

School of Design

An Investigation of Learning Game Design: Scoring Activity-Goal Alignment

Josh Whitkin

**This thesis is presented for the Degree of
Doctor of Philosophy
of
Curtin University**

November 2013

To the best of my knowledge and belief this thesis contains no material previously published by any other person except where due acknowledgment has been made. This thesis contains no material which has been accepted for the award of any other degree or diploma in any university.

Abstract

This research aims to improve the practice of designing educational video games (“learning games”). This thesis aims to both validate and extend Shelton’s theory of activity-goal alignment, which focuses on the relationship between a player’s activity and the designer’s intended learning goal in any learning game. The thesis develops and evaluates two novel tools. First, an autoethnographic account of a recent learning game project confirms Shelton’s prior findings that activity-goal alignment theory meets an important need in learning game design practice and that Shelton’s theory might be made more accessible to practising designers. The AGA Scoring Tool is developed, and both it and Shelton’s theory are evaluated through analytic discussions of designs of several existing learning games: activity-goal alignment theory is found useful, and scoring activity-goal alignment is argued to be clearer than Shelton’s narrative-based approach. Secondly, this thesis argues that there is need for improved tools for assessment of learning games. A critical review of existing assessment tools yields a list of criteria for any learning game assessment tool. A basis for a new learning game assessment tool is developed from three theories: Higher Order Learning theory, Gee’s principles of Deep Learning, and Shelton’s activity-goal alignment. These three theories are argued to comprise an important, prevailing position within the learning game design literature. A new tool, the AGA-Based Assessment Tool, is proposed and exercised in critical discussions of several learning games. Important gaps between learning game design practice and theory are revealed using the tool. The thesis concludes that scoring activity-goal alignment is useful to the learning game designer because it makes an important theoretical position from the learning game design literature clear and simpler to apply in practice.

Table of Contents

Abstract	3
Table of Contents	4
Acknowledgements	6
1 Introduction	7
1.1 Need.....	7
1.2 Introduction to Activity-Goal Alignment	9
1.3 Significance	10
1.4 Definitions of Terms and Concepts	14
1.5 Method	17
1.6 Findings.....	18
1.7 Limitations of This Research.....	18
1.8 Outline	20
1.9 Conclusion.....	22
2 Literature Review.....	23
2.1 Introduction	23
2.2 Conceptions of Learning.....	28
2.3 Play and Learning: A Fallacious Tension	35
2.4 Video Game Literature around the Concept of “Mechanic”	38
2.5 Hierarchies of Player Experience.....	43
2.6 Fidelity in Simulators.....	46
2.7 Learning Game Design: What Makes a Good Learning Game?	49
2.8 Deep Learning and Higher Order Thinking	51
2.9 Conclusion.....	55
3 Method	56
3.1 Introduction	56
3.2 Position within Design Research	57
3.3 Autoethnography	61
3.4 Analytic Emphasis	67
3.5 Ethical Considerations.....	69
3.6 Contribution to Literature.....	73
4 Scoring Activity-Goal Alignment	76
4.1 Introduction	76
4.2 The Story of <i>Battlefood</i> 's Development	76
4.3 Modes of Communication around Activity-Goal Alignment.....	84
4.4 Development of AGA Scoring Tool	88
4.5 AGA Scoring Tool V1.....	93

4.6	How to Use the AGA Scoring Tool	96
4.7	Discussion of AGA Scoring Tool Design	97
5	Using Activity-Goal Alignment Scores in Analysis	99
5.1	Introduction	99
5.2	Human Cellular Biology	100
5.3	Game Design	122
5.4	Discussion	134
5.5	Summary of Findings.....	135
5.6	Conclusion.....	139
6	Review of Learning Game Design Assessment	140
6.1	Introduction	140
6.2	Prevailing Advice in Learning Games: Aim Higher	140
6.3	Existing Assessment Tools.....	145
6.4	Application of Selected Existing Assessment Tools on <i>Math Blaster</i> series	157
6.5	Criteria for Assessment Tools for Amateur Designers.....	168
6.6	Conclusion.....	169
7	Activity-Goal Alignment-Based Assessment of Learning Games	170
7.1	Introduction	170
7.2	Development of the AGA-Based Assessment Tool	170
7.3	Activity-Goal Alignment-Based Assessment Of Three Math Games.....	179
7.4	Discussion	188
7.5	Conclusion.....	197
8	Discussion and Conclusion	198
8.1	Introduction	198
8.2	Significance	199
8.3	Questions for Further Research	204
8.4	Conclusion.....	205
9	Bibliography.....	206

Acknowledgements

As I have scrambled along the shoulders of giants in this doctoral journey, my path has been guided by experts and belayed by loved ones from many parts of my life. My heartfelt gratitude goes first to my advisor, mentor, colleague, and friend, Andrew Hutchison. It is no exaggeration to say that without his passionate and convincing argument for the feasibility of this challenge, I would not have even attempted this thesis. His advocacy and guidance has been priceless. Helen Merrick has earned my eternal gratitude for her disarming clarity, brilliance, and gentle guidance, somehow spearing pragmatism and vision from the stream of ideas that lay under this thesis. I am grateful to my many supporters and mentors in my unusual academic journey, notably including Suzette Worden, Philip Howard, Gina Neff, Ingrid Richardson, Mark Cypher, and Chris Smyth, as well too many others to name here have collectively given academics a good name. My parents Lee White and Kerry Silvey, my brother Joby White, and my parents-in-law Stephanie and Brian Jenkin have contributed steady encouragement, support, sound advice, and time, the most precious of resources to this breadwinner/student/parent. To Helen Whitkin, my dear lifelong companion, I say simply: thank you. Above all others, I would not be who I am without you.

1 Introduction

1.1 Need

In this research, I score activity-goal alignment to improve the design practice and assessment of learning games. I discovered the need for this research during my professional work as a designer of *Battlefood*, a video game that aimed to teach players about nutrition. I noticed that certain player activities were inseparable from the learning goal. In the literature, I found one theory that fit my observations: activity-goal alignment theory (Shelton 2007)¹. Prior to discovering Shelton's theory, I tacitly understood the idea of activity-goal alignment, but without this particular term ("activity-goal alignment"), conversations around this topic were entangled in various related ideas from the literature. As a practising designer reviewing our struggle as a team, I felt the development process of *Battlefood* would have been better if we had specifically discussed our game in terms of activity-goal alignment.

Shelton found that activity-goal alignment theory can benefit designers in their own game design practice. While I agreed, I envisioned an improved method of analysis. This led to my research question:

Q1: How might practising learning game designers and researchers find it useful to score activity-goal alignment theory of a learning game?

In this research I develop the "AGA Scoring Tool," a simpler, more visual form of analysis than Shelton's discussion-based approach. The acronym AGA refers to the term "activity-goal alignment." In chapter 5 I discover the tools' usefulness during critical analysis of several learning games, and conclude that scoring activity-goal alignment can help designers improve their understanding of learning games. As I developed the AGA Scoring Tool, I noticed that this research could address a larger need: improving methods of assessment of learning games.

As mentioned in the Definitions section of this chapter, I use the term "assessment" when I evaluate **the games, not their players**. This can be confusing because in learning game literature (D'Angelo,

¹ Other theorists have proposed similar ideas (B. Winn and Heeter 2008; Malone 1981) but I chose Shelton's theory because it was best related to foundational theory while being grounded in practice, and most clearly stated the importance of this particular relationship over other design issues (e.g. single vs multiple players). I compare similar theories in Chapter 2, and discuss the merits of Shelton's approach later in this chapter.

Rutstein, and Harris 2013; Squire and Jenkins 2003), the unqualified term “assessment” typically refers the **player’s** performance. This research does not address player assessment. In this research I always assess the quality of the **learning game**, not the abilities of its players.

1.1.1 The Need for Improved Assessment of Learning Games

What makes a good learning game? While other educational technologies have well-developed assessment tools and methods, Rice found learning games largely do not (Rice 2007). Rice’s finding was supported by De Freitas and Oliver who stated: "There is, at present, an over-reliance upon using available methods of evaluating leisure-based games" (S. De Freitas and Oliver 2006, 3). The need for improved methods of assessment can additionally be seen in Kissane’s difficulty in interpreting traditional critical reviews of learning games and software when evaluating the iPod Touch as a tool for mathematics education (Kissane 2011). Kissane noted that evaluative criteria were often not supplied, so teachers have no clear way to decide which game is better. Members of the IGDA’s Learning and Education Group recently discussed the need for improved assessment on the SIG’s private email list (IGDA LEG SIG 2012). This evidence makes it clear that improving learning game assessment is needed.

There are many varied existing methods of assessment, some of which I review in chapter 6, where I find important weaknesses. Specifically, a “list of attributes” approach is common, but I argue this approach has important drawbacks for the practising designer/researcher. In this research, I develop an alternative approach which is expressed in the thesis’s second research question:

Q2: How might combining the notions of scoring activity-goal alignment and Deep Learning offer useful new approaches to learning game assessment?

To address this question, I combine Shelton’s activity-goal alignment, Rice’s conception of higher order learning (Rice 2007), and Gee’s six principles of deep learning (Gee 2009a) in the “AGA-Based Assessment Tool,” which uses a 2 x 2 matrix to assign a learning game into one of four categories. Through discussion of both exemplary and flawed learning games, I show how this tool is useful to the practising designer/researcher in chapter 7. I explain how this tool’s design aims to improve on a common weakness among many existing assessment tools. Finally, in a broader discussion of the game design literature, I ask why certain learning game designers do not following well-established design advice and suggest that learning game literature, as a whole, might work better for the practising designer/researcher if it were simpler and clearer, even at the expense of

comprehensiveness. I argue that the AGA-Based Assessment Tool is an example of such an alternate approach.

