UNIVERSITY OF SOUTHAMPTON

Designing Immersive Serious Games

by

James Baker

A thesis submitted in partial fulfilment for the

degree of Doctor of Philosophy

in the

Faculty of Physical Sciences and Engineering

Department of Electronics and Computer Science

June 2017

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF PHYSICAL SCIENCES AND ENGINEERING
DEPARTMENT OF ELECTRONICS AND COMPUTER SCIENCE

Doctor of Philosophy

DESIGNING IMMERSIVE SERIOUS GAMES

James Baker

Serious games, designed for more than entertainment, can be used for educational purposes in order to more effectively teach certain ideas. One particular element which has not been investigated fully is the idea of immersion in educational games, in terms of how to engage learners with such games and to keep them engaged while playing. The subject has been approached from various perspectives, including education and gameplay, but the theories presented miss important aspects shown in others (such as designing gameplay towards learning outcomes), which makes further advancement in the field more difficult. To solve this problem, the Immersive Educational Games Model was proposed in this thesis, which will help the design of immersive serious games, in terms of providing engaging content and educational value. The model integrates key characteristics of instructional design, gameplay, immersion, and serious game theories to outline the key considerations in creating compelling educational gameplay.

To validate whether the model can be used to measure immersive qualities, a questionnaire-based instrument was created and tested through three experiments, utilising an established immersive serious game. While the results show a need for further investigation into measurements for immersive states, the instrument demonstrated that the model may be used to reliably identify aspects that make a serious game immersive.

Contents

List of Figures

List of Tables

Acronyms

ILO	Intended Learning Outcome
MMORPG	Massively Multiplayer Online Role-Playing Game
IEGM	Immersive Educational Games Model

Acknowledgements

I would like to thank my supervisors, Dr. Gary Wills and Dr. Ashok Ranchhod, for their constant support throughout the years. I would also like to thank my mum Liz, my step-dad Andrew, my brother Dan and my best friend Jo for bearing the brunt of my increasing stress and resultant madness, and yet never wavered in their love, support, and occasional kicks up the proverbial as required. I'm also immeasurably grateful for my friends and colleagues who have helped me with my research and sanity, including Vanissa Wannick, Nawfal Firas, Eamonn Walls, Noreena Lui, James Stallwood, and many others.

Lastly, for everyone who helped and participated in my experiments and research in any capacity – I could not have finished this without you, and you are all unequivocally awesome in my book. Literally, in fact, as this thesis will be a book of some description.

Chapter 1. Introduction

Over the years, computer games have developed from small hobbyist's projects into a global industry worth billions[1], with thousands of developers and millions of players from across many different cultures. As the popularity of computer games has grown, many companies and researchers are interested in the potential of games in non-entertainment fields. This has led to the creation of the genre of "serious games", which differ from purely entertainment-focused games. (Aldrich, 2009) (Freitas & Neumann, 2009) (Harteveld, et al., 2007) (Squire, 2006).

In particular, educators and researchers alike have shown great interest in serious games for education. When constructed well, these games can provide more engaging learning experiences and provide a greater understanding of the subject area than learning by more conventional means (e.g. textbooks) can. (Squire, 2006)

1.1 What are Educational Serious Games?

The definition of serious games varies between researchers. Most agree however that serious games are essentially computer games that are intended for more than just entertainment. These applications range from 'advergames', which act as promotional tools for particular products, to 'art games', which aim to convey particular artistic messages.

The field of 'educational serious games' arose to try bringing the popularity of computer games into classroom settings, or other learning situations. The idea is to provide games that teach certain subject matters, while also engaging learners and captivating their attention in ways that traditional teaching methods (e.g. lectures) cannot. There are commercial computer games that include educational content, to varying degrees (e.g. the *Civilization* games can subtly educate on politics and world history). However, as the main purpose of such games is to entertain, the educational value is incidental, and not likely to be learned or even identified by all players (Aldrich, 2009). With serious games, the primary aim is to inform and/or educate (Connolly, et al., 2012).

[1] http://www.mcvuk.com/news/read/global-games-market-worth-over-100bn/07021, 2012. *Global games market worth over $100bn.*, MCV

There is one important distinction to be made: serious games are not simulations. While simulations do educate, they do not focus on entertaining their users; as such, their virtual environments must reflect the real world, and the objectives for the user are far more precise. By contrast, serious games prioritise ‘fun’ as well as learning, and are not so constrained in their environments or rules. Indeed, the challenge of making a serious game is to integrate the educational and entertainment sides. If the serious game is fun but unrelated to the learning material it is designed to teach, it serves no purpose as educational material; if the game has strong links to the learning material but is not fun, it serves no purpose as a game (Hamalainen, et al., 2006), and therefore loses the engagement benefits that games provide.

1.2 What are the difficulties in developing educational games?

Unfortunately, creating games that offer these benefits is not a trivial task. Simply inserting educational content into any arbitrary game is not in itself a guarantee that it will be a valuable teaching tool or fun (Gunter, et al., 2006). Likewise, serious games are not equivalent to simulations (Aldrich, 2009), and require more subtle considerations for how to make its gameplay both entertaining and relevant. There have been several theories that attempted to clarify how engaging learning with serious games occurs. For instance, Malone (1982) proposed a set of design heuristics for enjoyable educational user interfaces, broadly including providing clear and satisfying challenges, familiar and appealing fantasies, and stimulating the user’s curiosity. Later examples include Freitas and Neumann (2009) and Kiili (2005), who each proposed cyclical models of in-game experimentation, where players are presented with particular problems in the game, and must draw upon and test their knowledge to solve them.

However, these theories are relatively diffuse, covering different areas of what makes educational games engaging, without a clear unifying theory that encompasses their core ideas. These approaches tend to neglect important aspects of general theories of games and engagement, such as the role and structure of narrative (Jenkins, 2004), or the preconditions of creating an absorbingly enjoyable experience (Norman, 1993). In addition, certain key aspects of educational theories are similarly neglected, such as how to structure educational content in ways that facilitate understanding (Gagne, 1970). As a result of these factors, it is so far unclear what the most vital aspects are to creating engaging and educationally valuable games, which encompass ideas from the fields of serious games, entertainment games, education, narratives and stimulating engaging experiences.

However, the idea of engagement is one potential means of helping to focus these aspects, in such a way as to facilitate the development of higher quality serious games. One of the most fundamental concepts of creating engaging experiences is '**flow**', coined by the psychologist Csikzentmihalyi (1990). 'Flow' describes the state of an 'optimal' immersive experience; active, exclusive concentration on a particular enjoyable activity, which is meaningful to the person undertaking it. When the person willingly ends this flow state, they reflect on their experiences, and how they have been affected by them.

Leading from this state is the idea of '**immersion**'. Immersion in this context is defined as the maximal state of engagement with a particular enjoyable activity. With respect to games, Brown and Cairns (2004) proposed three levels of immersion, each leading to the next: **engagement**, **engrossment** and **total immersion**. Engagement involves first getting involved in the game, via accessible controls and relevant gameplay. Engrossment is the point where the player's emotions become directly affected by the game. Finally, total immersion is the feeling of the player being fully engaged in the game, to the exclusion of all else (a phenomenon they call 'presence') – a state very similar to flow. However, they claim total immersion is a brief experience, in contrast to flow, which can last up to the entire length of the activity the person is engaged in.

From this perspective of varying levels of engagement, one can think of a player's engagement as being on a spectrum, with no engagement being the minimal point, and total immersion being the maximal point. The thesis uses this idea of immersion as a means of characterising engagement in a potentially measurable way. The objective is to help the creation of engaging serious games; whether the players reach the state of total immersion is less important than how consistently they engage with it.

It is important to note that, in this context, immersion does not strictly refer to the sensation of 'being' in a virtual world (such as virtual reality headsets). This research regards immersion as similar to the sensation of being absorbed in an enjoyable book or film; how engaged the person's mind is to a game, its world and its message. While virtual reality can represent one means of facilitating immersion, other types of activities can facilitate immersion simply by being in some way important to the person.

1.3 The Research Focus

This thesis explores how to make educational serious games engaging: this includes how people can be informed and instructed about a subject through educational serious games, and what aspects (both educational and gameplay) make them appealing for potential learners. Through this exploration, this thesis attempts to clarify the core aspects of creating serious games that are as immersive as possible.

While it is hoped that these core aspects will be applicable for any potential player, the target audience they are based around includes students between secondary and higher-level education students. This is because the more practical examples that form the basis of the research outcomes (from the literature review) focus on this target audience (Kelly, et al., 2007) (Freitas & Neumann, 2009), and it was deemed inappropriate to generalise without further testing among learners across other education levels.

In the context of the thesis research, the assumed purpose for educational games is for voluntary, independent learning, or as a supplemental classroom tool, as a way to help elaborate on particular topics. Due to the latter consideration, it is not assumed that the students are invested in the subject from the outset. The aim is not to replace classroom teaching, but to add an alternative option to help learners understand a subject.

1.4 Structure

The structure of how the thesis will unfold is outlined below, chapter by chapter.

Chapter 2 comprises of a literature review of serious games research. The review explores and evaluates various learning theories, what makes games immersive, and the role of narrative in games.

Chapter 3 introduces the research questions that arose from the literature review, and an overview of the methods used to investigate them.

Chapter 4 explores the initial step towards answering the research questions; refining and condensing the findings in the literature review into a model, elaborating the key aspects that impact immersion in educational serious games.

Chapter 5 details the first verification stage of the model, involving interviews with game experts on how well the initial model represents the issues of creating an immersive game.

Chapter 6 details the second verification stage, involving a questionnaire survey given to computer game players, in order to confirm the refinements of the model given by the experts.

Chapter 7 explores the third verification stage, an experimental study investigating the model's usefulness in assessing the immersive qualities of serious games, and the metrics derived from the model to measure them.

Chapter 8 details the final verification stage, an expansion of one of the experiments in Chapter 7.

Chapter 9 discusses what the findings from the verification stages indicate for researchers and developers of serious games.

Chapter 10 summarises the overall progression and outcomes of the research, and describes the future work to carry the research forward.

Chapter 2. Literature Review

In this chapter, the literature surrounding educational theories, serious games theories and applications, theories of immersion, and theories of narrative are explored. This process was undertaken to discover how games and teaching function both individually and together, and by doing so identify what areas need to be addressed or improved in teaching with games.

2.1 Learning Theories

Because educational serious games are first and foremost learning tools, understanding how people learn is a vital first step. This section explores a range of the theories about learning and instructional design, including theories centred on games and collaborative learning processes.

2.1.1 General Learning Theories

There are several different psychological approaches to how the mind processes and reacts to the world. Three of the more prominent approaches in instructional design are behaviourism, cognitivism and constructivism.

Behaviourism, as psychologist John Watson argued, is the perspective that looks at a person by their behaviour, and how it may be predicted, rather than a person's mental states, which are intangible and not scientifically measurable (Watson, 1913). This theory suggests that a person's actions are determined by their environment, with a person's behaviour changing and developing, adapting to environmental changes and developments, but not necessarily as a result of more developed internal reasoning. (Skinner, 1977) From this perspective, a person's learning can be judged by changes in the likelihood of exhibiting particular behaviours. The conditions affecting what behaviours are exhibited are influenced by two forms of conditioning: **operant conditioning** and **classical conditioning**. With operant conditioning, if a particular behaviour under particular conditions (e.g. drinking water when thirsty) has previously produced a 'reinforcing' result (e.g. no longer feeling thirsty), it is more likely to be exhibited again. B. F. Skinner argues that these 'reinforcers' are necessary for one's survival, in a natural selection sense, and thus why behaviour naturally gravitates towards producing reinforcing results (Skinner, 1974, p. 51). Classical conditioning, based upon the observations by Ivan Pavlov (Pavlov, c1927), involves the association of a stimulus with a particular response; after repeated associations, the person is conditioned to exhibit the response when presented with the stimulus.

In contrast to the behaviourist approach, Cognitivism focuses on a person's cognitive processes, such as how they think, reason and process information. This approach likens human thought and learning to a computer system; learning being measured by changes in the 'state' of a person's knowledge, and the mental connections they make between items of knowledge (Ertmer & Newby, 1993). While this approach can potentially accommodate for the complexities of human learning that behaviourism cannot (acknowledging a person's reasoning behind their behaviour, rather than treating it as simply a response to the environment), a person's knowledge-states are not tangible in the way observable behaviours are, which makes this approach more difficult to use when measuring a person's learning.

Both behaviourism and cognitivism seem to follow an objectivistic viewpoint, meaning the real world is separate from the learner. However, a more recent approach, Constructivism, suggests that a person's knowledge is derived from their subjective experiences of the world, and that learning occurs by imbibing new knowledge that the person finds meaningful (Vygotsky, 1978) (Jonassen, et al., 1995). In other words, constructivists believe that the understanding and worldview of each person is unique, based on how each person reasons about his/her own experiences. While this approach arguably introduces more complexity about how to measure learning than cognitivism, it does emphasise the importance of how dependent learning is on individual experience.

Bloom's taxonomy of educational objectives (Bloom, et al., 1956) has been used prominently in educational research to the present day. Of particular applicability to classroom and university teaching are his observations about the Cognitive Domain (describing intellectual skills). It was stipulated that all intended learning outcomes could be categorised into the six classes of the cognitive domain, with each class building upon the ones preceding it. These classes are:

- **Knowledge –** Behaviours that emphasise *remembering* of content. This includes knowledge of 'specifics' (i.e. terminology, facts), knowledge of organising and studying specifics (e.g. already-existing categories, criteria, methodologies), and knowledge of abstractions (theories, structures, principles)

- **Comprehension –** Behaviours emphasising the *understanding* of content 'Understanding' includes translation into different terms (e.g. write something in your own words), interpretation and extrapolation
- **Application –** Behaviour involving being able to use (*apply*) content in different contexts, having had knowledge and understanding of it
- **Analysis –** Involves breaking content down into constituent parts ('elements'), and analysing the relationships between them. Bloom in turn subdivides this issue into three levels: identifying/classifying elements, determining the relationships between the elements, and recognising the organisational principles that comprise the whole
- **Synthesis** – Involves combining elements in new ways, in order to form a new 'whole'. The potential products of this synthesis include unique communications of ideas, plans or proposed sets of operations to be carried out, and sets of abstract relations between elements
- **Evaluation –** Involves assessing the value of content, in terms of their accuracy, their effectiveness, etc. These evaluations differ from opinions in that they must adhere to clear criteria

In later years, Anderson (Anderson, et al., 2001) made certain expansions to Bloom's cognitive domain, creating a new taxonomy to help educators categorise learning objectives. This modified taxonomy adds an extra dimension, the **knowledge domain**, in addition to the cognitive domain. The cognitive domain includes the behaviours the student is expected to perform for particular learning objectives (e.g. 'remember', 'evaluate'). The knowledge domain on the other hand categorises the type of knowledge the student is expected to work with; the knowledge types include **factual**, **conceptual**, **procedural** and **metacognitive**. The idea is that learning objectives can be categorised according to the cognitive behaviour and knowledge type it expresses.

Interestingly, Merrill earlier proposed a similar theory to categorise learning objectives, known as the Component Display Theory (Merrill, 1994). He too proposed that objectives can be categorised into two domains;

- **Content –** The facts, concepts, procedures and principles that the teacher wishes to impart
    - **Fact –** A piece of information assigned to an arbitrary name

    - **Concept –** A group of facts, identified by common features which they all share
    - **Procedure –** A sequence of steps needed to achieve a particular objective
    - **Principle –** A cause-effect relationship to explain a particular event or process
- **Performance –** What the teacher wants the learner to do with the content. The four varieties proposed are:
    - **Remember (Generality) –** remember the general rule of particular content
    - **Remember (Instance) –** remember specific examples of the rule
    - **Using –** apply the content to a specific example
    - **Finding –** use what they know about the content to develop new content

Both Merrill's theory and Anderson's modified taxonomy suggest a means of structuring teaching through clear, precise definitions of knowledge types and instruction/assessment types; such definitions could be readily applied to games, wherein the very nature of their challenges requires demonstrating particular skills. One key difference between them is that the former assesses a learner's progress through their behaviour rather than their cognitive processes; for the purposes of testing, particularly through games (which are primarily interactive, requiring user interaction to progress), Merrill's CDT could offer a more practical approach to teaching through games, and gaining more assurance that the subject matter has been successfully imparted.

As a precedent for Merrill's CDT, Gagne proposed the Nine Events of Instruction (Gagne, 1970, pp. 303-319) to explain the conditions needed to ensure learning, retention and transferability for instruction on a particular subject matter. These include:

1. **Gaining/Controlling Attention –** Use novelty or surprise to get the learners' attention
2. **Inform learners of objectives –** Reduces anxiety in the learners
3. **Stimulate recall of relevant prerequisites –** Give the content context and relevance to the learners
4. **Present the stimulus situation –** Use teaching methods to present the content in small chunks (the methods depending on the 'type of learning' associated with the subject matter. These include learning the steps of a process, the differences between objects, the objects that make up a particular concept, and rules)
5. **Provide guidance –** Communicate with the learners, use visual aids, case studies, etc.

6. **Elicit performance from the learner –** Give learners opportunities to practice their skills, preferably through examples
7. **Provide feedback –** Comment on learner performance, but allow them to correct their own mistakes wherever possible
8. **Assess performance –** Assess whether the learners have satisfied the learning objectives
9. **Enhance retention and transfer –** Give the learners opportunities to apply their knowledge in a meaningful setting

He also theorised that there are eight types of learning, each of which builds upon the last. These types include:

1. **Signal Learning –** a conditioned response to a stimulus, the response being general, unfocused and emotional
2. **Stimulus-Response Learning -** a conditioned response which is precise, specific to the stimulus, which requires repetition and reward
3. **Chaining –** connecting two or more stimulus-responses in sequence, each being somehow related to the last, occurs suddenly once those conditions are fulfilled
4. **Verbal Association –** a sub-variety of chains, connecting different words together
5. **Discrimination Learning –** learning to discriminate stimuli, by identifying which aspects make a particular stimulus different from all others. Interference, or distractions, must be minimised for this to take place
6. **Concept Learning –** generalise an already-learned idea across multiple scenarios, using reinforcement. This learning process is gradual
7. **Rule Learning –** forming a chain of two or more concepts, which only need take place once. The teaching usually occurs by stating the rule verbally, provided the concepts have already been established in the learner
8. **Problem Solving –** solve problems by combining old rules to make new ones

The progression of learning suggested here strongly echoes the progression proposed in Bloom and Anderson’s work, as well as the implied progression of Merrill’s CDT; learners start by learning content at face value, then start to reflect on and understand the content, and finally experiment with the content they understand – making connections between different pieces, and being able to apply them to related, unfamiliar circumstances. In doing so, the learners gain mastery over the content they are to learn.

Laurillard's **conversational framework** shares certain aspects of this view (Laurillard, 2002). This is an iterative model, where a teacher conveys a concept, the learner expresses their understanding of it, and the teacher offers feedback correcting or supporting the learner's assertions. This process continues until the learner demonstrates that they understand the concept (e.g. via tests or coursework).

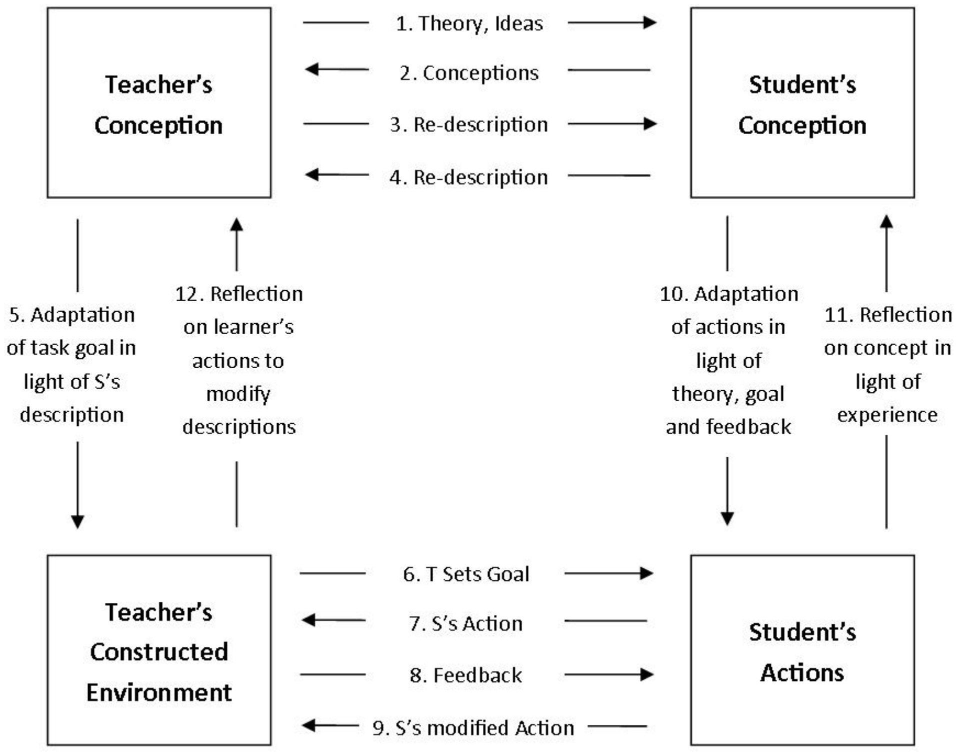


Figure 2-1: Laurillard's Conversational Framework. Sourced from (Laurillard, 2002)

While this is a model primarily targeted at helping students learn, in many ways it represents how learning occurs in games: the game system acts as a teacher, demonstrating what the player can and cannot do, and sets various challenges in the player's path to test their understanding of how the game world works. If the player fails to complete a particular objective, a system of in-game feedback and assistance helps the player to understand what they need to do, and suggest hints as to how to complete the objective. As such, one could think of games as an ideal medium to convey knowledge and encourage understanding, in a way that most standard tests cannot. A similar sentiment was shared by cognitive psychologist Donald Norman; he spoke of video games as an example of providing entertaining, skill-based experiences, while also encouraging reflection in order to find the best way to progress (Norman, 1993, p. 22). But while this iterative, reflective process shows how content can be reinforced, it does not account for how to promote the learner's interest in the subject matter, taking it as read that the learners will maintain a persistent interest before and throughout the process.

From his investigations into recurring themes in design theories, Merrill (2002) found that learning is promoted when:

- Learners are engaged in real-world problems
- Existing knowledge is employed as a basis for new knowledge
- The new knowledge is demonstrated to the learners
- The new knowledge is applied by the learners
- The new knowledge is integrated into the learners' life

Merrill asserts that the principles are true for all learning activities, and can be applied to any delivery system; this implies that they can also apply to the design of serious games. While he mentions himself his principles may not be appropriately substantiated, his findings are consistent with those of the other learning theorists mentioned in this report, that learners are more likely to learn when the content is made relevant, and there is the necessity of engaging with the material via experimentation and reflection.

However, it is considered very important for the learner to be motivated to learn, in order to engage with educational material at all. Gagne argues motivation as an important precondition to learning, which is largely affected by **social pressures** (the desire to please and avoid displeasure), and the desire for '**mastery**' over a particular topic, so they can use it independently in their lives. He also suggests that motivation comes intrinsically from the **satisfaction** of learning something else, provided it is taught effectively.

Keller argued further that motivation is "heart" of instructional design, as learners need to be personally motivated in order to keep engaged with the subject, and thus learn it well (Keller, 1983). He proposed the ARCS Model of Motivational Design to elaborate which factors affect a learner's motivation. These factors include:

1. **Attention –** Getting the attention of the learners from the start by using surprise, humour or getting learners to actively participate
2. **Relevance –** Ensuring the taught content is relevant to what the learners want to do, and what they already know
3. **Confidence –** Helping the learners to understand the extent of their abilities, encourage them with their learning, and offer feedback on their attempts. The learning objectives must be made clear for this to occur
4. **Satisfaction –** Ensuring the learners feel their newly learned skills are important and beneficial. The learners should also be kept challenged throughout the lesson

In examining the ways in which children learn, the psychologist Vygotsky theorised that children at first develop understanding socially, by asking questions of other people, and once this understanding is achieved, they are then able to reason about the subject internally (Vygotsky, 1962). Essentially, as in Laurillard's framework, development of ideas about the world first occurs through discussion, then reflecting and integrating the newly understood ideas into one's worldview. Vygotsky additionally proposed that people can solve any problem if they are assisted by those who already understand the subject, and that true understanding is the capacity to solve such problems unassisted (Vygotsky, 1978). The suggestion with both theories is that investigation and inquiry of more experienced people is a crucial element of learning.

The psychologist Skinner asserted that carefully-designed educational computer programs (or 'teaching machines') are an ideal method of educating learners. The primary reason is that they give the player focus and immediate feedback where a classroom scenario may not. In addition, he proposes that such machines allow the player to work through the taught material at their own pace, rather than one dictated by the teacher, which allows for more complete understanding of the material (Skinner, 1958). This self-paced approach is supported by the experiential nature of constructivism; indeed in the context of language learning, Flowerdew asserts that employing self-directed constructivist approaches to learning, where the students are expected to investigate and experiment on their own in practical exercises (with textual aids to consult as needed), can lead to deeper understanding of the presented material (Flowerdew, 2015). While Flowerdew points out the assertion is based on small-scale studies, and is more an indicator than conclusive proof of it, her findings seem concordant with the theories of Vygotsky, Laurillard and Skinner discussed earlier. This would indicate there is merit in the idea of aided-experimentation for deeper learning.

2.1.2 Learning Theories with Games

Gee proposed a set of principles for good learning games (Gee, 2005) (detailed in Appendix D). Firstly, these principles stressed that players should feel in control of their actions and learning in the game world, through **co-design** and **customising**. They further emphasise that the player should become invested in their in-game **identity**, and be able to easily experiment through **manipulation** of the game world. Gee argues the problems/challenges in the learning game must be **well-ordered**, to prepare the players for future challenges, the challenges themselves should be balanced, or '**pleasantly frustrating**' and should encourage the players to practice through the **cycle of experience**. He also feels that helpful information on how to progress should be provided '**on demand**' **and 'just in time'**. Furthermore, they should provide simpler **fish tank** levels, to help players understand the basic game rules, and broader **sandbox** levels to get the players safely acclimated to how the rest of the game works. Finally, Gee suggests that having the player use their acquired **skills as strategies** in proper context to accomplish a desired goal, encourage **system thinking** so they can understand how the game elements fit together, and using the player's experiences within the game to convey meanings and messages, or providing **meaning as action image**. In this way, the players can in turn become invested in, and develop their understanding of the material to be learned.

In addition, Sara De Freitas constructed a model to explain how learning occurs in serious games (Freitas & Neumann, 2009). The model is based on Kolb's Experiential Learning Cycle, which comprises of the following stages: get concrete experience, observe and reflect upon it, form abstract concepts, and test these concepts in new situations. De Freitas' model expands this idea into five cyclical stages:

- **Experience –** Gather experience with a particular concept, within a relatively safe virtual setting
- **Exploration –** Test boundaries and knowledge, form social connections
- **Reflection –** Understand their experiences, in order to facilitate the transfer of learning from virtual environments to abstract and real-world situations
- **Forming abstract concepts –** Use what they know to generate new ideas pertinent to the game world

- **Testing –** Practice the abstract concepts in virtual and real-world situations, including experimentation/reinforcement

This model is further influenced by conclusions drawn from Squire's work (Squire, 2006), which suggests that a compelling scenario should be established, the player roles and boundaries be made clear and useful feedback should be provided frequently. These traits are similarly described in the traditional learning theories explored, suggesting that learning theories can have considerable impact on the value of educational games.

Gunter et al. (2006) followed this viewpoint, claiming that sound educational practices needed to be integrated into the design of serious games from the start of development. As part of their argument, they created a serious game design paradigm RETAIN: the intention being to demonstrate that the level structure of an educational game can reflect the progression of motivation in Keller's ARCS model, the progression of learning in Gagne's nine events, and the order of knowledge presented in Bloom's taxonomy. The model provides an interesting outlook on how to incorporate educational elements into game design, demonstrating reasonable game analogues for Gagne's and Bloom's findings (e.g. 'Attention' to 'Scenario Exposition'). However, it carries several assumptions about the traditional structure of games and the nature of game challenges, which do not seem to be sufficiently explored.

To address the potential pitfalls in designing educational games for large groups of students, Villalta et al. created a set of design guidelines for making engaging classroom-based multiplayer games (Villalta, et al., 2011). These guidelines are divided into six main categories:

1. **Game Mechanics –** the placement of guidance and feedback, keeping interactions simple, and linking the mechanics to the learning objectives
2. **Game Progression –** making sure the narrative is clear, logical and immersive, and the game's difficulty gradually increases
3. **Methodology –** allow the teacher to act as a 'mediator'; encouraging reflection and discussion, and modifying the game to suit the player's interests and cognitive needs
4. **Collaboration –** ensuring the story and mechanics demand collaboration and discussion to proceed

5. **On-Screen Information –** keeping the language clear and concise, spacing activities and characters to make the most of the available virtual environment, allowing customisable characters to be emotionally invested in
6. **Holism –** making sure the design of the educational and gameplay ('ludic') aspects accommodates for including new sequences, which may take place in the virtual world and the real world

Dickey (2007) found knowledge in MMORPG settings could be divided into four categories in a similar manner to Anderson and Merril, with each area representing a different part of a person's understanding of the world:

- **Declarative knowledge** – Pieces of information that are taken to be true. These in turn are commonly split into four types of 'artifact'
  - Facts – Logically linked pieces of information
  - Concepts – Symbols, events and objects that share characteristics, and are identified by the same name
  - Procedures – Set of ordered steps needed to solve a problem, or achieve a goal
  - Principles – Rules and guidelines that explain cause-and-effect relationships
- **Procedural Knowledge** – The understanding of how to perform a particular task, action or process
- **Strategic Knowledge –** The ability to apply knowledge of a particular topic to different situations
- **Metacognitive Knowledge** – Reflection on one's thought processes during a particular activity

A further approach to serious games, coined by Sasha Barab as 'transformational play' (Barab, et al., 2010 (b)), focuses more on player interaction with the environment. His theory centred on three key ideas:

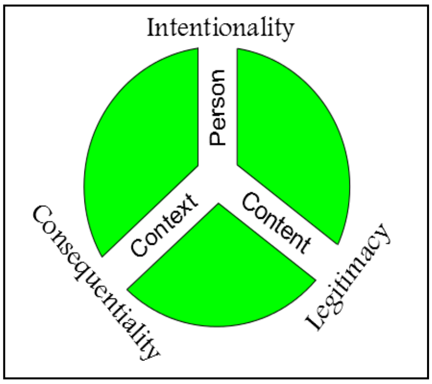


Figure 2-2: Transformational Play Model. Sourced from (Barab, et al., 2010 (b))

- **Player with Intentionality -** The player must know and apply the appropriate subject matter to proceed
- **Content with Legitimacy -** The player progresses through the story by making decisions
- **Context with Consequentiality -** The consequences of these decisions are tangible, impacting on the game world.

To validate this philosophy, he developed an educational MMORPG called *Quest Atlantis*, in which teachers could create their own small quests and game mechanics to explore certain subject areas. These quest areas in theory allow students to gain a greater understanding of the subject area by being able to directly influence the world (e.g. deciding the fate of Frankenstein's monster, in a quest area representing the village from the novel *Frankenstein*).

Indeed, there are several aspects of MMORPGs, which make them both appealing and valuable for educational purposes. One of the most appealing characteristics is the ability to customise one's in-game character, or 'avatar'. While it is not unique to MMORPGs, it has been described in literature as a key component of the genre. (Dickey, 2007) Typically, this includes various presentational aspects (e.g. clothes, facial structure or even species), but also includes gameplay attributes and skills. These aspects, in turn, introduce frequent decision-making on how one's avatar grows and changes.

Developing one's avatar into a unique and useful entity helps the player to network with other players, which in turn affects gameplay progression: for instance, a team may enlist a particular avatar if they have a useful skill the team does not possess. Because of these gameplay and social influences, the player has a greater emotional investment in how their character develops (Yee, 2006), with the player's avatar essentially acting as an extension of that player (a phenomenon Squire coins as a 'projective identity' (Squire, 2006)). At the same time, the virtual nature of the avatar provides a certain level of anonymity, which makes the player feel safer and less constrained by anxiety (Rankin, et al., 2008).

As well as the immediate benefit of expanded interactions and social-status motivation, it has been argued that MMORPGs can influence scientific thinking in players. In other words, players can analyse aspects of the world, create theories from this analysis, then test and evaluate these theories. Steinkuehler (Steinkuehler, 2008) found that forums for *World of Warcraft* not only contained such scientific discussions, but they made up the clear majority of discussions in the forum (about 86%). Examples range from discussions on the best way to complete certain quests to the levelling mechanics of different classes: indeed, some players even created informal scientific models and algorithms to validate their theories (see Fig. 2). Also, very importantly, most of these theories were left open to argument and evaluation. In short, many facets of scientific thinking are used and developed by *World of Warcraft* players. In addition, many gaming genres have similar forums, where similar strategic discussions occur (Squire, 2006): it could be inferred from this that if the scientific-thinking conditions of MMORPGs can be identified, they can be potentially applied across many genres, further expanding their educational possibilities.

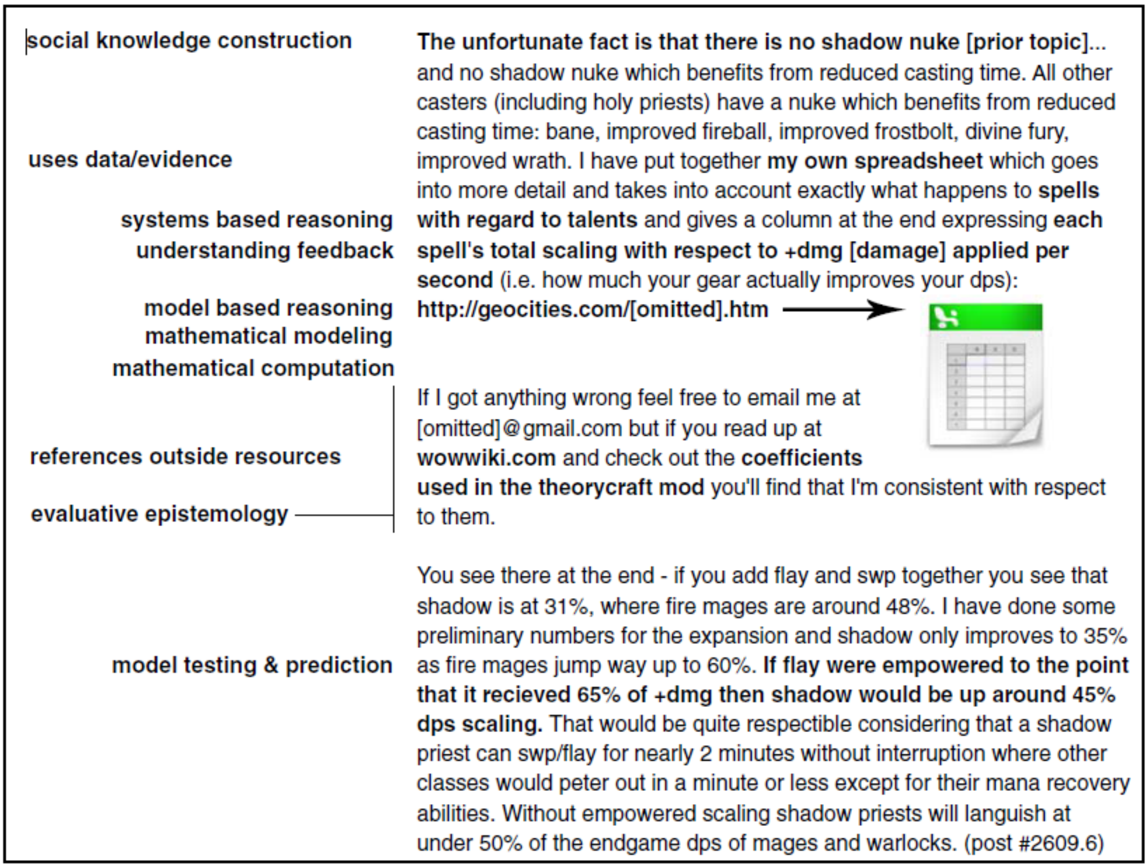


Figure 2-3: An example of model-based reasoning in World of Warcraft, analysed for its structure. Sourced from (Steinkuehler, 2008)

2.1.3 Collaborative Learning Theories

Since educational games are not all necessarily single-player, it is important to understand collaborative learning, and the additional considerations to educational games that it brings. At its simplest level, collaborative learning involves problem-solving as part of a group, increasing the understanding of each group member in a way they could not have done alone. As a result, it is regarded as a highly effective way of learning, and used in many educational contexts. This section explores the basic processes that allow collaborative learning to occur, and investigates guidelines for generating these situations in educational and serious games contexts.

To understand how collaborative teams work, it is necessary to understand the underlying factors in human behaviour that cause us to interact in the ways we do. The central point of defining human behaviour is somewhat obvious: everyone is different. Humans can vary in terms of personality, values, perspectives, motivations and experiences, and each of these factors in turn define how we react in different situations.

In this way, much the same as behaviourism, people communicate using a stimulus-response approach, implying communication is to some extent instinctive. People experiment with certain behaviours and communications and gauge the response - if the response is positive, they are more likely to perform such behaviour again. This idea has been elaborated in several theories of interaction, including **social exchange theory**, **equity theory** and **attribution theory** (Guirdham, 2002, pp. 78-79) (detailed in Appendix A).

