(16) = 2.65$, $p < 0.05$), pause time in the intrinsic group ($t(16) = 2.29$, $p < 0.05$), and division attempts in the extrinsic group ($t(17) = 2.29$, $p < 0.05$). Paired samples t-test comparisons for each group between the first and third quartiles revealed an additional significant difference for division attempts in the intrinsic group ($t(18) = 2.21$, $p < 0.05$).

Division attempts were therefore significantly lower in the intrinsic group in the early stages of the study, but the extrinsic group decreased, and the intrinsic group increased so that division attempts made by the two groups were no longer significantly different by the late stages of the study (see figure 7.14). TLIN scores between the groups were not significantly different in the early stages of the study but scores increased for the intrinsic group so that TLIN was significantly higher in the intrinsic group by the late stages of the study (see figure 7.15) The time spent paused was not significantly different between groups in the early stages of the study, but time increased for the intrinsic group so that the time spent paused was significantly higher in the intrinsic group by the late stages of the study (see figure 7.17).

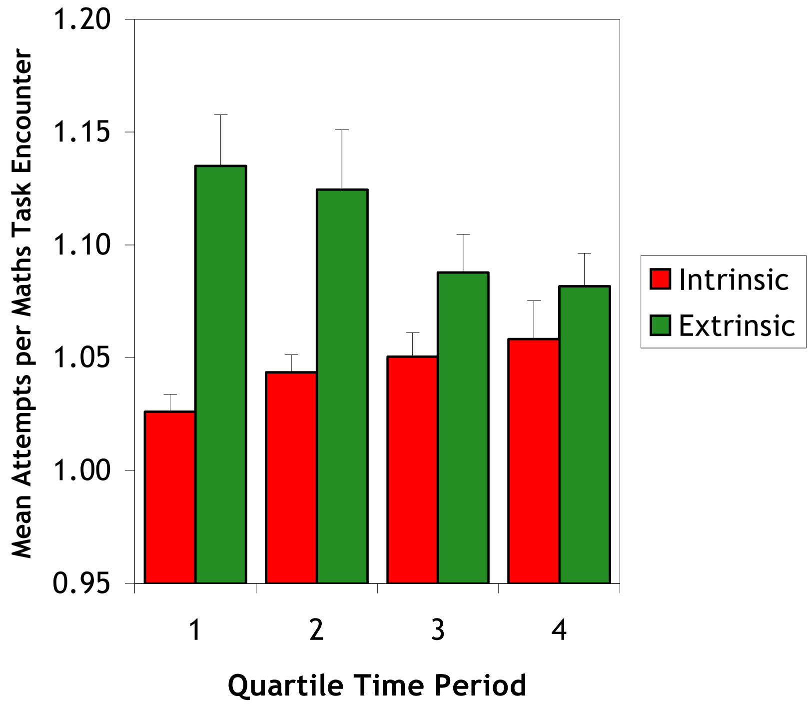


Figure 7.14. Mean number of division attempts by quartile periods.

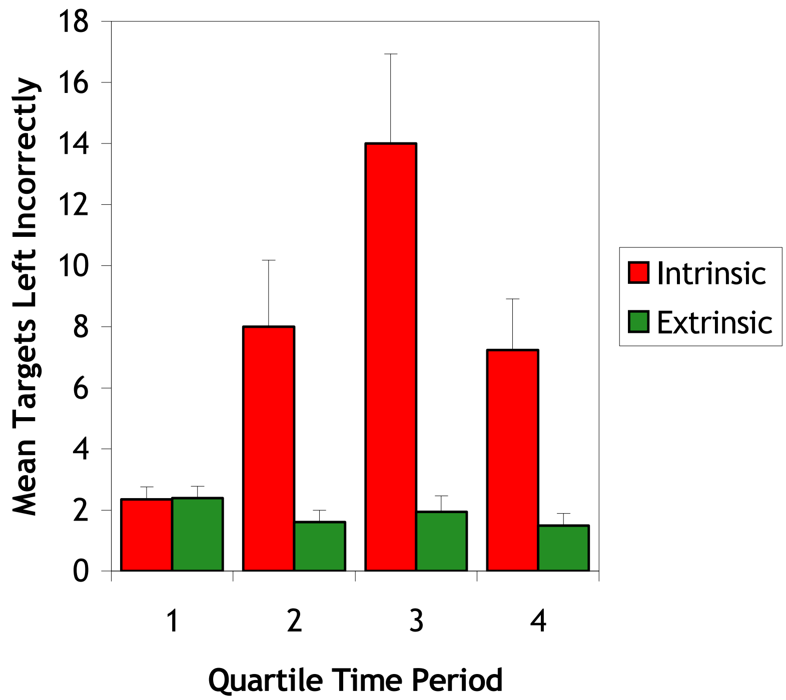


Figure 7.15. Mean targets left incorrectly by quartile periods.

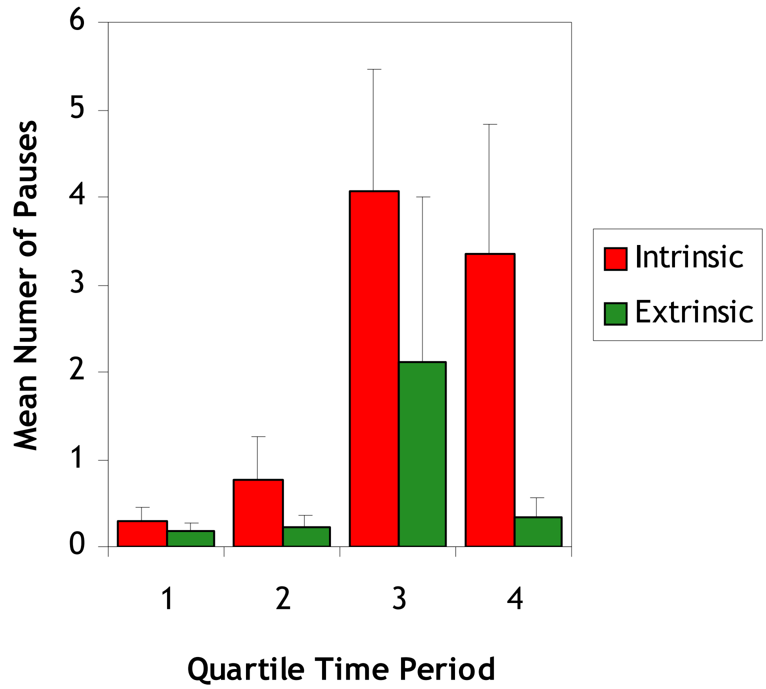


Figure 7.16. Mean number of pauses by quartile periods.

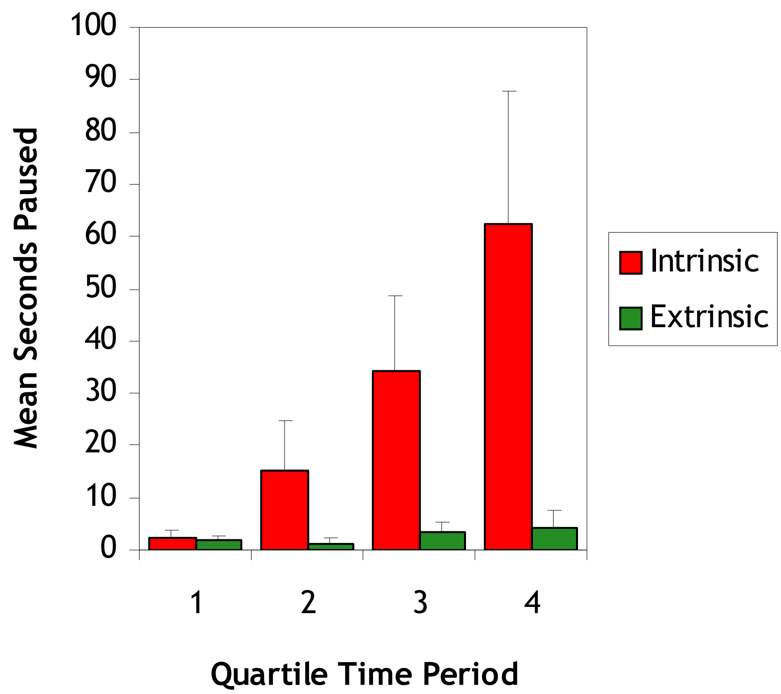


Figure 7.17. Mean time spent paused by quartile periods.

7.3.6 Gender

Although the role of gender is not considered a key part of the aims of this thesis, previous studies have revealed significant group by gender interaction effects on learning outcomes. Table 7.8 shows the numerical scores for each test by group and gender. These were examined using a three-way repeated measures ANOVA with three levels of the within-subjects factor 'test' (pre, post and delayed), three levels of the between-subjects factor 'group' (intrinsic, extrinsic and control) and two levels of the between-subjects factor, 'gender'. No significant interaction of gender was found for numerical scores.