1.2 Introduction to Activity-Goal Alignment

In this section, I present a summary of the central conceptual framework of the thesis: Shelton's theory of activity-goal alignment². To understand activity-goal alignment theory, it is useful to discuss its three component words: activity, goal, and alignment.

- **Activity** refers to video game player's action (e.g. driving, flying, shooting, and puzzle-solving) within the game³. All video games have some kind of activity.
- **Goal** refers to an educational purpose (e.g. arithmetic, systems thinking, history). All learning games have a learning goal.
- **Alignment** can be seen in the Venn diagram in Figure 1. The degree of overlap of the two circles illustrates how related the two entities are (Venn 1880). Alignment theory is discussed in detail in chapter 2.

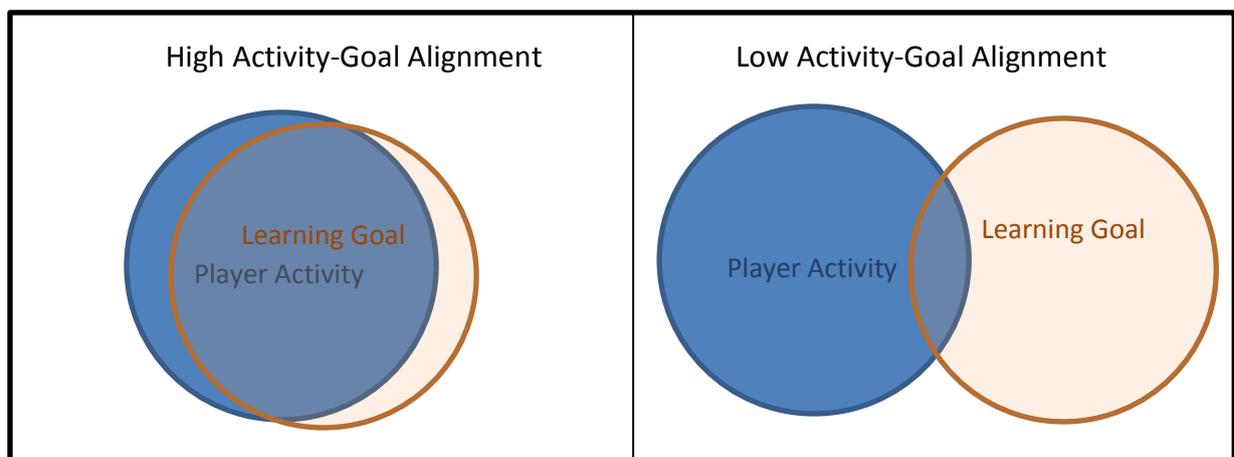


Figure 1: Activity-goal alignment theory states that any learning game's activity and learning goal may strongly overlap ("High") or weakly overlap ("Low").

Essentially, activity-goal alignment theory considers the alignment between the player's activity (e.g. driving, flying, shooting) and the learning goal (e.g. arithmetic, systems thinking, history).

² Activity-goal alignment is explored in detail in chapter 4.

³ Among video game designers, the term "core mechanic" is a common term that closely relates to the term "activity." This term "mechanic" is discussed later in this chapter and explored in more detail in chapter 2.

For example, imagine an instructional driving game, where the player learns to drive a car by playing a 3D simulation of normal driving on public roads. Such a game has high activity-goal alignment because the activity of driving the car in the game resembles a real-world driving experience. By contrast, a road safety quiz has low activity-goal alignment, because the activity of taking a test and recalling facts is not very similar to driving a car.

Shelton argued that activity-goal alignment is an important decision in any learning game's design. I agree, arguing that Shelton has done a commendable job focusing attention on this single, crucial relationship. In chapter 7 I argue that the practising designer faces a barrier: the typical modality of learning game design theory, including Shelton's theory, is text. To continue Shelton's work of clarifying and simplifying, this research aims to represent the activity-goal alignment of a learning game in a form other than textual discussion. Specifically, I score activity-goal alignment using the "AGA Scoring Tool," which is a simple, quick process that generates a clear, easy to understand Venn diagram and few paragraphs of text. In this way, tool is designed to suit the needs of busy professional learning game designers.

Returning to the driving example, it is important to note that a high AGA score does not always mean a more effective learning game. Both driving quizzes and simulators are valuable forms of education, depending on the learning goal. For example, if one needs to memorize stopping distances to pass a written driving test, a quiz may be more effective than a simulator. In other words, a game with a high AGA score is not always superior to a low scoring game. I discuss the danger of misinterpreting the meaning of the AGA score in chapter 5, and provide the AGA-Based Assessment Tool, which relates a game's AGA score to its quality, in chapter 8.

1.3 Significance

To show how this research is significant, I argue that effective learning games are needed, that learning game design research can improve the success of learning games, and that activity-goal alignment is an important approach.

The idea that video games today are a mainstream medium, on par with television or film, informs the significance of this research. The opportunity of learning games, as well as the importance of understanding video game design comes from their ubiquity. Around the world, rich or poor, young or old, most people play video games, mostly with other people (Entertainment Software Association 2011). 97% of US teens play video games (Lenhart, Project, and John 2008). \$24 billion was spent on video games in 2011. Seven out of the ten top-selling iOS apps are games (Burns,

2011). On Facebook in 2012, games have over 254 million monthly active users worldwide, far more than any other type of application (AppData 2012).

It is easy to forget how diverse video games are. Most people accept that games are common when they reflect on the “casual” gaming market, which is a significant part of most social media user experiences, and the top-selling category of applications on all major mobile phones. However, even if one focuses solely on dedicated gaming devices like the Sony PlayStation and Microsoft Xbox and gaming PCs, one might be surprised at the diversity in the top-selling game designs. Action games on these platforms get a lot of media attention, but consume less than half of playtime. Online games are not very different: Figure 2 shows how little play time online games consume.

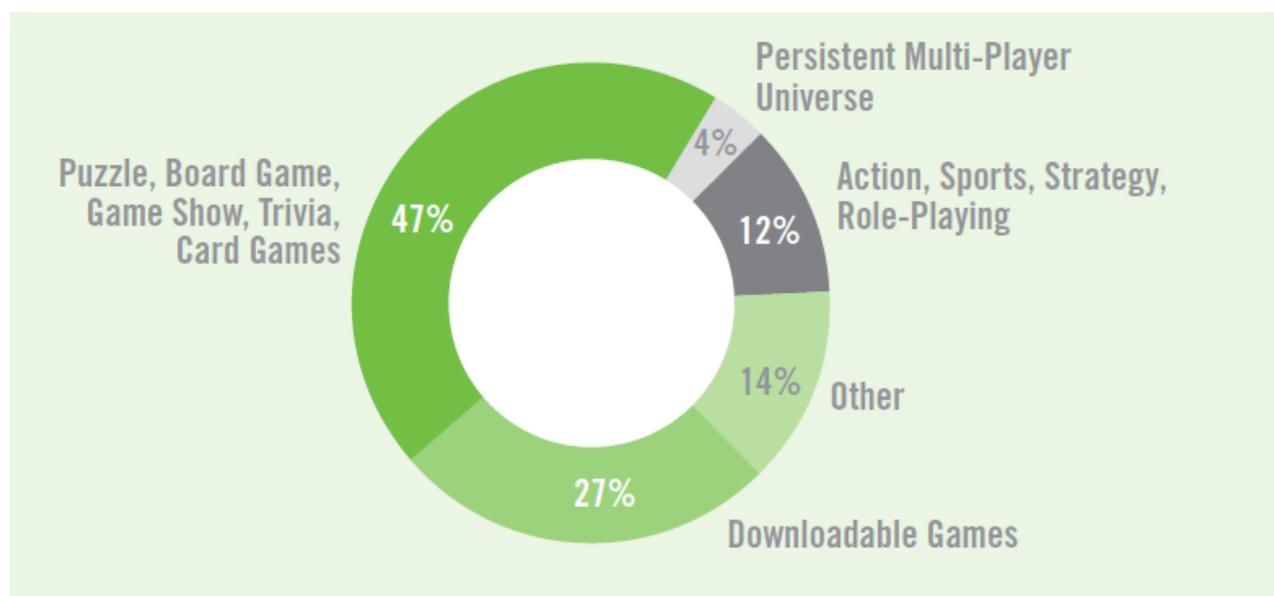


Figure 2: Types of online games played most often. (Entertainment Software Association 2011)

In 2012 the Obama Administration stated that “harnessing the power of interactive games” is part of their “Educate to Innovate” campaign, which aims to use educational technology to improve science, technology, engineering, and mathematics (STEM) learning outcomes (White House 2012).

Marketing consultant MC Milker notes that developers of educational games and toys comprise a \$1.1 billion consumer market (Milker 2010). This large, healthy market provides support for the idea that educational games are important to the typical modern parent. Additional evidence comes from sophisticated consumers of learning games who have found learning games useful in the field. For example, the CEO of a human rights organization found games are unique instrument for social change because they allow players to experience famines, disasters, refugee camps, and other scenarios. She argues that the development cost of a video game is not significant given games' ability to reach a broad audience (Oke, Dutt, and Treat 2012). Similarly, the World Food Programme

describes their 2005 learning game "Food Force" as hugely successful and continues to develop sequels (Deepend 2005).

In short, learning games are important: they are a large healthy market that is believed to be an important part of education by parents, scholars and national leaders.