One consistent idea that arises from these learning theories is that of ‘benefit’. An individual’s goals form the core of interactions, whether seeking to maximise them, equalise them with other people’s benefits, or simply to understand the ‘motives’ behind them. If the individual has unfulfilled goals, their tension is increased, which leads them to interact with others in order to fulfil them and reduce their discomfort. These goals range from the intrinsic (personal development, relationships) to the extrinsic (recognition, money). (Guirdham, 2002, pp. 83-87)

From these personal motivations for interaction comes the reason for groups and teams forming. Groups are formed of a number of people who interact, and are psychologically conscious of each other (Guirdham, 2002, p. 465). When such groups cooperate, trust, and structure themselves to achieve a common goal, the group can be thought of as a ‘team’. (Guirdham, 2002, p. 492)

Groups can be said to form over four stages (Tuckman & Jensen, 1977):

1. **Forming –** Getting to know one another
2. **Storming –** Expressing different opinions on approaching the task
3. **Norming –** Resolving these differences by finding common ground
4. **Performing –** Working together towards the goal more efficiently

During these stages, a structure arises in the group which reduces the risk of unpredictable behaviour. This structure is influenced by where the perceived ‘power’ lies, who exerts the strongest leadership skills, popularity and the role each person plays.

An important factor in one’s performance in a group is how much they value themselves, and how well they interact with others in the group. In turn, a considerable influence on an individual’s self-worth is their status within the group (i.e. how much power they hold in the group) (Guirdham, 2002, p. 64). Members can improve their status by moving to another group that places a greater value on their skills, or simply become part of a group that more closely matches their own preferences and ideology.

These group status mechanics are seen in many games, particularly in team-based shooting games (e.g. *Counter-Strike*) and in MMORPGs: in both, players join teams in order to increase their prestige, or because of shared ideologies. The idea of ‘social capital’ seen in MMORPGs can also be seen in group status mechanics, as a means of forming mutually-beneficial alliances. (Guirdham, 2002, p. 40)

The central motivation to collaboration appears to be the same as motivating individual learning: wanting to gain something you are personally invested in. The difference is that there must be a common objective between each collaborating person, the objective should be something each person cannot do on their own, and each person must be able to contribute in some way to achieving the objective.

However, the process of developing knowledge collaboratively is more difficult to uncover. The largest difficulty, pointed out by Stahl, is that new knowledge from collaboration is an emergent result from group interactions, meaning it is difficult to know where exactly from the process it comes (Stahl, 2004).

Stahl did attempt to formalise this process previously (Stahl, 2000) by dividing knowledge development into two domains: social discussion, and personal reflection. In this model, a person shares their understanding, develops their knowledge through discussion and compromise, and then imbibes their new understanding. In this way, social activity and personal cognition are interdependent.

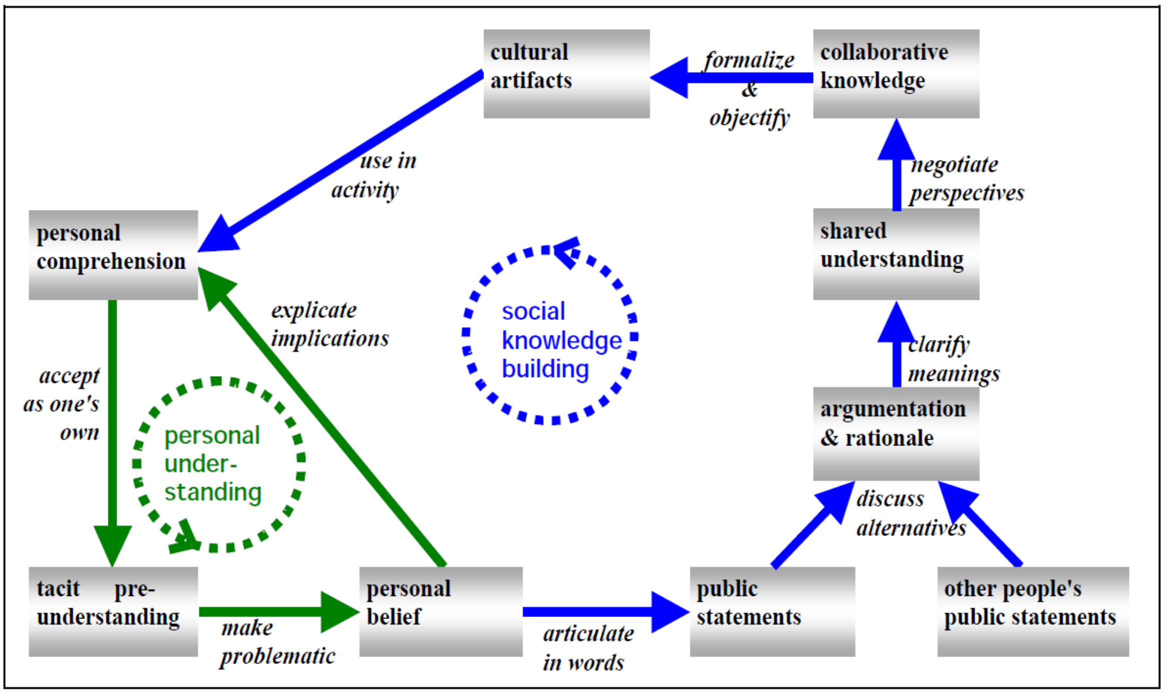


Figure 2-4: Stahl's Model of Collaborative Knowledge Building. Sourced from (Stahl, 2000)

However, as Stahl himself points out, this is a reductionist view of social knowledge construction, not coming close to covering all potential factors. Certainly its intangible nature makes it of limited help to construct learning activities around. But the ideas it presents are reflected in several learning theories, including Laurillard's conversational framework, and de Freitas' exploratory model, which in turn helps illuminate two important factors in developing a learner's understanding, collaboratively or individually; discussion and reflection.

Hamalainen (Hamalainen, et al., 2006) added that key considerations for a collaborative environment must also include curricular-specific tasks (requiring close teacher collaboration). He also noted the environment should convey work settings that would not be possible in traditional classroom teaching.

As part of his investigation into collaborative learning, (Dillenbourg, 1999) identified four key factors that affect how learning-effective a collaborative activity will be. These include the **setup of initial conditions**, **establishing roles**, the **rules of interaction** and **regulating interactions** (Appendix B).

An attempt to distil these ideas of collaborative learning into games was made by (Wendel, et al., 2012), who designed a serious game *Escape from Wilson Island* which exclusively involved collaborative tasks. The guidelines they operated under in its design included a **common goal**, **heterogeneous resources**, **refillable personal resources**, **tradable resources**, a **scoring system** and a **trading system** (Appendix C). These guidelines were intended to give purpose to the player's collaborations, and foster trust and responsibility amongst the collaborating group members.

2.2 Immersion, Flow and Challenge

One of the most appealing aspects of games is the immersion that they offer. Much like books, games can draw people into them for extensive periods, occupying the player's full attention in a positive experience. This is of particular interest from an educational standpoint, to ensure the learners pay close, voluntary, attention to the educational content being delivered.

The psychologist Csikzentmihalyi coined the term '**flow**' to describe the state of an 'optimal' immersive experience (Csikszentmihalyi, 1990). Flow is described as a state of active, exclusive concentration on a particular enjoyable activity, which is meaningful to the person undertaking it. When the person willingly ends this flow state, they reflect on their experiences, and how they have been affected by them.

Malone proposed a set of heuristics for designing 'enjoyable user interfaces', derived from his work investigating what makes computer games intrinsically motivating for education (Malone, 1982). He places these heuristics into three categories:

1. **Challenge**
    a. Clear goals for the activity, as well as feedback about how close the user is to achieving it
    b. Uncertainty about reaching the goal (i.e. multiple difficulty levels, progressively more difficult or increased number of level goals)
2. **Fantasy**
    a. Integrating fantasies (aspects that evoke mental images in the user of objects/situations that are not really there) into the system which emotionally appeal to the players
    b. Representing activity mechanics through 'metaphors' that the player is already familiar with

3. **Curiosity**
    a. Providing an ‘optimal level of information’
        i. The use of audio/visual presentation to decorate, enhance the fantasy of the game, and to represent systems
        ii. The use of randomness to enhance enjoyment, without making the activity’s tools unreliable
        iii. The use of humour to enhance enjoyment, without being inappropriate
    b. Drawing upon the user’s desire to have a complete, consistent and parsimonious knowledge base (e.g. when a user comes across a problem they cannot solve, the system introduces knowledge to solve it they were not aware of before)

Malone asserts that the ‘**Fantasy**’ and ‘**Curiosity**’ concerns are similar whether the system is designed as a ‘tool’ (to achieve some external outcome) or a ‘toy’ (with no external outcome). As for ‘**Challenge**’, while it is important for toys, he suggests while tools should be designed to be as easily usable and accessible as possible, having challenge and uncertainty can hypothetically increase the pleasure of performing mundane tasks. In terms of educational games (which can be thought of as a combination of tools and toys), one suggestion is that relevance towards the feelings and experience of the player is an important component.

When designing an activity to engender flow, it is important to balance the difficulty of the immersive activity with the skill level of the person; if the difficulty is too high, the person gets frustrated, and if the difficulty is too low, they lose interest. Game developers are similarly concerned with deciding how to balance the difficulty level according to the player’s demonstrated skills. Challenge balancing in games comes in two basic forms (Missura & Gartner, 2009):

1. **Static Balancing –** The level of challenge is pre-determined, based on repeated testing and altering of gameplay activities during development
2. **Dynamic Balancing –** The level of challenge alters during gameplay, based on the level of skill the player demonstrates during the game

With either approach, the principle of balancing is the same: keep the game from being too difficult, or too easy, for either case will result in players breaking from the state of flow, and thus stop engaging with the game. This view was certainly shared by Hamari et al (2016), whose study with game players demonstrated the positive effect challenge can have on immersion in game-based learning.

Brown and Cairns proposed three levels of immersion within games, each leading to the next: **engagement**, **engrossment** and **total immersion**. Engagement involves first getting involved in the game, via accessible controls and relevant gameplay. Engrossment is the point where the player's emotions become directly affected by the game. Finally, total immersion is the feeling of the player being fully engaged in the game, to the exclusion of all else (a phenomenon they call 'presence') – a state very similar to flow. However, they claim total immersion is a brief experience, in contrast to flow, which can last up to the entire length of the activity the person is engaged in (Brown & Cairns, 2004).

Norman (1993, pp. 34-35) proposed a set of high-level guidelines to create an optimal flow environment:

- Provide a high intensity of interaction and feedback
- Have specific goals and established procedures
- Provide motivation
- Provide continual challenge - not so difficult it is frustrating, and not so easy that it is boring
- Provide a sense of direct engagement with the activity and environment
- Provide appropriate tools to the user to fit each task, (i.e. to aid them, not distract)
- Avoid distractions that destroy the user's subjective experience

These guidelines highlight the importance of feedback, goals, motivation, providing a balanced challenge, minimising distractions, and allowing learners to interact directly with the environment and activity.

Csikszentmihalyi's 'flow' theory was additionally used to distil player enjoyment factors, in the GameFlow model by (Sweetster & Wyeth, 2005). These factors include concentration, challenge and player skills, control, clear goals, feedback, immersion and social interaction. While the theory was not conclusively demonstrated in Sweetser and Wyeth's study, there seem to be several parallels to learning theories present in these factors, including Keller's 'attention' and 'confidence' motivational factors, the progression of Gagne's nine events, and the social elements of Laurillard's conversational framework.

Kiili approaches the 'challenge' aspect of flow in educational games more specifically, in his Experiential Gaming model (Kiili, 2005), displayed in Figure 2-5. The model involves two cyclical components: in the first, the **ideation loop**, the students think of unstructured ideas to solve a particular challenge. In the second, the **experience loop**, the students experiment with these ideas, with clear goals in mind about what the ideas should accomplish. They then reflect upon the feedback they get from observing the experiments, and refine their findings into schemata, to further their understanding of how the game works. With this new understanding, the students then return to the challenge, and think of new, more structured ideas, based on what they have learned. However, as with Laurillard's conversational framework, the model describes the ideal process of students engaging with in-game problems, but takes it for granted that they would be willing to persist through the game throughout their difficulties in finding solutions. This in turn highlights the issue of how to motivate them to engage with this process.

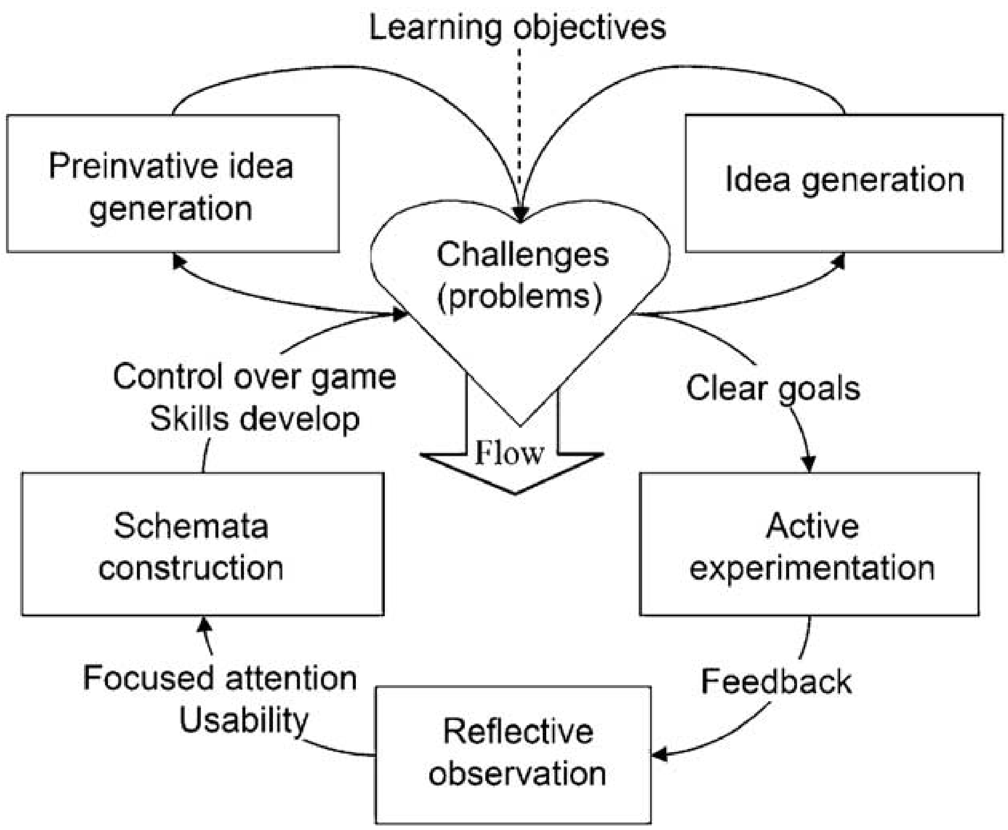


Figure 2-5: Kiili's Experiential Gaming Model (Sourced from (Kiili, 2005))

Expanding on the general immersive themes in games, Dickey (2007) identified five aspects of MMORPGs that contribute to intrinsic motivation. These include:

- **Choice –** How one's character is developed, and the choice of which small quests to partake in
- **Control –** The selection of quests, the order of completion, and the strategies employed to do so
- **Collaboration –** The communication tools available, the collaborative quests, and the 'social capital' of each character
- **Challenge –** Each quest is designed with a certain skill set and/or attribute level in mind
- **Achievement –** Increased character status, attributes and skills, as well as in-game items and currency

The first three can be seen as ways of immersing the player in the game world, while the latter two are ways of immersing the players in the game tasks themselves.

In addition, Dickey made observations into the 'small quests', which make up the majority of the gameplay in MMORPGs. While these naturally vary between games, she categorised the general types of quests into six different tasks, as well as identifying the knowledge types associated with them (Dickey, 2007). These categories include:

- **Collection Quests –** Requires a player to find one or more specific items, or to perform a specific task a number of times. Typically, these quests are intended to introduce players to new concepts or facts about the game world. As such, the knowledge type expressed here is **declarative**
- **Goodwill Quests –** Less formal quests, where a high-level (more experienced, more skilled, more powerful) player helps a low-level player. These express the **declarative** knowledge type for low-level players, and the **metacognitive** type for high-level players, who must reflect on what they know of the task, and reinforce their own knowledge

- **Fed-Ex Quests –** The player is required to deliver, collect or manipulate items by going to a different area. The knowledge type here is **procedural**, as the quests require learning how to overcome obstacles and find items in order to proceed
- **Messenger Quests –** The player simply find a non-playable character (NPC) and talks to them. These too express **procedural** knowledge, since the information conveyed by these NPCs usually pertain to new methods or processes for the player to observe, then recount
- **Bounty Quests –** The player must find and defeat a certain character or monster. Assessing enemy strengths and weaknesses, as well as devising strategies, requires the player to demonstrate **strategic** and **metacognitive** knowledge
- **Escort Quests –** Requires a player to escort an NPC to another location. The exploration and consideration of environmental factors means these too elicit **strategic** and **metacognitive** knowledge from the player

An advantage of the small quest structure is that the player gets exposed to various new areas and resources in the game, encouraging them to understand the game world more, and thus become more involved in it. Additionally, the small quests provide experience to advance character development, and foster collaborative and strategic thinking. The idea that can be drawn from these observations is that educational content can be taught through the player learning about the game world, and by participating in the role they play in-game.

One gameplay element of serious games that is promoted in several implementations and theories is experimentation, and their resultant consequences. In other words, the player's actions within a game should have a tangible impact on the game world, including triggering changes to the landscape, changing the behaviours of in-game characters, or changing the behaviour of enemies and obstacles. As discussed in Section 2.1.2, this is one of the cornerstones of Barab's Transformational Play theory (Barab, et al., 2010 (a)). Harteveld, in the design of his educational game, noted it as an important teaching tool, under the name of 'exploration' (Harteveld, et al., 2007), and Freitas and Neumann similarly emphasise the use of exploration in serious games (Freitas & Neumann, 2009). The idea is endorsed by Squire as the foundation of his theory that games are 'designed experiences', where players learn by 'doing' and 'being' (Squire, 2006).

In addition, there are factors regarding the gameplay and overall experience of games that affect a player's immersion. One such factor is that of uncertainty, the idea that there needs to be variation and unexpectedness in a game's progression (Harteveld, et al., 2007). This provides motivation for players to replay the game: otherwise players will be able to play through a game with minimal effort, and will not be required to think about their actions.

A further factor presented is that of 'fantasy', using creative license to offer metaphors or analogies of real-world processes. This abstraction can focus the player's attention, and in doing so helps facilitate immersion. (Garris, et al., 2002) For instance, having unfamiliar sensory input (i.e. sights, sounds) can be attention-grabbing, whilst also satisfying one's innate desire for unusual experiences. It also allows real-world processes to be presented to the players from different perspectives, and thus gain a deeper understanding of them.

There have been numerous attempts to integrate these immersion theories with educational theories for particular serious games. Among the more noteworthy of these attempts is the serious game *Immune Attack*, developed by a collaborative effort between the Federation of American Scientists (FAS), Brown University, and the University of Southern California. (Kelly, et al., 2007)[2]

The aim of *Immune Attack* is to teach introductory immunology concepts to college-level students. It takes the form of a 3D action/exploration game, similar to 1995 game *Descent*: essentially, the player controls a miniature craft inside a person's body, and must train immune cells to interact with the body and respond to threats. For instance, the first level requires the player to direct monocytes towards infected areas to encourage them to morph into macrophages. In designing the gameplay, the developers wanted to be as scientifically accurate as possible. For this reason, the idea for a First-Person Shooter game (among the more popular genres) was vetoed, as it did not reflect how the immune system worked.

[2] http://www.fas.org/immuneattack/, 2011. *Immune Attack*, Federation of American Scientists

Figure 2-6: Screenshot of Immune Attack, showing the blood vessel environment and an information box displaying details about a selected cell. Sourced from (Kelly, et al., 2007).

This interesting observation about the suitability of gameplay elements for educational ones was characterised by Harteveld as **harmony of gameplay elements**, or ensuring no element of gameplay feels out of place, with respect to other gameplay elements and the subject matter (educational or otherwise). The observation finds support from Gunter's assertion of needing to integrate educational practices into game design, and suggested from Frazer's observations of educational mini-games, where he found the presence of an irrelevant boating challenge as motivation in a pyramid-building game to be hindrance to learning and engagement (Frazer, et al., 2007).

One approach which investigates how to integrate educational content into gameplay to produce more effective serious games is the Learning Mechanics – Game Mechanics (LM-GM) model (Arnab, et al., 2015). The model comprises of a set of non-exhaustive design mechanics, split into two dimensions: **learning mechanics** (e.g. 'Guidance', 'Experimentation', 'Assessment'), derived from education literature and theorists, and **gameplay mechanics** (e.g. 'Role Play', 'Levels', 'Feedback'), which were derived from literature on game mechanics/dynamics. When applied to a particular serious game, one would construct a 'game map', using the elements of the model to illustrate how the learning mechanics and gameplay mechanics interact, if at all, in turn allowing game developers/assessors to identify where discrepancies are.

The model is still in its early stages, however, with one of their conclusions being that the model is too complicated for novices. A possible cause of this could be the large number of total mechanics (77), each of which lack precise definitions, possibly making it difficult to interpret how they should be applied. Nonetheless, the model was shown in Arnab's study to draw attention to how gameplay and educational aspects can be combined, and endorses the idea that is a key part of creating engaging serious games.

2.3 Narrative Theory with Games

In addition to gameplay, an important aspect of player immersion is the game's narrative. Many engaging games essentially tell stories, which gives context and meaning to the underlying gameplay (Qin, et al., 2009). By providing this context, players can get more invested in the game's activities, and thus have greater incentive to continue playing. As such, a convincing narrative essentially contributes to the state of 'flow' discussed in Section 2.2, and is thus an important consideration for educational serious games.

At the most basic level, each narrative is comprised of the 'story' and the 'discourse' (Hargood, et al., 2009). The 'story' consists of the narrative information, and the 'discourse' comprises of the way the information is conveyed. Creating a believable narrative requires that the story elements are convincingly combined in a discourse, to help the flow and understanding of the narrative. This problem is known as '**narrative cohesion**'. (Hargood, et al., 2011)

Hargood et al identified several variables that influence narrative cohesion. They are targeted at traditional narratives, but the ideas behind them can be applied to game-based narratives as well. These factors include:

- **Logical Sense –** The narrative content must make sense with respect to the story's progression. Traditionally, this is achieved by using connective language to explain the content, and measured by the use of connective terms, relative explanations and references to previously told information
- **Theme –** The concepts communicated implicitly through the course of the narrative. These are formed in narratives with 'motifs', which are atomic narrative elements (e.g. "the helpful beast")
- **Genre –** The presence of reoccurring features which contextualise the narrative

- **Narrator –** The presence of an identifiable storyteller communicating the narrative throughout
- **Style –** The way in which story elements are presented in the discourse

These variables are important because discrepancies within them can potentially take the person out of the story, and in terms of game narrative, can disturb their sense of flow and engagement.

Two agreed-upon narrative structures in games include (Salen & Zimmerman, 2004, pp. 383-384) (Jenkins, 2004):

1. **Embedded Narratives –** The story is pre-determined, providing players with specific overarching goals to achieve in order to complete the game. While the player can achieve the goals using potentially a variety of methods, the goal and overall outcome is still the same
2. **Emergent Narratives –** The story develops uniquely for each player, depending on what actions they take while playing the game

Both narrative types can offer immersive experiences for the player, but both have disadvantages. With embedded narratives, the overall story focus potentially weakens its immersion on repeat plays (as the player already knows how the story will turn out). By contrast, emergent narratives lack much focus at all, meaning the narrative may become boring, or possibly not develop at all (Aylett, 1999).

With that said, Qin et al. argue that an emergent approach makes more sense for a game narrative, in that the person playing a game feels they have more influence and involvement in the story progression than they would with books or films (Qin, et al., 2009). They further found six factors, which facilitate and keep a player in a state of flow using the game narrative. Their two antecedent factors for immersion were '**curiosity**' and the game's relative '**difficulty**'. Once the player is immersed, the factors important to maintain flow are '**concentration**' on the narrative, the sense of '**control**' the player has over the characters and game world (i.e. feel they are able to enact their own strategies for solving game problems), and '**familiarity**' with the story (how much they know about the game's story, and how it will progress). Finally, they describe '**empathy**' towards the game characters as the factor the player reflects on after exiting the flow state.

However, Salen and Zimmerman argue most game narratives ultimately combine these structures, providing overall goals, setting and characters (embedded narrative), while allowing the player to choose the means by which they achieve those goals and how they interact with the world (emergent narrative). With this in mind, they proposed a set of factors which contribute to a more engaging game narrative. (Salen & Zimmerman, 2004, pp. 385-399) These include:

- **Goals –** overarching objectives for the player to achieve by progressing through the game, embedded in the narrative and game challenges
- **Conflict -** impedances to the player’s progress through the narrative
- **Uncertainty** – the outcome of the game narrative should not be a foregone conclusion
- **Core Mechanics –** how the game mechanics the player uses to progress aid the narrative experience the designers are trying to convey
- **Narrative Space –** a narrative environment (i.e. game world) which presents compelling problems, supports creative experimentation to solve them and reacts to the player’s choices with “meaningful consequences”
- **Narrative Descriptors** – representations used to communicate the in-game story to the player

A recurring theme throughout these factors is, once again, the idea of player agency; where the player feels like an active participant in the game world, with their actions having a meaningful impact on the game environment and the story. The core mechanics factor additionally emphasises the importance of selecting fitting game mechanics to suit the game experience; a reiteration of the harmony of gameplay elements concern in Section 2.2.

Burn and Schott further emphasise attention to detail in a game’s narrative as a way of enhancing the story using branching details that are not necessarily mandatory to the core story to heighten a player’s agency. (Burn & Schott, 2004)

Game researchers have also identified the importance of plausibility and curiosity: both are deemed necessary to maintain the players' curiosity. Gunter’s RETAIN model (Gunter, et al., 2006) emphasises the need of curiosity and scene-setting, in the form of ‘game focus’ and ‘didactic focus’. Dickey asserts that it can be established in games through the interplay between characters, events and the environment. The context and the setting establish the boundaries of what is plausible in the game world (which, as discussed earlier, the player can discover through exploration) (Dickey, 2011). She suggests the types of curiosity present in games come in two forms:

- **Perceptual –** Curiosity fostered by novelty
- **Epistemic –** Curiosity driven by the desire to answer an intriguing question, and is only satisfied when an answer is found. Much like challenge, this too must be balanced according to the player's abilities

2.4 Summary

There are many issues to consider with creating engaging games, and engaging-whilst-being-educational learning activities. The primary element of immersive activities in general involves keeping a person in a state of ‘flow’; in terms of serious games, this largely consists of intrinsically satisfying, challenging gameplay that is appropriate to the ILOs conveyed in the game. Another important consideration is the inclusion of an engaging narrative that encourages learner involvement, having the game world tangibly affected by the player’s decisions. The multiplayer and collaborative elements further show the development of knowledge and engagement with learning activities to be linked to reactive environments that allow for player experimentation. Each of these issues can have an effect on how immersed a person feels in games and learning activities alike, but there is as yet no model or framework which considers all these issues together, particularly not from a serious games standpoint; this means serious game developers can potentially miss crucial considerations for getting players engaged, which can negatively impact the uptake of their games, or the educational value within those games. To provide a potential solution to this problem, the focus of this research is on the key aspects of what makes an educational game work in an engaging way.

Chapter 3. Methodology

Following the literature review in Chapter 2, it was decided to investigate the core elements that influence how engaging an educational game is. To that end, this chapter explains the overarching progression of the research, and an overview of the research methods employed.

3.1 Research Methods

This section outlines the research methods used in the research, exploring their strengths and weaknesses.

3.1.1 Quantitative and Qualitative Research

There are two paradigms for collecting data scientifically, which are the 'qualitative' and 'quantitative' paradigms. Each presents its own advantages and disadvantages for establishing the correctness of the model. This subsection briefly outlines each approach.

Quantitative research is a paradigm designed to gather factual, measurable data, such as numerical ratings. The research methods for this paradigm can include structured surveys and experiments (Adamson, 2005), which can have a large sample size, with data that can be interrogated using statistical methods, such as comparison or correlation (Morse, 1991). This approach allows the researcher to study a paradigm in an objective way, and thus in theory the data collection will be unaffected by the researcher's biases. (Sale, et al., 2002) (Creswell & Clark, 2011). Furthermore, the closed-ended nature of these studies supports a large number of participants, meaning their samples can be more representative of the target population.

On the other hand, the fact quantitative research is closed-ended means it can only be used to test the ideas the researcher already has; it does not support the emergence of new ideas and directions put forward by the participants. (Borg & Gall, 1983, p. 437)

Qualitative research is used to gather open-ended data on a particular research problem. While quantitative research focuses on what responses the participants give, qualitative research also investigates why they give such responses, delving into their opinions and feelings on the subject matter (Hove & Anda, 2005). Typically, this is achieved through semi-structured or unstructured interviews, which give the participants the opportunity to explain their point of view. One advantage of this open-ended approach is that it can reveal new ideas and directions the researcher did not consider. Furthermore, it permits the participants to explain their opinions more clearly and in more detail (Borg & Gall, 1983, p. 436) (Oishi, 2003). In the context of the model, this can provide deeper insights, which can be used to structure the model, or else understanding why certain errors exist in the model in the first place.

However, the data from qualitative research cannot be represented numerically, and so must be interpreted from the researcher's point of view, making the data analysis potentially subject to researcher bias (Creswell & Clark, 2011, p. 12). Furthermore, the in-depth, open-ended nature of qualitative studies means that they are potentially more time-consuming to conduct and to analyse, and can only realistically be performed with a relatively small sample size, which means the results are less generalizable to a wider population (Sale, et al., 2002).

3.1.2 Mixed Methods and Triangulation

As established in the previous subsection, qualitative and quantitative research each have advantages which make them suited for validating this model, but also each have disadvantages which limit how useful their feedback data can be. However, the advantages of quantitative research can compensate for the disadvantages of a quantitative study, and vice versa. For this reason, some researchers use a 'mixed methods' research methodology by using both quantitative and qualitative studies to research the same phenomenon (Jick, 1979) (Morse, 1991). Combining the use of these two paradigms allows the researcher to uncover as many issues as possible for the research problem they are investigating. In particular, it can be used to build upon a theoretical basis (such as a framework) by providing perspectives from qualitative and quantitative studies (Creswell & Clark, 2011, pp. 10-11).

One variant of mixed-methods is called 'triangulation'. In triangulation, a research problem is cross-validated using data gathered from at least two different sources. This can include multiple different methodologies (a mix of quantitative and qualitative methods), or else multiple uses of the same methodology (Adamson, 2005). The idea behind this method is that it allows the research problem to be considered from multiple perspectives, in turn allowing the researcher's conceptions of the research issues to be improved with greater accuracy. (Jick, 1979). The methodology of this study benefits from triangulation because the proposed model is new, and based primarily on the researcher's perspective on the issue; triangulation would allow the findings to be considered from alternative points of view within the games field, each identifying new points and discrepancies the other views cannot, to construct the most representative model of educational game immersion possible.

3.2 Interviews

Interviews are one way of eliciting qualitative data, and have two general forms: individual interviews and focus groups. Individual interviews take the form of a one-on-one, guided discussion between the participant and interviewer. Focus groups on the other hand involve a discussion between participants as a group, structured and guided by a trained moderator. Both methods involve questioning the participants about the research problem being investigated, and finding out their opinions regarding the issue in depth. (Kaplowitz & Hoehn, 2001) (Oishi, 2003, p. 172)

The advantage of focus groups is that the discussions between participants can help stimulate each other to put forward ideas, which the group can expand and build upon, providing development, which is not possible in a one-to-one setting (Kitzinger, 1995). On the other hand, it is possible for group effects to bias the results of the discussion, such as ‘group think’, where one participant dominates the discussion, leaving the others less willing to express their opinions. With appropriately conducted individual interviews, there is no pressure to conform to others, which allows the participant to more freely express their opinions, and because they are the focus, they can express their opinions in more detail (Kaplowitz & Hoehn, 2001).

3.3 Surveys

Questionnaires are a typical method of eliciting quantitative data from a large group of participants. One form which they can take is the self-administered questionnaire, where the participants are given the questionnaire by the researcher individually, and are left to complete them on their own, while being available to elaborate on any questions which the participants may not understand. Because this form has a relatively low cost compared with other methods (such as interviews), the larger sample size allows it to be more representative of the target population, and the questionnaires are simpler to distribute and potentially allows quicker responses from participants (Borque & Fielder, 2003).

3.4 GQM Approach

The GQM (Goal Question Metric) approach, created by (Basili, et al., 1994), is a process used to help create metrics for particular purposes in a piece of software (Fenton & Bieman, 1997). This initially involves establishing the overarching reason or ‘goal’ for measuring a particular aspect of the software. After the goal is established, ‘questions’ are devised which need to be answered to satisfactorily achieve the goal. Finally, in order to answer these questions in a measurable way, a set of 'metrics' is constructed for each question.

3.5 Game Immersion Measurements

While there was a lack of studies found regarding how to measure a person's sense of immersion whilst playing a game (not to mention a serious game), certain studies have explored particular promising approaches.

Three of these approaches were tested in a study conducted by Jennett et al. (2008), who investigated several methods of measuring immersion during and after playing a non-serious game. The three approaches described were the ones which they felt were most promising, showing statistically significant results in support of representing immersion:

3.5.1 Eye Tracking

In this approach, as the participant plays a particular game, their eye movements are tracked, focusing specifically on the number of 'fixation' periods (where the eye is focused on a particular point) the participant experiences over time. The theory is that less frequent fixations (thus more periods of particular focus) indicates that the participant becomes less distracted by external influences, and therefore becomes more immersed.

This approach makes use of a game to test for immersiveness, and a control game designed to be as unengaging as possible: in the case of Jennett's study, the testing game was the 1998 first-person shooter PC game *Half-Life*, and the control game was a simple application where the participants click on squares as they appear on screen.

3.5.2 External Task (Tangram Puzzle)

In this approach, the participant is given an external task to complete before and after playing a particular game; for Jennett et al.'s study, this task is a Tangram puzzle, where a selection of different-shaped blocks must be arranged to form a particular shape. They hypothesized that the difference in task completion times before and after playing would be smaller (i.e. improvement would be lesser) the more immersed the participant is after finishing playing. The reason is that immersion reduces one's external awareness, thus taking more time to adjust back to awareness of the real world.

To judge how immersed a participant was, they were given an immersion questionnaire to fill in after playing, consisting of Likert items about their experience playing the game (e.g. *I felt that I really empathised/felt for with the game*, *I did not feel any emotional attachment to the game*).

3.5.3 Immersion Questionnaire

This approach utilizes a general qualitative questionnaire about how immersed the participants felt while playing a particular game, issued when they had finished playing it. This questionnaire comprised of Likert items related to the feelings of immersion experienced (e.g. *To what extent did the game hold your attention?*, *To what extent did you notice events taking place around you?*).

3.6 Research Methodology

The flow of the methodology used in this research to explore engaging educational games is shown in Figure 3-1. The green triangle encompasses each of the stages and sources used to create and confirm an initial model based on it, while the blue rectangle encompasses the experimental stages used to validate the confirmed model in a practical setting. The following subsections explain each of the methodology stages in sequence, describing what methods were used for each stage of developing the model and why.

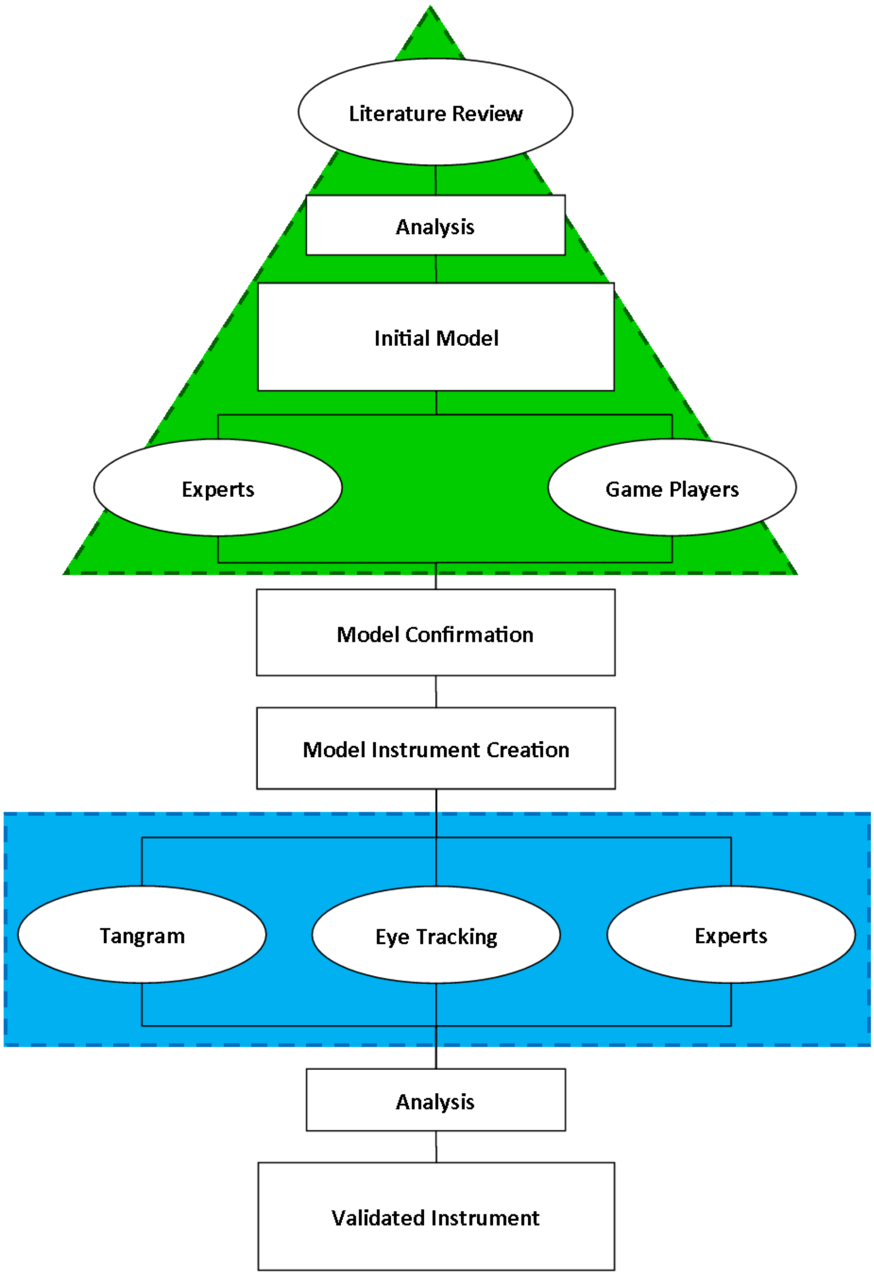


Figure 3-1: Research Methodology Diagram

3.6.1 Confirmation Methodology

Upon compiling the literature from Chapter 2, a gap was discovered in the research pertaining to exactly what are the most important elements of facilitating immersion in educational games. To this end, the first research question explored was:

RQ1: **What aspects affect a player's immersion in an educational game?**

Since the research questions would benefit from both qualitative and quantitative methods, in order to provide additional perspectives to the researcher's perspectives on educational game immersion, a mixed methods approach was employed. The idea was to balance the literature extrapolations with the opinions of field experts and a general population, in order to confirm the framework from a variety of relevant perspectives. In the context of this study, the data sources chosen to validate the model come from the literature review, experts in the game academic/ development field, and general game players (see Figure 3-2).