Table 7.8. Mean and S.D. of numerical scores by group, gender and test

	Intrinsic		Extrinsic		Control	
	Male (n=9)	Female (n=8)	Male (n=8)	Female (n=10)	Male (n=9)	Female (n=9)
Pre-test	19.00 6.16)	16.22 (6.63)	16.70 (6.27)	18.75 (7.32)	17.56 (8.34)	17.11 (5.47)
Post-test	25.75 (10.25)	23.78 (10.35)	23.20 (8.01)	26.13 (14.74)	21.78 (13.81)	23.22 (9.43)
Delayed-test	24.62 (13.16)	24.11 (8.49)	23.40 (6.96)	25.13 (14.58)	22.78 (12.51)	21.44 (8.96)

Table 7.9. Mean and S.D of percentage scores by group, gender and test

	Intrinsic (n=17)		Extrinsic (n=18)		Control (n=18)	
	Male (n=9)	Female (n=8)	Male (n=8)	Female (n=10)	Male (n=9)	Female (n=9)
Pre-test	33.56 (12.56)	36.46 (11.47)	35.51 (18.23)	30.08 (10.31)	31.43 (19.08)	32.32 (11.22)
Post-test	49.40 (23.27)	52.44 (17.83)	40.32 (15.41)	48.52 (24.32)	40.19 (23.65)	40.56 (18.61)
Delayed-test	53.45 (27.15)	62.82 (18.18)	39.45 (11.31)	50.50 (19.72)	40.46 (19.43)	40.67 (17.85)

Table 7.9 shows the percentage scores for each test by group and gender. These were also examined using a repeated measures analysis of variance (ANOVA) with three levels of the within-subjects factor 'test' (pre, post and delayed), three levels of the between-subjects factor 'group' (intrinsic, extrinsic and control) and two levels of the between-subjects factor, 'gender'. No significant interaction effect of gender was found for percentage scores.

7.4 DISCUSSION

Like study one, the central aim of this study was to evaluate our hypotheses about the educational effectiveness of intrinsic and extrinsic approaches, by comparing learning outcomes for a fixed amount of time-on-task. Unlike study one, this included a framing system as part of the game, and a structured session away from the game in order to encourage reflection. The results provide useful evidence for resolving our conflicting hypotheses about the value of intrinsic games. One hypothesis predicts greater learning gains in the intrinsic version, as a result of motivational engagement and a deeper connection with the learning content. However, the other two hypotheses predict smaller learning gains: the first as a result of embedded intrinsic learning content transferring less effectively than the extrinsic; and the second because integrating learning content within a game's flow-experience may impede reflection-in-action. The discussion section of this chapter will provide a summary of this study's findings and chapter eight will then go on to interpret these findings as they relate to the original hypotheses.

7.4.1 Learning Outcomes

Pre-test scores were generally quite low (mean 17.4 out of a maximum of 60), evading a potential repeat of the high pre-test scores observed in study one. Numerical scores showed a significant overall improvement for all groups (around 6 points), but no one group performed better than any other using this measure. Interaction effects should be interpreted with some caution as no correction was made for multiple comparisons (see results section). Nonetheless, the breakdown of numerical scores by question-type confirmed this overall increase, and showed that it was only significant for divisor-based and dividend-based questions (not conceptual). Graphs of the data (figure 7.5) show this improvement in the numerical scores of all groups – including the control. Some increase for the control group was expected, as they had attended a session designed to help them reflect on the mathematical content of the tests. However, it was thought that the control group would not gain as much from the reflection session as the other groups, because the mathematical content was not relevant to their version of the game. Nonetheless the teacher running the sessions reported that all three groups were very enthusiastic and attentive as a result of the children's excitement about their involvement with the game and study. It appears that this opportunity for reflection enabled the control group to produce significant gains in numerical scores that made their gains indistinguishable from those made by the other two groups.

Calculating scores as a percentage of the total questions answered, provides an alternative measure that can account for strategies which are slower but more accurate as a result (such as applying the multiplication-grid). Percentage scores also showed a significant overall improvement for all groups (around 15 percentage points), but this time there was a significant difference between groups as well (see figure 7.6). Again, interaction effects should be interpreted with some caution as no correction was made for multiple comparisons (see results section). Nonetheless, the intrinsic group's scores in the delayed-test were significantly higher than either of the other two groups. The breakdown of percentage scores by question-type confirmed this overall increase for all kinds of question (including conceptual) but showed that only the intrinsic group achieved significantly higher scores than the control group for dividend-based questions (the type included in the game – see figure 7.6b). Overall the intrinsic group clearly produced the best learning outcomes in this study, providing evidence supporting the effectiveness of the intrinsic game.

The breakdown of learning outcomes by divisor revealed no significant differences between the groups, but did show overall gains for most divisors. No significant gains were made for the divisor nine, but few children will have encountered it in the game, as it was only included in later levels. In fact it is more surprising that significant improvements *were* observed for the divisors six and eight, as only two children encountered the divisor 6 in-game and none encountered the divisor 8 (see table 7.7). This may suggest that some children were able to apply the multiplication grid as a general solution to division problems in the game. No significant gains were observed for maths tasks were the answer was 'none of these', but these involve several steps in order to eliminate the possibility of each suggested divisor, and so naturally provide more scope for making mistakes.

7.4.2 Challenge Levels

No significant difference was observed between percentage scores for comparable questions in the post-test challenge and the post-test. However concerns were raised during the study about the effect of children favouring speed over accuracy in the challenge levels in order to finish first. Therefore in the delayed-test challenge, the health-based scoring mechanism was replaced with a points-based scoring mechanism to encourage accuracy over speed. Percentage scores for the delayed-test challenge revealed a significant difference between mathematics tasks performed in the game and test situations for both intrinsic and extrinsic groups. Challenge scores in the delayed-test were significantly higher than the post-test, but there was no difference in the average time it took for children to complete the post-test and delayed-test challenges.

The challenge levels alone present a slightly confused picture of transfer between the game and test, particularly in light of the changes in scoring mechanism that were required to observe a significant effect. The points-based scoring mechanism was not a feature of the main game and so could theoretically be creating additional transfer issues that are specific to the challenge. Malone and Lepper's taxonomy (1987) identifies scoring mechanisms as providing both personal and inter-personal motivations for playing games. While health-based and points-based scoring mechanisms are programmatically similar, they are quite different motivational concepts. Health is largely irrelevant until it is depleted, whereas every point counts for something extra. Consequently, transfer was also examined directly between the game and tests to see if it provided a similar picture of transfer losses (see section 7.4.4). Nonetheless, however these results are interpreted, there is no evidence to suggest that learning content transfers less effectively from the intrinsic version of the game than the extrinsic.

7.4.3 Testing Process

The disparity between group differences in numerical and percentage learning gains may suggest that a speed / accuracy trade off is key to understanding the results of this study. Group differences were only observed in percentage scores, which can account for strategies that are slower, but more accurate. Interaction effects should be interpreted with some caution as no correction was made for multiple comparisons (see results section). Nonetheless, the test process data revealed that combined-group times per question in the delayed-test were significantly longer than in either the pre or post-tests. This suggests an overall change in strategy whereby children were taking longer to answer questions. The test process data also revealed a positive correlation between percentage scores and average question timings across all three tests. This shows that children who were spending longer on questions achieved higher accuracy scores. No significant differences were observed between groups in timing data, but these results do support the idea of an overall speed for accuracy trade off taking place as a result of the intervention.

7.4.4 Learning Process

<u>Game Task Outcomes</u>

The basic analysis of game task outcomes suggested that on average, children in the intrinsic group encountered a significantly larger total number of maths tasks, but were less successful at solving their first encounters with tasks. However, about 20% of all first encounters in the intrinsic group were classified as 'targets left incorrectly' (TLIN) – significantly more than the 5% in the extrinsic group. The heuristic used in the intrinsic version classifies all the skeletons in the player's current room as being 'left' when the player leaves that room. This is a 'best guess' heuristic, but is likely to overestimate the number of conscious rejections of skeletons: when the player is exploring; skeletons are out of view; or the player is being chased by another skeleton. By comparison the extrinsic data is precise about the number of conscious rejections of dividends – the player has no choice but to answer a question in order to proceed.

Therefore the higher value for TLIN in the intrinsic group might partly account for the lower first encounter accuracy, and the higher total number of maths tasks in the intrinsic group (because TLIN skeletons can be re-encountered again and again). Nonetheless, the analysis of breakdowns of accuracy by divisor showed that the extrinsic group performed significantly better than the intrinsic for the divisors 2, 3 and 5. Divisor accuracy scores are not affected by TLIN – only divisors that were actually used by the player: 'targets attacked correctly' (TAC), 'targets attacked incorrectly' (TAIN) and 'distractors attacked incorrectly' (DAIN). Therefore these results do suggest that, on average, the extrinsic group were more successful than the intrinsic at in-game mathematical tasks. This result is further discussed alongside other measures of in-game accuracy and their observed changes over time in section 7.4.5.

Game Task Divisors

The breakdown of divisors used in first encounters with maths tasks showed that the divisors 2, 5 and 10 were used around 55% of the time. These are the multiplication tables included in the national curriculum objectives for this age group, and so could be considered familiar to the children. This means that around 45% of the content could be considered new to this audience, and more likely to challenge their existing strategies for division. This is a large improvement over the 28% of new content in study two, but fairly similar to the 42% in study one. An even larger quantity of new learning content may be desirable in order to create the conflict necessary to shift the children from using familiar strategies (such as rote memorisation of tables) to the one presented by the game (applying the multiplication grid).