Broadly speaking, learning game research is significant. The field of learning game research has enjoyed renewed interest recently, including millions of dollars in funding worldwide from private donors, government grants, and private equity (Ferenstein 2010). There is broad support among advocates and critics alike for further learning game research. Respected institutions such as MIT (Klopfer, Osterweil, and Salen 2009) and noted scholars (Gee 2006) have argued that learning games have great potential that has not yet been met.

1.3.1 Is Design Important To Learning Game Success?

The previous argument established the importance of learning game research but did not specify a type of research. Since product design is a key determinate for any sort of product research (Bloch 1995), and learning games are obviously a product, it follows that learning game design research is significant.

A more difficult question is: Which problems in learning game design research are worthy of research? While others have convincingly argued that one particular problem in learning game design is assessment of learning games, and this is an important, unsolved problem that requires further research (Rice 2007). Another commonly discussed type of problem relates to activity-goal alignment, although prior theorists have used many other terms, as I explain in the literature review. Many have argued that this problem is very important to learning game design research, and I make this argument in the following section.

1.3.2 Is the Activity-Goal Relationship a Problem Worthy of Research Efforts?

Why should anyone care about activity-goal alignment theory, let alone my attempts to improve upon it? To answer this question, I argue three points:

- a. The relationship between activity and goal is an important problem
- b. Activity-goal alignment theory has improved the prior understandings of the activity-goal relationship
- c. This research can further improve on activity-goal alignment theory

I use the phrase "activity-goal **relationship**" to mean the relation between a player's activity within a game and the overall learning goal of the game. For example, a learning game might provide a player activity⁴ of exploring an abandoned cave, and a goal of teaching long division. How does a spelunking activity relate to the mathematical learning goal?

The established understanding of activity-goal relationship is not tidy or simple. In the literature review in chapter 2, I show how, from a variety of perspectives and through various research fields, the idea has evolved over twenty years. The basic concept is not recent: Piaget, an eminent psychology researcher, noted that, when developing curricula for children utilizing play, activity and goal must be balanced. Piaget stated that an over-emphasis on free activity can lead to aimless play, while an over-emphasis on the learning goal can lead to a curriculum that most students fail to understand (Piaget 1970). Thomas Malone, an important learning game researcher, used the lens of player choice⁴ to address Piaget's observation, and found that it did explain the relationship between activity and goal: "it is not clear how to structure educational environments in which free choice leads to productive learning" (Malone 1981, 339). As research continued, improved lenses such as fidelity (A. L. Alexander et al. 2005) were able to partially explain the relationship.

However, despite this literature and numerous calls to aim higher in learning game design (Gee 2009a; Devlin 2012a), many learning games continue to exhibit low activity-goal alignment. Some designers are intentionally choosing this, and some theorists are attacking this choice, as I discuss in chapter 7.

While most theorists agree that activity-goal alignment is important, a sceptical reader might note the scarcity of learning games with high activity-goal alignment and ask if the alleged relationship between activity and goal has sufficient value to justify continued research. I argue that our understanding of activity-goal relationship is far from complete, and while progress has been made, the topic merits further research. Learning game designer-researchers Klopfer, Osterweil, and Salen (2009) repeatedly advise designers to deepen their understanding of game activity in relation to learning outcomes. Of their fourteen points of advice to designers, three of them are relate to the activity-goal relationship:

- "Choose wisely" principle, p.28
- "Put learning and game play first", p. 31

⁴ The term "activity" relates closely to, but is not identical to, the concept of "game mechanic" from the game design literature. This relationship is explained in the Definitions section in this chapter, and the discussion of "mechanic" in the literature review in Chapter 2.

- “Find the game in the content”, p.31

Similar advice regarding this particular issue can be seen in other highly regarded learning game designer-researchers (Tate, Haritatos, and Cole 2009) and theorists (Gee 2009a).

The preceding literature fits with my own experience, as I discuss in chapter 4. For these reasons, I assert that research that usefully improves our understanding of activity-goal relationship is significant. Since Shelton’s activity-goal alignment theory specifically addresses this relationship, it follows that it is significant. If this research makes activity-goal alignment theory more useful, as I argue in this thesis, then this research is significant. Furthermore, Shelton himself argued the significance and worthiness of activity-goal relationship as a research topic in his literature review (2007), explaining how activity-goal alignment combines, builds on, and addresses related work in the field.

To summarize, this argument for the significance of this research has three key points:

- 1) Successful learning games are needed.
- 2) Learning game design research can improve the success of learning games.
- 3) The problem of activity-goal alignment is important to learning game design research.

If these three points are true, then improvements to activity-goal alignment are significant. In chapter 8, I discuss how this research usefully contributes to activity-goal alignment theory, which suggests that this research is significant. I also show how this thesis might contribute to other theory in, as well as practice of, learning game design.

1.4 Definitions of Terms and Concepts

Terminology is especially problematic in learning game research. Aldrich (2009) called it the “Babel problem,” noting the constantly evolving, overlapping definitions of existing terms as well as the frequent coining of terms that are not adopted beyond their originating papers. Possibly in recognition of the Babel problem, some notable researchers⁵ have begun intentionally using broad, imprecise definitions and coining new terms sparingly. In this research I follow this precedent.

Here are the terms and ideas most essential to this thesis.

⁵ Clear examples of thought leaders in learning games who use broad imprecise definitions and avoid coining terms are noted designer/researcher Jesse Schell, in *The Art of Game Design: A Book of Lenses* (Schell 2008), and activity-goal alignment authors Shelton and Scoresby (2010).

1.4.1 Game

In this thesis, a “game” or “video game” is any sort of digital or electronic game played on a variety of platforms (PC, console, mobile device, and arcade) for entertainment purposes. This definition includes single-player and multi-player games, played online and offline. It specifically excludes games that focus on betting or gambling real money such as slot machines or online poker. I call these “gambling games.” Also, board games, playground games, or other games played without a digital device are referred to as “non-digital games”.

This thesis does not address non-digital games since their purpose, context, and modality are so different from typical video games.

1.4.2 Learning Game

I use Klopfer et al’s term “learning game” throughout this research. Their definition follows:

Learning Games are differentiated from Games for Training in that they target the acquisition of knowledge as its own end and foster habits of mind and understanding that are generally useful or useful within an academic context. Learning Games may be associated with formal educational environments (schools and universities online or off), places of informal learning (e.g. museums), or self-learners interested acquiring new knowledge or understanding. Games in this space include MIT/Maryland Public Television’s math and literacy game Lure of the Labyrinth, and the Federation of American Scientist’s Immune Attack, as well as COTS games like Civilization, Rollercoaster Tycoon, and SimCity.” (Klopfer, Osterweil, and Salen 2009, 21)

In other words, a learning game is a video game that seeks to teach or inform the player. The term “learning game” is roughly equivalent in meaning to “educational games,” “edutainment,” “digital game-based learning,” “game based education,” or “games for education.” The term “learning game” is not identical to “serious game,” which Abt (2002) defined as a video game not played primarily for amusement. The term “games for training” refers to a type of learning game.

Learning games belong to a larger category of software defined as “e-learning,” which includes a wide variety of offline and online teaching tools. Khan Academy (Khan 2011) is an example of e-learning. This research is not equally relevant to all types of learning games. Here, the term “learning game” is intended to exclude certain types of games that might teach, as explained later in this chapter.

Shelton (2007) and others (Klopfer, Osterweil, and Salen 2009) note it is a common mistake to consider “learning games” as homogeneous when they are in fact very diverse. Breuer and Bente

(2010) suggest learning games can be understood by dividing them along important boundaries. While many such attempts have been made, using Venn diagrams (Martens, Diener, and Malo 2008), hierarchies (Aldrich 2009), and simple categorization (Prensky 2004), Sawyer's taxonomy of learning games (Sawyer and Smith 2008) provides specific examples and names for categories of learning games. These examples reinforce Shelton's point: learning games are highly dissimilar. Ideas and theory from one context, age range, and purpose often don't apply to other, a defining trait of design research (Cross 1982).

1.4.3 Core Mechanic

The term "mechanic" has been used for over thirty years in the game design literature by a variety of practitioners and theorists. In chapter 2 I present a variety of definitions, but generally in this thesis I use Hunicke et al's definition: "the various actions, behaviours and control mechanisms afforded to the player within a game context" (Hunicke, LeBlanc, and Zubek 2004). For example, betting is a mechanic in poker, jumping over obstacles is a mechanic in side-scrolling video games like *Mario*. The term "core mechanic" means the primary mechanic in a game. For example, there are a few word puzzle mechanics in adventure game X, but the core mechanic is exploration.

1.4.4 Alignment

In the education literature, the term "alignment" refers to the agreement between teaching objectives, activities, and assessments so they are mutually supportive (Tyler 1950). In Shelton's use of the term, the alignment component is simple. If the player's activity matches the learning goal, they are aligned (Shelton and Scoresby 2010).

1.4.5 Higher Order Thinking and Learning

Newmann (1990) described "higher order thinking" as "reasoning, problem-solving, critical thinking and creative use of the mind." This definition fits a number of related concepts, including metacognition (E-Line Media 2007) and situated learning (Gee 2008). Higher order learning refers to mastering a holistic understanding of complex systems (Rice 2007). Gee's "Deep Learning" is a similar concept, as I discuss in chapters 2 and 7. Assessment of higher order learning is often difficult, as mastery can be subtle, complex, and difficult to measure (Devlin 2012a). Lower order learning is the inverse of higher order learning, and refers to the ability to perform simple calculation and recall facts.

1.4.6 Assessment

In this research, I use the term “assessment” when I evaluate **the games, not their players**. This term can be confusing because in learning game literature (D’Angelo, Rutstein, and Harris 2013; Squire and Jenkins 2003), “assessment” often refers to measuring **player** performance. Player performance is beyond the scope of this thesis. Here, I assess the games, not their players.