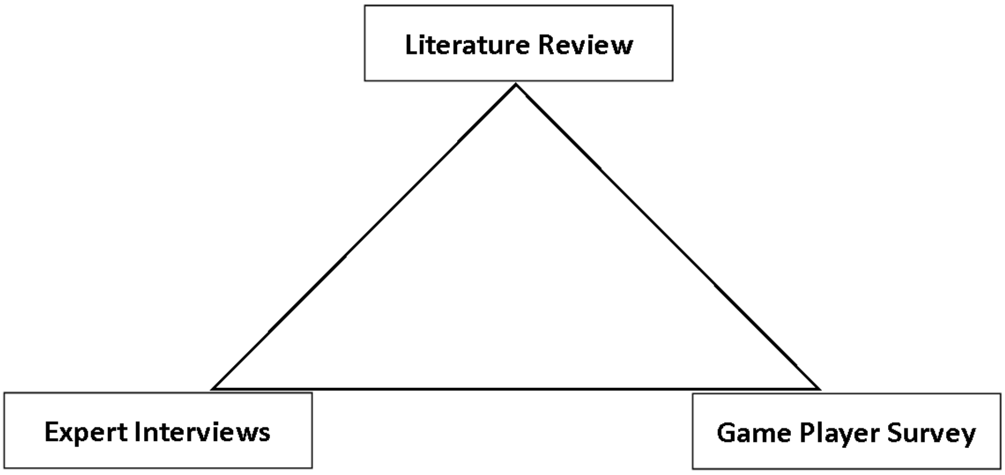


Figure 3-2: Triangulation of the research methods employed to confirm the model

Initial Model Development

The model was initially developed through synthesising the common themes arising from the literature explored in Chapter 2. The details of this process are explained in Sections 4.1 and 4.2.

Expert Interviews

To provide an initial evaluation of the model, in terms of how representative of the issues it was, the next stage of the research was to seek feedback on the model from experts in the field. Because detailed opinions were being sought in order to refine the initial model, and because hearing individual perspectives clearly was vital to gain as wide a perspective as possible, individual interviews were conducted with the experts.

To prompt the needed discussion points for the interviews (i.e. which aspects of the initial model the experts disagreed with), a questionnaire with 4-point Likert scale questions was chosen. Such a scale, (Strongly Disagree to Strongly Agree) omits the neutral option, making it a 'forced choice' scale (Allen & Seaman, 2007); this was deemed appropriate to use because the questionnaire was not used for the interview analysis, but to gauge the agreement for the aspects of the model in order to focus the interview questions, and it was not deemed likely in this context that field experts would require a neutral option to provoke discussion.

Ethical approval for the interviews was applied for, and subsequently granted, by the University of Southampton's ERGO (Ethics and Research Governance Online) committee (**Reference ERGO/FPSE/9132**).

Further details of the study and its analysis are given in Chapter 5.

Player Survey

After amending the model according to the experts' feedback, the next stage was to investigate whether the updated model is representative of immersive aspects in educational games, by providing it to a representative sample of game players with expertise on the merits of immersive games.

For this reason, self-administered questionnaires were chosen for the quantitative confirmation study.

For the questions themselves, 5-point Likert scale opinion statements were used, comprising of the items 'Strongly Agree', 'Agree', 'No Opinion', 'Disagree' and 'Strongly Disagree'. This 5-point scale was used, as it was a recommended minimum requirement for a Likert scale (Allen & Seaman, 2007).

Ethical approval for the surveys was applied for, and subsequently granted, by the University of Southampton's ERGO committee (**Reference ERGO/FPSE/10627**).

Further details of the study and its analysis are given in Chapter 6.

3.6.2 Model Instrument Creation

Upon creation and confirmation of the initial model from a theoretical standpoint, the next research question involved investigating whether the model could be applied in practice:

RQ2: **Can the Immersive Educational Game Model be used to measure the immersive qualities of an educational game?**

To try and answer this question, an instrument was created to measure immersive qualities in educational games, with its created metrics based on the confirmed model. This instrument was then tested in three experimental studies.

Model Instrument Creation

In order to provide a means of measuring an educational game's immersive qualities using the IEGM, the first step was to define a set of metrics, which could be used to test each aspect of the model. Since the intended purpose of the IEGM metrics was to be measure specific aspects of software (specifically, immersive qualities in games), the GQM approach was deemed an appropriate way to help generate the metrics.

3.6.3 Experimental Methodology

After creating the IEGM metrics, the next stage was to test whether they could accurately judge the immersive qualities of educational games in practice (and thus, whether the IEGM can be used for measuring immersive qualities). To achieve this, three experiments were conducted, using an already-existing immersive educational game to test on.

Ethical approval for each of the experiments was applied for, and subsequently granted, by the University of Southampton's ERGO committee (**Reference ERGO/FPSE/14087**).

Experiment 1: Tangram Task

The first experiment was to confirm that the used educational game elicits immersion from the participants on a physiological level. To this end, a Tangram task method similar to that described in Section 3.5.1 was chosen to assess how immersed the players were while playing the selected game. In this method, the participants were asked to complete a Tangram puzzle for a specific shape before and after playing the selected educational game.

The completion times for the 'before' and 'after' conditions were then compared with each other (as an initial assessment – if the completion time before playing is shorter than the completion time after playing, that would be a certain sign of immersion, rendering further analysis redundant). Following that, the time differences between the conditions were correlated with the results of the IEGM metrics questionnaire.

Further details of this experiment are given in Section 7.2.1.

Experiment 2: Eye-Tracking

As with the first experiment, the second experiment was used as a confirmation method to physiologically assess how immersed the players were in a particular game. This experiment made use of an Eye-Tracking method as described in Section 3.5.2, where participants would have their eye movements tracked while playing the selected immersive educational game, and while playing a control non-immersive game. The rates of fixations over time were then compared between the immersive and non-immersive conditions.

Further details of this experiment are given in Section 7.2.2.

Experiment 3: Questionnaire

The final experiment aimed to test the created metrics directly. To this end, a modified form of the Questionnaire study in Section 3.5.3 was conducted, where the participants would answer an immersion questionnaire after playing the selected immersive game. The difference in this experiment was that the questionnaire was about the model metrics, as opposed to being about the general state of immersion.

Further details of this experiment are given in Section 7.2.3.

An amendment to the experiment was later granted, to permit using a £1 bar of chocolate as an incentive to participate (**Reference ERGO/FPSE/17699**).

Chapter 4. Initial Development of the Immersive Educational Games Model

Following the literature explored in the previous chapter, this chapter explores the development of a model to address the problems in designing educational games to be immersive.

4.1 Initial Analysis of Educational Game Effectiveness

Previous theory work around game design has uncovered a wide variety of issues regarding entertaining gameplay and engaging worlds, and those focused on educational games have raised many issues about how to make instruction through games compelling via presentation of the educational content in the game (Arnab, et al., 2015).

However, while these issues are important, the view of the researcher is that each theory misses key issues raised by other theories. While it makes sense for studying a particular aspect of serious game design in isolation, it is not useful to have these issues so separate from each other when it comes to implementing an immersive serious game, particularly when the entertainment and educational aspects appear to benefit from each other's inclusion.

To address the disparity of these issues, the compiled literature was analysed in terms of what makes educational games 'effective'. 'Effectiveness' in this context is defined as the capacity of an educational game to demonstrate and provide instruction on its subject matter, in addition to providing enjoyable and engaging entertainment (for instance, how well it satisfies the conditions for an optimal 'flow' experience). The reason for this approach was to provide a complete yet concise overview of the issues surrounding educational game design, in such a way that it could be used by educators and game designers to identify the most important areas to address when creating an educational game.

To this end, as the literature was being explored, a preliminary design model was created, called the Effective Serious Games Model. This was created by evaluating the literature explored in Chapter 2, in terms of what the recurring terms and themes were in each of the articles and implementations, and the relative strengths and weaknesses of their ideas. These themes were included in the model if they appeared to recur in two or more sources pertaining to gameplay, educational theory, or both. The included themes were then subdivided into five categories, in order to help clearly represent the expected inter-relationships of each of the included themes:

- **Flow –** Representing the issues relevant to maintaining a sense of flow, as described by Csikszentmihalyi (1990), as this is vital to maintain the players' attention. The issues include:
    - **Balanced Difficulty**
    - **Consistent Rules**
    - **ILO Focus**
    - **Character Control**
- **Knowledge-Building –** Representing the issues related to how to provide instruction on the subject area the serious game addresses, and how to facilitate the best learning conditions within them. This covers the issues of:
    - **Contextualisation**
    - **Reflection**
    - **Abstraction**
    - **Feedback**
    - **Gameplay-Educational Harmony**
- **Exploration –** Representing the issues in creating an in-game environment which helps to convey the principles of the gameplay and subject matter. These issues include:
    - **Consequentiality**
    - **Boundaries**
- **Narrative –** Representing the issues in creating an engaging in-game narrative which encourages the player to become invested in the game environment and motivates them to continue playing. These issues include:
    - **Goals**
    - **Conflict**
    - **Uncertainty**
    - **Narrative Cohesion**
    - **Curiosity**

- **Multiplayer –** Representing the issues regarding the unique issues regarding multiplayer serious games (specifically involving collaborative tasks). While these are less general, in that they do not apply to single-player games, they were nonetheless regarded as important for learning outcomes which could be better conveyed through group activities. These issues include:
    - **Common Goals**
    - **Collaborative Tasks**
    - **Resource Sharing**
    - **Strategic Discussion**

While the creation of this model did help to provide an initial focus as to how one might best create and utilise serious games, on reflection, it was deemed to be too vague about its overall objective; ‘effective’ was not a used term in the literature, and in context could be easily interpreted in several different ways (e.g. ‘effective’ instruction, ‘effective’ at being entertaining). In addition, there still arose repetitions of aspects in amongst the five categories, such as ‘Common Goals’ and ‘Goals’, which only differed in terms of scope rather than context. In its current state, the model was thought more likely to create confusion for developers and educators, as opposed to assisting them with assessment and development.

In order to provide a more clear and concise guiding model, a new focal point was needed, which would be recognisable and identifiable for both educators and game developers. Immersion was chosen as that focus, as it is a defined term in game academia (Brown & Cairns, 2004), and its underlying concept of promoting voluntary engagement in an activity or subject matter recurs as a desirable end goal throughout educational and gaming literature, and indeed in literature concerned with creating optimally engaging environments (Laurillard, 2002) (Gee, 2005) (Csikszentmihalyi, 1990) (Barab, et al., 2010 (b)) (Norman, 1993). Since learning and enjoyment are better facilitated the more engaged a person gets, it made sense to consider designing educational serious games to facilitate that maximal stage of engagement, immersion.

This in turn led to the research question:

RQ1: **What aspects affect a player’s immersion in an educational game?**

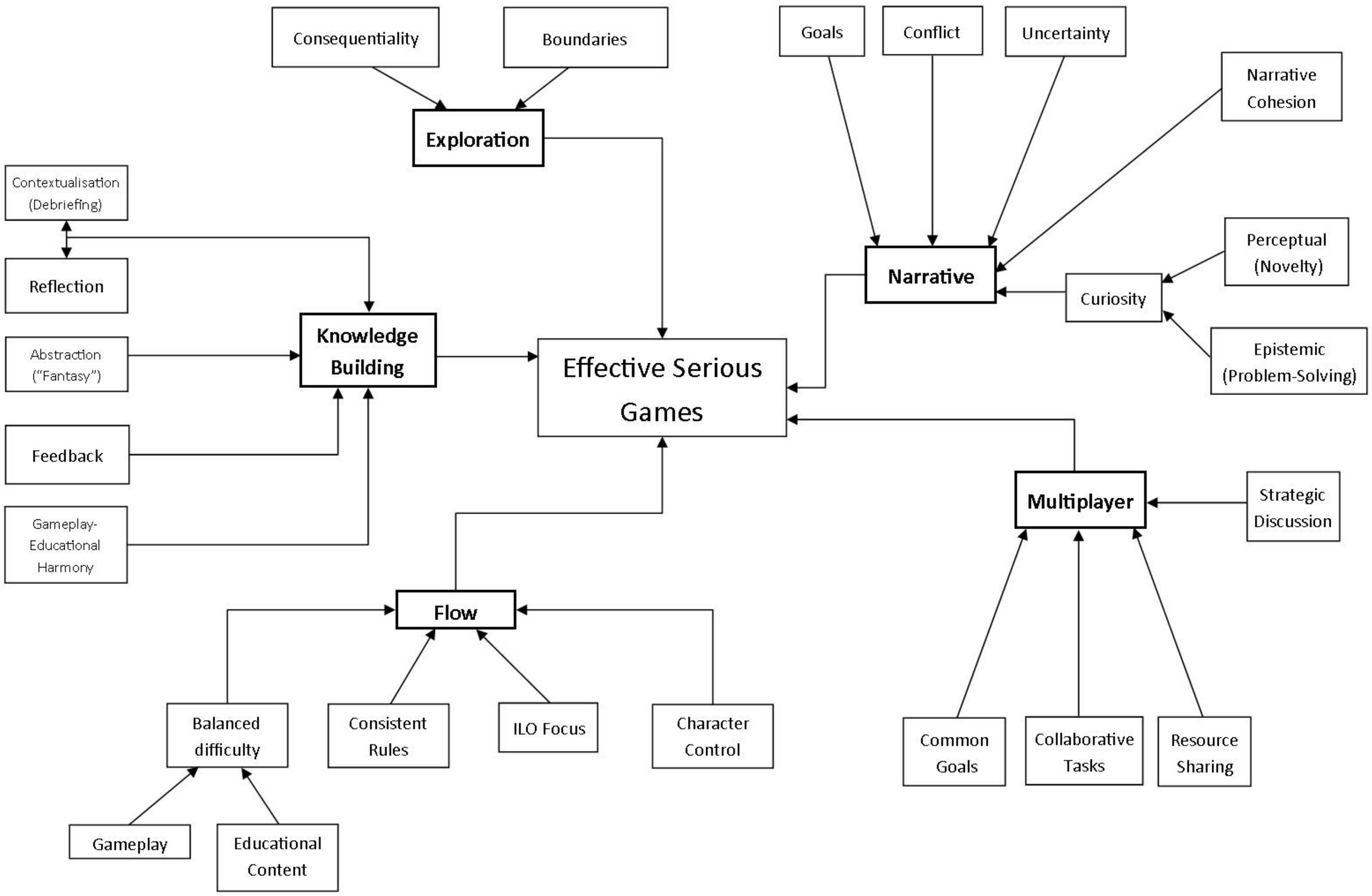


Figure 4-1: Effective Serious Games Model

4.2 Development of the Immersion Model

To answer **RQ1**, and by doing so, attempt to provide a unified view of immersion aspects in educational games, the prior Effective Serious Games Model was altered and refined, in order to create a new model, focused on 'immersion' as opposed to 'effectiveness'. Having this Immersive Educational Games Model would help to provide a useful blueprint for designing immersive educational games, to ensure they will be as effective at entertaining and conveying their messages as possible.

With the initial terms and themes established in the Effective Serious Games Model, the creation of the Immersive Educational Games Model involved filtering these terms and themes, both for repetitions (multiple different ideas arguing the same essential point), and for ideas that were less or not important towards generating an immersive or educational experience. To help accomplish this, the themes were grouped together into broad categories according to the central focus of the theme (*Gameplay*, *Education*, *Narrative*, and *Multiplayer*). The themes that fit into multiple categories were placed in a hybrid category of the two most predominant ones (e.g. *Gameplay/Education*). In this way, the patterns and recurrences between themes, as well as their place with regard to facilitating immersion, were easier to identify.

These filtered ideas were then synthesised into a small set of defined aspects, each representing a specific theme. The idea behind this was to present the core aspects that represent immersion in a clear fashion, which could in turn be used to focus future research and development on making educational games immersive.

Once these core aspects were identified, the remaining theme categories they were grouped in were renamed, to reflect how the contained core aspects were interrelated. These category names became *Education*, *Gameplay* and *Agency*.

4.3 The Immersive Educational Games Model

The resulting Immersive Educational Games Model groups the synthesised aspects regarding immersion into three categories, each representing a larger research concern linking them together (represented diagrammatically in Figure 4-2). The nature of the categories, the aspects, and the literature justification for their inclusion are explained in the following subsections.

4.3.1 Education

The first category contains the aspects involved in making educational material immersive in a game. This involves how to maximise the possibility of learning the material within the game, by emphasising why it should matter to the player in the real world, and by presenting it naturally within the progression of a game.

Relevance

Several educational researchers have theorised learning material is more likely to be retained long-term if it is made personally relevant to the learner. In terms of education, Gagne suggests the issue of transferability as one of his events of instruction (Gagne, 1970) (Gunter, et al., 2006). Keller similarly proposed that learners need to see value in the material for them, in order to motivate themselves to learn it in the first place: these values can include personal satisfaction, means to particular goals, or to coincide with the values of their peers (Keller, 1983). The idea of learner value is supported in Merrill's first principles of instruction, which proposes encouraging learners to draw connections between the material and their everyday life (Merrill, 2002). To this end, educational games should make clear what the benefits of their presented material are to the player, in order to help them become engaged in it.

ILO Focus

Ensuring that an educational game's intended learning outcomes are clearly integrated into the game's design is a vital concern. Neglecting either gameplay design or the educational components defeats the point of making an educational game (Hamalainen, et al., 2006). Kelly describes this concern in the design of *Immune Attack*, where its team initially thought to model the immune system using a First Person Shooter game; the suggestion was dismissed upon realising the immune system mechanics have nothing in common with the FPS genre (Kelly, et al., 2007). Certain educational game design researchers (Barab, et al., 2010 (a)) (Villalta, et al., 2011) have similarly insisted that educational content must be developed into the gameplay, so the players need to demonstrate knowledge of the content in order to proceed.

Presentation

As part of a serious game's design, the developers must consider how to most effectively convey their ILOs. This includes how the educational content should be paced in the game (including how frequently to introduce new challenges, and what order in which to present those challenges), and what in-game activities (challenges) are best suited to teach particular ILOs. Gagne notes in his nine events of instruction (Gagne, 1970) that the teaching methods used need to correspond to the type of learning that needs to occur. He also emphasises the importance of appropriate prerequisite knowledge to understand increasingly complex ideas – a sentiment echoed in Gee's game learning principles (Gee, 2005), and is the foundation of Bloom's cognitive domain classifications (Bloom, et al., 1956).

4.3.2 Gameplay

The second category outlines the aspects in gameplay that influence a player's immersion. This includes the way the game's challenges are designed, and the underlying feedback mechanisms they employ.

Balanced Challenge

In order to be continually immersed in a game, an important consideration is keeping the challenges difficult enough to be engaging, but not so hard as to be too frustrating. This idea has been frequently recited in games immersion, drawing inspiration from the Flow theory of optimal experience (Csikszentmihalyi, 1990): becoming completely engaged in an activity to the exclusion of all external stimuli. Csikszentmihalyi and Norman both identify that balancing difficulty according to a person's ability in an activity as an important precondition to generate this optimal experience (Norman, 1993). Additionally, balancing challenge has been an important component in educational game theories with respect to maintaining the 'flow' state through pleasantly frustrating challenges (Gee, 2005) (Kiili, 2005) (Harteveld, et al., 2007), and providing feelings of achievement (Brown & Cairns, 2004) (Dickey, 2007).

Feedback

A serious game needs to provide feedback to the player about how well they are progressing. This is important in order to help the player understand where they are going wrong, when to experiment with new strategies, and which strategies are working. This feedback-reconsideration cycle is an integral part of Laurillard's Conversational Framework for teaching (Laurillard, 2002), and Gagne's nine events of instruction (Gagne, 1970). Likewise it is applied in Kiili's Experiential Learning model for educational games (Kiili, 2005). It is also important for the developers to know when and how to place feedback in order to provide appropriate guidance to the player. (Skinner, 1958) (Keller, 1983) (Norman, 1993) (Gee, 2005) (Dickey, 2007) (Harteveld, et al., 2007)

Guidance

Several theories have proposed that appropriate guidance is needed to keep the players focused on the learning objectives, and provide assistance to keep them from getting too frustrated in the game's challenges. This in turn helps the players to gradually understand the presented content. The idea for this derives from Gagne's guidance event of instruction (Gagne, 1970, pp. 313-314), and is endorsed by Merrill's assessment that providing no guidance to the learner leads to more ineffective learning. (Merrill, 2002) Keller likewise suggests that suitable guidance should be provided if the task is unreasonably beyond the learner's abilities, in order to raise their confidence (Keller, 1987). One set of guidelines for designing educational games similarly notes the importance of guidance, and that it needs to be readily available (Villalta, et al., 2011).

Consequences

The game world should react to the player's actions as they engage in the game's challenges. This idea allows the game to provide guidance directly pertaining to how well the player understands the material. Barab argues this point as part of his transformational play model (Barab, et al., 2010 (a)), and the principle has been seen in other educational game instantiations (Harteveld, et al., 2007) (Freitas & Neumann, 2009). This is not to imply that strict real-world simulation is required for every facet of the educational game, but simply to suggest that there must be clear cause-and-effect relationships linking into the actions that the player can take in the game.

4.3.3 Agency

The third category focuses on aspects which let the player feel like an active, immersed participant in the virtual game world. In this way, the player's learning feels more immediately important and consequential.

Narrative

This aspect is concerned with the story progression of the game, including its overarching objectives, the characters and places in the game world, and how the tale progresses as the player proceeds. Establishing a compelling narrative can act as a strong motivator to play a game (Freitas & Neumann, 2009), providing a context for the game's actions and meaningful goals for the player. At the same time, the narrative needs to be flexible enough be influenced by the player's actions to keep them engaged (Keller, 1983) (Burn & Schott, 2004) (Jenkins,

2004) (Salen & Zimmerman, 2004) (Qin, et al., 2009) (Dickey, 2011). Part of this concern is about ensuring that the progression of a story makes logical sense, not only in terms of the sequence of plot points, but also keeping the characters' motivations believable and consistent with their personality. (Hargood, et al., 2011)

Curiosity

A further way to promote attention and motivation is to inspire curiosity in the players. This can be accomplished through mysteries and surprises in the narrative, by details within the environment, or by other stimulating cues. (Keller, 1983) (Dickey, 2011) The idea behind this is to provide a virtual world the players will want to explore; as they become more involved with the world, so too do they become involved with the learning content. (Salen & Zimmerman, 2004, p. 388) (Burn & Schott, 2004) (Harteveld, et al., 2007) (Kelly, et al., 2007) (Freitas & Neumann, 2009)

Fantasy

Having a fictional ("fantasy") setting to encapsulate the educational content is a further important consideration in a serious game. Fantasy settings can allow players to experience the educational content from different perspectives, allowing them to gain a deeper understanding of the material (Garris, et al., 2002). Fantasy settings can also benefit gameplay, as they do not need to mimic reality, and can thus be tailored to be less realistic but more engaging. (Hunicke, et al., 2004) Barab et al also argue that a fantasy setting is an important part of becoming immersed in the game world, precisely by setting up a new world with a new role to play in it (Barab, et al., 2010 (a)).

Identity Projection

The player should feel involved in their in-game role, by being able to control their virtual character's attributes and actions. By becoming invested in the virtual character, and the role they play in the game world, they also become invested in the material presented, if only to advance their in-game character (Gee, 2005) (Squire, 2006) (Yee, 2006) (Dickey, 2007) (Barab, et al., 2010 (a)), while at the same time feeling safe to engage with the game due to the anonymity the character provides (Rankin, et al., 2008).

Experimentation

Serious games should offer the player opportunities to experiment within the game world; encounter problems, formulate a potential solution, attempt it, and then try again if it does not work. This process can allow the player to directly engage with the taught material, as well as encourage the player to reflect on their actions (i.e. what strategy works in what situation). In this way, they may better understand the educational content, and where and how it can be applied. For this reason, experimentation is a key part of Laurillard's conversational framework (Laurillard, 2002), the analysis, synthesis and evaluation categories of Bloom's cognitive domain taxonomy (Bloom, et al., 1956), and of several serious game theories (Gee, 2005) (Kiili, 2005) (Barab, et al., 2010 (a)). Norman similarly notes the importance of experiential learning to creating an immersive activity (Norman, 1993, pp. 22-23). Indeed, in traditional games, it has also been observed that experimentation and sharing discoveries amongst a community helps to develop understanding in an academic fashion (Steinkuehler, 2008).

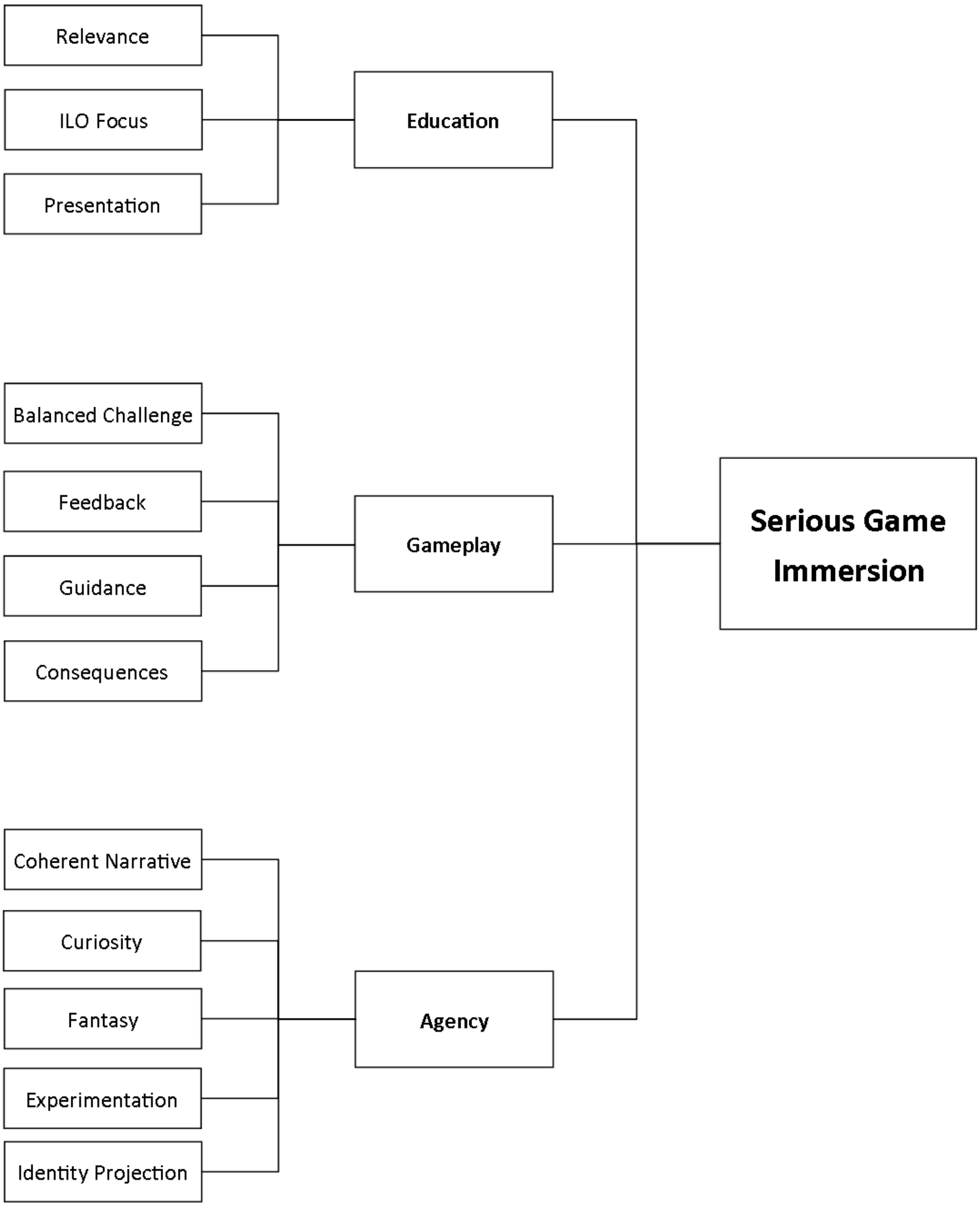


Figure 4-2: Immersive Educational Games Model

4.4 Summary

The model presented identifies the perceived core aspects of immersion in educational games, to help establish a structure to assist developers in making educational games with optimum teaching potential. By integrating and condensing aspects from a wide range of game design models, frameworks, guidelines and learning theories, its intended purpose is to provide a clearer representation of the various research fields that make games immersive in terms of education and gameplay.

Chapter 5. Model Verification Stage 1: Expert Interviews

While the model presented is derived from the literature review, the aspects included were ultimately synthesised by the researcher. Consequently, while certain aspects may be important in the researcher's opinion, they may be less so in the opinions of others working and researching in the educational games field; there could additionally be aspects and perspectives the researcher did not consider. In order to discover these potential new aspects and omissions, and adjust the model accordingly, it was decided to perform a pair of surveys collecting opinions on the factors the model presents. The purpose of these studies was to further answer the research question raised earlier:

RQ1: **What aspects affect a player's immersion in an educational game?**

Once the initial model was constructed using the literature review as a basis, a set of interviews were conducted with game experts in academia and design. The objective was to discern whether the model was truly representative of immersive aspects in educational games, in order to answer the research question.

5.1 Interview Process

Nine participants were interviewed, each one specializing in a particular field of games development or academia. While a large number of interviewees as possible would be preferable to evaluate the model, individual interviews are relatively time expensive to organize and conduct. Furthermore, according to Nielson's findings on heuristic evaluations, between 3 and 5 evaluators will be able to uncover approximately 75% of usability problems in a system; in addition, according to these findings, the amount of problems they can collectively identify becomes increasingly limited after 10 participants are used, and no amount of participants will realistically uncover all the potential problems (Rogers, et al., 2011, pp. 507-508). Taking the aspects of the model to simulate the 'heuristics' of the model (fitting because of the intention of the model to guide the design of educational games), the same principle can be applied to this study, as the aim is to uncover the problems present within the model, as is in the case of heuristic evaluation with system usability.

The experts included researchers in games for health, game theory, e-Learning and psychology, advergames, narratology, difficulty balancing, as well as game developers. The experts were selected from among academic staff and PhD candidates working at the University of Southampton and the Winchester School of Art.

The participants were requested to take part by e-mail, and interviewed individually at a place of their convenience. In each interview session, the participant was asked to read an information sheet explaining the purpose of the study, and asked to sign a consent form. The participants were then given a short questionnaire sheet, describing each of the initial aspects of the model, and asked to circle how they would rate each aspect on a four-point Likert scale (the handout, including the participant information and the questionnaire, is shown in Appendix E).

Upon completing this questionnaire, the participants were asked verbally, for each aspect not answered with 'Strongly Agree', the participants were asked why they gave the answer they did. The assumption behind this was that the questions not answered with 'Strongly Agree' (or not answered at all) would potentially indicate faults with the aspect the questions represented. Having asked about each of these aspects, the participants were then asked verbally if they had any additional comments regarding the model. The reason for this was to reveal any possible omissions in the model, or potentially better ways that the model could be structured, which may not have been covered in the responses to the questions about the individual aspects. The participants' verbal responses were recorded with an audio recording pen.

When each participant had finished answering this question, the recording pen was turned off, the participant was thanked for their participation, and their interview was concluded.

5.2 Interview Analysis

Thematic analysis was utilised to process the data obtained from the expert interviews. Thematic analysis involves the researcher identifying the common patterns (or 'themes') from a set of qualitative data. These themes are identified through interpretation of the data, based on the points the participants were attempting to convey through their perspective on the research problem (Aronson, 1994) (Braun & Clarke, 2006). The decisions on which themes are most important to the research are in turn filtered through the perspective of the researcher. The identified themes could then be used to inform the choices on improving the proposed model.

The interviews were first transcribed by the researcher, with sentences unrelated to the research questions omitted in order to ease the analysis process. The software NVivo 10 was used to assist with this thematic analysis. As there were two research questions to explore, two nodes were used to structure the collected data: the first being *Educational Game Immersion Factors*, comprising of data related to the aspects influencing educational game immersion, and *Factor Importance*, comprising of data related how important the proposed aspects are.

When starting the thematic analysis, thirteen codes were initially created; the first twelve of these were the aspects put forward by the initial model, *Relevance*, *ILO Focus*, *Presentation*, *Balanced Challenge*, *Feedback*, *Guidance*, *Consequences*, *Narrative*, *Curiosity*, *Fantasy*, *Identity Projection* and *Experimentation*, and the thirteenth code was *Other*, representing additional thoughts and aspects brought up in the interviews.

5.2.1 Lack of clarity of ILO Focus

While the idea of having a close integration between ILOs and gameplay mechanics was strongly-agreed with by 7 of the 9 experts, there were several instances during the interviews where it was asked what the term ‘ILO Focus’ meant. Indeed, Expert 8 while filling the questionnaire section out recommended changing the term entirely. This confusion was likely due to the fact that intended learning outcomes are not a term commonly used outside of educational contexts; this in turn could raise more difficulties when using the model or expanding it, as it makes it more difficult for those outside educational fields to quickly discern what exactly is meant by the abbreviation. On this basis, it was decided to change the title of the aspect from *ILO Focus* to ***Gameplay-Educational Integration***, as it uses no esoteric terms and is clearer what it refers to, while preserving the spirit of the original term (designing gameplay objectives to work with the intended learning outcomes).

5.2.2 Lack of clarity of Presentation

Another point of contention was the *Presentation* aspect, with Experts 4 and 5 confused by the term, associating it more with graphical presentation; given the specialism of both experts was in games and education, this indicates another confusion of terminology between game and educational fields, which is unhelpful when trying to intuit what exactly is meant by it in the model. In addition, Expert 8 posited that of the three issues raised by *Presentation* - teaching method, cognitive overload and sequencing – only sequencing was strongly agreed with. With regard to teaching methods, approaching it from the perspective of learning environment architectures proposed by R. C. Clarke, the expert suggested there is only one suitable architecture to construct a serious game from;

> E8: *"…realistically your game's going to be Guided, because it's not possible to make an Exploratory game, given the techniques and technologies we currently have. So in a certain sense, the question makes no sense; there is only one architecture for game-based teaching, and that's a Guided architecture*"

In terms of cognitive overload, the expert determined it to be irrelevant from a design and measurement perspective, as it entirely depends on the player, and is not something, which can be built in;

> E8: "*It's given to you by your ILO structure, it's given to you by your challenge structure. The degree to which we do or don't overload you is an **outcome**, it's a result of 'have we got the correct structure for you to be immersed in it?'*"

Expert 5 suggests a similar view with regard to sequencing being the priority, from the perspective of gameplay:

> E5: "*This also relates to feedback, something that breaks immersion is if they feel somebody is being presented with a challenge they do not feel they have the tools yet to overcome… You need to present things in the correct order so the state of your player is consistent*"

The message from both aspects is that *Presentation* in its initial form is an overabundance of considerations in this aspect, most of which are player-dependent and therefore out of the game's control. But the other message indicates that while an educational game (or any learning activity) cannot force anyone to learn, it can structure its content in such a way as to facilitate the best possible conditions for knowledge to be gained. With this in mind, in order to solve both the issue of filtering irrelevant aspects (in terms of serious game design) and avoiding confusing terminology, it was decided to reduce *Presentation* aspect to ***Sequencing***: this focuses solely on the order in which the educational content is presented, ensuring that each piece of content presented builds upon previous pieces of content encountered in the game.

5.2.3 Guidance only when Appropriate

A further recurring theme among the experts was the idea of *not forcing guidance upon the player*. Experts 7 and 8 indicated the potential for player annoyance if guidance is given unsolicited:

> E7: "*If you get to a point where you are stuck, that you want to experiment, try stuff out, I think it would be very difficult for the game to make a judgement on when that guidance should arrive, and you might end up frustrating the player when they could have figured it out, because the guidance maybe arrived too early. Or if there is guidance but it comes too late, there might be more frustration in like 'Hey, why didn't they tell me earlier?'*"

> E8: "*You can spend an awful lot of time giving guidance to a player, and the player doesn't need it, but you don't know that*"

The main points of contention seem to be guidance being presented when the player does not want it (and indeed would prefer to solve it on their own), and guidance being presented only at appropriate times, such as when a new game mechanic needs to be introduced. Because the general idea of guidance was approved of in the interviews, but with those two points of execution being the flaws, they were added as additional conditions for the Guidance aspect.

5.2.4 Consequences inherent to Games

A less frequent theme, but one deemed highly significant by the researcher, is the relevance of *Consequences*. The point put forward by Expert 7 was that consequences (actions having a tangible impact on the game world and/or progression) is an intrinsic part of every game, educational or otherwise, and so does not need to be included as part of the model.

> E7: "*I think it's a given that there's consequences in what you do – in every game, your action has some consequence to the world, whatever you do.*"

Expert 8 commented on this point when filling in the questionnaire section, declaring it is strongly agreed "by definition".