Game Task vs. Test Comparison

The challenge levels were specifically designed to examine the transfer of learning between the game and test situations, but the results were not conclusive in their own right. Although it is not possible to compare identical questions between the main game and test, accuracy can be compared separately for each divisor. In this way, the learning content actually encountered in game can be compared with similar content in the tests. Accuracy scores for the divisors, 'none of these', two, ten, five and three were compared between in-game (figure 7.13) and delayed-test (figure 7.9) for the intrinsic and extrinsic groups. This revealed significantly higher in-game scores for both versions, using the divisors two and three, and significantly higher in-game scores for the extrinsic version using the divisors five and 'none of these'. This data backs up the significant differences between game and test accuracy observed in the delayed-test challenge levels. In many cases average accuracy scores in game are in excess of twenty percentage points higher than those in the test – potentially representing a significant transfer issue. However both tests of transfer have shown a difference in scores that is not limited to the intrinsic group, and this new data also shows some differences that are exclusive to the extrinsic group.

7.4.5 Behavioural Process

Accuracy Over Time

The discrepancy between groups in the way that 'targets left incorrectly' (TLIN) are measured led us to construct a new measure of overall accuracy for maths tasks encountered in the game. This records the number of attempts made on each maths task for all encounters. This measure reduces the impact of high TLIN scores, because these only add one additional attempt to a maths task rather than discounting that task as being performed incorrectly. Our previous measure of accuracy also only included first encounters in order to be comparable with the tests that only allow one attempt at each task. Including all encounters in this new measure incorporates maths tasks that children were required to repeat in the game. This not only includes second and third tries on an individual task, but the possibility of having to repeat an entire level of tasks once three incorrect answers have been made. This could potentially provide a different, and arguably more complete picture of the overall accuracy of maths tasks in the game.

Analysing changes for this new measure of accuracy over time, revealed significant and interesting differences between groups (see figure 7.14). The source of interaction effects should be interpreted with some caution as no correction was made for multiple comparisons (see results section). Nonetheless, at the beginning of the study the intrinsic group needed significantly fewer attempts at each maths task than the extrinsic group. However as the study progressed the number of attempts required by the intrinsic group grew, while the number of attempts required by the extrinsic group fell. By the end of the study there was no significant difference between the groups in the average number of attempts required for each maths encounter.

In contrast to the old measure, this new measure suggests that sometimes children in the intrinsic group were actually more accurate than the extrinsic group at in-game maths tasks. Higher TLIN scores in the intrinsic group may partly explain this discrepancy, but previous comparisons of accuracy by divisor (which do not include TLIN) were also significantly higher in the extrinsic group (see section 7.4.4). Therefore, it seems more likely that this difference is a result of including all encounters in the analysis – not just first encounters. This now tests the ability of children to correctly repeat the same learning content, rather than just their ability to get it right first time.

The group difference in accuracy observed at the start of the study could therefore be interpreted as a difference in children's willingness to reengage with repeated learning content. It could be argued that requiring a child to repeat a quiz may create a more negative affective state than requiring a child to repeat a level of skeletons. This may then have a direct effect on their subsequent performance at the repeated maths tasks. Nonetheless, over time, the negative association of repeating a quiz could actually be creating a strong motivation to answer questions correctly (and get back to the game as quickly as possible). This could then explain the increase in accuracy over time in the extrinsic group, despite the increased difficulty of learning content over time as well.

To explain the decrease in accuracy for the intrinsic group, it is important to remember that both the mathematical and gameplay content of the intrinsic game gets more difficult as the game progresses. While both groups face increasingly difficult division problems, the intrinsic group must also contend with increasingly mobile and more aggressive skeletons at the same time. The combined effect could easily explain the decrease in overall accuracy observed in the data. Nonetheless, according to this measure the intrinsic group were no less accurate than the extrinsic group even by the end of the study, providing an alternative and insightful perspective on this issue.

TLIN and Pausing the Game

The process data was also used to examine changes in TLIN, pauses made, and time spent paused over the course of the study. These measures were of interest because they could potentially provide evidence to support or reject subjective observations made about the children's behaviour during the study. The first observation was that children in the intrinsic group increasingly engaged in reconnaissance-like behaviours as the difficulty of the gameplay increased. This usually involved opening a door to view the skeletons inside a room and quickly running out again to consider their (division) strategy. The second observation (based on similar intentions) was that children began to pause the game in dangerous situations in order to consider their strategy and consult the multiplication grid.

The graph of TLIN over time (figure 7.15) appears to support the observation about reconnaissance-behaviour. This behaviour was unnecessary at the start of the study because all skeletons on early levels were immobile. Accordingly, there were no group differences in TLIN scores at the start of the study. However by the end of the study the intrinsic group have significantly higher TLIN scores than the extrinsic. TLIN scores for the intrinsic group actually peak in the third quarter of the study, but this may be explained by children switching to the alternative strategy of pausing the game.

The time spent paused at the start of the study is the same for both groups (see figure 7.17), but the intrinsic group spend significantly more time paused as the study progresses. This supports the observation about children pausing the game and also suggests that pausing the game may become the dominant strategy over time. This seems plausible, as pausing the game is a much quicker way of achieving the same aims as the reconnaissance behaviour.

7.4.6 Gender

Gender differences have not been a key area of focus for this thesis, but study two did reveal significant gender differences in the effectiveness of the intrinsic approach. However, this study showed no significant difference between genders for learning gains in either numerical or percentage scores. Nonetheless, like study two, subjective observations suggest that boys were more reluctant than girls to engage with the instructive content included within the game. Comparing this study with study two highlights the possibility that it was the boys that benefited most from the inclusion of the reflection session away from the computer. Naturally, this apparent gender difference could actually be a result of differences in gaming experience, whereby regular game players are more reluctant to engage with instructive content that does not conform to their standard model of a game.

7.5 CONCLUSIONS

This study produced the largest overall learning gains of all the studies. Furthermore, the intrinsic group's percentage scores were significantly higher than either of the other two groups at delayed-test, and only the intrinsic group performed significantly better than the control group in percentage scores for dividend-based questions. The significant difference between groups at delayed-test – two weeks after the main intervention – supports the theory that the intrinsic game creates a greater connection with the learning content. However, there was no significant difference between versions for scores in conceptual questions, so it would be unfair to describe this as deeper learning. Nonetheless, it is evidence that intrinsic motivation does create more effective learning, and supports the value of the intrinsic approach to educational game design.

CHAPTER EIGHT

Conclusions and Reflections

This thesis has examined a range of hypotheses relating to the educational effectiveness of intrinsic games. Our main hypothesis predicted greater learning gains from the intrinsic approach, as a result of motivational engagement and a deeper connection with the learning content. However, the alternative hypotheses predicted smaller learning gains: the first as a result of embedded intrinsic learning content transferring less effectively than the extrinsic; and the second because integrating learning content within a game's flow-experience may impede reflection-in-action. This chapter discusses the empirical results of this thesis with respect to these hypotheses, and offers interpretations and conclusions based on both the quantitative and qualitative findings. It goes on to discuss some other aspects of this research that were not included in our original hypotheses, but have proved to be significant to the outcomes of this work.

8.1 SUMMARY

This thesis set out to empirically evaluate the relative effectiveness of intrinsic and extrinsic approaches to developing an educational game. The four experimental studies have shown that intrinsic games have the potential to create a) a higher level of motivational appeal and b) improved learning outcomes, over extrinsic equivalents. Concerns about the difficulty of transferring embedded learning content from intrinsic games have also been shown to be unfounded. Some concerns remain over the role of flow in inhibiting reflection-in-action in the intrinsic game, but no evidence was found that extrinsic equivalents are any better at promoting reflection.

Children were found to be highly resistant to the instructional content of a mathematical framing system introduced into the game. A teacher-led reflection session was much more successful at inducting children into the underlying mathematical concepts of the game. Furthermore, this reflection session appeared to be the catalyst for significant learning gains that were shown to be best for children in the intrinsic group. This provides clear evidence in the support of the value of an intrinsic approach and justifies continued research in this area.

8.2 CONCLUSIONS

Study four compared learning outcomes between versions of Zombie Division for a fixed amount of time-on task, and found a clear learning advantage for the intrinsic version of the game. This is in direct opposition to the hypotheses relating to transfer and reflection, which predicted smaller learning gains in the intrinsic version, and suggests that we should now reject these hypotheses. This does not necessarily mean that the intrinsic version is free from any issues relating to transfer or reflection, but offers evidence that these hypotheses do not provide an accurate overall representation of the intrinsic game.

The results of studies one and two did not produce the same differences in learning gains between the intrinsic and extrinsic groups. Any difference in study one would have been concealed by the limited scope for improvement from high pre-test scores, but study two only showed a significant difference for girls in the intrinsic group. However, study four employed a different methodology that included a reflection session and delayed-tests two weeks after the post-tests. Therefore the results are not contradictory, but do highlight the importance of these changes in observing a statistically significant result.