1.5 Method

This thesis is a form of design research (Cross 1982) that uses analytic autoethnography (L. Anderson 2006), a form of ethnography. Ethnography is a common choice by leaders in design research (Kelley and Littman 2001), and is advocated for future design research (Frankel 2009). Autoethnography is a form of ethnography that avoids the notorious "participant observation" problem (Tedlock 2005) and has the “Value of Inner Knowing” (Duncan 2008, 30), two benefits that are important and useful for this research. To mitigate the risks of autoethnography, especially the subjectivity typical in narrative autoethnography, I use Anderson’s more analytic approach: “a broad set of data-transcending practices that are directed toward theoretical development, refinement, and extension” (L. Anderson 2006). Autoethnography has been found well suited to formative, exploratory questions in interactive software design (Duncan 2008), which describes this research. I have chosen autoethnography because it gives access to "insider meanings" (L. Anderson 2006) which is essential to truth-finding in subtle, complex problems that are difficult to articulate. This method contrasts strongly with the attribute-isolating approach common in learning game research (Malone and Lepper 1987), as I discuss in chapter 3.

Autoethnography requires the researcher to “belong” to their topic, and in chapter 3 I show my credentials as a practising learning game designer and researcher by discussing my career as an academic and professional video game developer.

In chapter 8, I argue that my methodology might modestly contribute to the literature around analytic autoethnography for two reasons: analytic autoethnography is not a common method in learning game design research, and unlike many autoethnographers who have more broad, vague aims (L. Anderson 2006), I show that analytic autoethnography can address specific aims in design research.

1.6 Findings

This research broadly shows that scoring activity-goal alignment highlights the key points in a designer's necessarily complex understanding of the relation between activity and goal in a learning game. For the practising designer / researcher, I find that scoring activity-goal alignment clarifies the activity-goal alignment in any learning game, which can prevent errors in tacit understandings as well as textual discussions. I discuss this finding in chapter 5. I find that scoring activity-goal alignment usefully extends Shelton's conception of activity-goal alignment.

I also find that activity-goal alignment, combined with other theories, can be used to assess learning games. This approach contributes to the literature in two ways: it provides a new method of assessment, and suggests a new approach to assessment modality: I argue that the spare, diagram-based assessment has certain advantages over checklists and academic texts.

Finally, this research offers two modest contributions to the methodology literature: First, analytic autoethnography as an explicit method is rarely used in learning gamed design research, so any research that uses this method has some value to the field. Secondly, I use autoethnography to focus on a single theory. Because it is not typical that autoethnography is used for a single, specific purpose (Chang 2008), this research serve as a useful model for applying the method to specific research questions.

All of these findings are reviewed in more detail in chapter 8.

1.7 Limitations of This Research

Among the many diverse types of learning game research, this research is more applicable to some types than others. Specifically, this work focuses on the 'mass market' context: learning games played by ordinary people in ordinary settings such as classrooms, at home, or while travelling. I discuss simulation and training literature, but more as a literature source than a target for my findings. I do not discuss my findings for less common contexts such as professional vocational training like business school simulations or military training. Therefore my findings should be treated with caution by learning game designers in those subfields.

There is a blurry border between learning games that are tools of persuasion and learning games that aim to impart knowledge in the classic sense of the word "learning" (Bogost 2007a). This research is unlikely to apply well to purely persuasive games. If the game's learning goal is so broad and vague that it is difficult to relate to the in-game activity, then that game is not a great fit for

activity-goal alignment theory. For example, the following types of learning game research are less likely to benefit from this research:

- Some types of role-play-focused learning games (e.g. “you’re a marine biologist...”) or multiplayer scenarios (board room simulations used in business schools) have broad, vague learning goals, such as “learn how to think like an executive/scientist,” and do not define any goals or win/lose conditions more specifically. Instead they provide an open-ended experience, which is intended to let the player gather knowledge indirectly. These types of games are difficult to map to activity-goal alignment theory. However, if a vague learning goal can be broken down or defined more clearly, this research is applicable. In this thesis, the game “*GameStar Mechanic*,” which aims to teach “Systems Thinking,” is analysed and scored using the AGA Scoring Tool.
- This work may not be as useful for sandbox games, including games that have no particular win/lose condition such as the Sims. I also include toys which appear to have game-like behaviour, especially electronic toys such as Tickle Me Elmo. Toys do not have rigid rules for play or a win/lose condition (Koster 2005), and their player’s primary activity is often too vague or complex to define clearly. Because activity-goal alignment theory requires a clearly defined activity, this type of learning game may not fit the findings in this research.

1.7.1 Intended Audience

Though this research aims to be of use to the broader learning research community, it is intended to be particularly useful for an audience of learning game designers and researchers. I am particularly interested in two types. The first type I call amateur designers: learning game designers are not experienced enough to be sure which ideas from the literature are most important. As Klopfer, Osterweil, and Salen have stated: “Confusion arises as newcomers to the discipline of games and learning are often lead to believe that making a game is simply a matter of representing some idea, concept or place in a virtual world” (Klopfer, Osterweil, and Salen 2009, 34) This research aims to improve on these naïve views.

The second type I call “researcher/designer.” These are academic researchers who design learning games themselves as part of their research activity. Researcher/designers are distinct from “pure scholar” academics that play and discuss, but do not build, learning games. Researcher/designers come from a variety of disciplines including education, media studies, and computer science. A second type of researcher/designer is the subject matter expert, who belongs to specialized disciplines (chemistry, mathematics, history) and do not normally study learning or games, yet aim to use video games to teach their subject.

There are many other types of learning game designers. For example, commercial designers of children's toys that incorporate learning games, or educators who add learning games to their classroom activities might also find this research useful, even though it was intended primarily for the two types above.

1.8 Outline

This dissertation has eight chapters:

1. Introduction
2. Literature Review
3. Methodology
4. Scoring Activity-Goal Alignment
5. Using Activity-Goal Alignment Scores in Analysis of Games
6. Review of Learning Game Design Assessment
7. Activity-Goal Alignment-Based Assessment of Learning Games
8. Discussion

The second chapter is the literature review. I present Shelton's original conception of activity-goal alignment theory, and relate it to established learning game design theory (e.g. fidelity, mechanics, alignment) as well as important ideas from the fields of education, training simulation, and commercial video game design.

In the third chapter I present the method for this thesis. I begin by describing my practical activities within this research, and then present my rationale. I fit this research in the ontology of research methodology, and argue for a design-based approach. I argue for ethnography, and specifically analytic autoethnography, as the best fit for the constraints and goals of this research. I then discuss known weaknesses of the method, and address ethical considerations.

In chapter 4 I establish the need for scoring activity-goal alignment, and describe my development of the AGA Scoring Tool to score activity-goal alignment. I start with an autoethnographic narrative describing my attempt to achieve extremely high activity-goal alignment in the development of *Battlefood*, a learning game. I discuss examples where activity-goal alignment would have been useful, and note that Shelton's application of activity-goal alignment theory could be amended to better suit practising designers' needs. In accordance with my method, I expose my process and explain my design reasoning in my creation of the AGA Scoring Tool.

In chapter 5, I exercise the AGA Scoring Tool by using it to analyse five different learning games. I draw from a variety of data sources including prior analyses, my observations of others' play experiences, and my own experience. I discuss a broad range of games, including the well-regarded *Re-Mission* and *GameStar Mechanic* as well as more obscure and unsuccessful learning games. In the first section I discuss several games' activity-goal alignment using a discussion based on Shelton's approach, and then score their activity-goal alignment using the tool built in chapter 4. I find that scoring activity-goal alignment is useful. In the second section, I probe for weakness in both activity-goal alignment theory and the scoring tool. In the end, I warn designers that scoring activity-goal alignment does not shortcut the reflection-in-action of good design practice. I find that multiple, diverse activities are more difficult to represent clearly in the activity-goal alignment scoring diagram, and that scoring activity-goal alignment illuminates differences between two games about a single learning topic. Finally, I discuss practical issues I encountered scoring activity-goal alignment.

In chapter 6, I explore how activity-goal alignment might be used for assessment of quality, not just design investigation, in learning games. I discuss existing methods of assessment of learning games, from both commercial and academic sources. I then create a second new tool, the **AGA-Based Assessment Tool**. This tool combines activity-goal alignment and Higher Order Learning theory to create four categories of learning games. I discuss how this tool's unique approach could benefit learning game designer/researchers.

In chapter 7, I exercise the AGA-Based Assessment Tool developed in chapter 6 to categorize several learning games. I discuss examples that fail to respond to key advice from veteran learning game designers, as well as exemplary games. I relate the tool's output to important design advice in the literature, and find that the tool is useful within limitations. I argue that activity-goal alignment-based categorization guides designers towards better learning game design because the results are difficult to misinterpret, advising designers to avoid one type of game design, and target another.

In chapter 8, the conclusion, I discuss the implications of this research. I discuss how findings from chapters 4 and 5 suggest that scoring activity-goal alignment is a useful "rule of thumb" that is quicker and more accurate than textual discussion of activity-goal alignment, although it does have drawbacks. In discussing chapters 6 and 7, I first note that the literature effectively advises designers to aim for high activity-goal alignment, but low-activity-goal alignment games are still being built despite this advice. I argue against using low-activity-goal alignment game designs in research, and suggest that scoring activity-goal alignment might help designers stay focused on the important opportunities in learning game innovation. I discuss how this work contributes to the literature,

focusing on Shelton's original conception of activity-goal alignment. Finally, I consider directions for further research.