Upon reflection following these observations, it was determined that *Consequences* is strongly connected to another aspect, *Experimentation*. While *Consequences* views the player's decision-making in terms of gameplay progression (which, as Experts 7 and 8 point out, is self-evident), *Experimentation* views it in terms of engagement with the game's world and educational content; from an educational standpoint, the value of *Consequences*, especially in terms of immersion, can be said to derive from *Experimentation*.

5.2.5 Fantasy is a Distraction to Learning

The most recurring theme among the experts interviewed regarded *Fantasy*. Several experts regarded the aspect as conditional, namely that it is unsuitable for educational content where it is a more concrete representation of the real world (i.e. not an abstract concept). Expert 3 for example postulated;

> E3: "*If the content of the ILO relates to something that is a procedure, it would be reality. If it is not reality, it would be harder for the player to understand, I think.*"

Expert 3 further put forward an example of how fantasy could act as a point of confusion:

> E3: "*If the colour of a mixture is different, maybe they misunderstand that the mixture should be blue - but if in the real world it is red, it can make them misunderstand.*"

Expert 8 followed similar views; in terms of serious games, fantasy elements are completely counterproductive to the learning process, because they are meant to teach ideas which can only be applied to the real world.

> E8: "*If you've got ILOs, it's really difficult to imagine how they might be relevant to a fantasy world, because an ILO can only be relevant to the world in which we exist, the world in which we are – can only be relevant to the world (as you've said earlier on, right at the beginning), is it relevant to the player, in their life. And if it's educational, we're talking about the current world, the real world*"

Expert 6 approached the content-fantasy discordance more in terms of the fantasy elements themselves; that they can act as a distraction from the educational material that the game is trying to put across:

> E6: "*While there is an element of fantasy in most games, I almost question if you're trying to educate someone on something that is entirely real world based, whether the element of fantasy would be a distraction*"

> E6: "*If it's engaging enough an example, but is fantastical, they'll remember the fantasy version and won't have realised what lesson to take from that to the real world*"

Experts 5 and 7 took a slightly different approach, suggesting that while having an imagined scenario is important, the 'non-realistic' requirement is unneeded, as it has no bearing on immersion;

> E5: "*Having that sort of fantasy world, in the sense that it's imaginary is extremely important, but whether it's realistic or not is, I think, irrelevant*"

> E7: "*You could have different perspectives that would definitely give you the opportunity for experimentation and for new innovation, but they could still be founded in realism*"

Expert 9 similarly argues that approaching an issue from different perspectives does not guarantee better understanding:

> E7: "*I don't necessarily think that makes it more immersive, it just gives you more opportunity to experiment and to push the user out of their comfort zone, think about things slightly differently*"

> E9: "*Epistemic games require a contextually sound base. Fantasy is not precluded but new angles do not guarantee better understanding*"

The main point gleaned from these experts is that the fantasy elements can be detrimental to a player's understanding and applying of the educational material the game presents. The theme expressed instead is the idea of *context*, namely that the game world must tie in to the material being taught to engage the player and fulfil its purpose as an educational game; non-realistic-ness or 'fantasy' is at best an irrelevance, and at worst a hindrance.

With that in mind, it was decided to rework this aspect from *Fantasy* to *Scenario*; the idea being to emphasise the idea of a unique, less familiar setting (as opposed to non-realistic) which appeals to the learner and is informed by the context of the game's learning outcomes.

5.2.6 Curiosity and Identity Projection not being General Aspects

One theme drawn from the discussions on Curiosity was that it is not general to the design of all educational games.

Expert 3 felt educational games can be immersive without regarding curiosity:

> E3: "*In some games, the player doesn't have curiosity, but maybe that game can make the player immersed – such as an addition game – which doesn't have curiosity, but I still play*"

Expert 8 approached the issue from the perspectives of player motivation and the game's educational value, with designing for curiosity being pointless in both:

> E8: "*It's simply a matter of definition by saying an educational game should continually be surprising, because all you're saying is 'in an educational game, you're continually learning something you didn't know'*"

Identity Projection was similarly regarded as a niche aspect in serious game immersion, in the sense not all games support the idea yet can still be immersive:

> E1: "*It's important, but not as important as the other educational factors.*"

Experts 6 and 7 in particular regarded Identity Projection being context-dependent:

> E6: "*It is useful, but I think you can still enjoy a game even when you realise you're controlling something that isn't you – you might care about it, but do not necessarily project your identity onto it… Certain types of game players identify with different bits of games*"

> E7: "*The fantasy and identity projection are the more difficult ones – they can't be plugged in every time – there's a lot more background to that*"

Because *Curiosity* and *Identity Projection* do not generally apply to immersion in every type of game, and are not integral for engendering immersion in a serious games context, it was decided to remove them as independent aspects entirely.

5.3 Updated Model Summary

Based on the feedback from the expert interviews, the following changes were made to the proposed model:

- The name of *ILO Focus* was changed to *Gameplay-Educational Integration*
  - This was to address the concerns raised in Section 5.2.1 about *ILO Focus* being unclear to those outside the educational fields
- The name of *Presentation* was changed to *Sequencing*, and the concerns of chunking and teaching methods previously covered in the aspect were removed, to focus solely on the sequence of educational content
  - This was to address the concern raised in Section 5.2.2 about the elements of *Presentation* outside of content sequencing being irrelevant, as it is the only element which the game has any control over

- Two additional conditions were added to *Guidance*
    1. Any guidance provided automatically in-game should only be to inform the player on how to use new gameplay tools
    2. All other forms of guidance should only be given on the player's request
  - This was to discourage educational games which force guidance unnecessarily on the player, boring them, and thus being counterproductive to facilitating immersion (as expressed in Section 5.2.3)
- *Consequences* was removed entirely,
  - This was because, as raised in Section 5.2.4, the idea of player actions having an impact on the game world is inherent in the design of all games, and it was subsequently deemed that the sentiment behind the idea (seeing the outcome of the player's use of educational content) is better covered in *Experimentation*
- *Curiosity* and *Identity Projection* were also removed,
  - This was to address the concerns raised in Section 5.2.6 that these are not general aspects of immersive serious games, applying only to select cases. They were thus determined to be inappropriate for the IEGM, which is meant to cover the core aspects of immersion, applying to all educational games
- The name of *Fantasy* was changed to *Scenario*, and the focus was changed from non-realistic environments to unfamiliar environments
  - This was in response to the concern raised in Section 5.2.5 that a fantastical setting may be detrimental to learning ideas applicable to the real world. The main point of the aspect was to create an unfamiliar environment which players would be interested in exploring, and so the aspect was changed to reflect that

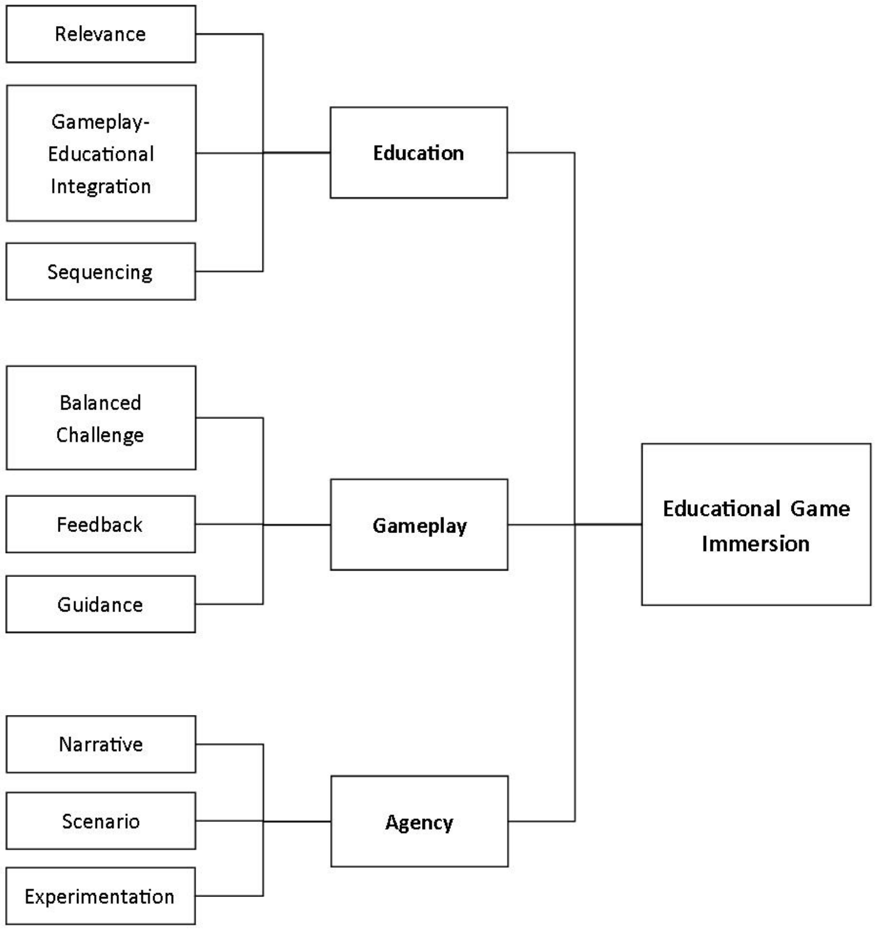


Figure 5-1: Immersive Educational Games Model (Revised, following the Expert interviews)

Chapter 6. Model Verification Stage 2: Player Survey

Once the qualitative data from the interviews was analysed, and used to alter the model as appropriate, the next study used a questionnaire with game players, based on the new version of the model. The objective of this study was to confirm the aspects of the model across a wider population of gaming experts.

6.1 Questionnaire Design

The first section of the questionnaire asks nominal questions about the participants' genders, age range, and the average range of time they spend playing computer/video games. The reason behind this was to determine what connection, if any, there was between the answers the participants give throughout the following sections, and their demographic information. In order to protect the privacy of the participants, nominal categories were used for questions about their age and average playing time. (Fink, 2003)

The following sections of the questionnaire corresponded to the categories of the updated model: *Education*, *Gameplay* and *Agency*; the purpose of this was to ensure a logical ordering of the questions asked, to avoid confusing the participants and potentially reducing the chance of the participants not completing it (Borque & Fielder, 2003). The questions in each section comprised exclusively of closed-ended questions, in order to elicit the required quantitative data to confirm the model. The questions are shown in Table 6-1.

Table 6-1: Player Survey Questionnaire Items (with Cronbach Alpha references)

Aspect	**Questions**	**Cronbach Alpha Codes**
	Education	
Relevance	I learn better if I can relate the experiences in an educational game to experiences in real life	RV_RelateReal
	I learn better if I do not find the objectives of a learning activity important to me	RV_REV_Important
Gameplay-Educational Integration	I find it distracting when an educational task does not feel related to what it is trying to teach	GE_RelateTeach
Sequencing	I learn better when each new piece of knowledge builds upon knowledge learned earlier	SQ_Builds
	Gameplay	
Balanced	I enjoy games that do not feel too easy	BC_Easy

Challenge	I enjoy games that feel too hard	BC_REV_Hard
Feedback	I find feedback on my actions in-game helps me to progress	FB_Feedback
Guidance	I find it easier to progress in a game if I receive guidance when I get stuck	GD_Stuck
	I prefer guidance to be given to me automatically, whether or not it is requested	GD_REV_Auto
	I find it frustrating when I get stuck on a task, with nothing to help me	GD_Frustration
	I prefer being told how to use new gameplay mechanics, rather than working it out on my own	GD_HowToUse
Agency		
Narrative	I prefer playing games which have clear goals to achieve	NV_ClearGoals
	I would be more likely to replay a game if the story is flexible (i.e. no two run-throughs of a game are exactly alike)	NV_StoryFlexible
	I become less interested in a game if its characters or world have inconsistencies	NV_Consistent
	I feel more engaged in a game if I care about the character I am playing as	NV_Character
	I feel I learn more when I am engaged in a role I play in a game	NV_Role
Scenario	I feel more comfortable playing games in unfamiliar settings	SC_UnfamiliarComfort
	I do not feel I learn better when ideas are presented in unfamiliar settings	SC_REV_UnfamiliarLearn
Experimentation	I feel I can understand a subject being taught to me if I can experiment with the ideas that are taught	EX_Understand
	I feel more engaged in games when using knowledge about the game's story and world to solve problems	EX_Engaged
	When I play games, I like my actions to have a noticeable impact on the game world	EX_Consequences

The questions in the *Education*, *Gameplay* and *Agency* sections were ordinal questions, based on the aspects of the model. They comprised of opinion statements, which asked the participants to rate their agreement with the statement, using a 5-point Likert scale, including 'Strongly Agree', 'Agree', 'No Opinion', 'Disagree' and 'Strongly Disagree' (Fink, 2003). Where possible, each aspect was represented by one question, but the aspects which had multiple facets were split into multiple questions to represent each facet, in order to avoid double-barrelled questions and thus collect as much precise data on the issues presented by the aspects as possible (Borque & Fielder, 2003).

6.2 Analysis Method and Sampling

When analysing and evaluating the survey data, it needed to be determined whether the participants agreed that the aspects of the model contributed to immersion, or not. For either case, the mean of the Likert question responses would be different from a selection of participants who universally selected ‘No Opinion’ for those questions.

To find out if these differences existed, the study was planned in order to use an independent means t-test to analyse the data. This test compares the means of two different groups of participants answering partaking in the same survey (Field, 2009). Each Likert item was assigned a numeric value (‘Strongly Agree’ : 5, ‘Agree’ : 4, ‘No Opinion’ : 3, ‘Disagree’ : 2 and ‘Strongly Disagree’ : 1[3]), and grouped by the model factors which they represent. The means of these grouped items was compared against their corresponding target value (representing answering ‘No Opinion’ for each item in the group), to identify whether each factor was agreed with or disagreed with.

The participants were self-selected, from those the researcher contacted in the Computer Games Society and game players from the Electronics and Computer Science department at the University of Southampton. While there is a possibility of self-selection bias, where certain people are more inclined to respond, thus skewing the results and making them less representative of the population (Whitehead, 1991), given the anonymity and relative lack of demographic information collected, it cannot be assumed; given both sample groups comprised of game players, the sample was deemed to be representative enough.

The sample taken included 16 participants; the minimum sample size calculated using the software G*Power Version 3.1.7 was 15. As the study’s purpose was finding the relationships between the model factors and educational game immersion, and was looking for broader, observable effects, a large effect size of 0.8 was chosen (Cohen, 1992). As the study is exploratory, an alpha value of 0.1 and a Power value of 0.8 were used (Cohen, 1992).

[3] With the exception of the reverse-worded questions, which had their values reversed – see Appendix E

6.3 Survey Analysis

The purpose of the questionnaire survey was to determine whether each of the aspects in the IEGM contribute to educational games immersion, according to general game players.

With this in mind, the Likert items were grouped into categories for each aspect they were representing; meaning the responses for each of the questions related to that aspect were averaged by the number of questions for that aspect, and the resultant average becoming the value for the aspect. For instance, the answer values for the statements "I learn better if I can relate the experiences in an educational game to experiences in real life" and "I learn better if I do not find the objectives of a learning activity important to me" were averaged to form the value for *Relevance* (see Table 6-1 for the full groupings of the questions according to the aspects).

The null hypothesis and alternative hypothesis being investigated were:

H_0: The mean ranking of each aspect is not different from the null value
H_1: The mean ranking of each aspect is different from the null value

Since these hypotheses require finding the differences between two different mean values, it was decided to use t-tests on each category.

Using this test, the averaged responses from the 16 participants for each aspect were compared against a predefined 'null' value. The null value was defined as the equivalent value of 'No Opinion' in the questionnaire items, which was equal to 3. If the participants' results for an aspect have no significant difference from the null value, the null hypothesis would be accepted. Since the null value can be considered a separate group of equal size, wherein all the hypothetical participants answered 'No Opinion' to each of the questions (as a means of directly comparing the participants' agreement to a neutral position), a t-test was deemed appropriate to use.

6.3.1 Normality

Because the t-test assumes that the sample distributions of the data are normal, it was first necessary to check the participants' data for normality. According to the central limit theorem, if the sample data is normally distributed, and the sample itself is large enough (generally over 30 (Field, 2009, p. 42)), then the sampling distribution will also be normally distributed; this in turn means that the sample data can represent the whole population.

To check for normality in the sample data, the Kolmogorov-Smirnov and Shapiro-Wilk tests were performed on the data (see Table 6-2). In both tests, *Relevance, Guidance* and *Narrative* had significant results ($p < 0.05$), meaning their distributions were not normal; since the t-test carries the assumption of normally distributed sample data, performing a t-test using the data may produce biased results (Field, 2009, pp. 343-345).

Table 6-2: Player Survey Normality Tests Output

	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	Df	Sig.
Relevance	.172	16	.200	.926	16	.209*
Gameplay-Educational Integration	.431	16	<.0001	.612	16	<.0001
Sequencing	.313	16	.000	.785	16	.002
Balanced Challenge	.251	16	.008	.862	16	.021
Feedback	.220	16	.038	.819	16	.005
Guidance	.188	16	.136	.954	16	.551*
Narrative	.166	16	.200	.955	16	.567*
Scenario	.250	16	.009	.829	16	.007

* Significant results ($p < 0.05$)

One potential solution to this problem is to expand the number of participants, as the central limit theorem suggests that sufficiently large samples should produce a normal distribution; however, this was not deemed practical in the study's timeframe, given the low response rate of the participants.

Alternatively, certain transformations can be used to 'correct' the data by adjusting the entire set of data values to fit a normal distribution (e.g. log transformation, square root transformation). While these can correct for issues in the data (e.g. unequal variances), their use can affect the analysis in profound ways, such as changing the hypothesis being tested (Field, 2009, pp. 154-156).

Another alternative is to use non-parametric tests to compare the data, which do not rely on the assumption of data normality as parametric tests (like t-tests) do. This is because, rather than testing the sample data directly, non-parametric tests mostly use a ranking-based system; each data value is given a rank relative to the rest of the values, and it is the resultant ranks that are compared (Field, 2009, p. 540). While non-parametric tests make fewer assumptions about the data, they can be less powerful than their parametric counterparts (meaning that they may be less likely to find an effect in the data, when it actually does exist). However, this presumption on its inferior power is only true if the data is normally distributed (Field, 2009, p. 551), which this sample data is not.

Given the difficulty of obtaining new responses in the available time, and the complexity involved in appropriately selecting a transformation method which would not violate the assumptions of the study, it was decided to use a non-parametric test.

The alternative non-parametric test selected was the Mann-Whitney test, due to its equivalent to the t-test (Field, 2009, pp. 540-542).

6.3.2 Comparing the Means of the Aspects

The rankings from the Mann-Whitney test are shown in Table 6-3. From this test, the participants' responses for each of the aspects differed significantly from each aspect's null value ($p<0.05$), with the mean ranks for each being higher than the null values. The exception is *Scenario*, which significantly differed from the null value, but had a mean rank lower than the null value. Consequently, the null hypothesis can be rejected, and the alternative hypothesis accepted.

Table 6-3: Player Survey Mann-Whitney Mean Ranks

Ranks

	Group	N	Mean Rank	Sum of Ranks
Relevance (Average)	Participants	16	20.50	328.00
	Null	16	12.50	200.00
	Total	32		
Gameplay-Educational Integration (Average)	Participants	16	23.50	376.00
	Null	16	9.50	152.00
	Total	32		
Sequencing (Average)	Participants	16	23.00	368.00
	Null	16	10.00	160.00
	Total	32		
Balanced Challenge (Average)	Participants	16	21.00	336.00
	Null	16	12.00	192.00
	Total	32		
Feedback (Average)	Participants	16	22.00	352.00
	Null	16	11.00	176.00

	Total	32		
Guidance (Average)	Participants	16	24.00	384.00
	Null	16	9.00	144.00
	Total	32		
Narrative (Average)	Participants	16	24.00	384.00
	Null	16	9.00	144.00
	Total	32		
Scenario (Average)	Participants	16	11.50	184.00
	Null	16	21.50	344.00
	Total	32		
Experimentation (Average)	Participants	16	24.00	384.00
	Null	16	9.00	144.00
	Total	32		

Table 6-4: Player Survey Mann-Whitney Significance Results

	Mann-Whitney U	Wilcoxon W	Z	Asymp. Sig. (2-tailed)	Exact Sig. [2*(1-tailed Sig.)]
Relevance (Average)	64.00	200.00	-2.78	.005	.015
Gameplay-Educational Integration (Average)	16.00	152.00	-4.86	<.0001	<.0001
Sequencing (Average)	24.00	160.00	-4.32	<.0001	<.0001
Balanced Challenge (Average)	56.00	192.00	-2.96	.003	.006
Feedback (Average)	40.00	176.00	-3.95	<.0001	.001
Guidance (Average)	8.00	144.00	-4.92	<.0001	<.0001
Narrative (Average)	8.00	144.00	-4.92	<.0001	<.0001
Scenario (Average)	48.00	184.00	-3.36	.001	.002
Experimentation (Average)	8.00	144.00	-4.92	<.0001	<.0001

6.3.3 Questionnaire Internal Consistency

To test the internal consistency of the questions themselves, a Cronbach's alpha test was performed on the 21 Likert-scale questions (see Table 6-4). Cronbach's alpha is used to estimate how likely a set of items is to provide data concerning individual preferences (Cronbach, 1951); in the context of this study, the value of alpha can help determine if the questionnaire items measure what they are meant to measure. The value of Cronbach's alpha was found to be 0.61. The table in Appendix G shows how the value of alpha changes with questionnaire item removed, but no item removed caused a great change in the value. While certain studies have suggested 0.7 is the minimum acceptable standard for internal consistency, because this particular study is exploratory, the value of 0.61 was deemed satisfactory.

Table 6-5: Player Survey Cronbach's Alpha Result

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.610	.606	21

6.4 Updated Model Summary

From the results of the Mann-Whitney test, there are significant differences between the participants' mean values for each aspect, and the target values for each aspect. Specifically, there eight of the nine factors are significantly higher than their target values, suggesting it is agreed that those aspects are contributors to educational game immersion. The exception is *Scenario*, which appears to be near-universally disagreed with. The implication for the model is that the *Scenario* aspect is not important for educational game immersion, and so it should be discarded from the overall model. However, given the importance of this aspect with regard to the literature review and the expert interviews, and the relatively small sample size of the survey, it was decided to retain the *Scenario* aspect. It was felt that unfamiliar *Scenarios* influence effective immersive designs and allow users varied and interesting situations to engage with.

Chapter 7. Model Verification Stage 3: Experimental Study

Based on literature concerning education, immersion and games, with findings confirmed by expert interviews and a questionnaire study with game players, the Immersive Educational Games Model (IEGM) was created, which identified the 9 major aspects that influence immersion in educational games. However, the model was only theoretical at the time of confirmation. In order to be of use as a means of judging the immersive qualities of educational games (thus being of benefit to game developers as well as educators seeking appropriate teaching games), the measurements of the IEGM needed to be defined more specifically, and then tested using example educational games.

As a result of this issue, the next research stage intended to answer the following research question:

RQ2: **Can the Immersive Educational Game Model be used to measure the immersive qualities of an educational game?**

To investigate this research question, a set of measurements was constructed in order to tangibly measure each aspect of the IEGM. Following this, a set of three experiments was constructed, involving having a selection of participants play an immersive educational game and a non-immersive educational game; the purpose was to measure both how immersed in the game they become, and to what degree this immersion is due to the aspects described in the model.

7.1 Creating the Instrument Metrics

In order to construct the instrument, the GQM software metrics approach was used (described in Section 3.6.2). Initially, the thinking was to use a unique measure for each 'goal', but such a quantity of measures would be difficult to test in a small number of experiments, particularly since each measurement method would require validation to demonstrate they were truly measuring the goal. For this reason, it was decided to use Likert items for each of the metrics, because they provide a quantitative way of measuring the opinions of players, and can be generally applied for any educational games. The full outcomes of the GQM process (i.e. the devised goals, questions and metrics) are shown in the Discussion, in Table 9-1.

The final outcome of this process was a set of 54 Likert items, with 6 items per IEGM aspect. To format these items in such a way as to be used by researchers and academics to gauge educational game immersive qualities, they were placed into a questionnaire including general demographic questions (age, gender and number of hours playing games weekly). The Likert items are described below (the full questionnaire can be viewed in Appendix H):

Relevance

- While playing, I learned about the subject I expected to learn about from the game
- While playing, I learned about a subject I did not expect to learn about from the game*
- I am interested in the subject the game is trying to teach
- I feel what I've learned would be useful to me in real life
- I do not feel like I learned much about the game's subject matter*
- I think this game would be valuable to people who wanted to learn about the subject matter

Gameplay-Educational Integration

- I feel the gameplay is out of place with what the game is trying to teach*
- I think the game mechanics reflect similar real-world situations
- I think the game challenges reflect similar real-world situations
- I think the subject matter is shown well through the game challenges
- I find the game to be both fun and educational
- The educational content made the game boring*

Sequencing

- I understood what kind of subject matter the game was attempting to teach
- I found knowledge from previous challenges useful in successive ones
- I felt with each game challenge, the game provided all the information I needed to complete it
- I think the order of the game challenges helped build up my skills
- There were parts of the subject matter I needed more information to understand*
- What I learned in previous challenges did not help me in successive challenges*

Balanced Challenge

- I found the game to be too easy*
- I found the game to be too difficult*
- I enjoyed playing the game
- I would not recommend this game for its gameplay*
- The game challenges felt fair
- The game felt frustrating to play*

Feedback

- I think the game's feedback helped me to understand how well I was doing
- I found the in-game feedback to be excessive*
- I found there was too little feedback in the game*
- The game let me know the important details I needed to know to progress
- I was confused throughout the game about how to play it*
- When I made mistake, I understood what I was doing wrong

Guidance

- I was able to find tutorials or hints to help me progress when I needed them
- Tutorials/hints were presented to me whether I wanted help or not*
- I found the game's guidance systems (tutorials, hints) helped me to progress
- I did not pay much attention to the tutorials*
- The game's tutorials got annoying*
- I feel the game should have let me discover more for myself*

Narrative

- I thought it was clear what I had to do to complete the game challenges
- The game's story makes sense to me
- I found myself interested in the characters
- I thought my actions in-game had a noticeable impact on the story
- I did not care about the story*
- I found the characters to be inconsistent*

Scenario

- I found the game's setting to be familiar to me*
- I found the game's setting similar to related real-world situations
- I felt the game's setting helped me to engage more with the game's subject matter
- The setting distracted from the subject matter*
- I was bored by the game environment*
- I felt involved in the role the game asked me to perform

Experimentation

- I felt able to experiment with the game mechanics introduced
- I was able to experiment with what I learned about the game's subject matter
- I felt I could experiment without permanently impacting the game environment
- I did not have much time to think about what I should do to complete the game challenges*
- The game gave me enough opportunities to retry challenges if my attempts did not work
- I was able to identify what strategies work in particular situations

* Reverse-worded questions

7.2 Experiment Descriptions

With the completion of this questionnaire, the next stage was to construct a set of experiments to determine how useful the IEGM metrics were for measuring immersive qualities.

To this end, there were three experiments decided upon, to measure the accuracy of the metrics and to determine the level of immersion in the game they are tested on. These experiments included:

- **Eye-Tracking (Level of Distraction)**
- **Tangram Tasks (Level of Immersion)**
- **Questionnaire (Level of Game Immersive Qualities)**

In Sections 7.2.1 - 7.2.3, the methodology for each of the three experiments is described; for each experiment, the purpose of each experiment is presented, as well as their variables, methods and hypotheses.

In addition, the target sample sizes are calculated for each experiment, based on the planned analytical methods used to investigate each hypothesis. The sample sizes were calculated using G*Power; a large effect size of 0.8 was chosen due to the exploratory nature surrounding the metrics and immersion measures, requiring a broader observable effect, and an alpha value of 0.05 and a Power value of 0.8 were chosen according to general practice (Cohen, 1992).

7.2.1 Experiment 1: Tangram Tasks

The first experiment was intended to investigate the 'level of immersion' experienced by the participants in a selected 'immersive' educational game.

To this end, the participants were asked to first complete a Tangram puzzle, requiring them to manipulate a set of different-shaped blocks to form a particular picture. The picture chosen for the puzzle in this experiment was a fox shape, shown in Figure 7-1. Following this, the participants were expected to play the 'immersive' educational game for 15 minutes, to guarantee that the time conditions for immersion are met – the theory being that if immersion will occur at all on a game, it will happen between 5-10 minutes of playing (Jennett, et al., 2008). After playing, the participants were presented with the same Tangram puzzle to complete again.

The measurement taken in the study is the time taken to complete the Tangram task before and after playing. If the participants completed the Tangram task more slowly after playing the game, it indicates that they experienced immersion while playing it. The reasoning behind this approach is that as a person becomes more immersed in a game, they gradually lose awareness of the real world, and the time it takes to adjust their awareness back to the real world becomes longer. Therefore, the more immersed a person gets, the longer it should take for them to complete an external task after being interrupted (Jennett, et al., 2008).

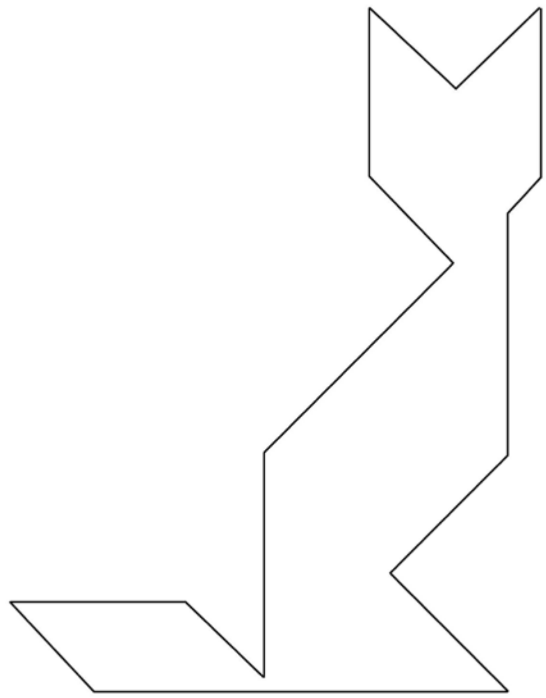


Figure 7-1: Experiment 1 Tangram Shape

It is important to note that the experiment, even in Jennett et al.'s initial conception, does not specifically require Tangram puzzles as the activity to complete (since the purpose is simply to provide an external task requiring effort, any activity that requires thought and attention will suffice). However, it was decided to use Tangram puzzles, and the same picture as used in the Jennett's study, in order to minimise complications that could unexpectedly influence the experiment outcomes; for example, finding a suitable external task that is comparable in attention requirements to Tangram puzzles would be difficult to qualify, and testing such an activity would exceed the time frame and bounds of the study.

Purpose

To measure the degree of immersion the participant experiences while playing the selected educational game.

Variables

The dependent variable is the degree of immersion the participant experiences.

The independent variable is whether the participant is being tested before playing the immersive game or after.

Method

At the start of the experiment, the participant will be instructed to complete a tangram picture from tangram pieces arranged into a square. Their time taken to complete it will be recorded. Afterwards, the tangram picture will be reset by the researcher into a square again.

After playing the educational game, the participant will be asked to complete the same tangram figure again, and their time taken to complete it will be recorded.

Hypotheses

$\mathbf{H_2}$: The time taken to complete the first tangram task will not be significantly different from the time taken to complete the second tangram task

$\mathbf{H_3}$: The time taken to complete the first tangram task will be significantly less than the time taken to complete the second tangram task

Statistical Analysis Method and Sample Size

The statistical analysis chosen for this measure was a one-tailed t-test, comparing the difference between two independent means.
The sample sizes were calculated using G*Power; a large effect size of 0.8 was chosen due to the exploratory nature surrounding the metrics and immersion measures, requiring a broader observable effect, and an alpha value of 0.05 and a Power value of 0.8 were chosen according to general practice (Cohen, 1992). The minimum sample size calculated for this measure was 21.

7.2.2 Experiment 2: Eye Tracking

The second experiment was designed to measure the 'level of distraction' the participants felt while playing a selected 'immersive' educational game. This involved investigating how their focus on the game changed as time passed, in particular as compared to a 'non-immersive' game.

To this end, the participants were asked to play the selected 'immersive' educational game for 15 minutes (again, to guarantee that the time conditions for immersion established by Jennett's study are met (Jennett, et al., 2008)), while having their eye movements tracked by an eye-tracker device. After this, the participants were asked to play a selected 'non-immersive' game for 15 minutes, with their eye movements tracked in the same way.

The measurement used to determine the participants' level of distraction were their eye movements during this period; specifically, the number of fixations over time. The theory behind this is that a person becomes less distracted as they become more immersed in a game, and become more focused on particular parts of the screen. This measurement approach was found by (Jennett, et al., 2008) to be a statistically significant indicator of immersion in entertainment games, and there was no reason to presume the case would be different for educational games, as it is the participant's immersive state that eye fixations are meant to determine.

Purpose

To measure the degree of immersion the participant experiences while playing the selected educational game.

Variables

The dependent variable is the degree of immersion the participant experiences.

The independent variable is whether the participant plays the immersive game or non-immersive game.

Method

Over the course of the experiment, the participant will play an immersive educational game and a non-immersive game.

They will play each game for 15 minutes (to guarantee that immersion occurs – the theory is that if immersion will occur at all on a game, it will happen between 5-10 minutes of playing (Jennett, et al., 2008)), after which they will be instructed to stop playing and move on the next task.

During each 15 minute period, the participant's eye movements will be tracked and recorded.

Hypotheses

H_0: The degree of immersion for the more immersive educational game will be no different to the degree of immersion for the less immersive educational game

H_1: The degree of immersion for the more immersive educational game will be higher than the degree of immersion for the less immersive educational game

Statistical Analysis Method and Sample Size

The statistical analysis chosen for this measure was a one-tailed t-test, comparing the difference between two independent means.

The sample sizes were calculated using G*Power; a large effect size of 0.8 was chosen due to the exploratory nature surrounding the metrics and immersion measures, requiring a broader observable effect, and an alpha value of 0.05 and a Power value of 0.8 were chosen according to general practice (Cohen, 1992). The minimum sample size calculated for this measure was 21.

7.2.3 Experiment 3: Questionnaire

The third experiment was to investigate the ‘level of immersive qualities’ the participants perceived in a selected ‘immersive’ educational game. The aim was to investigate to what degree each of the theorised immersive qualities in the IEGM instrument were present in an established immersive educational game, and thus how well the IEGM can be used to measure immersive qualities.

To accomplish this, the participants were asked to play the selected ‘immersive’ educational game for 15 minutes, and after their play session they were to answer a questionnaire about the immersive qualities of the game they played. The questions comprised of the set of metrics in the created IEGM instrument (the full questionnaire can be viewed in Appendix H).

While this approach is based on a similar questionnaire study of Jennett et al.’s (see Section 3.5.3), it differs in one major aspect: their questions were based on the feelings of immersion the participants experienced, whereas the questions from the IEGM metrics are based on the immersive qualities the participants observed in the game itself. This change was deemed necessary because this particular experiment is concerned with the game’s immersive qualities, rather than the immersive state of the participant (which Experiments 1 and 2 explore, from quantitative perspectives). However, the metric questions both relate to immersion and explore the subjective experiences of the participants while playing, as

Jennett's original questions do. On this basis, it was deemed an appropriate change that would not negatively impact the study.

Purpose

To measure the degree to which the IEGM immersion aspects are present in the game

Variables

The dependent variable is the degree to which the IEGM immersion aspects are present in the game

The independent variable is the participants' questionnaire responses

Method

To finish the experiment, the participant will be instructed to complete a questionnaire about the educational game they played. This questionnaire will be derived from the metrics created from the IEGM.

Hypotheses

H_4: The ranking of immersive qualities for the educational game will be significantly different to its predicted ranking

H_5: The ranking of immersive qualities for the educational game will not be significantly different to its predicted ranking

Statistical Analysis Method and Sample Size

The statistical analysis chosen for this measure was a one-tailed t-test, comparing the difference between two independent means.
The sample sizes were calculated using G*Power; a large effect size of 0.8 was chosen due to the exploratory nature surrounding the metrics and immersion measures, requiring a broader observable effect, and an alpha value of 0.05 and a Power value of 0.8 were chosen according to general practice (Cohen, 1992). The minimum sample size calculated for this measure was 21.

7.3 Method

Since the methodology of the proposed experiments can operate together (all three utilising the immersive and non-immersive game, and Experiments 1 and 2 utilising the same questionnaire), it was decided to attempt to integrate them into a single procedure, where the participants partake in all three experiments in one sitting. The advantage of this approach was expediency on the participants' part (as they will not have to be scheduled for three separate experiments needlessly), and because the methods of each do not interfere with each other, and each require the participants playing an educational game.

7.3.1 Participants

The participants for the experiment were selected from the University of Southampton and the Winchester School of Art. Since the IEGM instrument questionnaire was written in such a way as to be understood without knowledge of specialist concepts or terminology, and only Experiment 2 required special participant considerations (as certain eye conditions can bias the results from the eye-tracker), the only selection criteria for the participants were that they were aged 18 or over, and did not have astigmatism and cirrhosis.

7.3.2 Materials

The equipment used across these studies included:

- 1 x Tobii X120 Eye Tracker
- 1 x Desktop/Laptop Computer with Monitor
- 1 x Timer
- 1 x Tangram Puzzle set
- 1 x Printed Tangram Fox outline

The equipment was set up in a secluded, quiet area, in order to minimise external distractions as much as possible.

The experiments also required an immersive educational game, and a non-immersive educational game, for the participants to play. The following subsections describe the selection of the immersive educational game, and the creation of the non-immersive educational game.