These results support the hypothesis that motivational engagement and a deeper connection with the learning content will create greater learning gains for intrinsic games. Nonetheless, it does not necessarily prove that the intrinsic game was more motivating than the extrinsic, or that greater learning gains were a result of motivation or deep learning. Therefore this section will take each of these four elements in turn (motivation, deep learning, transfer and reflection) and discuss the additional evidence available to help support or reject their relevance to the effectiveness of intrinsic games.

8.2.1 Motivation

Study three compared the time children spent playing the intrinsic and extrinsic versions when they could freely switch between them, and found a strong preference for the intrinsic game. This result suggests that the level of motivational engagement produced by the intrinsic version of the game was greater than the extrinsic. It would be natural to assume that one of the benefits of this greater engagement would be to produce more time-on-task. However, study two found no significant difference for time-on-task between assigned versions when children could choose to play non-educational games instead. This shows that the motivational benefits of the intrinsic version over the extrinsic are not necessarily powerful enough to result in greater time-on-task. Clearly, children can also be strongly motivated by extrinsic approaches, but qualitative observations during study two suggested that the engagement produced by this motivation was focussed on the game rather than the (separate) learning content. Therefore the value of this motivation may lie within its ability to create a deeper connection with the learning content, rather than just the gameplay elements of the game (see section 8.1.2).

However, study two was unable to directly demonstrate a link between motivation and learning. Learning outcomes were not measured in study three, but study two found no correlation between time-on-task and learning gains. While differences between the intrinsic version and the control in study four clearly show learning gains as a result of playing the game, it did not include a measure of motivation that could be used to link the two measures. Therefore while the combined results provide evidence for both the increased motivational engagement of the intrinsic version and its improved learning outcomes, we were not able to collect any direct evidence of a link between the two.

One aspect of these motivational studies that is worthy of further consideration is the influence of school-contexts on studies of this kind. While both studies two and three provided a certain amount of free choice in playing Zombie Division, the competition for the children's attention was still carefully controlled. The relative ability of different versions of Zombie Division to compete with contemporary console games in a home environment could be very different. Children associate school environments with learning and may therefore be prepared to accept learning content and pedagogies that they may be more resistant to in their home leisure time. One teacher involved in these studies observed that children's engagement with very simple and dated educational software can be extremely high when the alternative is a maths lesson, but that this engagement rarely persists beyond the home-time bell. Encouragingly, many children involved in these four studies repeatedly pestered the researchers to allow them to take the game home (and even offered money!). Nonetheless the comparative success of intrinsic and extrinsic games in the home environment may be quite different and clearly this represents an interesting area for future research.

8.2.2 Deep Learning

The overall outcome of study four was predicted by the hypothesis that the motivational appeal of the intrinsic version creates a greater level of engagement, and a deeper connection with the learning content. This study provided two potential indications of the depth of learning: one through the delayed-test results (c.f. Biswas et al., 2004) and the other through scores for conceptual-based questions in the tests. The intrinsic group's percentage scores at delayed-test were significantly better than the control group, while the extrinsic group's scores were not. This could certainly be interpreted as an indication that a deeper learning had helped the intrinsic group retain their knowledge during the delay between playing sessions. However learning gains for conceptual questions alone were not significantly different between groups. Children demonstrated an overall improvement in percentage scores for conceptual questions, but the group difference appears to be a result of the intrinsic group's better performance at dividend-based questions, rather than the conceptual content of the tests. Therefore although deep learning is often linked to intrinsic motivation (e.g. Biggs, 1987; Chin & Brown, 2000) our results did not provide any evidence that the intrinsic version of the game created deeper learning. The delayed-test results may suggest a *greater* connection with the learning content, but this could not be justified as deep learning based on these results.

It is perhaps unsurprising that the intrinsic version failed to create any more improvement than the extrinsic in the children's conceptual understanding of the learning content[22]. Early versions of the game contained intrinsic features such as the giant skeletons, which were specifically designed to make children consider

[22] Study one appeared to demonstrate a specific improvement in conceptual learning, but this difference was attributed to high pre-test scores preventing an improvement in non-conceptual questions.

the concept of an inverse relationship between divisor and dividend. The intrinsic game also included skeletons that used weapons to defend against some divisors in order to make children consider multiple divisors. Unfortunately the further refinement of these, more conceptual, features was ultimately sidelined by the basic task of demonstrating learning gains for simple proceduralisation skills. Nonetheless, now that the intrinsic game has demonstrated the ability to improve procedural learning, future studies could concentrate on refining some of these conceptual features as well. Action-adventure games like Zombie Division rarely rely on a single game mechanic such as the one used in the prototype. However, this core mechanic could form the basic building block for a whole range of puzzles and challenges that pull together different mathematical concepts relating to division. Even the children in study three could see the potential for expanding upon the core game mechanics in order to create end-of-level bosses that could only be defeated by attacking them with every divisor that divides their dividend. Ideas like this would increase the scope of the game to improve conceptual learning as well as increasing the challenges and longevity of the gameplay.

8.2.3 Transfer

Studies one, two and four all provided some indication that accuracy scores in the game were significantly higher than those in the tests. However, contrary to our hypothesis, this does not appear to represent a specific problem transferring learning from the embedded context of the intrinsic game to the abstract context of the test. In most cases this difference was equally present in the accuracy scores of *both* groups, and in some cases it was actually significantly *larger* for the extrinsic group. This suggests that transferring learning between the game and test contexts may actually be a greater problem for the extrinsic version!

Nonetheless, both versions do appear to suffer from significant transfer problems, with average scores 27.5 percentage points higher in the game than in the delayed-test for study four. This difference is particularly puzzling for the extrinsic version, because the extrinsic quiz questions seem so similar to the tests. Nonetheless, there are a number of cognitive and affective differences between the quiz and test situations that could potentially be creating lower scores in the tests (see table 8.1). However, these same differences also exist between the intrinsic learning content and the test. If these reasons can create such a high difference in scores for the extrinsic quiz, then there is no reason why they shouldn't also create the same difference in the intrinsic version. Therefore there is no reason to expect that any additional transfer issues are present in the intrinsic version of the game.

Table 8.1. Differences between the extrinsic 'quiz' and computer-based tests.

Difference	Description	Type
1. Type of questions	The test includes divisor-based and conceptual questions in addition to the dividend-based questions used in the game.	Cognitive
2. Order of divisors	The test introduces divisors in a random order, whereas the game uses the same three divisors on each level, with harder divisors on later levels.	Cognitive
3. Task Feedback	The test provides no feedback on whether given answers are right or wrong. The game provides both audio and visual feedback.	Cognitive or Affective
4. Length	The test contains around three times more questions than a single level of the game.	Cognitive or Affective
5. Consequences	There are no obvious consequences for getting an answer wrong in the test. An incorrect answer in the game results in the loss of a 'life' and the threat of having to repeat maths tasks.	Affective
6. Challenge and Competition	The test does not provide clear, achievable goals or the ability to compare progress at those goals.	Cognitive or Affective
7. Stigma	The words 'test' and 'game' embody strong preconceptions about the relative pleasures of the two activities.	Affective

We will discuss each of these differences in turn and consider its potential for creating the large contrast in scores between the game and test contexts:

1+2. The additional types of questions found in the test (divisor-based and conceptual) provide one plausible reason why scores might be significantly higher in the game than in the tests. The children are only performing dividend-based maths tasks during the intervention and so naturally are not as well practiced at these other types. The random order of divisors could also be a potential factor contributing to lower scores in the test. If children are employing different strategies for solving problems with different divisors, then the extra cognitive demands of switching between strategies could be slower and more prone to mistakes. However, the challenge levels compared dividend and divisor-based questions in a random order between game and test contexts and still found a significant difference in scores in the delayed post-test challenge of study four. This finding suggests that neither of these factors can be the main cause of the observed difference between contexts.

3. The positive role that feedback can play in learning has deep founded roots (e.g. Thorndike, 1913). However, although there is a consensus on the instructional benefits of providing feedback to support self-moderated learning, there is some doubt as to whether feedback provides motivational benefits at all (Delgado & Prieto, 2003; Vollmeyer & Rheinberg, 2005). Feedback may offer longer-term cognitive benefits to learning by helping learners to moderate their own learning processes over time (J. Anderson & Schunn, 2000), but research suggests that it does not produce the short-term motivational benefits in a test situation like ours. Therefore it seems less likely that feedback on answers is the main cause of the difference between contexts.

4. The large number of consecutive maths tasks in the tests represents another potential difference between the game and test contexts. There were three times as many questions in the test than the average level. However, children were not told how many questions there were in the test or given feedback on their progress. So although they may have experienced a higher level of fatigue in the tests, they weren't able to make a conscious comparison. Furthermore any motivational difference must surely have been greatest for the intrinsic group, who were not used to answering questions in this way *at all*. As the intrinsic group did not

experience greater differences in scores between the two contexts it seems unlikely that this was a significant cause.