1.9 Conclusion

This introductory chapter has established the need and significance of this research, provided an introduction to the thesis and previewed the research activities chapter by chapter, laying the groundwork for the remainder of the dissertation. It has built toward one key point: this research identifies the important theory of activity-goal alignment in the existing learning game design literature, and in combination with other theory, presents it in a mode that designer/researchers can easily apply in practice.

2 Literature Review

This chapter aims to provide an understanding of the ideas related to activity-goal alignment (e.g. fidelity, mechanics, alignment), as well as discuss concepts used in this thesis to extend Shelton's activity-goal alignment. Additionally, it present certain fundamental theories of learning upon which this research is based.

2.1 Introduction

The term "learning game" has two components: learning, and game. Each of these two words signifies an important design element, yet their relation has been the subject of much research. Shelton's activity-goal alignment theory characterizes one of the most important relationships between "learning" and "game." In the process of developing activity-goal alignment theory, Shelton has repeatedly posited that commonalities exist between certain properties of learning games and theories of learning such as problem-based learning and activity theory (Shelton 2007; Walker and Shelton 2008; Shelton and Scoresby 2010). This chapter reviews ideas from three fields: commercial games, education, and simulators, aiming to:

- Present selected fundamental theories of learning upon which this research is based
- Provide an understanding of the ideas related to activity-goal alignment (e.g. fidelity, mechanics, and alignment).
- Discuss concepts used in this thesis to extend Shelton's theory of activity-goal alignment (e.g. gamification, Gee's "deep learning")

Next I discuss the scope of this literature review. The field of learning game research is highly interdisciplinary, drawing from video game design (Salen and Zimmerman 2003; Schell 2008), instructional design (Akilli and Cagiltay 2006), psychology (Csikszentmihalyi 1990), cognitive science (Chi 2008; Gentner, Holyoak, and Kokinov 2001), instructional design and education (Barab, Warren, and Ingram-Goble 2006; Squire et al. 2003; Herrington and Herrington 2005), and more. Only key ideas from a few of these fields are presented here. A more complete understanding of learning games and game theory can be found in a few useful texts. For a general text on commercial simulations and learning games, see Aldrich (2009). For game design theory, see Salen and Zimmerman (2003). For a survey of game studies topics beyond the scope of this review, such as gender, violence, and health, see Kirriemuir and McFarlane (2004).

2.1.1 Activity-Goal Alignment Theory Compared with its Ancestors

Shelton did not invent the ideas that comprise activity-goal alignment. For example, activity-goal alignment could be seen as an evolution of the endogenous / exogenous duality Malone and Lepper (1987) proposed. Their theory, in turn, adapted the terms from Kruglanski (1975). Kruglanski used the terms while arguing that the division of human behaviour is not only either internally or externally driven, as traditional attribution theory had posited. He instead suggested endogenous and exogenous motivations as a better distinction. In the context of learning games, exogenous games “separate learning content and game mechanics,” as Winn and Heeter explain:

Designers of exogenous educational games typically reuse successful game mechanics, such as hangman, a Jeopardy-style game show, or a Space Invaders-style shooter, inserting the content to be learned into the preexisting [sic] game structure and rules. (B. Winn and Heeter 2008, 1)

From this quote, Malone and Lepper’s discussion, and Halverson’s similar description (Halverson 2005), it is clear that exogenous theory is a predecessor of Shelton’s description of low activity-goal alignment.

There are important difference between the theories of activity-goal alignment and endogenous/exogenous theories. One key difference is in the approach: Where Shelton used a design-based approach to the design problem, Malone instead used a scientific approach. Malone apparently aimed to deconstruct the phenomenon of learning games into isolated components. The title of his paper, “A Taxonomy of Intrinsic Motivations for Learning,” suggests one can tidily categorize components in learning games as one might differentiate shellfish: measure, characterise, name, and assign to a neat box in a taxonomy. While there is nothing wrong with a scientific approach per se, Malone and Lepper stated a goal that has a designerly quality: “Our aim is to **synthesize** a number of competing theories into a framework that can guide not only further theorizing, but also the **design** of instructional environments” (Malone and Lepper 1987, 230).

Shelton’s approach to developing and applying activity-goal alignment was quite different. Shelton did not aim to deconstruct theory into attributes or taxonomies; instead he sampled existing theories, including constructionist building of knowledge, constructivist activity, problem-based learning, and project-based activity. Shelton then remixed them with a focus on keeping the theory connected to the actual play experience, then applied and refined it primarily through iterative, self-reflective practice, discussing the outcomes and impacts of his theory in the authentic context of a

learning game project (Shelton 2007). Shelton's approach reasonably fits common definitions of design research (Barab and Squire 2004)⁶.

2.1.2 Shelton and the Mad Hatter's Tea Party

Malone and Lepper had a relatively blank slate in 1987: they were pioneers of learning game research. Twenty years later, Shelton described the learning game literature as a "Mad Hatter's tea party" of "vastly different" opinions and approaches. Shelton noted that learning games have yet to "help people learn about complex relationships [...] across multiple contexts or at any reasonably large scale" (Shelton 2007, 104). Shelton argued that one reason for past problems with learning games is a lack of a framework that merges many disparate theories. He also believed that "most appropriate approach lies at the intersection of traditional game theory and instructional design" (p. 104), citing Squire's approach to studying the game *Civilization* (Squire et al. 2003) as sound.

Shelton discussed how he developed a "basic approach" to learning game design he named "activity-goal alignment" by addressing a potential problem in *River City*, a learning game he was developing:

Both Supercharged! and River City are examples of computer games designed for learning that have enjoyed some success for learning outcomes. However, the motivation attributes of each of these games were not necessarily designed to be balanced with the instructional activities, and some disconnect was reported with how students approached their designed learning activities and their motivation for playing. (p. 110)

Shelton did not discuss how he developed activity-goal alignment theory in detail, stating:

"Designing for activity-goal alignment ensures that a correct balance of game-like attributes are included for motivation, but that the activities within the game are meaningful, and therefore exist as more than just a means to an end" (p. 111).

Shelton also discussed the results of applying activity-goal alignment theory to the development of *Voices of Spoon River*, a text-based learning game for the classroom. He found activity-goal alignment helped game designers stay focused on the essential challenge of learning game design.

Shelton illustrated the need for activity-goal alignment by giving a contrasting example. He cited a learning game whose avatar was named after the first woman to graduate from MIT with a chemistry degree. The designers apparently chose this name with the intent of increasing motivation for female players, but as Shelton noted, these design choices "have the potential to distract from

⁶ I further define the difference and compare the merits of the scientific approach and designerly ways, arguing that the scientific approach to design problems can be a poor fit, in Chapter 3.

the learning activity" in part because they move the player's attention away from the learning goal. Shelton's observation suggests that the excitement and pleasure of learning game design can yield decisions based on questionable reasoning, which might be part of why designers can and do regularly fail to see the results they hoped for from their designs (Devlin 2012a).

Shelton argued that an activity-goal alignment approach can improve decision-making in learning game design:

Designing for activity-goal alignment ensures that a correct balance of game-like attributes are included for motivation, but that the activities within the game are meaningful, and therefore exist as more than just a means to an end. (Shelton 2007, 110)

Shelton believed in the importance of activity-goal alignment, stating that games "may provide more direct learning benefits if the beneficial potential that games offer can be effectively organized into the game activity itself" (Shelton 2007, 109)

According to Shelton, researchers have adopted a sceptical view of the value of learning game research, citing Cuban's (1986) "utter disappointment" in educational technology's impact since 1920. Reeves (1993) made a similarly harsh assessment, and I found related concerns in a more recent educational game literature survey (Kirriemuir and McFarlane 2004). In keeping with Shelton's advice to be sceptical, a reader might ask: Are learning games worth the research investment? The following section answers this by making clear the exciting potential of this type of research.

2.1.3 Benefits of Learning Games

Learning games have been studied since the dawn of personal computers (Loftus and Nelson 1985). Whether or not designed specifically for learning, video games have long been theorized to offer powerful educational opportunities (Malone 1981) despite weak historical evidence of overall efficacy (Wilson et al. 2008). While some recent projects have showed more convincing results (Tate, Haritatos, and Cole 2009; Merry et al. 2012), some say that learning game design is still largely an immature field (Gee 2009a; Devlin 2012c). Despite his sceptical approach, Shelton did not believe learning game research is hopeless: "I generalize that educational games may provide learning benefits if the beneficial potential games offer can be organized into the game itself, and into the social activity surrounding the game" (Shelton 2007, 109).

Broadly speaking, video games offer a new and unique form of interaction over other methods of learning, enabling learners to experiment and freely explore (Prensky 2009). Attempts have been

made to understand the nature of these benefits. Early literature reviews in this field (Cavallari, Hedberg, and Harper 1992) served as a viable starting point for more inclusive and thorough reviews (Kirriemuir and McFarlane 2004; D. S. Freitas and Maharg 2011). Methodologies varied widely: Wilson et al looked for evidence of efficacy (2008) while Egenfeldt-Nielson identified key tensions in the literature, such as learning versus playing, or freedom versus control (Egenfeldt-Nielsen 2006). The relation between benefits and game attributes are often theorized anew in each review, and resulting theories sometimes oppose, rather than build on, each other. However, many reviewers often agree on a few attributes, as Shearer observed (Shearer 2011). Commonly claimed benefits of learning games relevant to activity-goal alignment⁷ include:

- **Low-cost yet effective** (e.g., Belanich, Sibley, & Orvis, 2004; Driskell, & Dwyer, 1984; Rieber, 1996; Smith, Sciarini, & Nicholson, 2007, cited in Wilson et al. 2008).
- Increased **motivation** (Blunt and Annandale 2007; Prensky 2004; Dondlinger 2007)
- Improved higher-order thinking e.g. deduction and hypothesis testing (Rice 2007)
- Improved **self-efficacy** e.g. confidence and risk-taking in problem-solving (Gee 2006; Young et al. 2009)
- **Social** e.g. affiliation and teamwork (Gee 2004)

There is some disagreement even among this short list. Wilson felt that learning game research has over-focused on motivation while the impact of games on other learning outcomes has been lacking (K. A. Wilson et al. 2008).