7.3.2.1 Selection of the Immersive Educational Game

For the immersive educational games to be used in the study, two games were selected as potential candidates: *Re-Mission* and *Immune Attack*. *Re-Mission* is a game designed to instruct about the overall workings of cancer and its management through a 3D action game. It is aimed towards adolescents and young adults, especially those suffering from common forms of cancer (Kato, et al., 2008). *Immune Attack* is a similar 3D action game designed to instruct about the workings of the immune system, and is designed towards students at the high school/university level (Kelly, et al., 2007).

Both games satisfied most aspects of the IEGM, and were thus considered to be 'immersive', under the criteria proposed in the experiment. The ways they satisfied the immersive criteria are described as follows:

Relevance

The introductory levels of *Re-Mission* and *Immune Attack* introduce their broad subject areas (cancer and immune system mechanics respectively) in simple-to-understand terms, emphasising the problems associated with them, and how these problems are biologically addressed. In *Re-Mission*, this involves the identifying and treatment of cancer cells and bacteria; in *Immune Attack*, this involves how the immune system mechanics (e.g. macrophages) work, and why.

These problem areas were presented in simple terms, explained both through short descriptions, and practical in-game demonstrations. This in turn made the subject areas accessible to those that were unfamiliar with the subject, and able to understand it in a real-world context, thus fulfilling the *Relevance* requirements.

Gameplay-Educational Integration

The gameplay derives from the subject matter, with the behaviours of cells reflecting realistic patterns. While the delivery systems in the games (sentient nanobots for *Re-Mission*, and a miniaturised craft for *Immune Attack*) do not reflect real world situations, the principles being applied by the gameplay are reflected (e.g. through the use of a chemotherapy weapon to destroy cancer cells, and an antibiotic weapon to destroy bacteria). The gameplay mechanics in turn apply these principles in an engaging way, via third-person shooting for *Re-Mission* and third-person exploration and mechanic emulation (e.g. guiding macrophages) in *Immune Attack.*

Sequencing

In both games, the introductory levels start with the basic, overarching concepts, gradually introducing more specific elements framed in the context of earlier, broader elements (e.g. Hodgekin's Lymphoma as an extension from cancer for *Re-Mission,* and going from identifying macrophages to guiding macrophages in *Immune Attack*). Each challenge is based on a clearly specified objective, which build upon the educational content introduced at earlier points.

Balanced Challenge

The games assume nothing about the player's prior gaming abilities, introducing the basic controls first and providing simple movement tasks initially, before moving on to tasks that require increasingly quick thinking and skill.

Since the experiment made no assumptions about the participant's gaming experience, only the introductory levels were used, thus in theory none of the challenges proved to be too difficult. Under similar reasoning, a more experienced player is likely be able to get through these introductory parts quicker, reaching the more challenging parts more easily, precluding the chance of players finding the game too easy.

While the cutscenes both games offer may halt the progress of the player for certain periods (thus potentially getting experienced players stuck needlessly at the less challenging parts), it was considered acceptable given the cutscenes' importance towards the *Narrative* category of the IEGM. In addition, it was determined that in a 16 minute play session, cutscenes would occupy no more than 3-4 minutes, based on playing the opening missions of *Immune Attack* and *Re:Mission*.

Feedback

The games provide encouragement for performing desired actions (towards completing challenges), and cautionary messages for performing less desirable actions (counterproductive to completing challenges). Additionally, in both games, a helper instructs you on the objective of each challenge before it starts. In this way, the games allow the player to know how well they are performing, and whether or not they need to reconsider their strategies.

Guidance

Hints and encouragements are presented after a certain period of inaction or too much movement in the wrong direction. Similarly, if the wrong method is used to complete particular objectives, the helper explains that the method is ineffective, and hints at what needs to be done to progress. Guiding arrows are also presented to indicate the direction the player should move in to find the next challenge, or to complete the current one.

Narrative

The main characters and their overarching role are established at the start, giving the players a clear overarching purpose for their in-game actions. *Immune Attack*'s characters largely acting in supporting roles, without much major involvement in the narrative beyond providing objectives and general gameplay advice. *Re-Mission* meanwhile seems to have more developed characters with its impulsive protagonist Roxxi, and her older advisor Smitty, with the story focused on Roxxi's developing abilities and understanding of her duties as a cancer-fighting nanobot.

Immune Attack's narrative approach seems more emergent, because the actions of the player take more precedence, while *Re-Mission's* more embedded approach places more emphasis on the pre-defined story and character developments. But while their approaches may be different, the fact that both games have a consistent structure and characters indicates that both games can contribute well to immersion, according to the IEGM.

Scenario

The settings for the games were developed as a reasonably realistic depiction of the human body from a microscopic level, and presented from an unfamiliar perspective; for instance, *Immune Attack* frames the immune system's workings in the context of a microscopic craft using various technologies to bolster the immune system, while *Re-Mission* frames fighting cancer in terms of controlling a nanobot to directly shoot at cancer cells. In this way, the games present their subject areas from an unfamiliar role, and a less familiar environment, which would promote engagement with the subject matter.

Experimentation

The player is free to explore throughout the game environments as they wish – while they are encouraged to follow particular paths, there is freedom in when to follow them, and how. In

Re-Mission, the game affords opportunities to make mistakes in its action gameplay – ammunition pickups are frequent and refill when needed, and if the wrong cells are targeted or destroyed accidentally (e.g. as part of the blast radius of a chemo rocket), the game does not penalise, simply provides warnings. In *Immune Attack*, while several of the opening challenges are timed, failure simply resets that particular challenge without additional penalty; similarly, the time limits seem lenient enough in these challenges such that mistakes made (e.g. incorrect targeting of bacteria) do not guarantee failure. At the same time, in both games there are visible consequences to mistakes made, most notably in the propagation of cancer cells and bacteria that are not destroyed quickly enough.

While both games were found to be suitable candidates for the experiment, *Re-Mission* was chosen for the piloting due to issues in running *Immune Attack* at the same time as the Tobii X120 eye-tracker (rendering it unplayable past the opening menu).

7.3.2.2 Creation of the Non-Immersive Educational Game

For the non-immersive educational game, a game was needed to exemplify the opposite qualities integral to immersion. Assuming *Re-Mission* follows the IEGM's guidelines well, the non-immersive educational game should follow its guidelines as little as possible. Such a game would abide by the following guidelines:

- The non-immersive game should not be structured to teach its subject through its gameplay. Specifically:
    - The subject should not be shown in a way that will be useful to the players in real-world applications
    - The gameplay and educational elements should not relate to each other
    - The different pieces of information the game conveys should not build on or link to each other
- The gameplay must not be conscientious of the player's skill level:
    - The challenges should not be balanced, not being designed towards predicted or actual player ability
    - There should not be clear feedback on how the player is progressing
    - There should be either no available guidance/tutorial mechanisms, or an intrusive number of them
- A meaningful context for the gameplay should be absent, in the following ways:
    - There should not be a clear story, or clear goals, to incentivise the player

- There should not be a unique presentation of the subject matter
- There should be no opportunities to practice or experiment (ideally, challenges would give the players only one chance, regardless of experience level)

To this end, a game fitting this description was created by the researcher using the Game Maker Studio software (a screenshot of which is shown in Figure 7-2: Non-Immersive Game Screenshot. The game was themed around simple mathematics, where the player is expected to complete a presented equation; they are presented with a number of blocks containing numbers and mathematical symbols, and are expected to drag these blocks into slots with the mouse, such that the completed equation equals the presented answer block. When the player has completed the equation, the background turns briefly blue, the level is reset, and a different equation is generated.

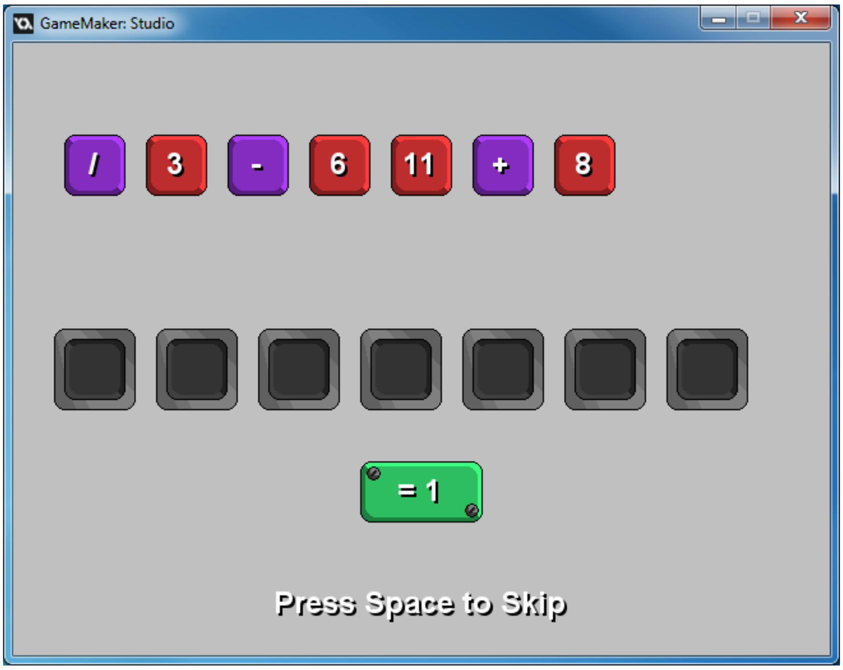


Figure 7-2: Non-Immersive Game Screenshot

The equations that the game asks the player to solve are randomly generated by an algorithm. The number of slots/blocks varies between 3 and 7 and does not take the equations previously solved by the player into account. There is no explanation given, in-context or otherwise, for why the player should be doing this, no indication (until solved) how close the player is to solving the equation, and no more in-game help given than can be helped (the 'Press Space to Skip' option was given in the event of an unexpected glitch in the random-equation algorithm). In short, it contradicts the IEGM aspects as far as possible while still remaining playable, and so was thought to be a sufficient counterpoint to *Re:Mission*.

7.3.3 Procedure

The procedure for this approach is described as follows:

Arrival

Each participant was told ahead of time when their experiment time was scheduled. When they arrived, they were invited into the experiment room and offered a drink of water. If the participant turned up earlier than their scheduled time, they were greeted, asked to wait and offered water.

Briefing

When the participants arrived, they were greeted and given a copy of the participant information, which includes the goal of the experiment and an overview of what they were expected to do at each stage of the experiment. They were told that they would be called into the experiment room one at a time to partake in the experiment, which would last approximately 45 minutes per person. They were asked to wait, and were provided with a selection of refreshments.

When the participant entered the experiment room, the participant information was read verbally by the researcher, after which the participant was asked if they have any questions.

When all questions have been answered, and the participant agreed to take part, they were asked to sign a consent form, and the researcher emphasised that they could withdraw from the experiment at any time, and after the experiment they could request for their data not to be used, and if this was the case, it was complied with.

Tangram Task (before playing the game)

The researcher presented a tangram square to the participant (set up before their arrival), and instructed them to construct a fox shape from the tangram shapes. They were then instructed on the ways in which the tangram shapes can be moved, and asked if they had understood or had any questions. When the participant's questions were answered and they confirmed verbally they understood the task, the researcher instructed them to start, and the timer was activated. When the participant completed the task, the researcher stopped the timer, recorded the time taken to complete the task, and informed the participant that they would be moving on to the next task.

Eye-Tracking (Immersive Game)

The researcher set up the game, starting at the first level, with the eye-tracking monitor in place. Once everything was in place and operating, the researcher asked the participant to play through the game. They were informed that their eye movements would be tracked for the duration of the play time, and they were asked to stop after 15 minutes. Once they understood the instructions, they were asked to play the game and the timer was started. While the participant was playing, the researcher reset the tangram picture back to a square.

After 15 minutes, the researcher stopped the timer and instructed the participant to stop playing.

Tangram Task (after playing the game)

After the participant stopped playing, the researcher directed them back to the tangram square, and asked them to create a fox shape again, as before. When this was clearly understood, the timer was activated, and the participant commenced the task. When the participant completed the task, the researcher stopped the timer, recorded the time taken to complete the task, and informed the participant that they would be moving on to the next task.

Metrics Questionnaire

The participant was then given the immersive qualities questionnaire and asked to complete it. They were informed that they would be able to ask any questions if they had any problems filling out the questionnaire, and then asked if they have understood this. When the participant verbally confirmed that they did, the researcher instructed them to start. While the participant was filling out the questionnaire, the researcher reset the tangram blocks into their original square formation.

When the participant completed the questionnaire, the researcher filed the questionnaire and informed the participant that he/she would be moving onto the next task.

Eye-Tracking (Less-immersive game)

The researcher set up the less-immersive game, starting at the first level, with the eye-tracking monitor in place. Once everything was in place and operating, the researcher asked the participant to play through the game. They were informed that their eye movements would be tracked for the duration of the play time, and they were asked to stop after 15 minutes. Once they understood the instructions, they were asked to play the game and the timer was started.

After 15 minutes, the researcher stopped the timer and instructed the participant to stop playing.

Experiment Conclusion

The researcher thanked the participant for their time. The researcher then reminded the participant that if they do not wish their data to be used, they could contact the researcher to request removal of their data with the researcher complying with this.

7.4 Piloting of the Experiments

To uncover the potential flaws of the proposed study, the combined experiment was piloted on 4 researchers in the field of games from the University of Southampton who fall under the proposed sample criteria for the study.

From the feedback collected from the participants, and the experience of the researcher, the following issues were uncovered from the piloting:

1. **Controls**

   The most consistent issue among the participants was the lack of clarity with regard to *Re:Mission*'s controls. While the key mappings of each in-game function are on the game's options menu (accessible from the title menu, or the pause menu), the key mappings are not made clear within the game itself; instead, the game refers to the keys by their function (e.g. 'the Turn-Left key'), which caused confusion among the participants, rendering them unable to continue until the controls were clarified by the researcher. A similar issue was found in the non-immersive game, where the controls were not elaborated upon at all.

   While this arguably indicates the importance of control clarity with regard to immersion (in turn an aspect of the 'Guidance' aspect), it is unhelpful in judging the other immersive aspects if the participants are incapable of continuing playing in the time limit without outside assistance. For this reason, it was decided that a print-out of the *Re:Mission* key mappings should be made available to the participants in the final experiment.

2. **Clarity of Game Instructions**

   In the non-immersive game, another issue across all the participants was the lack of clarity of what to do, not understanding the layout of the level until it was explained by the researcher.

   This was initially an intentional aspect of the game, to provide as little Guidance as possible. But since this aspect also renders the player incapable of continuing, it is equally unhelpful in measuring the other immersive aspects. Because of this, a brief header instruction “Drag the blocks into the slots to complete the equation” was included for the main experiment at the start of the game, in order to explain what they have to do and explain the basic controls.

3. **Tangram Picture**

   A potential issue discovered with the tangram task was that for each participant, the second tangram task took noticeably less time than the first. This is likely due to initial inexperience with tangram puzzles (the two participants who were more familiar with tangram puzzles were able to solve them far quicker), coupled with the fact the tangram figure used in each task was identical, thus the participants had more experience solving it. With this in mind, one preferable direction in the final experiment could be to give a preliminary tangram task, to ensure that they understand how they work, thus avoid a potential bias in the experiment times. However, due to potential complications in selecting a new Tangram picture for testing, including time constraints, it was decided against constructing a practice Tangram picture. It was decided that asking whether the participants tried Tangram puzzles before, and explaining if not, would be sufficient.

4. **Game/Level Choice**

   While the tutorial sections of *Re:Mission* were chosen because they establish the central story, it was pointed out that the missions spend a great deal of the 15 minutes playing time simply establishing the basic controls, as opposed to engaging with the subject matter. This is certainly the case in the first tutorial, which focuses primarily on basic movement. However, starting from the main missions would not be ideal, since the story and context of the game are outlined in the tutorials, which are important aspects of the Narrative aspect of the IEGM. Given that there was confusion regarding the controls from all of the participants, it is likely that providing the controls print-out (as elaborated in the “Control” point) would help to reduce this time spent on working out which keys to press.

5. **Headphones/Speakers option**

   One participant turned down the option of earphones (which were the only means of receiving sound), and became confused with the instructions in *Re:Mission*, as they were displayed in subtitles at the bottom of the screen, which the participant was unaware of (possibly because the eye-tracker partially obscured the bottom portion of the screen). For this reason, it was decided to provide headphones for the final experiment, in order for the participants to receive the necessary information to proceed.

6. **Eye-tracking Signal Issues**

   Due to issues with lighting, which differed between participants, it was difficult to calibrate the eye-tracker to each participant. Some more time needed to be allotted in the final experiment to accommodate for this (the process did not however take longer than 5 minutes). A more severe problem was found with participant movements during the eye-tracking segments; the eye-tracker would lose track of the participants’ movements if they moved their head too close to the screen. To attempt to circumvent this, the participants were instructed before they started playing to try not to lean forward while playing. In addition, the eye-tracker was moved to the underside of the monitor to reduce the risk of lost eye movement data from the participant accidentally leaning in too close.

7. **Timing Issues**

   Some issues arose in the timekeeping of the eye-tracking and tangram sections of the experiment. One was due to the researcher sometimes being too late starting the timer on the eye-tracking section, resulting in inaccurate measurements. Another was due to the participants interpreting differently when to start the tangram task, likely due to them being unaware they were being timed, or that they would need to start when the researcher told them to. For this reason, the researcher needs to be more diligent with the timing of the eye-tracking sections, and should preface the tangram task instructions with a clear indication of when they should start (e.g. 'When I tell you to start…')

8. **Questionnaire Issues**
   When filling out the questionnaire, one participant was unclear as to what 'challenge' was referring to in the context of the *Re:Mission* game, unaware it referred to the gameplay tasks. In the final questionnaire, a clarification of this term will be included: "A task within the game that requires gameplay skill to complete/overcome".

7.5 Analysis and Discussion

In this section, the findings for each experiment are presented, in addition to discussions of the findings, including why they may have occurred, and their overall significance to answering RQ2.

Table 7-1: Stage 3 Demographic Descriptive Statistics

	N	Range	Minimum	Maximum	Mean		Std. Deviation	Variance
					Statistic	Std. Error		
Gender	30	1	0	1	0.33	0.088	0.479	0.23
Age	30	4	1	5	2.47	0.184	1.008	1.016
HoursPlaying PerWeek	30	4	0	4	0.43	0.19	1.04	1.082

7.5.1 Experiment 1: Tangram Tasks

The purpose of Experiment 1 was to determine whether a game considered immersive, according to the IEGM, can immerse players in the time span of 15 minutes (in which immersion is said to occur, if it is going to happen at all). The measurement chosen to assess this was the difference between the times taken to complete a tangram task, before and after playing the immersive game.

The null hypothesis and alternative hypothesis being investigated were:

H_0: The time taken to complete the first tangram task will not be significantly different from the time taken to complete the second tangram task

H_3: The time taken to complete the first tangram task will be significantly less than the time taken to complete the second tangram task

The results for the sample are shown in Table 7-2; 'Before_Game_Time' represents the tangram completion time before playing the educational game, while 'After_Game_Time' represents the completion time after playing the game, both measured in seconds. For the sample, the mean of After_Game_Time is greater than the mean of Before_Game_Time (shown in). Indeed, in the sample there were only four participants (2, 16, 17 and 24) whose value of After_Game_Time was less than their value of Before_Game_Time.

Since the investigated hypotheses require finding the differences between two different mean values, it was decided to use t-tests on each category.

Table 7-2: Experiment 1 Completion Time Comparisons

Participant	Before_Game_Time (seconds)	After_Game_Time (seconds)
1	39.11	34.9
2	72.8	84.65
3	27.2	20.18
4	81.53	35.9
5	20.88	13.69
6	393.1	20.29
7	19.41	16.77
8	573.31	33.59
9	32.2	24.34
10	46.07	41.44
11	30.43	14.74
12	96.8	20.71
13	31.81	17.36
14	357.46	45.62
15	392.17	73.45
16	24.13	26.07
17	65.79	73.17
18	88.35	22.52
19	241.86	21.9
20	45.4	22.53
21	65.49	23.88
22	61.1	24.65
23	62.37	35.58
24	69.71	109.82
25	25.99	16.79
26	100.9	30.51
27	33.54	20.66
28	86.04	24.59
29	43.39	28.42
30	100.55	27.05

Table 7-3: Experiment 1 Descriptive Statistics

Descriptive Statistics

	N	Range	Minimum	Maximum	Mean	Std. Deviation	Variance
Before_Game_Time	30	553.90	19.41	573.31	110.9630	137.25414	18838.700
After_Game_Time	30	96.13	13.69	109.82	33.5257	22.68091	514.424
Valid N (listwise)	30						

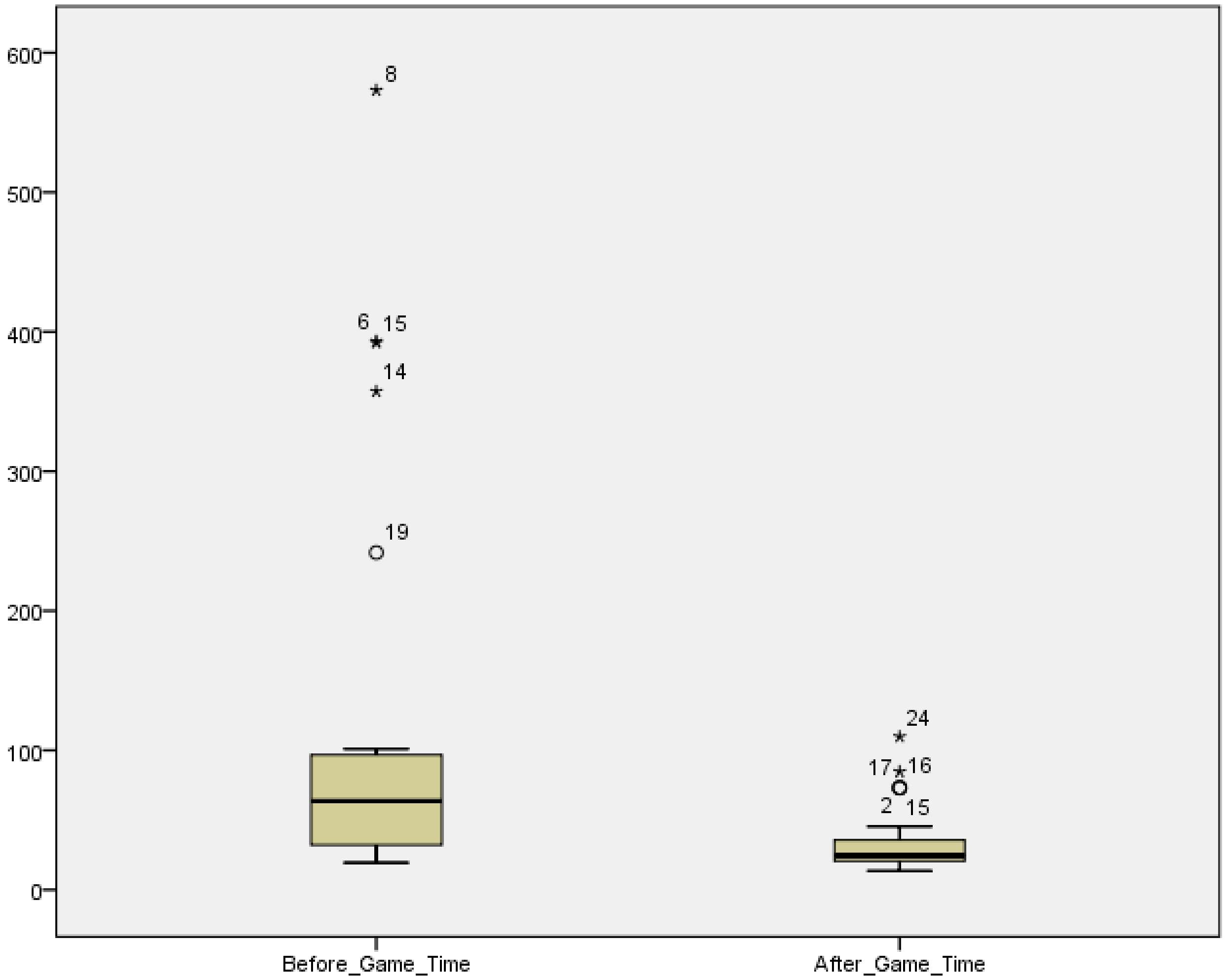


Figure 7-3: Experiment 1 Descriptive Box Plots

7.5.1.1 Normality

Because the t-test assumes that the sample distributions of the data are normal, it was first necessary to check the participants' data for normality. To accomplish this, the Kolmogorov-Smirnov and Shapiro-Wilk tests were performed on the data (see Table 7-4). In both tests, the means of Before_Game_Time and After_Game_Time had insignificant results ($p < 0.05$), meaning their distributions were normal, and therefore performing a t-test using the data would be appropriate to use. (Field, 2009, pp. 343-345).

Table 7-4: Experiment 1 Normality Tests Output

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Before_Game_Time	.363	30	<0.001	.640	30	<0.001
After_Game_Time	.258	30	<0.001	.723	30	<0.001

a. Lilliefors Significance Correction

7.5.1.2 Paired t-test

The paired-samples t-test was used to compare the means of Before_Game_Time and After_Game_Time, the output of which is shown in Table 8-2. From this test, the participants' times differed significantly from each other ($p<0.05$), and consequently the null hypothesis can be rejected. The t-value $t(29)=3.134$, in the context of the experiment, indicates that the participants took less time to complete the tangram task after playing the game than it did to complete the tangram task before playing. As a result, the alternative hypothesis can be rejected.

Table 7-5: Experiment 1 Paired t-test Output

Paired Samples Test

		Paired Differences					T	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	Before_Game_Time – After_Game_Time	77.43733	135.33903	24.70941	26.90091	127.97376	3.134	29	.004

7.5.1.3 Correlation with Immersion Scores

One explanation of the mean completion times before playing being significantly larger than the mean completion times after playing is that the player will have already learned how to complete it the first time, and so will not need to take as long the second time to complete it. But while the participants may have learned how to solve the puzzle, if they have become immersed from playing the game, their time taken to solve the puzzle again will be closer to the original time than if they are not immersed (Jennett, et al., 2008).

To determine whether this occurred, the participant's differences in completion time between the two tangram tasks (TimeDifference) were correlated with the total score given in the IEGM questionnaire (TotalImmersionScore); this TotalImmersionScore was the sum of the participant's responses to the Likert items on the questionnaire, according to the numerical values given in Appendix H. The rationale was that the greater the value of TotalImmersionScore (which would indicate the participant was immersed), the smaller the value of TimeDifference, since the distraction effects of coming out of immersion would slow down the participant's completion time.

To decide whether a parametric correlation test would be appropriate, TimeDifference and TotalImmersionScore were first tested for normality using the Kolmogorov-Smirnov and Shapiro-Wilk tests. TotalImmersionScore produced a significant result ($p<0.05$) for the Shapiro-Wilk test, indicating that its distribution was not normal, thus unsuitable for a parametric test.

Table 7-6: Experiment 1 Correlation Normality Tests

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	Df	Sig.
TimeDifference	.337	30	<0.001	.663	30	<0.001
TotalImmersionScore	.165	30	.036	.962	30	.338

a. Lilliefors Significance Correction

The Spearman's Rank-Order correlation was chosen to analyse the data, as it was non-parametric, making no assumptions on the data's normality. There appeared to be a weak, positive correlation (r_s=0.173), as can be seen in the scattergraph in Figure 7-4. However, the result was insignificant ($p>0.05$); as such, the little correlation between immersion and tangram times for this sample are likely to have occurred due to chance. Thus the null hypothesis $\mathbf{H_2}$ can still be accepted.

Table 7-7: Experiment 1 Spearman's Rho Output

Correlations

		TimeDifference	TotalImmersionScore
TimeDifference	Correlation Coefficient		.173
	Sig. (1-tailed)		.180
	N		30
TotalImmersionScore	Correlation Coefficient	.173	
	Sig. (1-tailed)	.180	
	N	30	

7.5.1.4 Regression with Immersion Scores

To investigate more closely into the relationship between the TimeDifference and TotalImmersionScore variables (i.e. whether the total score of the IEGM questionnaire can predict the difference in completion times of the Tangram task), linear regression was conducted on the two variables.

Regression is a technique to predict the value of an 'outcome' variable from one or more 'predictor' variables, using a model to determine a line of best fit for the available data (Field, 2009, p. 198). The first stage is establishing this line of best fit, going through, or near, as many data points as possible; this is achieved through the method of least squares, which finds the line that produces the least 'sum of squared differences' (SS), which – as the name implies – is the sum of positive and negative differences of the data points from their predicted place on the line (with the differences squared to circumvent the problem of positive and negative differences cancelling each other out) (Field, 2009, pp. 200-201). The second stage involves assessing how well this line fits the actual data; this can be assessed by a value called 'R^2'. This value is calculated by taking the model sum of squares (SS_M, the difference between the sum of squared differences from the mean and the sum of squared differences from the predicted model line), and dividing it by the sum of squared differences from the mean (SS_T) (Field, 2009, p. 202). In other words;

$$R^2 = SS_M/SS_T$$

The value of R (obtained from the square root of R^2) indicates the correlation between the two variables, while R^2 identifies how much variance the predictor variable can account for in the outcome variable (Field, 2009, pp. 206-207).

One further important stage is calculating the F-ratio of the model, which measures how much the model improves the prediction of the outcome variable, when compared to the model's inaccuracy; a more representative model will have a greater value (at least 1, if not more) (Field, 2009, pp. 203-204).

Table 7-8: Experiment 1 Regression Model Summary (R and R Square values)

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.052[a]	.003	-.033	138.83371

a. Predictors: (Constant), TotalImmersionScore

b. Dependent Variable: TimeDifference

Table 7-9: Experiment 1 Regression ANOVA (F-Ratio and Significance Values)

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1439.320	1	1439.320	.075	.787[b]
	Residual	539694.374	28	19274.799		
	Total	541133.694	29			

a. Dependent Variable: TimeDifference

b. Predictors: (Constant), TotalImmersionScore

Table 7-10: Experiment 1 Regression Model Coefficients

Coefficients[a]

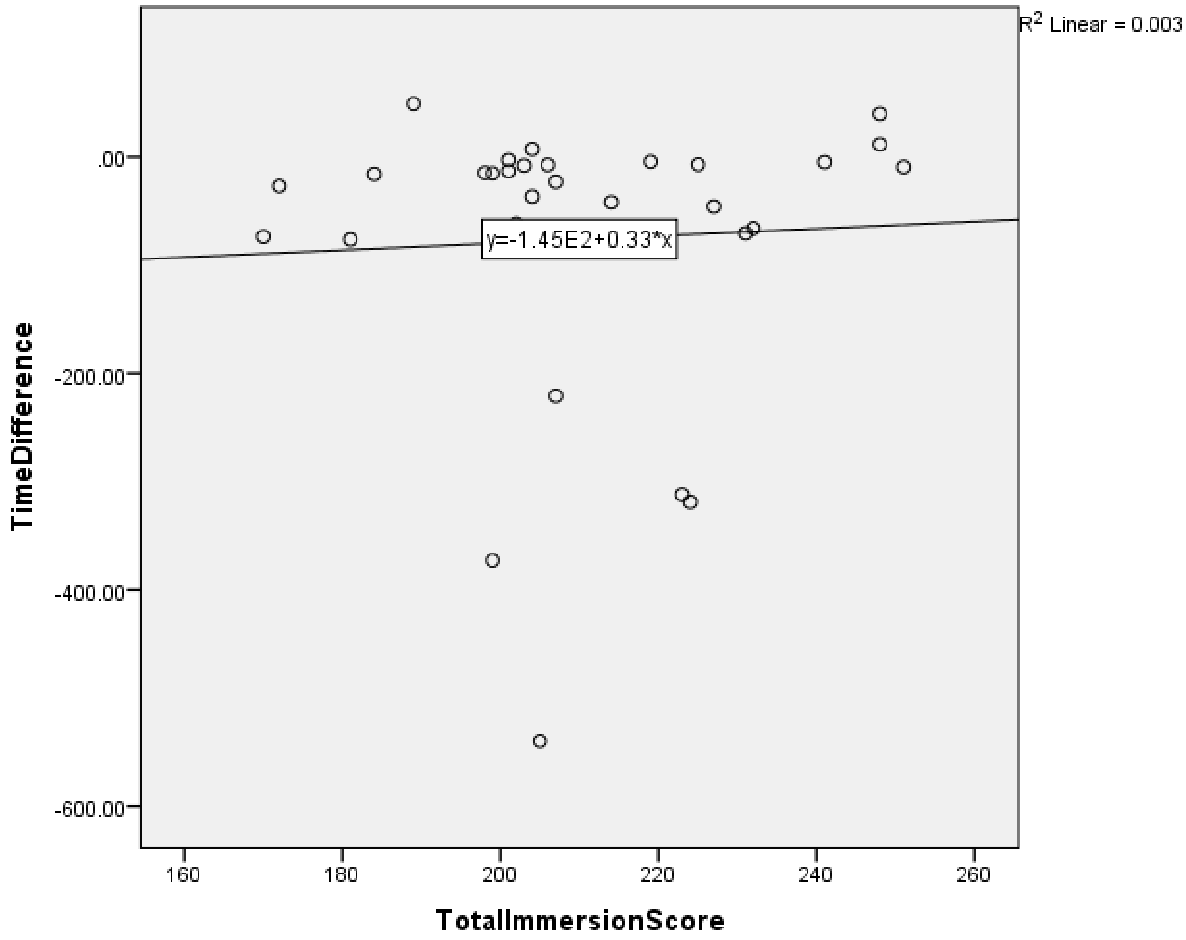
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-145.342	255.459		-.569	.574
	TotalImmersionScore	.330	1.208	.052	.273	.787

a. Dependent Variable: TimeDifference

From the linear regression results conducted (with TimeDifference as the outcome variable and TotalImmersionScore as the predictor variable), it appears from the value of R^2 that TotalImmersionScore can account for only 0.3% of the variation in TimeDifference (see Table 7-8), thus suggesting that the IEGM questionnaire findings are not a significant predictor of the level of immersion (defined in this experiment as the Tangram task relative completion times). The value of the F-ratio as 0.075 substantiates this (see Table 7-9), as it is less than 1, and therefore according to Field not a model that produces accurate predictions; in addition, the F-ratio is not significant ($p > 0.05$).

In short, the linear regression suggests that the IEGM questionnaire cannot predict the differences in the Tangram task completion times (and thus, the level of immersion after playing an immersive game). Given this information, the null hypothesis $\mathbf{H_2}$ can still be accepted.

Figure 7-4: Scatter Graph of TotalImmersionScore against TimeDifference, with Regression Line



7.5.1.5 Discussion

The results of Experiment 1 seem to indicate there is no connection between a person's immersion level, and level of distractedness in performing an external task. However, the study by (Jennett, et al., 2008), which formed the basis of the experiment, had contrary findings, indicating a significant correlation to the participant's immersive level and the size of time differences between the two tangram tasks.

One likely explanation is with the questionnaire given in the experiment. The questions were derived from the aspects of the confirmed IEGM, which ask about particular immersive qualities in the game itself. The assumption behind its usage was that if the participants would rate these qualities highly, it would indicate the participants were immersed while playing the game. On the other hand, Jennett et al's questionnaire asked about the feelings of immersion in the participants themselves (e.g. “I did not feel any emotional attachment to the game”). These questions may provide a more direct indicator of the participant’s immersion, without the need for assumption of the game being immersive.

A further biasing factor could be how varied the completion times are before playing the game; the standard deviation of Before_Game_Time being over five times greater than After_Game_Time. This could be an indicator of the problem described in the piloting feedback (see Section 7.4), that the participants required different lengths of time to understand how to complete the tangram puzzles, which would not be required for the second task, as the puzzle was exactly the same (hence the variance for After_Game_Time would be considerably less). While it was decided not to run a preliminary, different tangram task as it was not thought the variances between each variable would be that different, this was likely a mistaken assumption.

Another possible issue could have been with the experimental conditions; as the experiment took place simultaneously with Experiment 2 (involving having their eyes tracked), and the participants were observed throughout the play session by the researcher. The combination of these may have restricted how immersed they could become in the game.

If this study were to be repeated, the inclusion of questions that address the immersed feelings of the participants would be useful in gaining a more complete idea of how immersed a participant is, and thus potentially provide more significant results.

7.5.2 Experiment 2: Eye Tracking

The purpose of Experiment 2 was to investigate the degree of immersion a player can experience when playing the chosen immersive educational game, Re-Mission, in order to help establish whether the qualities identified in the IEGM do lead to player immersion.

The null and alternate hypotheses being investigated are:

$\mathbf{H_0}$: The degree of immersion for the more immersive educational game will be no different to the degree of immersion for the less immersive educational game

$\mathbf{H_1}$: The degree of immersion for the more immersive educational game will be higher than the degree of immersion for the less immersive educational game

To measure this, it was decided to compare the correlations of the eye movements of the participants from playing the chosen ‘immersive’ and ‘non-immersive’ games. Specifically, it was decided to use a method employed by a similar experiment utilising eye-tracking (Jennett, et al., 2008); for both the ‘immersive’ and ‘non-immersive’ conditions, the mean number of eye fixations per second (MeanFixations_I and MeanFixations_NI respectively) were correlated with the progression of time in the experiment (Time). The theory is that the Time-MeanFixations_I correlation should be less than the Time-MeanFixations_NI correlation, since this would suggest the participants have fewer eye movements as time goes on in the ‘immersive’ condition, in turn indicating that they are immersed.

7.5.2.1 Normality

To decide whether a parametric correlation test would be appropriate, MeanFixations_I and MeanFixations_NI were first tested for normality using the Kolmogorov-Smirnov and Shapiro-Wilk tests. Both MeanFixations_I and MeanFixations_NI had significant results ($p<0.05$) for both tests, indicating that their distributions were not normal, therefore were not suitable for using in a parametric test.

Table 7-11: Experiment 2 Normality Tests

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
MeanFixations_NI	.052	900	<.001	.962	900	<.001
MeanFixations_I	.061	900	<.001	.985	900	<.001

a. Lilliefors Significance Correction

7.5.2.2 Non-Parametric Correlation Test

The Spearman’s Rank-Order correlation was chosen to analyse the data, as it was non-parametric, making no assumptions on the data’s normality. MeanFixations_I showed a negative, very weak correlation (r_s=-0.054), but was an insignificant result, whereas MeanFixations_NI showed a significant, but very weak negative correlation (r_s=-0.077).