5+6. The three different versions of the challenge levels provide some interesting insights into the potential role of consequences, challenge and competition on the difference between scores in game and test contexts. The original challenge levels used in study two did not include lives, so that they would be comparable with the tests. However, making the player invulnerable by removing consequences from the game seemed to change the way that children approached maths tasks. Furthermore, no significant difference was observed between these challenge levels and the post-test of study two.

Consequently, the concept of lives was re-introduced into the post-test challenge levels of study four. However, although these looked like lives, it was impossible for the player to run out of lives and restart the challenge as this would prevent a fair comparison with the tests. Unfortunately, the children's behaviour still appeared to be different from the 'normal' game and there was no significant difference between scores in these challenges levels and the post-test. This was put down to their focus on the competitive goal of finishing the challenge levels first, rather than their performance at maths tasks (which could not be compared). However, the timing data could not support this theory, and it is possible that the children simply saw through the concept of the fake lives.

Finally, a points-based scoring mechanism was introduced into the delayed-test challenge of study four. This provided a new way for children to compare their success that focussed on performance. While there was no points mechanism in the 'normal' game, there were consequences to children's performance at maths tasks which directly affected their progress in the game (by sending them back to the start). Both the challenge with a score and the 'normal' game thereby include the means for creating a competitive challenge based on children's performance at maths tasks. Furthermore, results showed a significant difference between scores in these challenge levels and the delayed-test of study four.

Therefore the findings would seem to support the idea that it was the role of consequences, challenge and competition that were responsible for the significant difference between scores in the game and test contexts. The extrinsic quiz questions may appear more similar to the tests than to the intrinsic game, but both the quiz and the intrinsic game share these common motivational elements that are missing from the test. The extrinsic quiz distinguishes itself from the test by using these basic elements of games (Malone & Lepper, 1987) to create a more motivating context for eliciting answers from the children.

7. The role of children's preconceptions about 'tests' and 'games' is another factor that could be responsible for differences between scores in the two contexts. Nonetheless, the tests were also computer-based and performed in identical conditions to the game, so although it cannot be ruled out as a contributing factor, the methodology should have minimised its effect.

In conclusion, the results of this study lead us to reject the hypothesis that transfer issues will lead to smaller learning gains in intrinsic games. We have shown that in the case of Zombie Division, the intrinsic version produced larger learning gains and demonstrated no additional transfer issues from the extrinsic version. Furthermore, our discussion suggests that the main cause of the differences between game and test scores is the result of a motivational difference arising from gameplay elements that are not included in the test. We found no evidence to suggest that there are any additional issues relating to the transfer of embedded learning content from the intrinsic game.

8.2.4 Reflection

Our final hypothesis predicted that integrating learning content within the flow of the intrinsic game would prevent reflection-in-action and lead to smaller learning gains. None of our results supported the outcome of this hypothesis: studies one and two showed no significant difference in learning gains between versions, and study four showed significantly higher learning gains for the intrinsic version. Despite this, the inclusion of a 'reflection session' in study four was a tacit acknowledgement that children were not reflecting effectively on the game's mathematical 'framing system'. This highlights an important theoretical distinction between the reflection-in-action provided by the game and reflection-on-action provided by the reflection session[23] (Schön, 1983). Reflection-in-action may still have been a significant problem for the intrinsic game, but merely offset in study four by the opportunity provided for reflection-on-action.

Although a direct comparison with previous studies is not possible, it is clear that the reflection session was greatly beneficial to the learning gains observed in study four. This was primarily supported by the significant gains produced by all three groups between pre and post-test. Learning gains observed in the control

[23] Schön actually suggests that all reflection is reflection-in-action while the reflector can still influence the outcome of the event they are reflecting on. Therefore this distinction depends on viewing each skeletal encounter as a separate event rather than the intervention as a whole.

group were not significantly different from the other groups, despite the absence of any learning content in the control version of the game. However, all three groups were exposed to learning content in the reflection sessions, so the gains between pre and post-test must be largely attributed to the these sessions. The notable attentiveness of all groups in the reflection sessions raised concerns about potential Hawthorne effects, and the unrepresentative transfer of motivation from the control version of the game. Nonetheless, while this does question external validity, the reflection session did seem to be more successful at delivering the framing system's learning content than the framing system itself.

This appears to support the premise that the flow (Csikszentmihalyi, 1988) produced by games like Zombie Division is in conflict with reflective learning processes. However, if the relationship was this straightforward then the extrinsic version should also have benefited from having its learning content placed in the more reflective context of a quiz. Yet unlike the intrinsic group, the extrinsic group was indistinguishable from the control in terms of their learning gains between the post and delayed-test.

In fact the analysis of changes in behaviour over time suggested that players in the intrinsic version were creating their own reflective opportunities in the game. This included both reconnaissance and pausing behaviours that provided more time for children to consider their mathematical strategies. If children were indeed reflecting during the game then perhaps it was not the reflective opportunities provided by the reflective session that were critical to the observed learning gains. Many children actively ignored or avoided the framing system within the game, but they enthusiastically engaged with the same content when it was delivered by a teacher away from the game. For these children the reflection session was less about reflecting upon the framing system, than being effectively exposed to it for the very first time.

One final observation was made about the children's natural reflection-on-action as part of the study. Conversations about the game in all groups appeared to focus on level progress and fighting skeletons. So perhaps the integrated nature of the learning content in the intrinsic game meant that children were discussing division strategies between interventions as a form of reflection-on-action. Conversely the same conversations for the extrinsic group revolved around an arbitrary combat system that does not benefit their learning.

Therefore, although it is impossible for these studies to rule out the possibility that the intrinsic game inhibited reflection-in-action, they do allow us to reject the hypothesis that this would lead to smaller learning gains. Furthermore, there is reasonable evidence to suggest that reflective behaviours are naturally taking place in and around the intrinsic game. The concept of when and how reflection takes place for all digital games (not just educational ones) is a fascinating one and is certainly worthy of further study. Nonetheless, this study indicates that players may even be prepared to bend the rules of the game to create their own opportunities for reflection.

8.3 REFLECTIONS

This section considers concepts that were not part of our original hypotheses about the effectiveness of intrinsic games, but nonetheless turned out to have a significant influence on the outcome of these studies. The relative motivational power of competition and production values was drawn into focus in studies two and three. The learning gains from study four also highlighted the role of conflict and new learning content in the educational effectiveness of Zombie Division. These areas are examined and their implications on this and future research are discussed.

8.3.1 Production Values and Competition

Studies two and three clearly highlighted the motivational power of Zombie Division's production values. These studies suggested that the motivation produced by the game's production values could actually be greater than any potential motivational distinction between intrinsic and extrinsic approaches. Although it may not create ideal conditions for learning, the extrinsic approach can clearly be powerful enough to make children participate in a learning activity. This challenged our implicit assumption that the central benefit of the motivation produced by intrinsic games would be to produce greater time-on-task. Group competition also had a huge impact on all of the studies in this thesis. The results of the challenge levels in study four even suggested that it was the competitive challenge presented by the games (both intrinsic and extrinsic) that were responsible for consistently higher scores in the game than in the tests.

However, production values and competition appeared to have such a significant impact on these studies that it is difficult to know whether our results would have been the same without them. For example, it is interesting to consider whether the production values could be made so low that their use as an extrinsic reward is not powerful enough to motivate children to perform the quiz. Many children did seem to view the extrinsic game as a balance of 'pain' and 'reward', protesting about the necessity of the mathematical content. In contrast the intrinsic game seemed to have value in its own right, and while children objected to the instructional content of the framing system, they never objected to the mathematical content of the gameplay. This line of argument suggests that lowering the production values may actually have a greater impact on the appeal of the extrinsic game than the intrinsic.

Indeed, if you turn the 'pain' vs. 'reward' perspective around then you could also contemplate how much more (mathematical) pain the children in the extrinsic group would tolerate before the reward of the game was not enough. Consider what might have happened if we had included 100 or 500 maths tasks on each level rather than around 20. It would seem reasonable to expect that far fewer children in the extrinsic group would have got very far with the game, whereas the intrinsic group would be largely unaffected. In other words the nature of the extrinsic approach may be sensitive to schedules of reinforcement (Ferster & Skinner, 1957) in a way that the intrinsic version is not.

It is likely that removing the competitive group environment would also have a significant impact on the outcome of these studies. However, children playing games at home can (and do) still compete, by comparing their progress in the school playground or other social contexts. In fact, the growing prevalence of network gaming means that it is becoming increasingly easy for children to compete against their friends even when they're not in the same place. Competition clearly represented a significant motivation for children in this study and it is one aspect of game design that educational game designers should not ignore. Even if the game does not encourage competition, children in a classroom context will find a way to compete. It is better to design for this eventuality and include competition on your own terms than have children subvert your educational goals in order to create their own form of competition.

8.3.2 Encouraging New Strategies

Study four produced both the most convincing learning gains, and the most convincing group difference for the intrinsic version of the game. However this difference emerged between the post and delayed-tests rather than the pre-and post-tests. In fact, the intrinsic group's percentage score gain of 7.4 points between post and delayed-test is far greater in proportion to the amount of extra playing time (35 minutes), than their gain of 15.9 points between pre and post-test for the main intervention (100 minutes + 30 minute reflection session). Significant overall learning gains were also observed for divisors that were not even encountered in the game, suggesting that children were able to apply the multiplication grid as a general solution to division problems.