Perhaps the most important benefit of learning games relates to the idea that learning is an intrinsic pleasure for humans. The idea that pleasure comes from the act of learning is broadly accepted among key learning game scholars (Gee 2004; Resnick 2004; Prensky 2004; Osterweil 2006), and has support from a biological point of view: learning can be seen as a chemical phenomenon, using dopamine as a reinforcing signal that is critical in the initial formation of stimulus-response associations (Byrne 2008, 364).

Learning is obviously a crucial concept in learning games. The following section presents an overview of theories of learning relevant to learning games, and discusses key texts and concepts from the literature.

⁷ Other benefits have been commonly claimed, including literacy improvements (Shaffer 2006; Gee 2004; Squire 2008), improved hand/eye coordination, visual and spatial processing (K. A. Wilson et al. 2008), but they are not as relevant to AGA.

2.2 Conceptions of Learning

There are many various way to understand learning. For example, Kolb (1984) defines learning as a process composed of four phases (concrete experience, reflective observation, abstract conceptualization, and active experimentation). Others focus on the conscious and unconscious aspects of learning as a result of real life situations, simulations, or imagination (Botkin, Elmandra, & Melitza, 1979, as cited in Thatcher, 1990). This section discusses the major approaches to learning in the learning game literature, and their relevance to activity-goal alignment.

Prensky (2004) uses a sociocultural lens for understanding learning. Prensky reviewed learning modalities over time, to argue that game-based learning is not a revolution, but a return to the roots of education. He reviewed the progress of learning since prehistoric times. Apprenticeship, the most ancient form of learning (e.g. “grab a rock and throw it at prey”) is used today in sports and music. Instructive pictures (e.g. sand drawings, cave paintings) are used today in textbooks. Spoken language enabled verbal descriptions. Storytelling gave “teaching stories” (e.g. parables), and Socratic methods are still used today in universities. Writing and the printing press enabled big changes to education: standardization (dividing learners into age groups, knowledge into disciplines), and linear thinking.

Prensky used this perspective to argue that, prior to writing, learning was a highly interactive experience. Writing is an inherently linear and non-interactive medium. Schools train students to think linearly, along the lines of the reasoned, linear arguments made in books, and de-emphasizing the need to reason on the fly. The historically recent computer revolution empowers other modalities of knowing, which challenges the traditional schools and pedagogy at its foundations.

Other scholars’ arguments support much of Prensky’s argument. Our traditional formal education is famously linear and didactic, more the same than different since the 16th century (Chou et al. 1993). Other forces, including a new “creative class” comprising 30% of the workforce (Florida 2002) have been said to align with the need for educational revolution (Laursen and Thiry 2008). However, school is a deeply rooted, conservative socio-cultural tradition. The semiotic meaning of a university degree goes far beyond its economic value, signalling achievements of social class and identity (Archer, Hutchings, and Ross 2003). Change does not come naturally or easily to such powerful and important social institutions (Klopfer et al. 2009).

Among the many approaches to understanding learning, certain approaches are often cited in the learning game design literature. These common approaches are presented in the following section.

2.2.1 Cognitive Science

Cognitive science is a large, well-funded field of research containing an array of approaches to understanding learning. Traditionally cognitive scientists took a positivist “hard science” approach, where the human brain is conceived as a sort of biological machine (Pinker and Norton 1998). They advocate for the hard science approach to understanding learning by noting that the rest of the body has been understood through the science of medicine and physiology, greatly increasing our understanding. Similarly, the brain itself is being reverse-engineered and when this occurs, they argued, a great leap forward in our understanding of the learning process will occur. This breakthrough has not yet occurred. For example, Alexander and Brown (2011) theorized that the medial prefrontal cortex region is “concerned with learning and predicting the likely outcomes of actions, whether good or bad” but noted there are many competing theories, none of which match the data, so it appears that that particular question is still open.

Learning game researchers often cite a variety of cognitive science research in their work. However, most cognitive science researchers caution against overextending their findings, and Cross, a leading design theorist, argued against a naïve faith on a “hard science” approach to designerly problems (Cross 1982). For example, Gee cites Elman that “work in cognitive science has shown that people need to be presented with problems in a fruitful order, getting initial problems that set up good generalizations for later problems” (Gee 2004, 2). This and other discussions suggest that Gee aligns more with the a more holistic and multidisciplinary to approach to understanding learning known as “learning science” (R. K. Sawyer 2005).

The difference between learning scientists and “hard science” cognitive scientists is significant. For example, Bransford (2000) found five principles common in learning science. The first principle is:

Students come to the classroom with preconceptions about how the world works. If their initial understanding is not engaged, they may fail to grasp the new concepts and information that are taught, or they may learn them for purposes of a test but revert to their preconceptions outside the classroom. (p. 14)

In this principle we can see conceptions of learning that include conceptual change, a concept that some regard to be outside of the strictly 'hard science' subdomain of cognitive science (Chi 2008). I discuss learning science’s approach more in the “deep learning” discussion later in this chapter.

2.2.2 Constructivism

Constructivism (Piaget and Cook 1955) is a commonly referenced concept in learning research. Not to be confused with Seymour Papert's Constructionism (Harel and Papert 1991), constructivism suggests one learns by experience to create one's own knowledge. Constructivism stands opposed to traditional didactic methods of teaching (L. P. Rieber 1992).

Constructivism takes many forms, some of which include an element of playfulness (Perkins 1999). Constructivism is opposed to objectivism, but in instructional design "they coexist and are used when appropriate," like quantitative and qualitative research methods (Vrasidas 2000).

Constructivism is often advocated today by learning researchers. Perkins (1992, 55) stated that "virtually all contemporary approaches to teaching and learning have a constructivist cast." A literature review of best practices of improving science education through technological intervention (Laursen and Thiry 2008) found many constructivist goals:

- The use of hands-on, minds-on, inquiry-based teaching and learning strategies
- The importance of involving a range of stakeholders in projects from the beginning
- The importance of explicitly addressing dissemination and sustainability of good work

Paiget's constructivism has been developed to include Vygostky-influenced social constructivism (Richardson 1997) and other forms. Despite some well-formed arguments against it (Özdemir and Clark 2007), constructivism has begat well-regarded innovations in education such as authentic activity (J. Herrington et al. 2004).

Constructivism is relevant to activity-goal alignment in that game-based learning interventions are often, but not always, designed around constructivism (Prensky 2006). For example, some constructivists believe that the teachers' role must change significantly: teachers should be guides to the students learning journey, as opposed to Godlike entities pouring knowledge into passive learners (Strommen and Lincoln 1992). This change validates learning game research trends, which have long sought to make learning games so effective that supporting material, structured environments, and human teachers are not required (Squire and Jenkins 2003).

2.2.3 Motivation

I aim to build tools that make it easier for designers to build intrinsically motivating games in this research. It is therefore important to understand the literature around the idea of "intrinsic motivation." The term motivation has a long history in educational literature. Dewey introduced the

term to common usage in 1905 “yet the spirit of motivation was fully anticipated by Rousseau, Pestalozzi, Froebel, and, in fact, almost every educational writer of note in modern times” (G. M. Wilson 1917). Motivation is commonly discussed in both learning game (Gee 2004) and commercial game research (Salen and Zimmerman 2003). Numerous scholars have studied motivation in learning games (Chin et al. 2008; Garris, Ahlers, and Driskell 2002; Lepper and Cordova 1992; Malone and Lepper 1987; Paras and Bizzocchi 2005). Motivation is often found to be a key benefit in both specific studies of learning games (Knowledge Adventure 2012a; HopeLab 2012) and surveys of learning game efficacy (Vogel et al. 2006).

There is no widely agreed definition of motivation. While motivation has been characterized as a single, general characteristic (Moen and Doyle 1978), many scholars today consider motivation to be divided into two types: intrinsic and extrinsic (Lepper, Greene, and Nisbett 1973). In psychology, the term “intrinsic motivation” refers to an person’s intrinsic need to be competent and self-determining (Deci and Ryan 1985, 5); however, the game design literature is usually (although certainly not always) focused on improving the design of the game, as opposed to using games as a way to understand players. Malone set this precedent early, as can be seen in his choice of title: “Towards a Theory of Intrinsically Motivating Instruction” (Malone 1981). Instruction, not motivation, is the topic. This research follows Malone’s precedent in this regard.

Malone is a notable pioneer in his advocacy of intrinsic motivation in learning games. Malone argued that noted constructivist scholars Piaget and Bruner were opposed to extrinsic motivation. Malone stated that these notable psychologists argued for the importance of “intrinsically motivated play-like activities for many kinds of deep learning” (Malone 1981). Malone then stated his position: “If students are intrinsically motivated to learn something, they may spend more time and effort learning, feel better about what they learn, and use it more in the future.” While I critique Malone’s method in chapter 3, I generally argue in this thesis that the prevailing view of learning game literature today shares Malone’s advocacy for intrinsic motivation.