Table 7-12: Experiment 2 Correlation for the 'Immersive' Game

Correlations

			Time	MeanFixations_I
Spearman's rho	Time	Correlation Coefficient		-.054
		Sig. (1-tailed)		.052
		N		900
	MeanFixations_I	Correlation Coefficient	-.054	
		Sig. (1-tailed)	.052	
		N	900	

Table 7-13: Experiment 2 Correlation for the 'Non-Immersive' Game

Correlations

			Time	MeanFixations_NI
Spearman's rho	Time	Correlation Coefficient		-.077*
		Sig. (1-tailed)		.010
		N		900
	MeanFixations_NI	Correlation Coefficient	-.077*	
		Sig. (1-tailed)	.010	
		N	900	

*. Correlation is significant at the 0.05 level (1-tailed).

Since there would appear to be a difference between the correlations, the null hypothesis $\mathbf{H_0}$ can be rejected. However, the correlation for the ‘immersive’ game was higher than the correlation for the ‘non-immersive’ game, indicating that the participants were more immersed in the ‘non-immersive’ game than the ‘immersive’ game. Consequently, the hypothesis $\mathbf{H_1}$ can also be rejected.

7.5.2.3 Discussion

The rejection of both hypotheses implies that the participants became more immersed in the ‘non-immersive’ game than the ‘immersive’ game.

One explanation for this finding could be due to a certain methodological issue; for the ‘immersive’ game, the participants were given a printed list of controls, which would have necessitated looking away from the screen to reference them, and their fixations consequently not being identified by the eye tracker. This could have affected the participants more in the early stages of the game, where the tutorial requires them to use the controls for the first time, meaning that the participants would have fewer fixations registered during the early game.

Alternatively, the finding could indicate important qualities for immersion that the IEGM overlooks, or does not emphasise enough. For instance, the controls for the 'immersive' game (as mentioned earlier, requiring a printed list) were far more complex than the drag-and-drop controls of the 'non-immersive' game. As a result, it could be that it was easier for the participants to achieve the state of 'flow' in the 'non-immersive' game, as they did not have to spend more time understanding the basic controls, and could focus on the task at hand. Likewise, the more frequent progress feedback and clear, singular goal of the 'non-immersive' game could have more easily facilitated immersion, as clarity of progress and goals is an important aspect of enjoyable experiences (Malone, 1982).

7.5.3 Experiment 3: Questionnaire

The purpose of Experiment 3 was to determine whether the given immersive game (*Immune Attack*) supports and contains the immersive aspects described by the IEGM, according to the people who played the game. It was decided to measure this using a questionnaire with 5-point Likert scale questions describing the points of each aspect.

To accomplish this, the Likert items were grouped into categories for each aspect they were representing (as in Chapter 6); meaning the responses for each of the questions related to that aspect were summed, and the resultant sum becoming the value for the aspect. The full groupings of the questions according to the aspects are shown in Appendix J.

The null hypothesis and alternative hypothesis being investigated were:

H_0: The ranking of immersive qualities for the educational game will not be significantly greater than their null values

H_5: The ranking of immersive qualities for the educational game will be significantly greater than their null values

Since these hypotheses require finding the differences between two different mean values, it was decided to use t-tests on each category.

7.5.3.1 Normality

To see if the response data satisfied the normality assumption for t-tests, the Kolmogorov-Smirnov and Shapiro-Wilk tests were performed on the data (see Table 7-14). In the Shapiro-Wilk test, all of the aspects produced significant results ($p<0.05$), while in the Kolmogorov-Smirnov test, the aspects of *Gameplay-Educational Integration*, *Sequencing*, *Feedback* and *Narrative* each produced significant results; this indicates that none of the aspects' distributions were normal, and so performing t-tests on the data would produce biased results. (Field, 2009, pp. 343-345). Consequently, it was decided to use an equivalent non-parametric test, as it would not assume a normal distribution.

Table 7-14: Experiment 3 Normality Tests Output

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Relevance	0.167	30	0.033	0.954	30	0.214
GE_Integration	0.126	30	0.200	0.943	30	0.112
Sequencing	0.126	30	0.200	0.943	30	0.112
BalancedChallenge	0.171	30	0.026	0.946	30	0.135
Feedback	0.125	30	0.200	0.977	30	0.732
Guidance	0.191	30	0.007	0.951	30	0.175
Narrative	0.137	30	0.158	0.946	30	0.130
Scenario	0.176	30	0.019	0.935	30	0.068
Experimentation	0.214	30	0.001	0.939	30	0.085

a. Lilliefors Significance Correction

7.5.3.2 Non-Parametric Means Test

The Mann-Whitney test was selected as an alternative, due to its equivalent to the t-test (Field, 2009, pp. 540-542).

Using this test, the responses from the 30 participants for each aspect were compared against a null value. The null value was calculated from the sum of 'No Opinion' for each of the Likert values representing a particular aspect. Since 'No Opinion' on the Likert scale has the value 3, and each aspect is represented by 6 questions, the null value was calculated as 3*6 = 18. Thus in terms of the experiment, if the participants' results for an aspect are significantly less than or equal to the null value, the null hypothesis would be accepted.

The ranking output of the Mann-Whitney test is shown in Table 7-15, and the significance statistics are shown in Table 7-16. According to these results, the participants' responses for each of the aspects differed significantly from each aspect's null value ($p<0.05$), with the mean ranks for each being higher than the null value. Consequently, the null hypothesis can be rejected, and the alternative hypothesis accepted.

However, since multiple comparisons are being conducted, there is the danger of Type I errors accumulating, thus increasing the risk of finding a significant result where there is none. One common way of dealing with this is to perform a Bonferroni adjustment/correction, where the alpha value of the experiment is divided by the number of tests conducted, thus providing a more stringent alpha value, and so reduce the risk of a false positive result (Hair, 2010) (Pallant, 2010). Since 9 tests were performed in this analysis, the Bonferroni adjustment was applied to the original alpha value of 0.05, producing a new alpha value of 0.006 (4 s.f.). Since the significance values for each factor (Table 7-16) was less than 0.001 – thus satisfying the newly-adjusted alpha value – the conclusions remain unchanged.

Table 7-15: Experiment 3 Mann-Whitney Mean Ranks

Ranks

	Group	N	Mean Rank	Sum of Ranks
Relevance	Participants	30	44.50	1335.00
	Null	30	16.50	495.00
	Total	60		
GE_Integration	Participants	30	45.50	1365.00
	Null	30	15.50	465.00
	Total	60		
Sequencing	Participants	30	45.50	1365.00
	Null	30	15.50	465.00
	Total	60		
BalancedChallenge	Participants	30	45.50	1365.00
	Null	30	15.50	465.00
	Total	60		
Feedback	Participants	30	44.50	1335.00
	Null	30	16.50	495.00
	Total	60		
Guidance	Participants	30	42.50	1275.00
	Null	30	18.50	555.00
	Total	60		
Narrative	Participants	30	42.50	1275.00
	Null	30	18.50	555.00
	Total	60		
Scenario	Participants	30	42.50	1275.00
	Null	30	18.50	555.00
	Total	60		
Experimentation	Participants	30	45.00	1350.00
	Null	30	16.00	480.00
	Total	60		

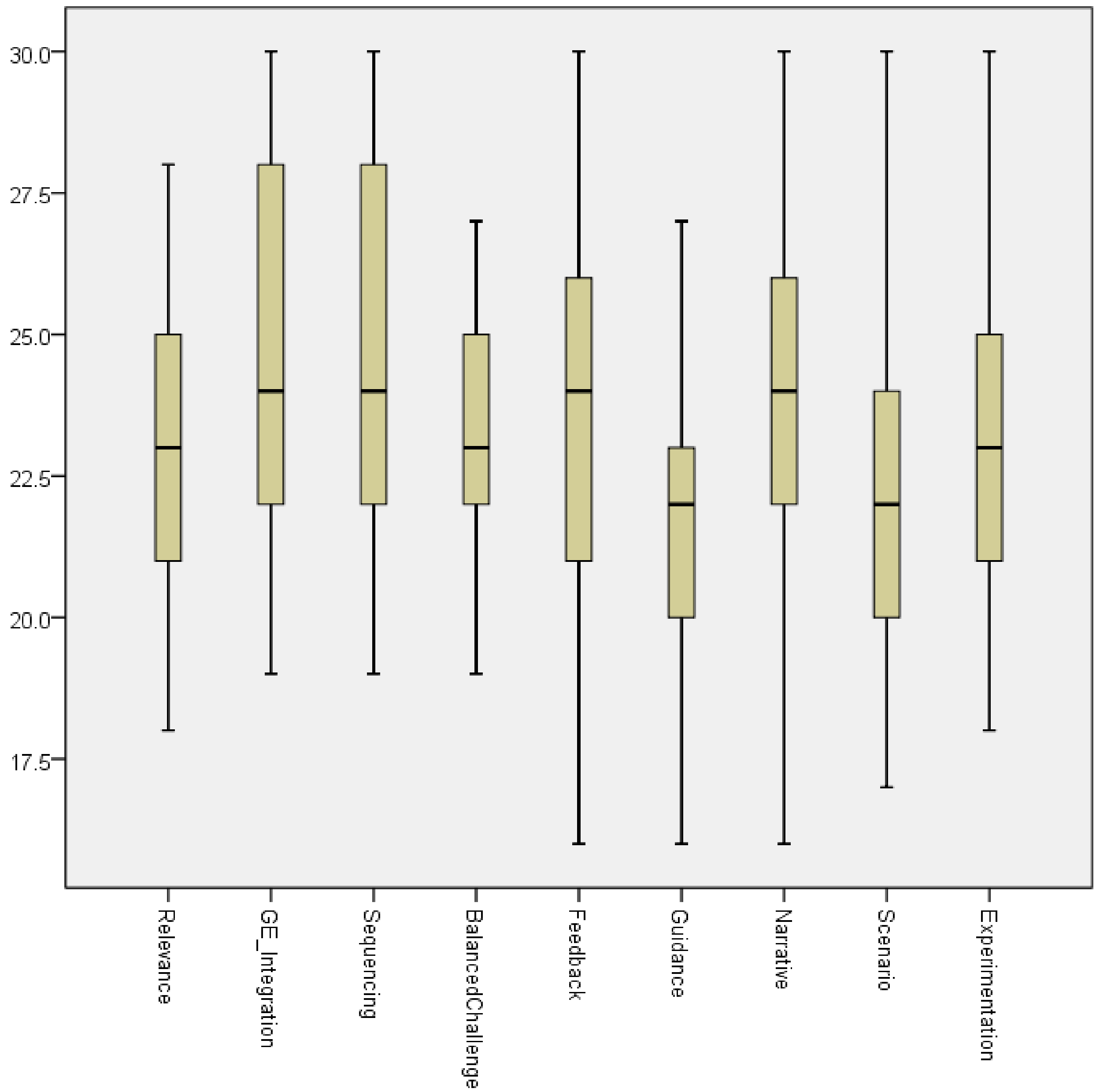


Figure 7-5: Experiment 3 Descriptive Box Plots

Table 7-16: Experiment 3 Mann-Whitney Significance Statistics

	Mann-Whitney U	Wilcoxon W	Z	Asymp. Sig. (2-tailed)
Relevance	30.00	495.00	-6.752	<0.001
GE_Integration	0.00	465.00	-7.117	<0.001
Sequencing	0.00	465.00	-7.117	<0.001
BalancedChallenge	0.00	465.00	-7.124	<0.001
Feedback	30.00	495.00	-6.645	<0.001
Guidance	90.00	555.00	-5.796	<0.001
Narrative	90.00	555.00	-5.696	<0.001
Scenario	90.00	555.00	-5.700	<0.001
Experimentation	15.00	480.00	-6.937	<0.001

7.5.3.3 Questionnaire Internal Consistency

To test the internal consistency of the questions themselves, a Cronbach's alpha test was performed on the 54 Likert-scale questions (see Table 7-17). The value of Cronbach's alpha was found to be 0.927, which indicates a good internal consistency in the questionnaire for the collected sample, and thus that the response data gathered is reliable.

Table 7-17: Experiment 3 Cronbach's Alpha Output

Reliability Statistics

Aspect	Cronbach's Alpha	N of Items
All Questions	**.927**	**54**

7.5.3.4 Discussion

The findings from Experiment 3 do seem to indicate that the IEGM metrics can accurately identify the immersive qualities in an immersive educational game; the high internal consistency of the questionnaire would also indicate that it is a reliable measure. As a result, the experiment does seem to support the assertion of RQ2 that the IEGM can be practically used to measure immersive qualities in a game.

However, there are limitations on the experiment which may leave its findings to be challenged. One such limitation is that the IEGM metrics were only tested on a game judged to be 'immersive'; it is not known from the study how the participants' responses would differ if the questionnaire was about a 'non-immersive' game.

7.6 Summary

In order to investigate whether the IEGM can be used to practically measure immersive qualities in educational games, a set of subjective metrics was created, in the form of 54 Likert items, with 6 items representing each of the 9 aspects of the IEGM. To test these metrics, three experiments were created. The first was to investigate whether the metrics could be used to accurately identify an immersive educational game as 'immersive' for each of the IEGM aspects. The second and third were to determine how immersed the players were in the immersive game, thus substantiating the results of the first experiment. Since the experiments did not interfere with each other, and in order to be more expedient, the experiments were planned to be carried out in conjunction with each other. For use in the experiments, an immersive game (*Re:Mission*) was selected, and a deliberately non-immersive mathematics game was created.

The three experiments were piloted on 4 University of Southampton researchers in the games field. Based on their feedback, the following alterations were made to the experiments:

- A print-out of the *Re:Mission* key mappings was created for the participants to refer to
- The instruction "Drag the blocks into the slots to complete the equation" was added to the non-immersive game to ensure the players were able to continue playing unassisted
- Headphones would be provided for the final experiment
- To help prevent eye-tracking detection issues, the participants would be instructed before they start playing to try not to lean forward while playing, and the eye-tracker will be moved from the front of the monitor to the underside
- A clarification of the term 'challenge' would be included in the questionnaire, defined as "A task within the game that requires gameplay skill to complete/overcome"

The paired-samples t-test and Spearman's correlation for the Tangram experiment indicates no correlation between immersion and tangram times for this sample. Similarly, the Spearman's correlations for the 'immersive' and 'non-immersive' games in Experiment 2 indicates that the participants experienced less immersion in the ‘immersive’ game than the ‘non-immersive’ game.

The findings of the Mann-Whitney test in the Questionnaire experiment show that the mean ranks for each item on the questionnaire were significantly higher than the null value, indicating that the ‘immersive’ game was deemed to be immersive according to each component of the IEGM. However, the results of the factor analysis (grouping each of the nine components in one category) indicated that the subcategories of *Education*, *Gameplay* and *Agency* were not needed.

Chapter 8. Model Verification Stage 4: Player Questionnaire

The non-parametric test results from Experiment 3 (Section 7.5.3) indicate that the IEGM questionnaire can be used to measure immersive qualities in serious games. However, to completely answer RQ2, there are two other issues which should be addressed to help ensure that the IEGM is as clear a measurement tool as possible. The first issue involves whether the groupings in the original model are appropriate; namely, are the 3 subcategories of *Education, Gameplay* and *Agency* appropriate divisions of each of the 9 aspects? It may be that, from the questionnaire results, the scores for *Relevance* correlate highly with the scores of *Balanced Challenge*, which means dividing them by category is less useful to developers and assessors who wish to understand the links between the aspects. The second issue is the weightings of each of the aspects; some aspects may contribute more to the overall immersion score than others, which is important for a measurement tool as an indicator of which aspects to prioritise.

To address these two issues, it was decided to conduct factor analysis on the questionnaire responses. Factor analysis is a statistical method used to reduce a large number of variables into groups (factors), each representing an underlying variable of what is being measured, and also to describe how the variables relate to each other and what is being measured. For this reason, it was deemed appropriate to the above issues, as it provides information about the factor groupings, as well as how far each contributes to their groupings. For this experiment, factor analysis was used on the 9 current groupings the IEGM, in order to compare the way it groups the questions with the current groupings under the IEGM factors. In this way, it can be seen how closely the IEGM factors reflect the experiences of the sample.

8.1 Method

Since the data gathered for this experiment are the responses to the IEGM immersive qualities questionnaire – an expansion of Experiment 3 with more participants, in other words – the method remains similar to that described in Section 7.3.

8.1.1 Participants and Data Collation

The participants for the experiment were selected from the University of Southampton and the Winchester School of Art, and additional UK-based contacts of the researcher. Since the IEGM instrument questionnaire was written in such a way as to be understood without knowledge of specialist concepts or terminology, the only selection criteria for the participants were that they were aged 18 or over.

Because this matched the selection criteria of Experiment 3, it was deemed appropriate to collate the sample 30 from Experiment 3 with the 60 of this stage (for a total of 90), in order to provide a sufficient number of participants to perform a reliable factor analysis on; the sample size was selected on the basis of the 'common rule' regarding factor analysis being at least 10-15 participants per factor (Field, 2009, p. 647). While Experiment 3's sampling process also precluded those with astigmatism and cirrhosis (in order to avoid issues with the eye-tracker in the concurrent Experiment 1), it was not believed that removing those constraints for Stage 4 would negatively impact the collated sample.

8.1.2 Procedure

The procedure for this approach is described as follows:

Arrival

Each participant was told ahead of time when their experiment time was scheduled. When they arrived, they were invited into the experiment room and offered a drink of water. If the participant turned up earlier than their scheduled time, they were greeted, asked to wait and offered water.

Briefing

When the participants arrived in the experiment room, they were greeted and given a copy of the participant information, which includes the goal of the experiment and an overview of what they were expected to do at each stage of the experiment. After reading the participant information, the participant was asked if they have any questions.

When all questions were answered, and the participant agreed to take part, they were asked to check the box on a consent form (no signature was required as no sensitive data – i.e. eye movements – was being collected), and the researcher emphasised that they could withdraw from the experiment at any time, and after the experiment they could request for their data not be used, and their request would be complied with.

Playing the Game

The researcher set up the game, starting at the first level, and asked the participant to play through the game for 15 minutes. Once they understood the instructions, they were asked to play the game and the timer was started. After 15 minutes, the researcher stopped the timer and instructed the participant to stop playing.

Metrics Questionnaire

The participant was then given the immersive qualities questionnaire and asked to complete it. They were informed that they would be able to ask any questions if they had any problems filling out the questionnaire, and then asked if they have understood this. When the participant verbally confirmed that they did, the researcher instructed them to start. While the participant was filling out the questionnaire, the researcher reset the tangram blocks into their original square formation.

When the participant completed the questionnaire, the researcher filed the questionnaire and informed the participant that he/she would be moving onto the next task.

Experiment Conclusion

The researcher thanked the participant for their time. The researcher then reminded the participant that if they do not wish their data to be used, they could contact the researcher to request removal of their data with the researcher complying with this.

8.2 Analysis and Discussion

In this section, the analysis of Stage 4's collected data is presented. This includes a redone mean-comparison test from Experiment 3 with the questionnaire results from the additional 60 participants of Stage 4, and a factor analysis of the questionnaire results for all 90 participants. The section concludes by discussing the findings and their implications for the IEGM's refinement.

In the analysis, four responses were excluded due to missing responses, making the total number of participants 86.

Table 8-1: Stage 4 Demographic Descriptive Statistics

	N	Range	Minimum	Maximum	Mean		Std. Deviation	Variance
					Statistic	Std. Error		
Gender	86	1	0	1	.38	.053	.489	.239
Age	86	5	0	5	2.22	.140	1.296	1.680
HoursPlaying PerWeek	86	4	0	4	.88	.138	1.278	1.633

8.2.1 Normality

To see if the response data satisfied the normality assumption for t-tests, the Kolmogorov-Smirnov and Shapiro-Wilk tests were performed on the data (see Table 7-14). In both tests, the aspects of *Sequencing*, *Feedback* and *Experimentation* produced significant results ($p<0.05$), indicating that their distributions were not normal, and so performing t-tests on the data would produce biased results. (Field, 2009, pp. 343-345). Consequently, it was decided to use an equivalent non-parametric test, as it would not assume a normal distribution.

Table 8-2: Stage 4 Normality Tests Output

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Relevance	.148	86	<0.001	.960	86	.009
GE_Integration	.147	86	<0.001	.966	86	.022
Sequencing	.120	86	.004	.978	86	.143
BalancedChallenge	.145	86	<0.001	.943	86	.001
Feedback	.106	86	.018	.980	86	.213
Guidance	.142	86	<0.001	.968	86	.032
Narrative	.153	86	<0.001	.950	86	.002
Scenario	.159	86	<0.001	.943	86	.001
Experimentation	.119	86	.004	.984	86	.374

a. Lilliefors Significance Correction

8.2.2 Non-Parametric Means Test

The Mann-Whitney test was selected as an alternative, due to its equivalent to the t-test (Field, 2009, pp. 540-542).

Using this test, the responses from the 86 participants for each aspect were compared against a null value. The null value was calculated from the sum of 'No Opinion' for each of the Likert values representing a particular aspect. Since 'No Opinion' on the Likert scale has the value 3, and each aspect is represented by 6 questions, the null value was calculated as 3*6 = 18. Thus in terms of the experiment, if the participants' results for an aspect are significantly less than or equal to the null value, the null hypothesis would be accepted.

The ranking output of the Mann-Whitney test is shown in Table 7-15, and the significance statistics are shown in Table 7-16. According to these results, the participants' responses for each of the aspects differed significantly from each aspect's null value ($p<0.05$), with the mean ranks for each being higher than the null value. Consequently, the null hypothesis can be rejected, and the alternative hypothesis accepted.

However, as with Experiment 3, since multiple comparisons are being conducted, there is the danger of Type I errors accumulating, thus increasing the risk of finding a significant result where there is none (Hair, 2010) (Pallant, 2010). Since 9 tests were performed in this analysis, the Bonferroni adjustment was applied to the original alpha value of 0.05, producing a new alpha value of 0.006 (4 s.f.). Since the significance values for each factor (Table 8-4) was less than 0.001 – thus satisfying the newly-adjusted alpha value – the conclusions remain unchanged.

Table 8-3: Stage 4 Mann-Whitney Mean Ranks

Ranks

	Group	N	Mean Rank	Sum of Ranks
Relevance	Participants	86	123.00	10578.00
	Null	86	50.00	4300.00
	Total	172		
GE_Integration	Participants	86	119.00	10234.00
	Null	86	54.00	4644.00
	Total	172		
Sequencing	Participants	86	128.00	11008.00
	Null	86	45.00	3870.00
	Total	172		
BalancedChallenge	Participants	86	118.00	10148.00
	Null	86	55.00	4730.00
	Total	172		
Feedback	Participants	86	122.00	10492.00
	Null	86	51.00	4386.00
	Total	172		
Guidance	Participants	86	108.00	9288.00
	Null	86	65.00	5590.00
	Total	172		
Narrative	Participants	86	117.50	10105.00
	Null	86	55.50	4773.00
	Total	172		
Scenario	Participants	86	120.00	10320.00
	Null	86	53.00	4558.00
	Total	172		
Experimentation	Participants	86	119.50	10277.00
	Null	86	53.50	4601.00
	Total	172		

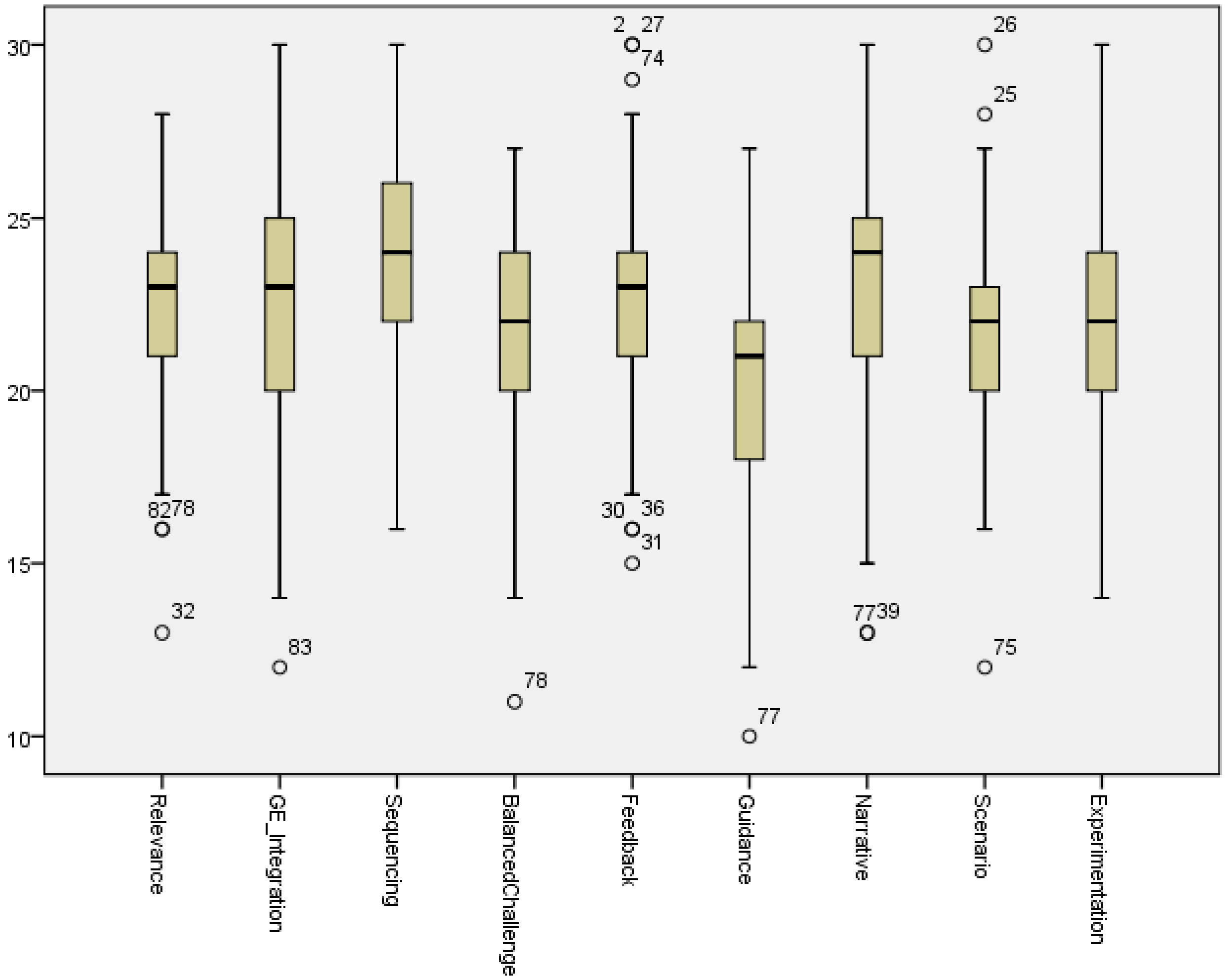


Figure 8-1: Stage 4 Descriptive Box Plot

Table 8-4: Stage 4 Mann-Whitney Significance Statistics

	Mann-Whitney U	Wilcoxon W	Z	Asymp. Sig. (2-tailed)
Relevance	559.000	4300.000	-10.427	<0.001
GE_Integration	903.000	4644.000	-9.230	<0.001
Sequencing	129.000	3870.000	-11.725	<0.001
BalancedChallenge	989.000	4730.000	-8.900	<0.001
Feedback	645.000	4386.000	-10.031	<0.001
Guidance	1849.000	5590.000	-6.178	<0.001
Narrative	1032.000	4773.000	-8.782	<0.001

Scenario	817.000	4558.000	-9.479	<0.001
Experimentation	860.000	4601.000	-9.451	<0.001

8.2.3 Questionnaire Internal Consistency

To test the internal consistency of the questions themselves, a Cronbach's alpha test was performed on the 54 Likert-scale questions (see Table 7-17). The value of Cronbach's alpha was found to be 0.928, which indicates a good internal consistency in the questionnaire for the collected sample, and thus that the response data gathered is reliable.

Table 8-5: Stage 4 Cronbach's Alpha Output

Reliability Statistics

Aspect	Cronbach's Alpha	N of Items
All Questions	.928	54

8.2.4 Factor Analysis

To find the groupings of the IEGM aspects, based on the questionnaire data, two approaches can be employed; Factor analysis, and Principal Component Analysis.

Both PCA and factor analysis are concerned with reducing a given number of variable combinations into a smaller number, while representing the variance in the patterns of their correlations (Pallant, 2010). While both approaches can produce similar results, they each utilise different techniques to estimate the factors. On the one hand, PCA simply reduces the original variables into a smaller set of linear combinations, retaining all the variance of the original variables. On the other hand, factor analysis uses a mathematical model to estimate the resultant factors, retaining only the shared variance between the original variables. (Pallant, 2010, p. 182) As such, PCA only establishes what groupings (components) exist in the data, and which items contribute to each (acting essentially as a 'summary' of the data (Pallant, 2010, p. 182)), whereas factor analysis can estimate the underlying factors, albeit requiring stricter assumptions (Field, 2009, p. 638).

Original Factor Groupings: Method

The PCA method was chosen to find the groupings, since this study simply intends to find out what groupings exist, and whether the original assumption of three overarching components (*Education*, *Gameplay* and *Agency*) is supported, which PCA can provide without the added complications that arise from the calculations necessary for factor analysis (Field, 2009, p. 638). While it can produce different results to factor analysis techniques, this generally occurs only when PCA is used on fewer than 20 items, with low communalities (<0.4) (Field, 2009, p. 638); while the former is true for the data in this study, having only 9 items, the latter is not (see Table 8-6), and thus PCA was thought a suitable method to be used.

Table 8-6: Stage 4 PCA Communalities Table

Communalities

	Initial	Extraction
Relevance	1.000	.679
GE_Integration	1.000	.753
Sequencing	1.000	.727
BalancedChallenge	1.000	.707
Feedback	1.000	.778
Guidance	1.000	.549
Narrative	1.000	.675
Scenario	1.000	.627
Experimentation	1.000	.667

Extraction Method: Principal Component Analysis.

Additionally, it was decided to use the varimax orthogonal rotation, in order to ensure that the way the items are grouped is easier to interpret (Field, 2009, p. 644). Orthogonal rotation was chosen because it keeps the factors independent of each other (Field, 2009, p. 642), and the point of the IEGM factors is that they are independent, in order that they only measure one aspect of immersion.

Original Factor Groupings: Results

The PCA was performed with varimax orthogonal rotation on the 9 original groups, with the sample size of 90. The Kaiser-Meyer-Olkin measure was used to verify the sampling adequacy for the analysis, producing a result of KMO = 0.907 (which (Field, 2009, p. 659) describes as a 'superb' value, thus confidently indicating an adequate sample size).

Furthermore, Bartlett's test of sphericity produced a significant value $\chi^2(36) = 413.921$ for $p < 0.001$, indicating that the inter-item correlations were appropriately large for PCA.

Table 8-7: Stage 4 KMO Measure and Bartlett's Test

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy		.905
Bartlett's Test of Sphericity	Approx. Chi-Square	405.609
	Df	36
	Sig.	<.001

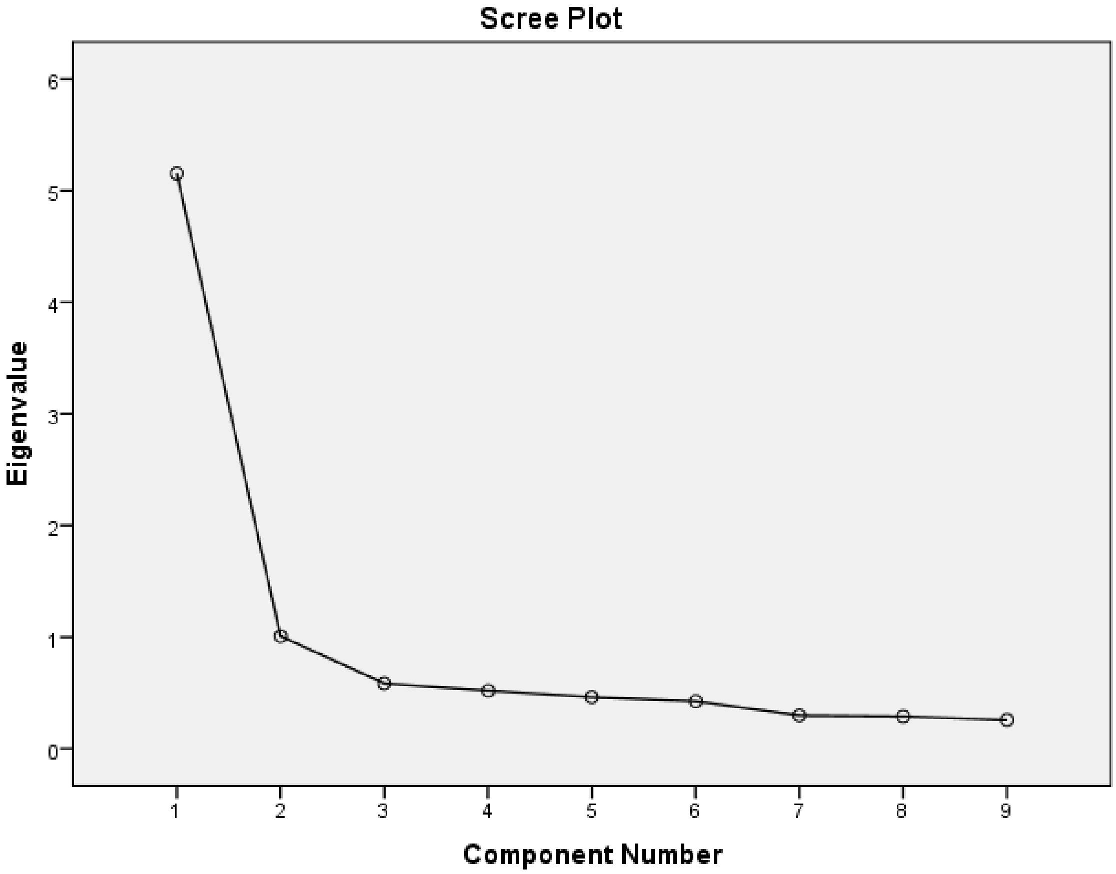
When performing the PCA, the components were extracted based on those whose eigenvalues were greater than 1, based on Kaiser's recommendation (Kaiser, 1960). The result was then compared with a visual representation of the eigenvalues in the form of a scree plot, to check that it represents when the contribution of the components to the overall variance became minimal, and thus when to stop extracting.

Interestingly, the PCA identified only 2 components, as opposed to the original 3 groups (*Education*, *Gameplay* and *Agency*). The implication is that the correlations between each of the original 9 factors were significant enough that the original 3 groups were not true subcategories of educational game immersion; if this is taken to be true, they should simply belong to the overarching category of *Serious Game Immersion*.

Table 8-8: Stage 4 PCA Component Extraction

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sum of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	5.164	57.379	57.379	5.164	57.379	57.379	3.571	39.674	39.674
2	1.012	11.245	68.624	1.012	11.245	68.624	2.605	28.949	68.624
3	.580	6.440	75.064						
4	.519	5.766	80.830						
5	.456	5.067	85.898						
6	.425	4.720	90.617						
7	.297	3.300	93.917						
8	.289	3.207	97.124						
9	.259	2.876	100.000						

Table 8-9: Stage 4 Scree Plot



Having established whether the original groupings of the factors were accurate, the next stage was to investigate the questions themselves; does each question belong in their respective factor?

<u>Question Groupings: Method and Results</u>

To find out how well the questions were grouped into their respective factors, the Cronbach's Alpha value was calculated for each of the original factors, in addition to the change in Cronbach's Alpha if each of the questions were deleted. These values are displayed in

Table 8-10.

For the overall Cronbach's Alpha of the IEGM questionnaire (α = 0.927), the questions that lower its value are described below, or have no impact at all, in addition to what the Alpha value would be if the question was removed:

- While playing, I learned about a subject I did not expect to learn about from the game (0.932)

- I found the game to be too easy (0.928)
- I found the game to be too difficult (0.927)
- Tutorials/hints were presented to me whether I wanted help or not (0.929)
- I found the game's setting to be familiar to me (0.932)
- I felt I could experiment without permanently impacting the game environment (0.927)

The majority of the questions, when individually removed, had a negative impact upon the Alpha value, which indicates that their grouping is appropriate. The degree of positive impact the above questions has on the Alpha value when removed was considered negligible; since the value was over 0.8 whatever individual question was removed, the reliability estimate was still 'acceptable', according to the suggested criteria by (Field, 2009). On this basis, it was decided to retain the questions, because the alternative was to remove questions that are important to measuring particular aspects.

8.2.5 Discussion

Experiment 3 seemed to indicate that the questionnaire could be a viable measure of the immersive qualities of a particular educational game, thus supporting the research question. However, there are questions present which seem to produce unreliable responses, on the basis of the used sample. These questions are as follows (the statistics table for the item totals can be found in Appendix J):

1. "While playing, I learned about a subject I did not expect to learn about" (RV1)
2. "I found the game to be too easy" (BC1)
3. "Tutorials/Hints were presented to me whether I wanted help or not" (GD1)
4. "I found the game's setting to be familiar to me" (SC1)

The reliability issue of RV1 could be due to the differing expectations of the individual participants; they may not have been familiar with the cancer subject matter, and thus less clear on what exactly they were meant to be learning from the game, and more confused about how to answer the question. On the other hand, if this was true, the preceding question ("While playing, I learned about a subject I expected to learn about") should have been impacted as well. An alternative possibility is RV1's positioning to its similarly-worded predecessor. It is also possible that whether a player learns something unexpected has no significant bearing at all on immersion – or indeed that it can actually contribute to immersion for some. Regardless, until the reason can be more clearly established, it could be better to remove the question.