This is an interesting finding and an analysis of the learning content encountered during this final playing period may suggest a mechanism behind its success. Over the course of the study most children encountered maths tasks using the divisor three, but not the divisor four (see figure 7.9), which means they finished somewhere between level four (where the divisor 3 is introduced) and level six (where the divisor four is introduced). The average time taken to complete a level was around twenty minutes so it is not unreasonable to suggest that the majority of children were using the divisor three, for most (if not all) of their last 35 minutes playing-time. All the divisors encountered in the game before the divisor three (two, five and ten) are included in the school curriculum for children of this age. Therefore these divisors were potentially already familiar to the children, while the divisor three is not.

This creates ideal conditions for a Piagetian (1950) style of cognitive conflict in the final period of playing time. Children who may have successfully applied memorisation and heuristics to solve division problems with the divisors two, five and ten, can no longer apply these methods for the divisor three. This could make them more receptive to the multiplication-grid method that works across all divisors, and transfers most effectively to the test. Indeed, Siegler and Svetina (2006) suggest that newly discovered strategies are most readily adopted when they provide a large improvement in accuracy over previous strategies. If children are familiar with their two, five and ten times tables then the multiplication grid is unlikely to be any more accurate until they encounter the divisor three.

Unfortunately there is no hard evidence to support this theory, but it is a compelling explanation which points towards the potential importance of progressing to the unfamiliar learning content within the game. If correct, reducing familiar learning content to an absolute minimum could have significantly improved learning gains observed in these interventions.

8.4 FUTURE RESEARCH

8.4.1 Collaborative Reflection

Children in these studies (particularly the boys) were reluctant to pay attention to the instructional content of the framing system, yet were eager to engage with the same learning content explained by a human being. Competition was clearly a large motivation for children to play the game, but competition can create both positive and negative motivations (Malone & Lepper, 1987). Adapting the game to support collaborative play between players could potentially provide solutions to both of these problems: maintaining the positive motivations of social play, and using paired or group work to promote reflection between players. Indeed modern console systems are ideally suited to this style of game (see figure 8.1), allowing players to effortlessly link up with other nearby players via local wireless or anywhere on the Internet via a network router.

Figure 8.1: Zombie Division as it might look on the Nintendo DS

8.4.2 Age Group and Conceptual Learning

The experience of these studies was that it was extremely difficult to encourage seven and eight year olds to adopt new mathematical strategies. Consequently the majority of effort was spent refining the procedural learning content of the game, rather than the potential for conceptual content. Unfortunately the one study involving ten year olds did not measure learning gains, but they appeared to be far more open to adopting the new mathematical strategies presented by the game. This was almost certainly because less instruction was necessary, and the children required relatively little prompting from the framing system in order to appreciate the mathematical basis of the game. Nonetheless, their appreciation was such that they were even able to suggest intrinsic game mechanics, which incorporated aspects of their own conceptual understanding of mathematical division. Children with this level of mathematical experience may therefore represent a more appropriate audience for the game than their younger peers. Such a game could then concentrate on providing a mathematical playground of interrelated game mechanics that made conceptual links between the children's procedural understanding of mathematics.

8.5 THE WIDER CONTEXT

8.5.1 Methodological Approaches for Evaluating Games

This thesis has successfully employed a range of methodological approaches relevant to the evaluation of educational digital games. Quantitative measures of pre to post-test learning gains were augmented with detailed analyses of process data and qualitative interview transcripts. Informal qualitative observations have also proved essential to guiding the iterative development of the project from the outset. Some of the most formative of these qualitative observations (e.g. those in chapter 5) have resembled informal microgenetic studies (Siegler & Svetina, 2006) where the process of individual children's learning was observed over time.

This has demonstrated that a multi-method approach can be appropriate for classroom evaluations of games like Zombie Division. Nonetheless, the process data in particular provided some of the most valuable insights for interpreting the results of these studies. Without the process data we would have had no knowledge of the very high in-game accuracy scores, the proportion of familiar learning content encountered during play (and at different stages of play), or the development of reflective behaviours by the children. The main findings of this thesis may have come from pre and post test learning data, but these findings would have been impoverished without the picture of the process taking place in-between. Furthermore, many games used in learning contexts (e.g. K. D. Squire, 2004) focus on learning content based around systems and processes which are more difficult to measure using pre and post test data. Indeed, for some within the field of games and learning, it is specifically the potential of games to provide an understanding of real-time systems and processes (that are difficult to teach and test for using traditional methods) which provides their real value to education. Therefore process data could be even more important for evaluating these kinds of game. Player's understanding of systems and processes could conceivably be evaluated through the collection and analysis of data on the users' interactions and progress within the game. In theory, if this process data was detailed enough, it could be used both to provide quantitative group data as well as individual data that could be fed into a more formal microgenetic approach.

Unfortunately in order for researchers to collect detailed process data in this way it is necessary to have access to the source code of the game under evaluation. Obviously this is rarely possible with commercial gaming titles, and creating games specifically for the purposes of research takes a great deal of time and effort. The average commercial gaming title may have taken anything between 6 and 40 man-years of development time. Nonetheless, it is not impossible to achieve within a research context and hobbyist game development systems such as Blitz 3D (for 3D games such as Zombie Division) or Game Maker (for simpler, 2D games) can offer simple and cheap ways of creating bespoke games. There are also a host of hobbyist tools for creating all the 3D assets, animation, sound effects and music required for your games. The Zombie Division prototype represents something in the region of 6 months of full-time programming work by the author, and 3 man-months of asset generation by games industry professionals. This certainly adds a significant extra burden to any research project, but this project has certainly demonstrated that it is possible to achieve.

8.5.3 Implications for Using Games in the Classroom

One conclusion which could be implied from the outcomes of this research is that face-to-face teaching is more successful than game-based instruction. After all, the teacher-led sessions away from computers had the greatest impact on learning gains in these studies. However, it should be remembered that while the control group were not exposed to the learning content of the game, they were certainly exposed to its motivational effect. Even the teachers involved in this teaching attributed the control group's tangible enthusiasm for the face-to-face session to their involvement with the game and study. It therefore fair to say that the game was acting as a motivational anchor for the children's learning and face-to-face teaching alone would not have had the same effect. Nonetheless it is also apparent that the face-to-face teaching was a critical part of obtaining real value from using Zombie Division in a classroom context.

Unfortunately, educational computer games have been traditionally used by classroom teachers as a 'hands-free' mode of teaching: an individual reward for completing work, or simply just a way of keeping a class occupied while attending to other priorities. This research would seem to imply that such approaches may be far less effective than employing structured teaching around the learning content of the game itself. Yet this makes the use of educational games in the classroom far more time consuming for teachers. A game such as Zombie Division teaches only a tiny section of the curriculum, and the technical and organisational overheads for both children and teachers would almost certainly be prohibitive to its use. Moreover, Zombie Division teaches learning content which teachers can already teach effectively, so there would be little motivation for teachers to make the additional investment required to use it. Any future for games in the classroom almost certainly lies in teaching learning content which is difficult to learn in any other way: such as an appreciation of systems and processes, or in working with learning content which would otherwise be too dangerous or expensive to experience. Only in situations which provide something beyond traditional teaching can the investment of classroom time really be justified.

This research does point towards another possible use for games in the classroom as a *testing*, rather than a teaching tool. One of the most striking findings of these studies has been that children consistently perform better at the learning content presented within a gaming context than within a traditional testing context. So perhaps the real value of games in the classroom is for the assessment of learning. Even if games turned out to be less effective at teaching in a direct comparison with face-to-face work (the results of this study cannot say), then they could still have value in the classroom as a means of augmenting formative and summative assessment of children. Games could offer a significant boost to the fatigue and apathy created by the frequency of testing in the education system, as well as providing more detailed feedback on children's consistency, speed and accuracy.

However, the real educational potential of games is almost certainly outside of the classroom, where the choice is not between face-to-face or game-based teaching, but where game-based teaching can augment the teaching experienced during the school day. The motivational appeal of games means that many children will willingly choose to engage with them in their own free time, and these studies have demonstrated the superior appeal of intrinsic games. Games like Zombie Division could therefore represent motivating homework assignments, completed outside of school, but still providing formative assessment data to their teachers.

8.5.2 Implications for Game Designers

This research has significant theoretical implications to game designers for creating more effective digital learning games. However, there are practical and economic factors that present commercial barriers to the application of this research. Intrinsic games may be both more motivating and more educational than their extrinsic equivalents, but they are also more difficult and more expensive to develop. 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. This makes it difficult to justify a business case for developing intrinsic games, and so it is likely that most educational game developers will continue to adopt largely extrinsic approaches in the short term.