Many scholars have continued Malone’s work, developing a broad range of theories around motivation. For example, Moen and Doyle (1978) considered motivation as an underlying measure of other player emotions, noting that traditional teaching methods tended to use anxiety as negative motivation (e.g. use of bad grades by teachers) as a method of getting students to work. Moen and Doyle found that a focus on positive motivation has produced greater effectiveness with learning, and theorized that anxiety interferes with performance of complex tasks. Recent work has built on this idea. “Eustress” roughly means positive stress. This term has been used to explain why gamers choose to challenge themselves with stressful activities in play (McGonigal 2011). Some scholars

argued that motivation should be considered in a social context, since learning is a shared activity (Tuzun 2004). Bartle identified four types of players of multiplayer video games by dividing four types (achievers, explorers, socialisers and killers). Yee found three types of player behaviour in multiplayer games. Radoff (2011) generalized these prior findings to include any type of video game, not just multiplayer, and identified four types of motivation: achievement, competition, immersion, and cooperation. Radoff intended this theory to guide game designers: “It’s better to have a simple game that starts out focused on achievement and grows into a fantastic multiplayer competitive experience than one that becomes a mediocre combination of both cooperation and competition” (Radoff 2011)

Many learning game researchers share Malone’s simple intrinsic/extrinsic division of motivation in their consideration of the more complex models and behaviours. For example, Ryan et al considered competition a form of extrinsic motivation “because it encourages the performer to win and beat others, not to enjoy the intrinsic rewards of the activity” (Deci and Ryan 1985, 56).

While many learning game researchers regard intrinsic motivation as important and valuable in a learning game, this position is not universal. At least one theorist has argued that intrinsic motivation is not as important as the literature suggests (Becker 2007). However, this position is rare, and I did not find Becker’s argument convincing, as I discuss in chapter 7. Many modern scholars do not consider intrinsic and extrinsic motivation to be opposed. Instead they argue motivation is not a duality between intrinsic and extrinsic (Vallerand et al. 1993). Recent game design researcher/practitioners suggest holistic approaches that combine intrinsic and extrinsic motivation. For example, the rise of the term “gamification”⁸ has brought the issue of motivation to the foreground. Game studies scholars have argued that gamification is fundamentally opposed to intrinsic motivation (Bogost 2011b), and some advocates of gamification disagree (Mullany 2012). In recent discussions of gamification, Proto stated: “Gamification, by contrast, doesn’t rely on internal motivation...Gamification assumes that the player isn’t especially motivated — at least at the beginning — and then provides barrels of incentives to ramp up that motivation” (Proto 2011). As with the psychology literature, Proto essentially argued that intrinsic and extrinsic motivation can be combined in game design. However, psychologists suggest intrinsic motivation is inherently more effective than extrinsic motivation. Ryan and Deci convincingly argued for a continuum of motivation from the least effective, amotivation (no motivation) to the most effective, intrinsic motivation. They placed extrinsic in the middle of the continuum (Ryan and Deci 2000, 72).

⁸ This discussion is about motivation, and uses gamification only as an example. For those not familiar with the term, I present and discuss it in the following section of this chapter.

While there is a broad array of views regarding motivation theory, many experienced and successful designers assume or state that intrinsic motivation is crucial to creating excellent learning games. For example, Klopfer et al advised designers to “find the fun in the learning” (Klopfer, Osterweil, and Salen 2009). In their discussion of this advice, it is clear that they were advising designers to create intrinsically motivating game mechanics. Malone, Gee, and numerous other notable scholars offer similar advice. This position is well supported in the psychology literature, as I discussed here. In recognition of this prevailing view of the literature, I aim to build tools that make it easier for designers to build intrinsically motivating games in this research.

2.2.4 Alignment

In the education literature, the term “alignment” refers to a mutually supportive relationship between teaching objectives, activities, and assessments (Tyler 1950). Alignment is prized in educational design. For example, one particular, very influential (Walsh 2007) theory based on alignment is Constructive Alignment (Biggs 1996). Constructive Alignment combines Constructivism, in which the learner’s activities create meaning, and instructional designers’ alignment of the objectives of a course with its assessment.

In Shelton’s use of the term, the alignment component is simple. If the teaching activity matches the learning goal, they are aligned (Shelton and Scoresby 2010).

2.2.5 Transfer

One idea common to most conceptions of learning is “transfer” (Baldwin and Ford 1988). Transfer is the intuitive idea that material studied has some effect on performance. For example, if a driver stops at yield signs after reading an instructional driving pamphlet, but did not previously, then the knowledge in the pamphlet has transferred to the reader. The term “transfer” is commonly used in the simulation and training literature, and I discuss it more in that context later in this chapter.

2.2.6 Models in Learning

Gee has stated that one of the deepest and most important properties of games for learning is “using models to make learning from concrete experience more general and abstract.” (Gee 2009a). He argues:

Players may be experiencing a game’s virtual world, which might be quite graphically detailed, but very often they are using and thinking in terms of models. Models are crucial for good learning (diSessa, 2000, 2004; Lehrer & Schauble, 2000, 2005, 2006; Nersessian, 2002). They

help bridge between concrete experiences and more abstract and systematic understandings. Models are crucial to games and gaming, as well. (Gee 2009a, 74)

Like simulators, learning games often embody a model of their topic. For example, a driving game contains a set of equations that model tire cohesion, engine power curves, and other physical aspects of vehicle dynamics. Gee argued that the player's intention, desires, and goals are also modeled and expressed through the game's design and algorithms. The player typically must interact with the model and understand it to win—which means the player has learned the model (Gee 2009a). Some learning games intend the player's mental model to be informed by the game's model.

Although Gee recognized the value of aesthetics as part of the narrative, he devalued the role of aesthetics in modelling:

*[World of WarCraft] **simulates (models) a world of mountains, lakes, roads, buildings, creatures, and so forth, which, while fantasy, is meant to resemble aspects of the real world. However, players for the most part pay very little attention to this modeling aspect of World of WarCraft, because it usually plays no important role in game play. Rather, players concentrate on the embodied experiences of play, problem solving, and socialization that World of WarCraft offers. (Gee 2009a, 73, my emphasis)***

A game's story, mechanics, economics, social aspects, or aesthetics might each or all be aligned to its learning goals. While one may merge them to a single variable, as some do with fidelity (Hochmitz and Yuviler-Gavish 2011), the preservation of these differences is important, some would say critical (Shelton and Scoresby 2010), to good learning game design.

Models are often used in science learning games, but it is not always better to use a model in a science learning game. Squire and Patterson discussed how models are useful in teaching science:

Modeling involves the recursive process of observing phenomenon and building representations to illustrate those core ideas (also called abductive inquiry, see Peirce 1877/1986). Models such as Lotka-Volterra are constructed by scientists. Scientists engage in cycles of data collection, model building, and model testing. (Squire and Patterson 2009, 7)

However, video games have important differences from typical teaching simulations:

In contrast, games are generally constructed by experts trying to communicate ideas to novices. Educational games seek to teach the player the model's rules and emergent

properties through playing them (Gee, 2003; Squire, 2005). However, this mode of learning by which players learn is also abductive, in that they hold models about how the world works, and then are forced to amend those understandings as they encounter new experiences of them. (Squire and Patterson 2009, 7)

From Squire's discussion, it is clear that, although games might include simulation elements, they are often used for very different purposes from simulations in classroom education.

Squire's discussion is a useful framing for more specific theories that relate mental models to video game design. For example, many designers have cited Papert's concept of "microworlds" (Papert 1993). Microworlds theory means a small but complete subset of reality that, through personal discovery and exploration, enables one to learn a particular domain. This concept fits well with activity-goal alignment: In microworlds terminology, the activity is the learning goal, which suggests high activity-goal alignment, and thus learning games can target higher-order thinking as a learning goal by building microworlds. Theories such as microworlds fit with the prevailing advice regarding the most important opportunities for video game research: "Playing video games [...] helps us make or critique the systems we live in When we learn to play games with an eye toward uncovering their procedural rhetorics, we learn to ask questions about the models such games present" (Bogost 2007b). In chapters 6 and 7, I investigate further this general idea of using higher-order learning, including models, in games, and offer a tool to assess games against this advice.

2.3 Play and Learning: A Fallacious Tension

Games are a form of play, defined here as a complex multipurpose phenomenon fundamental to humans (Huizinga 2003) and many animals (Burghardt 2005, p93). Play rules are fundamental, social and become negotiable as players progress in mastery (Piaget 1932). Fundamental types of play include social (foursquare; wrestling; face-making), imaginative/fantasy, narrative/storytelling, and transformative/creative (Institute for Play 2012). Play in video games has been similarly categorized (Bartle 1996).

Psychologists Singer, Golinkoff, and Hirsh-Pasek convincingly argued that play is learning in their book "Play Equals Learning" (2006). However, they found that parents are confused about the value of play: while they believe time spent playing is learning, they also felt that more time should be given to cognitive tasks at the expense of free time when play occurs (p6). Some argue that the message that "play isn't really valuable" can be seen in learning games such as *Math Blaster*, where the player must solve a math problem to proceed with an action game. While these games can teach

effectively and can be fun (Becker 2008, p30), their design has been criticized as “chocolate-covered broccoli” (Osterweil 2006; Resnick 2004) because they don’t realize games’ full potential. By dividing play and learning, these games imply two messages to their players:

1. Play is not really learning. Play is “candy” - an extrinsic reward for learning.
2. Learning means official-looking school work

Learning games must be both playful and useful. Historically, designers wrongly considered these two properties to be opposing (Klopfer, Osterweil, and Salen 2009). To correct this, one might ask of play research: How serious is play? Studying children playing in the Holocaust, Eisen (1990) said this question is perhaps the most mystifying thing about play. Play is supposed to be disengaged from reality in a “magic circle” (Huizinger 1939), yet play is also credited with real-life functions: play is “the highest level of preschool development” (Vygotsky 1978a, p102) and helps form a “zone of proximal development” in learning (Vygotsky 1978b).