Using the tutorial levels exclusively could have influenced the inconsistency in BC1 and GD1; different participants have different skill levels with regard to games in the genre *Re:Mission* occupies, which could influence how easy they find the game, and how intrusive the tutorials are (as more skilled players may find the tutorials easy to the extent of being boring).

The reason behind the inconsistency in SC1 is less apparent. It is possible the question was not clear enough, that a more specific definition of what was meant by the game's setting was needed.

Additionally, the following questions reduced the internal consistency of the question group for certain aspects:

- "I feel the gameplay is out of place with what the game is trying to teach" (Gameplay-Educational Integration)
- "I was able to find tutorials or hints to help me progress when I needed them" (Guidance)

While the experiment can be more refined, the results support the idea that the IEGM can be used to measure the immersive qualities of an educational game. As such, it provides an important basis for defining tangibly how to make an educational game immersive.

Table 8-10: Experiment 3 Question Grouping Cronbach Alpha

Factor	Factor Cronbach's Alpha	Questions	Overall Cronbach's Alpha if Item Deleted
Relevance	.495	RV_ExpectedLearn	.927
		RV_REV_NotExpectedLearn	.933
		RV_InterestedInSubject	.925
		RV_LearnedUseful	.926
		RV_REV_NotLearnedMuch	.926
		RV_ValuableForSubject	.926

Gameplay-Educational Integration	.722	GE_REV_GameplayOutOfPlace	.926
		GE_MechanicsReflectRealWorld	.927
		GE_ChallengesReflectRealWorld	.928
		GE_ChallengesShowSubject	.926
		GE_FunAndEducational	.925
		GE_REV_EducationalBoring	.926
Sequencing	.676	SQ_UnderstoodSubjectMatter	.927
		SQ_PrevChallengeKnowledgeUseful	.926
		SQ_NecessaryInfoProvided	.926
		SQ_ChallengeOrderBuiltSkills	.926
		SQ_REV_SubjectNotEnoughInfo	.926
		SQ_REV_PrevChallengesNotHelpful	.927
Balanced Challenge	.553	BC_REV_TooEasy	.929
		BC_REV_TooDifficult	.928
		BC_EnjoyedPlaying	.925
		BC_REV_NotRecommendGameplay	.926
		BC_ChallengesFair	.927
		BC_REV_Frustrating	.926
Feedback	.656	FB_UnderstandProgress	.927
		FB_REV_Excessive	.927
		FB_REV_TooLittle	.927
		FB_ImportantProgressionDetailsProvided	.927
		FB_REV_ConfusedHowToPlay	.927
		FB_UnderstoodMistake	.926
Guidance	.668	GD_TutorialsHintsFound	.926
		GD_REV_PresentedWhenUnwanted	.930
		GD_HelpedProgress	.925
		GD_REV_PaidLittleAttention	.926
		GD_REV_Annoying	.926
		GD_REV_MoreDiscoveryBySelfNeeded	.927
Narrative	.789	NV_ChallengeObjectivesClear	.927
		NV_StoryMadeSense	.926
		NV_InterestedInCharacters	.925
		NV_ActionsImpactedOnStory	.926
		NV_REV_DidNotCareAboutStory	.925
		NV_REV_CharactersInconsistent	.926
Scenario	.436	SC_REV_SettingFamiliar	.933
		SC_SimilarToRealWorld	.927
		SC_HelpedEngageWithSubject	.927
		SC_REV_DistractedFromSubject	.927

		SC_REV_Bored	.925
		SC_InvolvedInRole	.925
Experimentation	.711	EX_AbleToExperimentMechanics	.926
		EX_AbleToExperimentSubject	.927
		EX_NoPermanentImpact	.928
		EX_REV_NotEnoughTimeToPlan	.926
		EX_EnoughRetries	.927
		EX_AbleToIdentifyStrategies	.927

8.3 Summary

The findings of the Mann-Whitney test in the Questionnaire experiment show that the mean ranks for each item on the questionnaire were significantly higher than the null value, indicating that the ‘immersive’ game was deemed to be immersive according to each component of the IEGM. However, the results of the factor analysis (grouping each of the nine components in one category) indicated that the subcategories of *Education*, *Gameplay* and *Agency* were not needed.

Chapter 9. Discussion

In this chapter, the results of the IEGM metrics experiments are discussed in relation to the research question, the methodology and the issues encountered when conducting them. Following this, the implications of the IEGM, as well as its experiments for confirmation and metrics testing, are explored in terms of their significance for research and applications.

9.1 The Problem

Serious games for education have been of great interest to educators and researchers, due to their potential to provide engaging learning environments that the players are intrinsically motivated to play and learn from. However, creating such games is not as simple as forcing educational content into an otherwise unrelated game, and require greater consideration of how to appeal to their audience and maintain their interest. (Gunter, et al., 2006) (Aldrich, 2009).

In order to determine how serious games can be made educational and entertaining, various theories of education, serious games design and general game design were investigated. One element which has been less investigated is the idea of immersion in educational games: how to engage learners with an educational game, and keep them engaged while playing. While the subject has been approached from the perspectives of general activities (Csikszentmihalyi, 1990), entertainment games design (Malone, 1981) (Brown & Cairns, 2004) (Gee, 2005) and instructional design (Gagne, 1970) (Keller, 1987) individually, no one perspective covers the full benefits that the others can offer. Likewise, while theories of serious games have explored the ideas of immersion and motivation, they do so in isolation from each other, and so miss potentially considerations that can enhance how well a learner can be immersed in an educational game.

9.2 Development and Confirmation of the IEGM

To attempt to solve this problem, the Immersive Educational Game Model was constructed from the literature review, to provide a unified model for immersion in educational games. The model encompassed twelve aspects, divided into the three categories *Education*, *Gameplay* and *Agency*, to reflect the contexts that the subject of immersion was covered from in the literature review. The aim of this model was to provide a clear and concise basis for both determining how immersive particular educational games are, and designing an educational game to be as immersive and educationally valuable as possible.

In order to confirm that the model was an accurate representation of immersion in educational games, a triangulation research approach was used, utilising two complementary studies, expert interviews and a questionnaire survey, in addition to the literature review. The results of this study would be used to answer the research question:

RQ1: **What aspects affect a player's immersion in an educational game?**

To this end, the model was first presented to nine experts in game research and design, in order to collect qualitative feedback about whether the aspects presented were agreed with, and investigate whether there were considerations that were missing from the model. On the basis of this feedback, three aspects were removed altogether (*Consequences*, *Curiosity* and *Identity Projection*), and the remaining aspects underwent additional changes, including added conditions, refocusing, and name changes. While these changes made deviations from terms established in the literature (e.g. removing the term 'ILO'), it was felt the underlying points and assumptions of each of the aspects remained the same; the changes simply made the aspects easier to provide tangible measures for, and more accessible for those who were not necessarily experts in either education or game design.

Once these alterations were completed, a questionnaire based on the altered model was presented to a selection of 16 game players, to collect quantitative feedback on whether the factors represent the point of view of the gaming community. From this study, it was found the only aspect which was not agreed with was the *Scenario* factor, suggesting it may not be a relevant influence either on player immersion or player learning. However, since the sample size was small, thus not necessarily population-representative, and both the literature review and expert feedback indicated its importance to immersion, it was decided to retain *Scenario*.

9.3 Investigating IEGM Metrics

While the IEGM was confirmed by the studies described and analysed in Chapter 5 and Chapter 6, the question remained as to whether it could be used to practically measure the immersive qualities of educational games. To this end, the next step was to investigate the following research question;

RQ2: **Can the Immersive Educational Game Model be used to test the immersive qualities of an educational game?**

To accomplish this, the first stage was to develop an instrument from the model, which in turn could be used to measure the immersive qualities of educational games. The GQM software approach was used to create the initial instrument, resulting in the creation of a set of questions measuring the aspects in the IEGM.

The progression of Goals, Questions and Metrics is shown in Table 9-1. The Goals were created to be representative of each of the 9 aspects of the IEGM, since they were the overarching reasons to measure the game in how well it facilitates immersion. The Questions in turn were derived from what was considered the practical information needed to meet their encompassing Goal; as such, some of the more player-centric, opinion-based Goals required comparatively few Questions (e.g. *Relevance*, requiring two), while other Goals involving more detailed aspects of the game design itself in turn required more Questions (e.g. *Gameplay-Educational Integration*, which required four).

As mentioned in Section 7.1, the initial approach to creating the Metrics themselves was to create unique measures for each question, based on the smallest possible pieces of information required to answer the Question. The logic behind this decision was that different Metrics require different types of data to specifically analyse and compare them; with *Gameplay-Educational Integration*, for example, the Metrics required to measure the applicability of the game's mechanics can be considered different to those required to measure the applicability of ILOs (the former requiring game genre information, the latter requiring ILO structure information).

While it was felt these questions identified the main requirements for measuring immersive qualities, based on the model, testing each of these unique measurements for each question was deemed too large a task to accomplish in the permitted time. To overcome this issue, it was decided to create a set of Likert items based on the questions; from this, a questionnaire was created, comprising of 54 five-point Likert questions, which could be given to players to retrospectively assess the immersiveness of a particular educational game.

Table 9-1: IEGM Metrics GQM Table

G : Metric applied to the Game P: Metric applied to the Player			
Goal	**Questions**	**Metrics**	**Questionnaire**
<u>Relevance</u> Evaluate how relevant the educational game's content will be to its player	What do the players expect to be able to do after playing the game?	G: Expected ILO	
		P: Perception of Learning	While playing, I learned about the subject I expected to learn about from the game
			While playing, I learned about a subject I did not expect to learn about from the game
	How far can the player's knowledge and understanding of the game's content be applied to subject-related situations in the real world?	G: ILOs of real world case studies	
		P: Perception of whether they feel it would be useful to them in real life	I am interested in the subject the game is trying to teach
			I feel what I've learned would be useful in real life
<u>Gameplay-Educational Integration</u> To evaluate how connected the educational game's gameplay mechanics are to the subject matter of the educational game, so it is	What subject matter is the game attempting to teach?	Overarching ILO	I understand what the game is trying to teach
		Subordinate LOs towards the overarching ILO	<RW> I feel the gameplay is out of place with what the game is trying to teach

appealing as both a game and an educational tool	What mechanics does the game employ?	Gameplay Genres	
	How accurately do the game's mechanics reflect subject-related situations in the real world?	ILOs of real world case studies	I think the game mechanics reflect similar real-world situations
	How accurately do the game's challenges reflect subject-related situations in the real world?	Tasks of real-world case studies	I think the game challenges reflect similar real-world situations
<u>Sequencing</u> Evaluate how well the presented subject matter builds upon itself (from basic concepts to complex concepts)	What specific intended learning outcomes is the game attempting to convey?	Overarching ILOs	I understood what kind of subject matter the game was attempting to teach
	What immediate prerequisite knowledge is required to learn the ILOs?	Subordinate LOs towards the overarching ILO	I found knowledge from previous challenges useful in successive ones
	In what order are the ILOs presented in the game's progression?	Learning type of ILOs	
	What are the in-game challenges that are intended to convey these ILOs?	Skill types of the challenges	
	What prerequisite knowledge is expected of the player to	ILOs of "immediate" prerequisites (i.e. ILOs that directly connect to	I felt with each game challenge, the game provided all the

	complete the in-game challenges?	the lowest-level ILOs of the game)	information I needed to complete it
	In what order are the in-game challenges presented in the game's progression?	Sequence of ILOs in game	I think the order of the game challenges helped build up my skills
Balanced Challenge Evaluate how well-balanced the educational game's challenges are in relation to the player's skills	What are the game's challenges?	Skill types of the challenges	
	How difficult does the player find the game's challenges?	Likert scale of difficulty	I found the game to be too difficult
			I found the game to be too difficult
	Does the player find the challenge enjoyable?	Likert scale of enjoyability	I found the game enjoyable to play
Feedback Evaluate how appropriately the educational game delivers feedback to the player, to help them understand what they are doing in-game	What type of feedback does the player receive?	Taxonomy according to LO Learning type • Remember • Find • Use	
		Typology of Formative Assessment • Weaker feedback • Feedback • Weak Assessment • Moderate Assessment • Strong Assessment	

	When in the game does the player receive the feedback?	Timestamp • After challenge attempt • After game action performance	
		Task flow point • After challenge attempt • After game action performance	
	Does the player find the feedback useful to understand their progress?	Likert Scale of usefulness	I think the game's feedback helped me to understand how well I was doing
	Is the amount of feedback provided by the game enough without being excessive?	Comparative feedback standard, according to skill type of challenge	I found the in-game feedback to be excessive
			I found there was too little feedback in the game
<u>**Guidance**</u> Evaluate how appropriately the educational game delivers guidance to the player, to help them progress	What type of guidance is available to the player?	Taxonomy of guidance • Remember • Find • Use	
	What is each piece of guidance made accessible to the player?	Education – ILO (or learning type)	
		Gameplay – Skill type	
	When is each piece of guidance made accessible to the player?	ILO Tree – guess what the player is attempting from the challenge ILO and the LOs of the corresponding game actions they are	I was able to find tutorials or hints to help me progress when I needed them

		performing	
		Number of attempts at challenge	
	Is the guidance presented automatically or not?	Yes/No question for each piece of guidance	Tutorials/hints were presented to me whether I wanted help or not
	Does the player find the guidance useful to progress in the game?	Likert Scale of usefulness	I found the game's guidance systems (tutorials, hints) helped me to progress
Narrative Evaluate how engaging the narrative of the educational game is to the player	What type of story is the game telling?	Taxonomy of Literature Genres	
		Taxonomy of Gameplay Genres	
	How clear are the game's objectives to the player?	Inclusion of ILOs in narrative flow	I thought it was clear what I had to do in the game challenges
	How consistent is the narrative, in terms of characters and plot progression?	Narrative Cohesion Variables • Logical Sense • Theme • Genre • Narrator • Style	The game's story makes sense to me
			I felt involved in the role the game asked me to perform
	What impact can the player's actions have on the story's progression?	Ratio of 'branching paths' (where plot point changes depending on the player's actions) to fixed plot points	I thought my actions in-game had a noticeable impact on the story

Scenario Evaluate how engaging the setting of the educational game is to the player, while being unfamiliar enough to allow deeper exploration of the subject	What is the setting of the game?	Country/town the player lives in	
		Occupation/Key Stage	
	How familiar is the setting to the game's audience?	Country/town the player lives in	
		Occupation/Key Stage	
		Likert Scale of familiarity	I found the game's setting to be familiar to me
	How does the setting differ from that of subject-related situations in the real-world?	Comparison with case studies	I found the game's setting similar to related real-world situations
			I felt the game's setting helped me to engage more with the game's subject matter
Experimentation Evaluate how much challenge-pertinent experimentation the educational game permits with the content it presents	When can the player experiment with what they have learnt?	Checklist of in-game challenges which have an area to experiment (i.e. can practice with the in-game actions in a sandbox context)	I felt able to experiment with the game mechanics introduced
			I was able to experiment with what I learned about the game's subject matter
	What are the in-game risks to the player as they are experimenting?	List of in-game actions, and on which challenges these actions have an irreversible impact on	I felt I could experiment without permanently impacting the game environment

		the game environment	
	How many opportunities does the player have to solve a game's challenge?	Number of maximum attempts possible for each challenge	
	How many of the game's mechanics are available to the player in each experimentation situation?	For each challenge, a list of the possible game actions	

Once the instrument was created, it needed to be confirmed in order to ensure it does accurately measure educational game immersion. To this end, three studies were conducted to investigating how well the instrument questions are suited to measure the IEGM. Each study involved a group of participants playing a particular 'immersive' educational game, *Re:Mission*, (selected based on its perceived compliance with the IEGM properties).

The first study explored the level of immersion in an 'immersive' educational game. In this study, a group of 30 participants were asked to play *Re:Mission*, and to complete a Tangram puzzle before and after playing it. The idea was that if the participants completed the task slower after playing the game, it would indicate that they were immersed in it (Jennett, et al., 2008).

The second study explored the level of distraction in an 'immersive' educational game. In this study, a group of 30 participants were asked to play *Re:Mission*, with their eye fixations being recorded by an eye-tracker. The theory was that the more immersed the participants were in the game, the fewer eye fixations per second they would have over time, since their focus would be less distracted from the game (Jennett, et al., 2008).

Finally, the third study explored the level of immersive qualities in an 'immersive' educational game. To investigate this, a group of 90 participants were asked to play *Re:Mission*, and then complete the IEGM instrument questionnaire (86 of whom included in the analysis, since four responses had missing data). Since questionnaires were deemed an appropriate way to measure immersion (Jennett, et al., 2008), if the questionnaire results rank each of the immersive qualities highly, it would indicate the instrument can accurately identify educational games as being immersive. Combined with the other two studies (which investigate the immersiveness of *Re:Mission* itself), it would validate the instrument as a measuring tool of core immersive elements.

The third study indicated that the instrument questionnaire could be used to reliably identify the qualities that make an educational game immersive, and thus validate that the IEGM could be used to measure how immersive educational games are, thus satisfying RQ2. In particular, the high internal reliability findings (see Section 7.5.3.3) indicated each of the Likert items were important to retain. However, the findings of the first and second studies were less conclusive, calling into question whether *Re:Mission* was immersive, and if so whether eye fixations and distracting tasks are reliable measures of immersion across games in general (since both measures were initially tested on entertainment games (Jennett, et al., 2008)).

9.4 Final Version of the Immersive Educational Games Model

The final version of the IEGM, after analysing the results of all verification stages, is presented in Figure 9-1. It is designed to act as a synthesis, clarification and refinement of the most important, generalisable elements of making educational serious games immersive, while being framed in terms of specific aspects used to structure and design educational games.

The reasoning behind this is that prior theories surrounding serious games were largely too disparate, treating engagement and learning within such games only through the lens of one particular field; examples include De Freitas and Kiili's respective Experiential Learning approaches (Kiili, 2005) (Freitas & Neumann, 2009), which only consider the knowledge experimentation and consequence aspects, or Dickey and Norman's observations, which focus on the overarching structures of world-building and engaging activities respectively (Dickey, 2007) (Norman, 1993), but do not incorporate the other, despite them both being important in keeping players interested. This is likewise a limitation of the general learning theories when generically applied to games; for example, while Keller's ARCs model of motivation and Laurillard's Conversational framework offer great insight into facilitating optimal learning conditions from particular activities (Keller, 1983) (Laurillard, 2002), they do not take into account other aspects that are important appeals of games (e.g. concerning world-building).

Conversely, theories such as Malone's user interface heuristics (Malone, 1982), which attempt a more inclusive approach, can be too vague or too abstract to be useful for developers to design immersive games with.

A model like the IEGM, which integrates each of these perspectives, and reframes them in terms of what would be most important for both game developers and game players, is therefore vital for research into serious games going forward. It provides a concise and inclusive approach which prioritises the engagement/immersion of the player, while also providing a way to clearly identify how much impact on a player's engagement each aspect of the game makes.

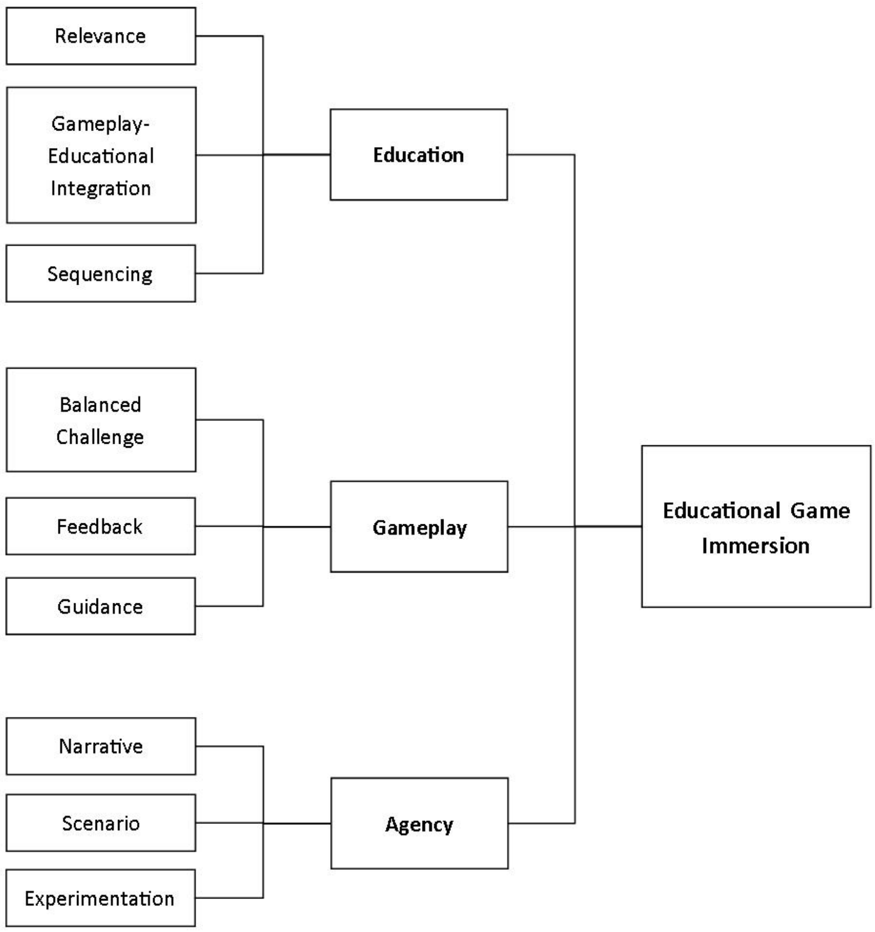


Figure 9-1: Immersive Educational Games Model (Final)

9.5 Implications

From the research involved in constructing the IEGM, this section explores the implications of the findings for researchers and developers in the field.

9.5.1 Educational/Serious Games Field

The literature of this research has demonstrated many disparate approaches to using games to teach, how to make games engaging or immersive, or both. The IEGM helps to unify these disparate ideas into a single model, which highlights the core immersive qualities of educational games. Since the model is derived from a wide range of theories in education, game design, narrative and immersion, as well as practical examples of serious games, it can offer a cohesive and substantiated perspective on immersion, which can be applied to a general range of educational games.

Furthermore, with its confirmation from field experts and game players, as well as its validation as a potential measurement tool, the IEGM presents a theoretically sound and practically substantiated basis for future research directions into educational games immersion. The separation of each immersive quality into small, more clearly defined components allows for more specific use in particular studies.

9.5.2 Game Researchers

For researchers in games, the IEGM questionnaire has been shown in the research as a reliable tool for measuring immersive qualities. Such a questionnaire could be useful for assessing serious games for their suitability in immersion research; for example, a study requiring a highly immersive game could be found by using the IEGM questionnaire after playing the particular game for 15 minutes, allowing for a relatively rapid and clear assessment of the game's immersive aspects. In addition, having the individual elements of each aspect elaborated can allow researchers to determine how immersive specific parts of the game are.

Furthermore, the issues encountered in the validation experiments help to highlight the difficulties of measuring immersion; specifically, the issues of genre and pacing of the game, as well as familiarity with the external task used to measure distraction. These issues would indicate that an important research direction in games would be to find a practical measurement of immersion that is applicable across multiple game types.

9.5.3 Education Researchers/Teachers

The confirmed IEGM can be valuable to education researchers because of its fundamental grounding in educational theories, including Laurillard's conversational framework, Bloom's cognitive domain categories (and Anderson's refinements), Gagne's events of instruction and many others. As such, it can provide substantial basis as to how games and education theory naturally coincide, in turn providing some more focused directions as to how to advance research into educational games; specifically, ways in which to present subjects through games in an interesting, immersive way, which students can reliably learn from. Having a strong educational basis when addressing these issues will only become more important as computer games become more pervasive.

Furthermore, the positive results of the Questionnaire experiment indicate that the IEGM can be helpful for teachers, as a tool for measuring the immersive qualities of educational games. With such a tool, they would be able to more reliably judge how suitable a game will be to help teach a particular topic, in terms of how well it integrates the topic, and how engaging the students are likely to find it. The IEGM can also help educators working in conjunction with game developers, as with *Immune Attack* (Kelly, et al., 2007), across all stages of development, as a way of clearly structuring the issues and considerations needed to make a game immersive, in turn potentially saving time by removing more trial-and-error based approaches from the design process.

Conversely, the inconclusive results of Experiments 1 and 3 demonstrate that measures of immersive states are not, at this time, fool-proof; both eye fixations and distracting tasks can be potentially subject to a number of biasing factors – the methodology alone requires a precise approach that is unlikely to be representative of typical relaxed behaviour when playing a game. This is not to say the approaches are inherently flawed (Jennett's study certainly suggests there is a certain reliability to these methods, under certain entertainment contexts), or that a way cannot be found for them to work in more generalised situations, but it is an area that requires more investigation. As such, it would be inadvisable for educators to use these methods until more reliable iterations of them have been developed.

9.5.4 Game Developers

For developers wishing to create educational games, as with education researchers, the confirmation and validation of the IEGM through the initial triangulation study and the

Questionnaire experiment suggests that it can be a reliable means of identifying the core immersive components of educational games, which accounts for the need to integrate educational content. In addition, the model's categories are localised enough that elements of it can be applied towards entertainment games as well (namely the 'Gameplay' and 'Agency' categories, highlighting the components that make gameplay progression and the game's world engaging).

Chapter 10. Conclusions and Future Work

This section describes the conclusions from the research, including the identification of the main research problem, the proposed solution, and the experiments used to validate the use of the solution. The section then concludes by exploring possible future directions to follow from the research.

10.1 Contributions

The research has produced two contributions towards the fields of educational games and immersion, which are described in this subsection.

10.1.1 The Immersive Educational Games Model (IEGM)

The first main contribution of the research is the creation of the Immersive Educational Games Model (IEGM), which incorporated the following contributions:

- A critical analysis of appropriate literature in education, games and narrative, including models and frameworks for the structuring of teaching content and serious games
- The synthesis of disparate theories of serious games structuring into a cohesive whole, as relates to encouraging player immersion, compressing very similar ideas and incorporating isolated but important ideas, in order to provide a more complete idea about the core immersive elements of educational serious games
- The integration of educational theories and entertainment games theories to provide a comprehensive, concise basis about what makes games and educational techniques engaging
- A list of nine core immersive qualities of educational serious games, confirmed by experts in the fields of games and e-learning and a selection of game players, providing a clear foundation of how to potentially build immersive educational games, or to measure the immersive qualities within them

10.1.2 Using the IEGM as an Instrument

The second main contribution of the research is the development of a questionnaire instrument to measure immersive qualities in educational games, which includes the following contributions:

- A demonstration that the IEGM can be used to derive a measurement tool for immersive qualities in educational games
- An application of the instrument with an established immersive educational game, helping validate the use of the IEGM as a measurement tool for reviewing completed educational games
- An exploration of the issues in using measures of immersive states for general scenarios, and the considerations that need to be taken into account for their use in future research in the field

The instrument can be used by game developers to help test their educational game in development by providing more focused feedback as to what parts of their game are facilitating immersion. Furthermore, it can also help educators and researchers to assess particular educational games on their immersive qualities, and thus how suitable the game is, or will be, as an instructional tool. This in turn demonstrates not only that the IEGM can be used to generate guidelines for serious game developers to design towards, but also guidelines for educators and researchers to help select educational games for particular purposes.

10.2 Future Work

The IEGM represents a clear identification of the central aspects that make educational games immersive, and the research has, to an extent, demonstrated that it can be practically applied to the measurement of immersive qualities of such games. However, the model is still at a very early stage; the instrument demonstrates it can be used as a subjective, retrospective measurement, but the findings from its complementary studies suggest that there may be additional considerations as to what makes educational games immersive, and in particular how one measures immersion in the first place. To that end, this subsection explores possible future work to be undertaken to address these issues.

10.2.1 IEGM Component Weightings

While the IEGM presents the core immersive qualities from a general perspective, it is not yet established how far each quality contributes to immersion. Certain qualities may have a greater impact than others, and specific aspects of those qualities may contribute more than others. For this reason, an important future step would be to conduct further research into the strength of relationship between each of the IEGM's nine components and immersion. This would in turn benefit future measurement tools using the model, by indicating what weighting each component would likely have on how immersive an educational game can be.

10.2.2 Immersion Measurements

The contentious aspects of using eye fixations and distracting tasks as immersion measures in this study would indicate further research needs to be carried out in refining these measurements. One direction could be to use these proposed immersion measurements across multiple immersive educational games with the same genre and subject, and compare their indications for each game. This would allow potentially more focused and useful insight into when these measurements are applicable, and uncover potential refinements that can be made to ensure greater reliability; this can then be broadened to multiple genres to further refine these measurements. In addition, there may be other immersion measurements that can provide more reliable results than either, if only in particular contexts; research by (Nacke & Lindley, 2008) suggest that physiological measurements can show promising indicators of immersion.

Through the refined immersion measurements that can arise from this, a more reliable validation base can be provided for future measures of immersive qualities, and thus help facilitate the creation of more immersive games in future.

10.2.3 Closer investigation of the IEGM Components

The high-level descriptions of the IEGM components means that using it to derive more specific measures for each component may be more challenging than necessary. For this reason, more investigation should be done into defining each of the IEGM components, in terms of the exact ways in which they contribute to immersion. This could include, for instance, defining the relative conditions of ‘too hard’ or ‘too easy’ for the *Balanced Challenge* component. Through this closer investigation, more reliable measurement tools can be constructed, with clearer guidelines, in turn helping to develop educational games to trigger immersion more reliably.

10.2.4 Consideration of other Educational Theories

The priority of the IEGM was to assess the immersive impact of educational games, primarily from the perspective of game theories. While educational theories do form a partial basis for the *Education* category, there may be alternative viewpoints not explored in the literature, which can provide a deeper insight into the underlying teaching mechanics and how they may inform the conditions of immersion. For example, one of the instructional design focuses in this research was De Freitas' experiential learning model, based on Kolb's instructional design perspective; the integration of other instructional design models (such as Dick and Carey's model (Dick, et al., 2008), or the Kemp design model (Kemp, 1977)) could be used to further explore how instructional goals are constructed, how learner characteristics are established, and how performance objectives and assessment instruments can be devised. This in turn can help expand the IEGM's *Education* category in terms of what important learner characteristics may not have been considered, and indeed how to structure the construction of an educational game compliant with the IEGM's metrics.

10.3 Final Conclusion

Creating engaging educational games is a difficult task, requiring far more than the arbitrary insertion of subject matter into unrelated gameplay. However, fulfilling this task allows people to learn and understand subjects in such a way that they enjoy it, and become intrinsically motivated to continue learning. The aim of the research conducted in the thesis was to find out the most important elements of creating such engagement in educational games.

There have been many theories describing how to make educational activities, games, and even everyday tasks engaging, as described in this thesis. While these theories address particular aspects of engagement, including the overarching concept of complete engagement, or 'immersion', there is a certain disparity between the approaches for educational games. As a result there lacked a clear, generalisable way of determining how immersive a particular educational game was, or indeed clear guidelines on how to construct an educational game to be immersive.

The creation of the IEGM, as a result of the research in the thesis, acts as a way to solve this issue. The model provides a unified and focused structure, which indicates the core components across all educational games that facilitate immersion. Through the development of an instrument based on it, the model further demonstrates that it can be used as a measurement tool to determine how immersive a particular educational game is.

There is much work still to be done to fully realise the aim of the research; the model components need further investigations in terms of their specific measures and strength of contributions to immersion, and there is further study required to develop reliable measures of immersive states. However, the IEGM has shown promise as a grounded representation of immersive qualities, and as a result can help developers, teachers and researchers alike to create deep, enjoyable games from which people want to learn.

References

Adamson, J., 2005. Combined qualitative and quantitative designs. In: A. Bowling & S. Ebrahim, eds. *Handbook Of Health Research Methods: Investigation, Measurement And Analysis : A Practitioner's Guide.* New York: Open University Press, pp. 230-245.

Aldrich, C., 2009. Virtual Worlds, Simulations and Games for Education: A Unifying View. *Innovate: Journal of Online Education,* 5(5).

Allen, I. E. & Seaman, C. A., 2007. Likert Scales and Data Analyses. *Quality Progress,* 40(7), pp. 64-65.

Anderson, L. W. et al., 2001. *A taxonomy for learning, teaching and assessing: A revision of Bloom's Taxonomy of educational outcomes: Abridged edition.* 1st ed. New York: Longman.

Arnab, S. et al., 2015. Mapping learning and game mechanics for serious games analysis. *British Journal of Educational Technology,* 46(2), pp. 391-411.

Aronson, J., 1994. A Pragmatic View of Thematic Analysis. *The Qualitative Report,* 2(1), pp. 1-3.

Aylett, R., 1999. Narrative in Virtual Environments - Towards Emergent Narrative. *Proceedings of the AAAI Fall Symposium on Narrative Intelligence.*

Barab, S. A., Gresalfi, M., Dodge, T. & Ingram-Goble, A., 2010 (a). Narratizing Disciplines and Disciplinizing Narratives: Games as 21st Century Curriculum. *International Journal of Gaming and Computer-Mediated Simulations,* January-March, 2(1), pp. 17-30.

Barab, S. A., Gresalfi, M. & Ingram-Goble, A., 2010 (b). Transformational Play: Using Games to Position Person, Content and Context. *Educational Researcher,* 37(7), pp. 525-536.

Basili, V., Caldiera, G. & Rombach, H. D., 1994. The Goal Question Metric Approach. *Encyclopedia of Software Engineering,* Volume 2, pp. 1-10.

Bloom, B. S. et al., 1956. *Taxonomy of Educational Objectives Handbook 1: Cognitive Domain.* 1st ed. Ann Arbor: Longmans.

Borg, W. R. & Gall, M. D., 1983. *Educational Research: An Introduction.* 4th ed. Broadway, New York: Longman Inc.

Borque, L. B. & Fielder, E. P., 2003. *How to Conduct Self-Administered and Mail Surveys.* 2nd ed. Thousand Oaks, California: SAGE Publications.

Braun, V. & Clarke, V., 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology,* pp. 77-101.

Brown, E. & Cairns, P., 2004. A Grounded Investigation of Game Immersion. *CHI'04 extended abstracts on Human factors in computing systems,* 24-29 April, pp. 1297-1300.

Burn, A. & Schott, G., 2004. Heavy hero or digital dummy? Multimodal player-avatar relations in Final Fantasy 7. *Visual Communication,* 3(2), pp. 213-233.

Cohen, J., 1992. A Power Primer. *Psychological Bulletin,* 112(1), pp. 155-159.

Connolly, T. M. et al., 2012. A systematic literature review of empirical evidence on computer games and serious games. *Computers & Education,* 59(2), pp. 661-686.

Creswell, J. W. & Clark, V. L. P., 2011. *Designing and Conducting Mixed Methods Research.* 2nd ed. Thousand Oaks, CA: SAGE Publications.

Cronbach, L. J., 1951. Coefficient Alpha and the Internal Structure of Tests. *Psychometrika,* 16(3), pp. 297-334.

Csikszentmihalyi, M., 1990. *Flow: The Psychology of Optimal Experience.* New York: Harper and Row.

Dickey, M. D., 2007. Game Design and Learning: A Conjectural Analysis of How Massively Multiple Online Role-Playing Games (MMORPGs) Foster Intrinsic Motivation. *Educational Technology Research and Development,* June, 55(3), pp. 253-273.

Dickey, M. D., 2011. Murder on Grimm Isle: The impact of game narrative design in an educational game-based learning environment. *British Journal of Educational Technology,* 42(3), pp. 456-469.

Dick, W., Carey, L. & Carey, J. O., 2008. *The Systematic Design of Instruction.* 7th ed. Boston: Pearson.

Dillenbourg, P., 1999. What do you mean by collaborative learning?. In: P. Dillenbourg, ed. *Collaborative-learning: Cognitive and Computational Approaches.* Oxford: Elsevier, pp. 1-19.

Ertmer, P. A. & Newby, T. J., 1993. Behaviorism, Cognitivism, Constructivism: Comparing Critical Features From an Instructional Design Perspective. *Performance Improvement Quarterly,* 6(4), pp. 50-72.

Fenton, N. & Bieman, J., 1997. *Software metrics: a rigorous and practical approach.* 2nd ed. London: International Thomson Computer Press.

Field, A., 2009. *Discovering Statistics Using SPSS.* 3rd ed. London: SAGE Publications.

Fink, A., 2003. *How to Ask Survey Questions.* 2nd ed. Thousand Oaks, California: SAGE Publications.

Flowerdew, L., 2015. Data-driven learning and language learning theories. In: A. Lenko-Szymanska & A. Boulton, eds. *Multiple Affordances of Language Corpora for Data-driven Learning.* Amsterdam: John Benjamins Publishing Company, p. 15.

Frazer, A., Argles, D. & Wills, G., 2007. *Assessing The Usefulness Of Mini-games As Educational Resources,* Southampton: University of Southampton.

Freitas, S. d. & Neumann, T., 2009. The use of ‘exploratory learning’ for supporting immersive learning in virtual environments. *Computers & Education,* Volume 52, pp. 343-352.

Gagne, R. M., 1970. *The Conditions of Learning.* 2nd ed. London: Holt, Rinehart & Winston.

Garris, R., Ahlers, R. & Driskell, J. E., 2002. Games, motivation and learning: A research and practice model. *Simulation & Gaming,* 33(4), pp. 441-467.

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

Guirdham, M., 2002. *Interactive Behaviour at Work.* 3rd ed. Harlow, England: Pearson Education.

Gunter, G., Kenny, R. F. & Vick, E. H., 2006. A Case for a Formal Design Paradigm for Serious Games. *The Journal of the International Digital Media and Arts Association*, pp. 93-105.

Hair, J. F., 2010. *Multivariate data analysis : a global perspective.* 7th ed. London: Pearson.

Hamalainen, R., Manninen, T., Jarvela, S. & Hakkinen, P., 2006. Learning to collaborate: Designing collaboration in a 3-D game environment. *The Internet and Higher Education*, 9(1), pp. 47-61.

Hamari, J. et al., 2016. Challenging games help students learn: An empirical study on engagement, flow and immersion in game-based learning. *Computers in Human Behaviour,* 54(2016), pp. 170-179.