Nonetheless, the mainstream entertainment software industry is looking to reach new audiences and is already starting to explore popular learning, with titles such as Brain Training and Big Brain Academy (Nintendo). Intrinsic games offer the potential for authentic gaming experiences that provide players with the benefits of learning, without ruining the motivational experience of a game. Players of the future could find themselves learning a foreign language whilst playing an absorbing adventure game, physics while conquering the solar system, or even mathematics while fighting skeletons. The intrinsic approach could help to show audiences that learning is a natural part of what makes games fun, and that it is only uninspiring pedagogies – not learning content – that stand in the way of learning games having mass-market appeal.

REFERENCES

Ainsworth, S., & Loizou, A. T. (2003). The effects of self-explaining when learning with text or diagrams. *Cognitive Science,* 27, 669-681.

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

Anderson, C. (2002). Violent video games and aggressive thoughts, feelings and behaviours. In S. L. Calvert, A. B. Jordan & R. R. Cocking (Eds.), *Children in the Digital Age* (pp. 101-119). Westprot, CT: Praeger Publishers.

Anderson, C., & Ford, C. M. (1986). Affect of the game player: Short-term effects of highly and mildly agressive video games. *Personality and Social Psychology Bulletin,* 12(4), 390-402.

Anderson, J. (2005). *Cognitive psychology and its implications* (6th ed.). New York: Worth Publishers.

Anderson, J., & Schunn, C. D. (2000). Implications of the ACT-R learning theory: No magic bullets. In R. Glaser (Ed.), *Advances in Instructional Psychology Volume 5* (pp. 1-33). Mahwah, NJ: Erlbaum.

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

Baker, R. S. J. D., Corbett, A. T., Koedinger, K. R., Evenson, S., Roll, I., Wagner, A. Z., et al. (2006). *Adapting to when students game an intelligent tutoring system.* Paper presented at the 8th International Conference on Intelligent Tutoring Systems.

Berylne, D. E., Craw, M. A., Salapatek, P. H., & Lewis, J. L. (1963). Novelty, complexity, incongruity, extrinsic motivation, and the GSR. *Journal of Experimental Psychology,* 66(6), 560-567.

Biggs, J. B. (1987). *Student approaches to learning and studying*. Melbourne: Australian Council for Educational Research

Biswas, G., Leelawong, K., Belynne, K., Viswanath, K., Schwartz, D., & Davis, J. (2004). *Developing learning by teaching environments that support self-regulated learning.* Paper presented at the 2004 conference on Intelligent Tutoring Systems

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

Bransford, J. D., Brown, A. L., & Cocking, R. R. (2000). *How people learn: Brain, mind, experience and school*. Washington, DC: National Academy Press.

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). New Jersey: Lawrence Erlbaum Associates.

Braun, C. M. J., & Giroux, J. (1989). Arcade video games: Proxemic, cognitive and content analyses. *Journal of Leisure Research,* 21(2), 92-105.

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

Bryce, J., & Higgins, D. (2000). Optimal experience: A framework for understanding the phenomenology of computer use. *International Simulation and Gaming Research Yearbook,* 8, 197-205.

Burn, A. (2006). Reworking the text: Online fandom. In D. Carr, D. Buckingham, A. Burn & G. Schott (Eds.), *Computer Games: Text, Narrative and Play* (pp. 88-102). Cambridge, UK: Polity Press.

Caillois, R. (1961). *Man, Play and Games* (B. Mayer, Trans.). New York: Free Press.

Callahan, J. S., Brownlee, A. L., Brtek, M. D., & Tosi, H. L. (2003). Examining the unique effects of multiple motivational sources on task performance. *Journal of Applied Social Psychology,* 33(12), 2515-2535.

Carraher, D., & Schliemann, A. D. (2002). The transfer dilemma. *The Journal of the Learning Sciences,* 11(1), 1-24.

Chin, C., & Brown, D. E. (2000). Learning in science: a comparison of deep and surface approaches. *Journal of Research in Science Teaching,* 37(2), 109-138.

Chou, T.-J., & Ting, C.-C. (2003). The role of flow experience in cyber-game addiction. *CyberPsychology and Behavior,* 6(6), 663-675.

Crawford, C. (1982). *The art of computer game design*. Retrieved 23rd May, 2007, from http://www.erasmatazz.com/free/AoCGD.pdf

Csikszentmihalyi, M. (1975a). *Beyond bordem and anxiety*. San Francisco: Jossey-Bass.

Csikszentmihalyi, M. (1975b). 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, UK: 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.

Dadds, M. R., Perrin, S., & Yule, W. (1998). Social desirability and self-reported anxiety in children: An analysis of the RCMAS lie scale. *Journal of Abnormal Child Psychology,* 26(4), 311-317.

Daniel, T. L., & Esser, J. K. (1980). Intrinsic motivation as influenced by rewards, task interest, and task structure. *Journal of Applied Psychology,* 65(5), 566-573.

Dawes, L., & Dumbleton, T. (2001). *Computer games in education project*. Retrieved 29th May, 2007, from http://partners.becta.org.uk/index.php?section=rh&rid=11207

de Freitas, S. (2006). *Learning in immersive worlds*. London: Joint Information Systems Committee

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.

deCharms, R. (1968). *Personal causation: The internal affective determinants of behavior*. New York: Academic Press.

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

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

Delgado, A. R., & Prieto, G. (2003). The effect of item feedback on multiple-choice test responses. *British Journal of Psychology,* 94, 73-85.

Dempsey, J., Lucassen, B., Gilley, W., & Rasmissen, K. (1993). Since Malone's theory of intrinsically motivating instruction: What's the score in the gaming literature? *Journal of Educational Technology Systems,* 22(2), 173-183.

Dewey, J. (1910). *How we think*. Boston: D. C. Heath.

Egenfeldt-Nielsen, S. (2005). *Beyond edutainment: Exploring the educational potential of computer games.* IT University of Copenhagen, Copenhagen.

Ellis, D. (1984). Video arcades, youth, and trouble. *Youth and Society,* 16(1), 47-65.

Ellis, H., Heppell, S., Kirriemuir, J., Krotoski, A., & McFarlane, A. (2006). *Unlimited learning: Computer and video games in the learning landscape*. London: Entertainment and Leisure Software Publishers Association.

Fabricatore, C. (2000). Learning and videogames: An unexploited synergy. *The International Conference of the Association for Educational Communications and Technology*.

FAS. (2006). *Harnessing the power of video games for learning*. Washington, DC: Federation of American Scientists.

Ferster, C. B., & Skinner, B. F. (1957). Schedules of Reinforcement. *Psychological Reports,* 3(4), 695-695.

Frasca, G. (2003). Simulation versus narrative: Introduction to ludology. In M. Wolf & B. Perron (Eds.), *The video game theory reader* (pp. 221-236). New York: Routledge.

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

Gee, J. P. (2005). Learning by design: Good video games as learning machines. *E-Learning,* 2(1), 5-16.

Ghani, J. A., & Deshpande, S. P. (1994). Task characteristics and the experience of optimal flow in human-computer interaction. *The Journal of Psychology,* 128(4), 381-391.

Green, C. S., & Bavelier, D. (2006a). The cognitive neuroscience of video games. In L. Humphreys & P. Messaris (Eds.), *Digital media: Transformations in human communication* (pp. 211-224). New York: Peter Lang Publishing.

Green, C. S., & Bavelier, D. (2006b). Effect of action video games on the spacial distribution of visuospatial attention. *Journal of Experimental Psychology: Human Perception and Performance,* 32(6), 1465-1478.

Greenburg, B. (2004). *Children spend more time playing video games than watching TV*. Retrieved 6th August, 2004, from http://newsroom.msu.edu/site/indexer/1943/content.htm

Greer, B. (1992). Multiplication and division as models of situations. In D. A. Grouws (Ed.), *Handbook of Research on Mathematics Teaching and Learning* (pp. 276-295). New York: MacMillan.

Griffith, J. L., Voloschin, P., Gibb, G. D., & Bailey, J. R. (1983). Differences in eye-hand motor co-ordination of video-games users and non-users. *Peception and Motor Skills,* 57, 155-158.

Griffiths, M. D., & Dancaster, I. (1995). The effect of Type A personality on physiological arousal while playing computer games. *Addictive Behaviors,* 20(4), 543-548.

Habgood, M. P. J., Ainsworth, S., & Benford, S. (2005a). *The educational content of digital games made by children.* Paper presented at the 2005 conference on Computer Aided Learning, Bristol, UK.

Habgood, M. P. J., Ainsworth, S., & Benford, S. (2005b). Endogenous fantasy and learning in digital games. *Simulation and Gaming,* 36(4).

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

Hays, R. T. (2005). *The effectiveness of instructional games: A literature review and discussion*. Orlando, FL: Naval Air Warfare Center Training Systems Division.

Hiebert, J., & Lefevre, P. (1986). Conceptual and procedural knowledge in mathematics: An introductory analysis. In J. Hiebert (Ed.), *Conceptual and Procedural Knowledge: The Case of Mathematics* (pp. 1-27). Hillsdale, NJ: Lawrence Erlbaum Associates.

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

Inkpen, K. M., Booth, K. S., Klawe, M., & Upitis, R. (1995). *Playing together beats playing apart, especially for girls.* Paper presented at the 1995 conference on Computer Support for Collaborative Learning, Bloomington, IN.