Hargood, C., Millard, D. & Weal, M., 2009. *Investigating a thematic approach to narrative.* Southampton: University of Southampton.

Hargood, C., Millard, D. & Weal, M., 2011. *Measuring narrative cohesion: A five variables approach.* Southampton: University of Southampton.

Harteveld, C., Guimaraes, R., Mayer, I. & Bidarra, R., 2007. Balancing pedagogy, game and reality components within a unique serious game for training levee inspection. *Technologies for E-Learning and Digital Entertainment*, pp. 128-139.

Hove, S. E. & Anda, B., 2005. Experiences from Conducting Semi-structured Interviews in Empirical Software Engineering Research. *Software Metrics,* 11th IEEE International Symposium(IEEE), p. 23.

Hunicke, R., LeBlanc, M. & Zubek, R., 2004. MDA: A Formal Approach to Game Design and Game Research. *Proceedings of the AAAI Workshop on Challenges in Game AI.*

Jenkins, H., 2004. Game design as narrative architecture. *Computer,* Volume 44.

Jennett, C. et al., 2008. Measuring and defining the experience of immersion in games. *International Journal of Human-Computer Studies,* 66(1), pp. 641-661.

Jick, T. D., 1979. Mixing qualitative and quantitative methods: Triangulation in action. *Administrative science quarterly,* pp. 602-611.

Jonassen, D. et al., 1995. Constructivism and Computer-Mediated Communication in Distance Education. *American Journal of Distance Education,* 9(2), pp. 7-26.

Kaiser, H. F., 1960. The Application of Electronic Computers to Factor Analysis. *Educational and Psychologcal Measurement,* 20(1), pp. 141-151.

Kaplowitz, M. D. & Hoehn, J. P., 2001. Do focus groups and individual interviews reveal the same information for natural resource valuation?. *Ecological Economics,* 36(2), pp. 237-247.

Kato, P. M., Cole, S. W., Bradlyn, A. S. & Pollock, B. H., 2008. A Video Game Improves Behavioral Outcomes in Adolescents and Young Adults with Cancer: A Randomized Trial. *Pediatrics,* 122(2), pp. e305-e317.

Keller, J. M., 1983. Motivational Design of Instruction. In: C. M. Reigeluth, ed. *Instructional Design Theories and Models: An overview of their current status.* Hillsdale, New Jersey: Lawrence Erlbaum Associates Inc, pp. 382-434.

Keller, J. M., 1987. Development and Use of the ARCS Model of Motivational Design. *Journal of Instructional Development,* 10(3), pp. 2-10.

Kelly, H. et al., 2007. How to build serious games. *Communications of the ACM,* 50(7), pp. 44-49.

Kemp, J., 1977. *Instructional design: a plan for unit and course development.* 2nd ed. Belmont: Fearon Publishers.

Kiili, K., 2005. Digital game-based learning: Towards an experiential gaming model. *Internet and Higher Education,* 8(1), pp. 13-24.

Kitzinger, J., 1995. Qualitative Research: Introducing focus groups. *BMJ,* Volume 311, pp. 299-302.

Laurillard, D., 2002. *Rethinking University Teaching: A conversational framework for the effective use of learning technologies.* 2nd ed. London: Routledge.

Malone, T. W., 1981. Toward a Theory of Intrinsically Motivating Instruction. *Cognitive Science,* 5(4), pp. 333-369.

Malone, T. W., 1982. Heuristics for Designing Enjoyable User Interfaces: Lessons from Computer Games. *Proceedings of the 1982 conference on Human factors in computing systems.*

Merrill, M. D., 1994. The Descriptive Component Display Theory. In: *Instructional Design Theory.* Englewood Cliffs, New Jersey: Educational Technology, pp. 111-157.

Merrill, M. D., 2002. First principles of instruction. *Educational technology research and development,* 50(3), pp. 43-59.

Missura, O. & Gartner, T., 2009. Player Modeling for Intelligent Difficulty Adjustment. *Proceedings of the 12th International Conference on Discovery Science,* pp. 197-211.

Morse, J. M., 1991. Approaches to Qualitative-Quantitative Methodological Triangulation. *Nursing Research,* 40(2), pp. 120-123.

Nacke, L. E. & Lindley, C. A., 2008. Flow and immersion in first-person shooters. *Proceedings of the 2008 Conference on Future Play Research, Play, Share - Future Play '08,* pp. 81-88.

Norman, D., 1993. *Things That Make Us Smart: Defending human attributes in the age of the machine.* New York: Basic Books.

Oishi, S. M., 2003. *How to Conduct In-Person Interviews for Surveys.* 2nd ed. Thousand Oaks, California: Sage Publications.

Pallant, J., 2010. *SPSS Survival Manual.* 4th ed. Berkshire, England: McGraw-Hill.

Pavlov, I. P., c1927. *Conditioned reflexes: An investigation of the physiological activity of the cerebral cortex.* 1st ed. London: Oxford University Press.

Qin, H., Rau, P.-L. P. & Salvendy, G., 2009. Measuring Player Immersion in the Computer Game Narrative. *International Journal of Human-Computer Interaction,* February, 25(2), pp. 107-133.

Rankin, Y. A., McNeal, M., Shute, M. W. & Gooch, B., 2008. User Centered Game Design: Evaluating Massive Multiplayer Online Role Playing Games for Second Language Acquisition. *Proceedings of the 2008 ACM SIGGRAPH symposium on Video games*, pp. 43-49.

Rogers, Y., Sharp, H. & Preece, J., 2011. *Interaction Design: Beyond Human-Computer Interaction.* 3rd ed. Chichester, West Sussex: John Wiley & Sons, Ltd.

Sale, J. E. M., Lohfeld, L. H. & Brazil, K., 2002. Revisiting the Quantitative-Qualitative Debate: Implications for Mixed-Methods Research. *Quality & Quantity,* 36(1), pp. 43-53.

Salen, K. & Zimmerman, E., 2004. Games as Narrative Play. In: *Rules of play: Game design fundamentals.* Cambridge, MA: The MIT Press, pp. 420-459.

Skinner, B. F., 1958. Teaching Machines. *Science*, 128(3330), pp. 969-977.

Skinner, B. F., 1974. *About Behaviorism.* London: Penguin Books.

Skinner, B. F., 1977. Why I Am Not a Cognitive Psychologist. *Behaviorism,* 5(2), pp. 1-10.

Squire, K., 2006. From Content to Context: Videogames as Designed Experience. *Educational Researcher,* 35(8), pp. 19-29.

Stahl, G., 2000. A model of collaborative knowledge-building. *Proceedings of the fourth international conference of the learning sciences,* pp. 70-77.

Stahl, G., 2004. Building Collaborative Knowing. In: J. Strijbos, P. A. Kirschner & R. L. Martens, eds. *What we know about CSCL.* Boston: Kluwers Academic Publishers, pp. 53-86.

Steinkuehler, C. A., 2008. Scientific Habits of Mind in Virtual Worlds. *Journal of Science Education and Technology,* 17(6), pp. 530-543.

Sweetster, P. & Wyeth, P., 2005. GameFlow: A Model for Evaluating Player Enjoyment in Games. *Computers in Entertainment,* 3(3).

Tuckman, B. W. & Jensen, M. A. C., 1977. Stages of small-group development revisited. *Group & Organization Management,* 2(4), pp. 419-427.

Villalta, M. et al., 2011. Design guidelines for Classroom Multiplayer Presential Games (CMPG). *Computers & Education,* Volume 57, pp. 2039-2053.

Vygotsky, L., 1962. *Thought and Language.* 1st ed. Cambridge, Massachusetts: Martino Publishing.

Vygotsky, L. S., 1978. *Mind in Society: Development of Higher Psychological Processes.* 1st ed. Cambridge, Massachusetts: Harvard University Press.

Watson, J. B., 1913. Psychology as the Behaviorist Views it. *Psychological Review,* 20(2), pp. 158-177.

Wendel, V., Gutjahr, M., Gobel, S. & Steinmetz, R., 2012. *Designing Collaborative Multiplayer Serious Games for Collaborative Learning.* Porto, Portugal, CSEDU.

Whitehead, J. C., 1991. Environmental interest group behavior and self-selection bias in contingent valuation mail surveys. *Growth and Change,* 22(1), pp. 10-20.

Yee, N., 2006. The Psychology of MMORPGs: Emotional Investment, Motivations, Relationship Formation and Problematic Usage. In: R. Schroeder & A. Axelsson, eds. *Avatars at Work and Play: Collaboration and interaction in shared virtual environments.* Netherlands: Springer, pp. 187-207.

Appendices

Appendix A : Theories of Interaction

- **Social Exchange Theory** – Proposes all meetings between people are 'exchanges', seeking some benefit. We aim to maximise our benefit while minimising input – any behaviour that produces this result is more likely to be repeated ((Guirdham, 2002), p. 78)
- **Equity Theory** – Whilst partaking in 'exchanges', we evaluate our own contributions against the received benefit. If they are almost equal, the exchange is regarded as 'fair'. If the exchange is not fair, it can cause distress in the individual ((Guirdham, 2002), p. 78)
- **Attribution Theory** – We associate causes to our own, or each other's behaviour. We use 'internal attributions' (e.g. personality) to explain other people's behaviour, and use 'external attributions' (e.g. environment) to explain our own ((Guirdham, 2002), pp. 78-79)

Appendix B : Dillenbourg's Factors of Effective Learning in Groups

Dillenbourg proposed that there are five considerations necessary to get the best result from group learning (Dillenbourg, 1999):

- **Setup of Initial Conditions** – Establish the optimal conditions to promote learning and interaction (e.g. group size, gender/culture mix, philosophical differences). Different tasks have different optimal settings
- **Role-based Scenario –** Providing a problem, or series of problems, that require different types of knowledge to solve
- **Interaction Rules –** Specifying how the group members should interact (e.g. allow members to communicate freely, or restrict their communication to certain statement types)

- **Monitoring and Regulating of Interactions –** The extent that a ‘facilitator’ redirects discussions to make them more productive, or the tools the members have available for self-regulation

Appendix C : Wendel’s Collaborative Game Components

In their collaborative serious game, Wendel et al observed a set of important components for collaborative gameplay (Wendel, et al., 2012):

- **Common Goal/Success** – Participation from all players is needed to succeed
- **Heterogeneous Resources** – Each player having a unique, useful action to contribute
- **Refillable Personal Resources** – Resources that deplete over time, or when the player acts dangerously. These can be replenished by the player, or other players
- **Collectable/Tradable Resources** – Resources that are necessary to win, and are tradable between players
- **Scoring** – Assigning scores based on player performance (e.g. frequency of helpful interactions). Perhaps an archaic game feature, but can provide quantitative data for assessment
- **Trading System** – Having a shared pool of resources that can be deposited in or accessed at any time, by any player. Wendel argues this is a mechanism for facilitating trust

Appendix D : Gee’s Good Game Learning Principles

Gee proposed thirteen principles for facilitating good learning in games:

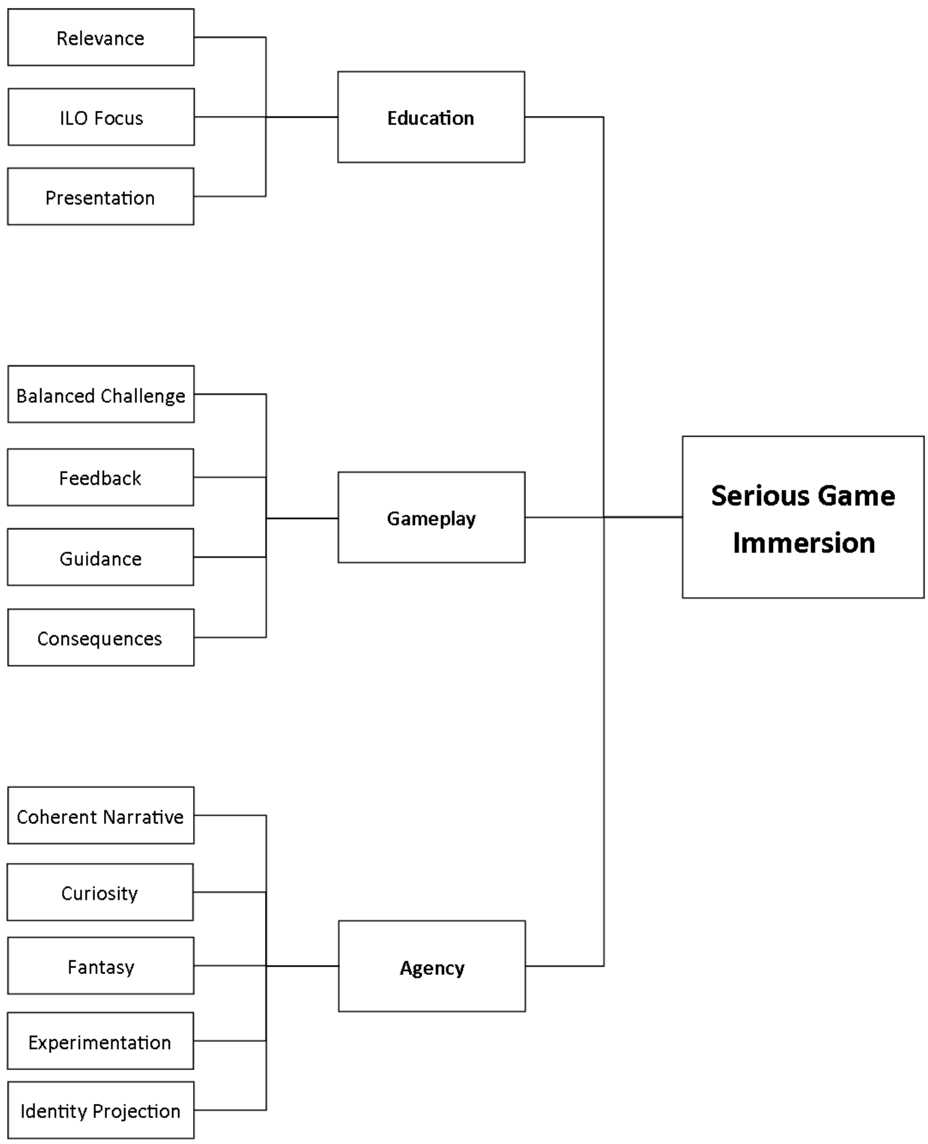
- **Co-design –** The players should feel like active agents, or ‘producers’ in the game world
- **Customise –** The players should be able to make their own decisions about how their learning should work, and be encouraged to experiment with different learning styles
- **Identity –** Involves the player becoming invested in their in-game role, as that investment can extend to the material presented
- **Manipulation and Distributed Knowledge –** Involves the player easily manipulating objects in the game world, through their in-game character

- **Well-ordered Problems –** Guiding the player early on to form strategies that will help them face challenges later in the game
- **Pleasantly Frustrating –** Ensuring the game's challenges are balanced enough
- **Cycle of Experience –** Having players learn through practicing, gaining experience and mastery of skills, and being tested on them
- **Information 'on demand' and 'just in time' –** Providing information when the player feels they need it, and while they are in a situation where they could use it
- **Fish tanks –** Providing simpler tutorial levels to ensure the players understand the basic game rules
- **Sandboxes –** Providing expanded tutorial levels that play like the actual game, but without the pressure of failure, to increase the player's morale before they tackle the real challenges
- **Skills as Strategies –** Having the players use the skills they acquire in context, in order to accomplish a goal that they desire
- **System Thinking –** Having the players understand how each element of the game fits into the larger game genre
- **Meaning as Action Image –** Using the player's experiences within the game to convey meanings and messages

Appendix E : Stage 1 - Expert Interviews Participant Information and Questionnaire

Participant Information

My name is James Baker, and I am doing a PhD in educational computer games. I am currently researching the factors which make educational games immersive. Based on the background research I have found, I have created a model which illustrates the 12 factors that I think contribute most to a player's immersion in a serious game. These factors are divided into 3 categories; **Education, Gameplay** and **Agency**.



The purpose of this interview is to collect feedback from you about my model, whether you agree or disagree with any of the factors, what you would categorise differently, or anything I may have left out.

If you wish to withdraw from the interview at any time, you may do so, and your interview data will not be used. Additionally, you may contact me at any time at jb29g08@soton.ac.uk following the interview to request that your data not be used, and it will be removed.

Questionnaire

Education

The following factors relate to keeping players immersed through the educational content of the game.

Please circle one answer for each factor.

Relevance Players feel more immersed in an educational game if its educational content is made relevant to the player, i.e. the player feels that they can use the content as part of their own life	Strongly Disagree	Disagree	Agree	Strongly Agree
ILO Focus An educational game is more immersive when its educational content is integrated into the gameplay, so the two do not feel distinct from each other while playing	Strongly Disagree	Disagree	Agree	Strongly Agree
Presentation It is important to present the educational content appropriately throughout an educational game in order to keep the player's interest; this includes ensuring that the right teaching method is used to present each piece of content, making sure the content is presented in order of prerequisites, and making sure the player does not feel overloaded with information at any point while playing	Strongly Disagree	Disagree	Agree	Strongly Agree

Gameplay

The following factors relate to keeping players immersed through the gameplay.

Please circle one answer for each factor.

Balanced Challenge A player can become more easily immersed in an educational game if the challenges it presents are balanced to the player's abilities (it does not feel too hard or too easy)	Strongly Disagree	Disagree	Agree	Strongly Agree
Feedback To prevent the player getting too frustrated with an educational game, it is important to provide frequent, helpful feedback on how well they are progressing through the game	Strongly Disagree	Disagree	Agree	Strongly Agree
Guidance In addition, there needs to be appropriate guidance provided to the player when they get stuck, in order to reduce their frustration and keep them focused on the challenge	Strongly Disagree	Disagree	Agree	Strongly Agree
Consequences The player's actions in-game should have a tangible, appropriate impact on the game world	Strongly Disagree	Disagree	Agree	Strongly Agree

<u>Agency</u>

The following factors relate to keeping players immersed through interest and involvement in the game world itself.

Please circle one answer for each factor.

Narrative To immerse the player in an educational game, and keep them engaged, the game must feature a compelling narrative. This includes establishing clear goals for the player, explaining the context of the game world, making sure the story keeps consistent as it progresses, and keeping the story flexible enough so that the player to be able to influence it	Strongly Disagree	Disagree	Agree	Strongly Agree
Curiosity Players become more immersed in a game when the game world and the story stimulates curiosity (e.g. through interesting details within the world, or mysteries and surprises in the narrative)	Strongly Disagree	Disagree	Agree	Strongly Agree
Fantasy Educational games are more immersive if they have a non-realistic ('fantasy') setting, because they can more readily disengage from the real world, and can approach the educational content presented from a new perspective	Strongly Disagree	Disagree	Agree	Strongly Agree
Identity Projection The player becomes more engaged with an educational	Strongly Disagree	Disagree	Agree	Strongly Agree

game if they feel in control of their in-game character, and feel invested in the role they play in the game				
Experimentation Educational games are more engaging when they allow players to experiment with the ideas they are trying to teach (i.e. allow the players to use the educational content to solve challenges, encourage them to try different ideas and understand why they work, or not)	Strongly Disagree	Disagree	Agree	Strongly Agree

Appendix F : Stage 2 - Player Survey Questionnaire

What is your gender?

☐ Male ☐ Female ☐ I prefer not to say

What is your age?

☐ Under 20 ☐ 20-25 ☐ 26-30 ☐ 31-35
☐ 35-40 ☐ Over 40 ☐ I prefer not to say

How many hours a week do you spend playing games, on average? :

☐ Fewer than 5 ☐ 5-10 ☐ 10-15
☐ 15-20 ☐ More than 20

Have you played an 'educational game' before?

☐ Yes ☐ No

Please evaluate the following statements

	Strongly Disagree	Disagree	No opinion	Agree	Strongly Agree
I learn better if I can relate the experiences in an educational game to experiences in real life	1	2	3	4	5
* I learn better if I do not find the objectives of a learning activity important to me	5	4	3	2	1
I find it distracting when an educational task does not feel related to what it is trying to	1	2	3	4	5

teach					
I learn better when each new piece of knowledge builds upon knowledge learned earlier	1	2	3	4	5
I enjoy games that do not feel too easy	1	2	3	4	5
* I enjoy games that feel too hard	5	4	3	2	1
I find feedback on my actions in-game helps me to progress	1	2	3	4	5
I find it easier to progress in a game if I receive guidance when I get stuck	1	2	3	4	5
* I prefer guidance to be given to me automatically, whether or not it is requested	5	4	3	2	1
I find it frustrating when I get stuck on a task, with nothing to help me	1	2	3	4	5

Please evaluate the following statements

I prefer being told how to use new gameplay mechanics, rather than working it out on my own	1	2	3	4	5
I prefer playing games which have clear goals to achieve	1	2	3	4	5
I would be more likely to replay a game if the story is flexible (i.e. no two run-throughs of a game are exactly alike)	1	2	3	4	5
I become less interested in a game if its characters or world have inconsistencies	1	2	3	4	5

I feel more engaged in a game if I care about the character I am playing as	1	2	3	4	5
I feel I learn more when I am engaged in a role I play in a game	1	2	3	4	5
I feel more comfortable playing games in unfamiliar settings	1	2	3	4	5
* I do not feel I learn better when ideas are presented in unfamiliar settings	1	2	3	4	5
I feel I can understand a subject being taught to me if I can experiment with the ideas that are taught	1	2	3	4	5
I feel more engaged in games when using knowledge about the game's story and world to solve problems	1	2	3	4	5
When I play games, I like my actions to have a noticeable impact on the game world	1	2	3	4	5

* Reverse-worded questions

** The numeric values of the categories were not visible to the participants

Appendix G : Stage 2 - Cronbach's Alpha for Questionnaire Items if Deleted (SPSS Output)

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
RV_RelateReal	75.06	27.263	.141	.607
RV_REV_Important	75.38	28.917	-.064	.641
GE_RelateTeach	74.94	28.463	.111	.607
SQ_Builds	74.75	24.867	.456	.562
BC_Easy	74.94	25.663	.444	.570
BC_REV_Hard	75.50	26.800	.126	.613
FB_Feedback	74.94	25.529	.392	.573
GD_Stuck	74.69	28.896	-.001	.618
GD_REV_Auto	75.31	27.029	.215	.597
GD_Frustration	74.88	26.650	.265	.591
GD_HowToUse	75.63	26.517	.313	.586
NV_ClearGoals	74.81	27.763	.101	.611
NV_StoryFlexible	74.94	23.796	.451	.556
NV_Consistent	75.06	26.329	.458	.575
NV_Character	74.75	25.000	.438	.565
NV_Role	74.94	29.663	-.130	.644
SC_UnfamiliarComfort	76.38	27.317	.141	.607
SC_REV_UnfamiliarLearn	76.38	31.317	-.358	.655
EX_Understand	74.81	25.629	.449	.569
EX_Engaged	74.88	29.717	-.135	.633
EX_Consequences	74.56	23.596	.643	.535

Appendix H : Stage 3-4 - IEGM Metrics Questionnaire and Values

What is your gender?

☐ Male ☐ Female ☐ I prefer not to say

What is your age?

☐ Under 20 ☐ 20-25 ☐ 26-30 ☐ 31-35
☐ 35-40 ☐ Over 40 ☐ I prefer not to say

How many hours a week do you spend playing games, on average? :

☐ Fewer than 5 ☐ 5-10 ☐ 10-15
☐ 15-20 ☐ More than 20

	Strongly Agree	**Agree**	**No Opinion**	**Disagree**	**Strongly Disagree**
While playing, I learned about the subject I expected to learn about from the game	5	4	3	2	1
* While playing, I learned about a subject I did not expect to learn about from the game	1	2	3	4	5
I am interested in the subject the game is trying to teach	5	4	3	2	1
I feel what I've learned would be useful to me in real life	5	4	3	2	1
* I do not feel like I learned much about the game's subject matter	1	2	3	4	5
I think this game would be valuable to people who wanted to learn about the subject matter	5	4	3	2	1

	Strongly Agree	Agree	No Opinion	Disagree	Strongly Disagree
* I feel the gameplay is out of place with what the game is trying to teach	1	2	3	4	5
I think the game mechanics reflect similar real-world situations	5	4	3	2	1
I think the game challenges reflect similar real-world situations	5	4	3	2	1
I think the subject matter is shown well through the game challenges	5	4	3	2	1
I find the game to be both fun and educational	5	4	3	2	1
* The educational content made the game boring	1	2	3	4	5

	Strongly Agree	Agree	No Opinion	Disagree	Strongly Disagree
I understood what kind of subject matter the game was attempting to teach	5	4	3	2	1
I found knowledge from previous challenges useful in successive ones	5	4	3	2	1
I felt with each game challenge, the game provided all the information I needed to complete it	5	4	3	2	1
I think the order of the game challenges helped build up my skills	5	4	3	2	1
* There were parts of the subject matter I needed more information to understand	1	2	3	4	5
* What I learned in previous challenges did not help me in successive challenges	1	2	3	4	5

	Strongly Agree	Agree	No Opinion	Disagree	Strongly Disagree
* I found the game to be too easy	1	2	3	4	5
* I found the game to be too difficult	1	2	3	4	5
I enjoyed playing the game	5	4	3	2	1
* I would not recommend this game for its gameplay	1	2	3	4	5
The game challenges felt fair	5	4	3	2	1
* The game felt frustrating to play	1	2	3	4	5

	Strongly Agree	Agree	No Opinion	Disagree	Strongly Disagree
I think the game's feedback helped me to understand how well I was doing	5	4	3	2	1
* I found the in-game feedback to be excessive	1	2	3	4	5
* I found there was too little feedback in the game	1	2	3	4	5
The game let me know the important details I needed to know to progress	5	4	3	2	1
* I was confused throughout the game about how to play it	1	2	3	4	5
When I made mistake, I understood what I was doing wrong	5	4	3	2	1

	Strongly Agree	Agree	No Opinion	Disagree	Strongly Disagree
I was able to find tutorials or hints to help me progress when I needed them	5	4	3	2	1
* Tutorials/hints were presented to me whether I wanted help or not	1	2	3	4	5
I found the game's guidance systems (tutorials, hints) helped me to progress	5	4	3	2	1
* I did not pay much attention to the tutorials	1	2	3	4	5
* The game's tutorials got annoying	1	2	3	4	5
* I feel the game should have let me discover more for myself	1	2	3	4	5

	Strongly Agree	Agree	No Opinion	Disagree	Strongly Disagree
I thought it was clear what I had to do to complete the game challenges	5	4	3	2	1
The game's story makes sense to me	5	4	3	2	1
I found myself interested in the characters	5	4	3	2	1
I thought my actions in-game had a noticeable impact on the story	5	4	3	2	1
* I did not care about the story	1	2	3	4	5
* I found the characters to be inconsistent	1	2	3	4	5

	Strongly Agree	Agree	No Opinion	Disagree	Strongly Disagree
* I found the game's setting to be familiar to me	1	2	3	4	5
I found the game's setting similar to related real-world situations	5	4	3	2	1
I felt the game's setting helped me to engage more with the game's subject matter	5	4	3	2	1
* The setting distracted from the subject matter	1	2	3	4	5
* I was bored by the game environment	1	2	3	4	5
I felt involved in the role the game asked me to perform	5	4	3	2	1

	Strongly Agree	Agree	No Opinion	Disagree	Strongly Disagree
I felt able to experiment with the game mechanics introduced	5	4	3	2	1
I was able to experiment with what I learned about the game's subject matter	5	4	3	2	1
I felt I could experiment without permanently impacting the game environment	5	4	3	2	1
* I did not have much time to think about what I should do to complete the game challenges	1	2	3	4	5
The game gave me enough opportunities to retry challenges if my attempts did not work	5	4	3	2	1
I was able to identify what strategies work in particular situations	5	4	3	2	1

* Reverse-worded questions

** The numeric values of the categories were not visible to the participants

Appendix I : Stage 3 - Experiment 1 & 3 Questionnaire Codes

Factor	Questionnaire Code	Question
Relevance	RV_ExpectedLearn	While playing, I learned about the subject I expected to learn about from the game
	RV_REV_NotExpectedLearn	While playing, I learned about a subject I did not expect to learn about from the game
	RV_InterestedInSubject	I am interested in the subject the game is trying to teach
	RV_LearnedUseful	I feel what I've learned would be useful to me in real life
	RV_REV_NotLearnedMuch	I do not feel like I learned much about the game's subject matter
	RV_ValuableForSubject	I think this game would be valuable to people who wanted to learn about the subject matter
Gameplay-Educational Integration	GE_REV_GameplayOutOfPlace	I feel the gameplay is out of place with what the game is trying to teach
	GE_MechanicsReflectRealWorld	I think the game mechanics reflect similar real-world situations
	GE_ChallengesReflectRealWorld	I think the game challenges reflect similar real-world situations
	GE_ChallengesShowSubject	I think the subject matter is shown well through the game challenges
	GE_FunAndEducational	I find the game to be both fun and educational
	GE_REV_EducationalBoring	The educational content made the game boring

Sequencing	SQ_UnderstoodSubjectMatter	I understood what kind of subject matter the game was attempting to teach
	SQ_PrevChallengeKnowledgeUseful	I found knowledge from previous challenges useful in successive ones
	SQ_NecessaryInfoProvided	I felt with each game challenge, the game provided all the information I needed to complete it
	SQ_ChallengeOrderBuiltSkills	I think the order of the game challenges helped build up my skills
	SQ_REV_SubjectNotEnoughInfo	There were parts of the subject matter I needed more information to understand
	SQ_REV_PrevChallengesNotHelpful	What I learned in previous challenges did not help me in successive challenges
Balanced Challenge	BC_REV_TooEasy	I found the game to be too easy
	BC_REV_TooDifficult	I found the game to be too difficult
	BC_EnjoyedPlaying	I enjoyed playing the game
	BC_REV_NotRecommendGameplay	I would not recommend this game for its gameplay
	BC_ChallengesFair	The game challenges felt fair
	BC_REV_Frustrating	The game felt frustrating to play
Feedback	FB_UnderstandProgress	I think the game's feedback helped me to understand how well I was doing
	FB_REV_Excessive	I found the in-game feedback to be excessive

	FB_REV_TooLittle	I found there was too little feedback in the game
	FB_ImportantProgressionDetailsProvided	The game let me know the important details I needed to know to progress
	FB_REV_ConfusedHowToPlay	I was confused throughout the game about how to play it
	FB_UnderstoodMistake	When I made mistake, I understood what I was doing wrong
Guidance	GD_TutorialsHintsFound	I was able to find tutorials or hints to help me progress when I needed them
	GD_REV_PresentedWhenUnwanted	Tutorials/hints were presented to me whether I wanted help or not
	GD_HelpedProgress	I found the game's guidance systems (tutorials, hints) helped me to progress
	GD_REV_PaidLittleAttention	I did not pay much attention to the tutorials
	GD_REV_Annoying	The game's tutorials got annoying
	GD_REV_MoreDiscoveryBySelfNeeded	I feel the game should have let me discover more for myself
Narrative	NV_ChallengeObjectivesClear	I thought it was clear what I had to do to complete the game challenges
	NV_StoryMadeSense	The game's story makes sense to me
	NV_InterestedInCharacters	I found myself interested in the characters
	NV_ActionsImpactedOnStory	I thought my actions in-game had a noticeable impact on the story

	NV_REV_DidNotCareAboutStory	I did not care about the story
	NV_REV_CharactersInconsistent	I found the characters to be inconsistent
Scenario	SC_REV_SettingFamiliar	I found the game's setting to be familiar to me
	SC_SimilarToRealWorld	I found the game's setting similar to related real-world situations
	SC_HelpedEngageWithSubject	I felt the game's setting helped me to engage more with the game's subject matter
	SC_REV_DistractedFromSubject	The setting distracted from the subject matter
	SC_REV_Bored	I was bored by the game environment
	SC_InvolvedInRole	I felt involved in the role the game asked me to perform
Experimentation	EX_AbleToExperimentMechanics	I felt able to experiment with the game mechanics introduced
	EX_AbleToExperimentSubject	I was able to experiment with what I learned about the game's subject matter
	EX_NoPermanentImpact	I felt I could experiment without permanently impacting the game environment
	EX_REV_NotEnoughTimeToPlan	I did not have much time to think about what I should do to complete the game challenges
	EX_EnoughRetries	The game gave me enough opportunities to retry challenges if my attempts did not work

	EX_AbleToIdentifyStrategies	I was able to identify what strategies work in particular situations

Appendix J : Stage 4 - Cronbach Alpha for Questionnaire Items if Deleted (SPSS Output)

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
RV_ExpectedLearn	195.71	474.726	.455	.927
RV_REV_NotExpectedLearn	197.02	501.882	-.308	.933
RV_InterestedInSubject	195.94	464.291	.633	.925
RV_LearnedUseful	196.35	463.806	.547	.926
RV_REV_NotLearnedMuch	196.37	466.919	.505	.926
RV_ValuableForSubject	195.81	471.518	.511	.926
GE_REV_GameplayOutOfPlace	196.28	465.380	.559	.926
GE_MechanicsReflectRealWorld	196.59	467.068	.439	.927
GE_ChallengesReflectRealWorld	196.56	471.732	.326	.928
GE_ChallengesShowSubject	195.99	467.376	.590	.926
GE_FunAndEducational	195.87	461.595	.647	.925
GE_REV_EducationalBoring	195.90	468.401	.549	.926
SQ_UnderstoodSubjectMatter	195.52	479.994	.301	.927
SQ_PrevChallengeKnowledgeUseful	195.85	472.718	.475	.926
SQ_NecessaryInfoProvided	195.85	468.859	.501	.926
SQ_ChallengeOrderBuiltSkills	195.67	472.857	.510	.926
SQ_REV_SubjectNotEnoughInfo	196.98	464.847	.464	.926
SQ_REV_PrevChallengesNotHelpful	195.84	474.538	.423	.927
BC_REV_TooEasy	196.86	478.216	.183	.929
BC_REV_TooDifficult	196.17	481.275	.169	.928
BC_EnjoyedPlaying	196.03	464.034	.668	.925
BC_REV_NotRecommendGameplay	196.52	464.958	.452	.926
BC_ChallengesFair	196.08	475.511	.410	.927
BC_REV_Frustrating	196.45	460.651	.541	.926
FB_UnderstandProgress	195.94	476.032	.384	.927
FB_REV_Excessive	196.72	470.321	.378	.927
FB_REV_TooLittle	196.13	477.595	.320	.927
FB_ImportantProgressionDetailsProvided	195.85	476.600	.369	.927
FB_REV_ConfusedHowToPlay	196.19	472.271	.349	.927
FB_UnderstoodMistake	196.33	468.081	.486	.926
GD_TutorialsHintsFound	196.24	464.963	.528	.926
GD_REV_PresentedWhenUnwanted	197.63	488.072	-.014	.930
GD_HelpedProgress	195.85	466.883	.621	.925
GD_REV_PaidLittleAttention	196.27	466.881	.479	.926
GD_REV_Annoying	196.56	460.414	.534	.926
GD_REV_MoreDiscoveryBySelfNeeded	196.87	469.172	.412	.927

NV_ChallengeObjectivesClear	195.79	478.897	.331	.927
NV_StoryMadeSense	195.79	469.273	.528	.926
NV_InterestedInCharacters	196.29	458.420	.641	.925
NV_ActionsImpactedOnStory	196.52	466.817	.501	.926
NV_REV_DidNotCareAboutStory	196.27	457.281	.641	.925
NV_REV_CharactersInconsistent	195.99	472.506	.502	.926
SC_REV_SettingFamiliar	197.07	497.642	-.204	.933
SC_SimilarToRealWorld	196.81	469.494	.404	.927
SC_HelpedEngageWithSubject	196.00	474.212	.464	.927
SC_REV_DistractedFromSubject	196.06	473.561	.430	.927
SC_REV_Bored	196.03	463.352	.663	.925
SC_InvolvedInRole	196.10	463.389	.683	.925
EX_AbleToExperimentMechanics	196.10	467.036	.568	.926
EX_AbleToExperimentSubject	196.31	471.324	.431	.927
EX_NoPermanentImpact	196.20	479.737	.201	.928
EX_REV_NotEnoughTimeToPlan	196.58	464.270	.546	.926
EX_EnoughRetries	196.08	474.452	.424	.927
EX_AbleToIdentifyStrategies	196.10	473.836	.392	.927

Appendix K : Stage 4 - Factor Analysis: PCA Component Matrix and Correlation Matrix

Component Matrix

	Component	
	1	2
Relevance	.741	-.361
GE_Integration	.828	-.267
Sequencing	.767	.372
BalancedChallenge	.777	-.320
Feedback	.667	.578
Guidance	.741	.064
Narrative	.797	-.200
Scenario	.778	-.146
Experimentation	.708	.413

Rotated Component Matrix

	Component	
	1	2
Relevance	.805	.176
GE_Integration	.816	.304
Sequencing	.371	.767
BalancedChallenge	.809	.230
Feedback	.166	.867
Guidance	.542	.509
Narrative	.749	.336
Scenario	.702	.367
Experimentation	.300	.763

Correlation Matrix

	RV	GE	SQ	BC	FB	GD	NV	SC	EX
Relevance (RV)	1.000	.611	.491	.577	.306	.478	.596	.581	.382
GE_Integration (GE)	.611	1.000	.487	.673	.398	.574	.685	.635	.521
Sequencing (SQ)	.491	.487	1.000	.475	.649	.479	.527	.550	.592
BalancedChallenge (BC)	.577	.673	.475	1.000	.408	.547	.629	.584	.359
Feedback (FB)	.306	.398	.649	.408	1.000	.505	.437	.387	.544
Guidance (GD)	.478	.574	.479	.547	.505	1.000	.503	.483	.492
Narrative (NV)	.596	.685	.527	.629	.437	.503	1.000	.536	.471
Scenario (SC)	.581	.635	.550	.584	.387	.483	.536	1.000	.514
Experimentation (EX)	.382	.521	.592	.359	.544	.492	.471	.514	1.000