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: Lawrence Erlbaum Associates.

Kafai, Y. B. (2001). *The educational potential of electronic games: From games-to-teach to games-to-learn*. Retrieved 1st January, 2004, from http://culturalpolicy.uchicago.edu/conf2001/papers/kafai.html

Kerawalla, L., & Crook, C. (2005). From promices to practices: The fate of educational software in the home. *Technology, Pedagogy and Education,* 14(1), 107125.

Kirriemuir, J., & McFarlane, A. (2004). *Literature review in games and learning*. Bristol: NESTA Futurelab.

Klinger, E. (1971). *Structure and functions of fantasy*. New York: Wiley.

Kolko, D. J., & Rickard-Figueroa, J. L. (1985). Effects of video games on the adverse corollaries of chemotherapy in pediatric oncology patients: A single-case analysis. *Journal of Consulting and Clinical Psychology,* 53(2), 223-228.

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

Laurillard, D., Stratfold, M., Luckin, R., Plowman, L., & Taylor, J. (2000). Affordences for learning in a non-linear narrative medium. *Journal of Interactive Media in Education,* 2.

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge: Cambridge University Press.

Lepper, M. R. (1985). Microcomputers in Education - Motivational and Social-Issues. *American Psychologist,* 40(1), 1-18.

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(3), 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). Hilsdale, NJ: Erlbaum.

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

Malone, T. W. (1980). *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, Pal Alto, California.

Malone, T. W. (1981). Toward a theory of intrinsically motivating instruction. *Cognitive Science,* 5(4), 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). Hilsdale, NJ: Erlbaum.

Mansvelder-Longayroux, D. D., Beijaard, D., & Verloop, N. (2007). The portfolio as a tool for stimulating reflection by student teachers. *Teaching and Teacher Education,* 23(1), 47-62.

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., & Schwartz, D. L. (2005). Physically distributed learning: Adapting and reinterpreting physical environments in the development of fraction concepts. *Cognitive Science,* 29(4), 587-625.

Matchett, M. J. (2005). *The reflective teacher*. Bangor, UK: Education and Training Inspectorate.

McAlpine, L., Weston, C., Beauchamp, J., Wiseman, C., & Beauchamp, C. (1999). Building a metacognitive model of reflection. *Higher Education,* 37, 105-131.

McFarlane, A., Sparrowhawk, A., & Heald, Y. (2002). *Report on the educational use of games*: TEEM.

Michael, A., Klee, T., Bransford, J. D., & Warren, S. F. (1993). The transition from theory to therapy: Test of two instructional methods. *Applied Cognitive Psychology,* 7, 139-153.

Michael, D., & Chen, S. (2005). *Serious games: games that educate, train, and inform* Boston, MA: Thompson Course Technology.

Miller, C. S., Lehman, J. F., & Koedinger, K. R. (1999). Goals and learning in microworlds. *Cognitive Science,* 23(3), 305-336.

Mitchell, A., & Savill-Smith, C. (2004). *The use of computer and video games for learning*: Learning Skills and Development Agency.

Moreno, R., & Mayer, R. E. (2005). Role of guidance, reflection, and interactivity in an agent-based multimedia game. *Journal of Educational Psychology,* 97(1), 117-128.

Nunes, T., & Bryant, P. (1996). *Children doing mathematics*. Oxford, UK: Blackwell Publishing.

Nunes, T., Schliemann, A. D., & Carraher, D. (1993). *Street Mathematics and School Mathematics*. New York: Cambridge University Press.

Orosy-Fildes, C., & Allan, R. W. (1989). Psychology of computer use: xii. Videogame play: Human reaction time to visual stimuli. *Perceptual and motor skills,* 69, 243-247.

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

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

Papert, S., & Talcott, J. (1997). The children's machine. *Technology Review,* 96(5), 29-36.

Parker, L. E., & Lepper, M. R. (1992). Effects of fantasy contexts on childrens learning and motivation: Making learning more fun. *Journal of Personality and Social Psychology,* 62(4), 625-633.

Pearce, C. (2004). Story as play space: narrative in games. In L. King (Ed.), *Game on: The history and culture of video games* (pp. 112-116). London: Lawrence King.

Perkins, D. N., & Salomon, G. (1988). Teaching for transfer. *Educational Leadership,* 46(1), 22-32.

Piaget, J. (1950). *The psychology of intelligence* (M. Piercy & D. E. Berylne, Trans. 2001 ed.). New York: Routeledge.

Prensky, M. (2001). *Digital game-based learning*. New York: McGraw-Hill.

Reiber, 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.

Rittle-Johnson, B., Siegler, R. S., & Alibali, M. W. (2001). Developing conceptual understanding and procedural skill in mathematics: An iterative process. *Journal of Educational Psychology,* 93(2), 346-362.

Rollings, A., & Morris, D. (1999). *Game architecture and design*. Scottsdale, AZ: Coriolis

Rouse, R. (2001). *Game design: Theory and practice*. Plano, TX: Wordware.

Ryan, R. M., & Deci, E. L. (2002). Overview of self-determination theory: An organismic dialectical perspective. In E. L. Deci & R. M. Ryan (Eds.), *Handbook of self-determination research* (pp. 3-33). Rochester, NY: University of Rochester Press.

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

Sandberg, J., & Barnard, Y. (1997). Deep learning is difficult. *Instructional Science,* 25, 15-36.

Sandford, R., Ulicsak, M., Facer, K., & Rudd, T. (2006). *Teaching with games: Using commercial off-the-shelf computer games in formal education*. Bristol, UK: Futurelab.

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

Schnotz, W., & Bannert, M. (2003). Construction and interference in learning from multiple representations. *Learning and Instruction,* 13, 141-156.

Schön, D. A. (1983). *The reflective practitioner: How professionals think in action*. New York: Basic Books.

Schott, G., & Kambouri, M. (2006). Social play and learning. In D. Carr, D. Buckingham, A. Burn & G. Schott (Eds.), *Computer Games: Text, Narrative and Play* (pp. 119-132). Cambridge, UK: Polity Press.

Sedighian, K. (2007). Toward operationalization of 'flow' in mathematics learnware. *Computers in Human Behavior,* 23, 2064-2092.

Sedighian, K., & Westrom, M. M. (1997). *Direct object manipulation vs direct concept manipulation: Effect of interface style on reflection and domain learning.* Paper presented at the HCI Conference on People and Computers XII, San Francisco, CA.

Selnow, G. W. (1984). Playing videogames: The electronic friend. *Journal of Communication,* 34(2), 148-156.

Shaffer, D. W. (2004). Pedagogical praxis: The professions as models for postindustrial education. *Teachers College Record,* 106(7), 1401-1421.

Shaffer, D. W. (2006). *How computer games help children learn*. New York: Palgrave Macmillan.

Siegler, R. S., & Jenkins, E. A. (1989). *How Children Discover New Strategies*. Mahwah, NJ: Lawrence Erlbaum Associates.

Siegler, R. S., & Svetina, M. (2006). What leads children to adopt new strategies? A microgenetic/cross-sectional study of class inclusion. *Child Development,* 77(4), 997-1015.

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

Squire, S., & Bryant, P. (2003a). Children's models of division. *Cognitive Development,* 18, 355-376.

Squire, S., & Bryant, P. (2003b). Children's understanding and misunderstanding of the inverse relation in division. *British Journal of Developmental Psychology,* 21, 507-526.

Steinkuehler, C. A. (2004). *Learning in massively multiplayer online games* Paper presented at the 2004 conference of the Learning Sciences Mahwah, NJ.

Thorndike, E. L. (1913). *Educational psychology volume 1: The original nature of man*. New York: Columbia University Teachers College.

Thorndike, E. L., & Woodworth, R. S. (1901). The influence of improvement in one mental function upon the efficiency of other functions: i. *Psychological Review,* 8, 247-261.

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(4), 389-401.

Tubbs, M. E., & Trusty, M. L. (2001). Direct reports of motivation for task performance levels: Some construct-related evidence. *Journal of Psychology,* 135(2), 185-204.

Twidale, M. B., Wang, X. C., & Hinn, M. D. (2005). *CSC: Computer supported collaborative work, learning and play.* Paper presented at the 2005 conference on Computer Supported Collaborative Learning.

Vallerand, R. J., & Ratelle, C. F. (2002). Intrinsic and extrinsic motivation: A heirachical model. In E. L. Deci & R. M. Ryan (Eds.), *Handbook of self-determination research* (pp. 3-33). Rochester, NY: University of Rochester Press.

Van Eck, R., & Dempsey, J. (2002). The effect of competition and contextualized advisement on the transfer of mathematics skills in a computer-based instructional simulation game *Educational Technology Research and Development,* 50(3), 23-41.

Vollmeyer, R., & Rheinberg, F. (2005). A surprising effect of feedback on learning. *Learning and Instruction,* 15(6), 589-602.

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

Wood, D. (1998). *How children think and learn* (Second ed.). Oxford, UK: Blackwell Publishing.

Wood, H., & Wood, D. (1999). Help seeking, learning and contingent tutoring. *Computers & Education,* 33(2-3), 153-169.