Recent play research suggests that, instead of balancing the two, play must be prioritized: “The integrity of play is perverted whenever it is made to serve social functions” (Rodriguez 2006). Learning game theory also prioritizes fun: “Find the Fun in the Learning,” is how Klopfer, Osterweil, and Salen (2009) express it. This principle also rejects the imagined conflict between fun and learning.

In arguing that play and learning are in opposition, some people say learning is not always fun (Cothran and Ennis 1998), and that trying to make school fun is “enormously destructive” (Daastol and Jago 1995). The problem lies in the term “fun.” Fun as in entertaining is opposed to the deep satisfaction of grasping a new concept. If “fun” is replaced with “play,” the opposing argument dissolves. In other words, play is always “playful” but not necessarily fun. Gee states that game play and learning are closely related as they both are difficult, sometimes frustrating, and require persistence to succeed, and generate deep satisfaction. So, if play and learning are related, what sort of relationship is it?

Some offer metacommunication as a good approach to understand play’s relation to learning (Salen and Zimmerman 2003, p449). Play requires metacommunication: a dog’s playful nip says both “this is just play” and “this is serious” (Bateson 1987). Similarly, learning games might simultaneously create a playful interaction around a serious idea, using methods “so playful and delightful that children experience in them the same pleasure” as play (Parker and Temple 1923).

One important difference between play and required learning was noted by Piaget (1932):

It is somewhat humiliating, in this connection, to see how heavily traditional education sets about the task of making spelling enter into brains that assimilate with such ease the mnemonic contents of the game of marbles. But then, memory is dependent upon activity, and a real activity presupposes interest.

Piaget is observing how freedom of effort (Osterweil 2006) can unlock incredible learning. Freedom, central to the constructivist view of learning, is foundational to a number of important motivational factors (Zepke and Leach 2010). Freedom is a crucial principle of learning game design, and perhaps the most important difference between play and learning.

A playful approach is a methodology used in a variety of fields including industrial design (Kelley and Littman 2001), genetics research (Page et al. 2003), and business strategy (Jacobs and Heracleous 2007). For example, traditional software design approaches assume efficiency is a universal goal: “a clear, efficient solution to a practical problem” (Gaver 2002). This assumption is proven unwarranted by a playful game feature: Nintendo’s “Help Cat” runs away from the cursor, making it harder, not easier, to get hints (Cook 2006b). Since users assume winning yields valuable information, Help Cat winners are predisposed to value the factual knowledge delivered (hints about using the hardware). Help Cat reveals the unwarranted assumptions in many learning games like *Math Blaster*, which assume that gameplay should be the reward, not the task.

The playful approach is common in the learning game research community. For example, it has been proposed that “Bad simulations educate better than good simulations. They raise more questions” (Schell 2012). This playfully suggests designers intentionally build bad software, which is patently silly, but at the same time profound. Schell presumably means that if designers wish students to realize that simulations are only a model of reality, we might as well show them a broken model and ask them to fix it, rather than trying to “make it correct”. Fixing broken things is a common game design puzzle pattern (Kreutzmann and Wolter 2011) that may engage students. As the value of fidelity in learning is questionable anyway (as discussed earlier), this is a useful approach.

2.3.1 Gamification

In chapter 7 I name one category of learning games “pointsification” which is closely related to the term “gamification.” Gamification is a controversial trend in serious game design. Its proponents state that gamification is not superficial in that it addresses genuine human desires such as reward, status, achievement, self-expression, competition, and altruism (Bunchball Inc 2011), but notable game designers oppose it because they see it as failing to engage deeply at the mechanic level, and are used merely as “an affirmation of existing corporate practices” (Bogost 2011b). Many notable

learning game researchers have publicly endorsed Bogost's blog post titled "gamification is bullshit" (Bogost 2011b) in the comments. A similar critique was leveled at edutainment, which has been portrayed as a failure because it sought to "gamify" dull educational tasks (Resnick 2004), overly relying on extrinsic motivation⁹.

Some have been more forgiving to gamification, suggesting that the movement may be just one phase of evolution in a still growing field. Hopson, in reviewing ten years of game design theory progress, described how gamification is following an evolutionary path similar to behaviorism in psychology:

*Before behaviorism, psychology was an extremely subjective field, driven primarily by opinion and introspection...Radical behaviorism was overly simplistic, but it laid necessary groundwork for later, more complex approaches such as cognitive psychology. I believe that **something similar is happening in games**... In a few years, the industry will move on and the topic will be taken for granted, but we will have permanently shifted towards a more empirical approach to game design, and our players will benefit from that. (Hopson 2012, my emphasis)*

2.4 Video Game Literature around the Concept of "Mechanic"

As this research is focused on the design of video games, the literature around game design is obviously important to this research. Of particular interest is the literature on "mechanics," a concept closely related to the "activity" in activity-goal alignment.

The game design literature is vast. The practice of game design has probably existed for thousands of years, and became more formalized in the era of board games (Woods 2004). Video gaming began to develop its own theory starting in the late 1970s (C. Crawford 1984), and a critical discourse was reasonably well established by 2003 (Salen and Zimmerman 2003). Today, many media and methods are commonly used by video game design theorists. Discussions are typically informal and practiced-based, moderately rigorous and precise. Theory is typically developed in conversation with a variety of game design stakeholders, including users, developers, publishers, retailers, academics, and journalists. Theory is disseminated via many diverse outlets, including industry magazine articles and columns, formal lectures, panels, and roundtables at industry and academic conferences, animated

⁹ Extrinsic motivation has long been known to be problematic to harness as a form of motivation (Kerr 1975) as discussed earlier in this chapter

video presentations, cartoons, written texts with illustrated diagrams presented on blogs, email lists, online forums, local community group meetings, and one-to-one personal discussions via email.¹⁰

Most game design theorists state that great game experiences require interdependence between many varied factors, including mechanics, content design (animations, narrative/writing, sound), player psychology, social context, and many others. By dwelling on mechanics, I do not mean to suggest that these other elements of the play experience are less important. However, this thesis does argue that mechanics are especially important and problematic in learning game design practice today. I argue the relation between game mechanics and efficacy in chapter 7. Broadly, this thesis aims to focus designers' attention on improving the mechanics in learning games.

Definitions of “mechanics” vary somewhat in the literature. In this research I use a reasonably commonplace definition: “mechanic” is the fundamental moment-to-moment player activity in a game. In this definition, mechanic is closely related to player activity, but the two terms are not identical. Examples of a game mechanic include sword combat, block pushing, boomerang throwing, swimming, and button-based puzzles (Stout 2010). Player activity is a broader term that could include selecting items from a list. This may or may not be part of the game’s mechanic. However, in most cases in activity-goal alignment theory, activity and mechanic are nearly synonymous.

The term “mechanic” has been used for over thirty years in the game design literature by a variety of practitioners and theorists. By discussing a few varying definitions, I hope to give the reader a reasonably well-rounded definition that matches how game designers today use the term.

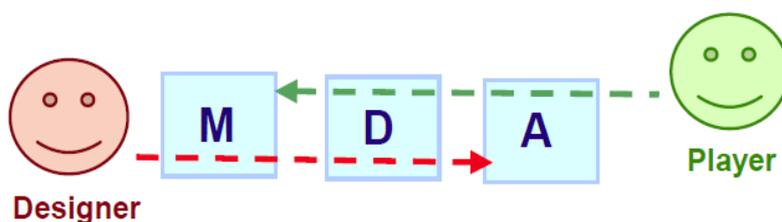


Figure 3: MDA stands for Mechanics, Dynamics, and Aesthetics. In MDA theory the Designer creates mechanics, which, through player action, create dynamics. The Player only perceives the aesthetics directly: the game’s dynamics, and thus mechanics, are experienced indirectly. (Hunicke, LeBlanc, and Zubek 2004)

¹⁰ This diversity of modality contrasts with **learning** game design theory, in which ideas are usually expressed in more traditional academic channels. These channels’ modality typically is academic prose, not suited to young players, publishers, and other non-academic stakeholders. In chapter 7 I argue that this limited modality may be partially responsible for the limited impact of learning game design theory on practice.

MDA: A formal approach to game design and game research (Hunicke, LeBlanc, and Zubek 2004) seeks to improve “systematic coherence” throughout all elements of a game design, and offers a theory called “MDA,” which the authors expressed in the diagram shown in Figure 3. The M in MDA theory stands for “Mechanics” which are procedural, created directly by the designer, in contrast to Dynamics, which are experienced by the player. Hunicke et al gave the example of bluffing versus betting: bluffing is a Dynamic, as it is something players decide to do and cannot be dictated by a design document, whereas betting is a Mechanic, because designers explicitly allow or forbid it in their creation of the game’s software. The term “mechanics” was defined as “the various actions, behaviors and control mechanisms afforded to the player within a game context” (p. 3). This definition was adopted by noted designer Raph Koster (2006).

Another text, *Rules of Play: Game Design Fundamentals* (Salen and Zimmerman 2003) is the field’s most comprehensive summary of fundamental concepts and language. While some of the concepts are more conceptual than practical, it is an important text that can broadly inform a designers’ understanding of games. Salen and Zimmerman did not use the term “mechanic” often, but the idea is present throughout the text. For example, the authors defined “descriptive meaningful play” as what “emerges from the relationship between player action and system outcome” (p. 37).

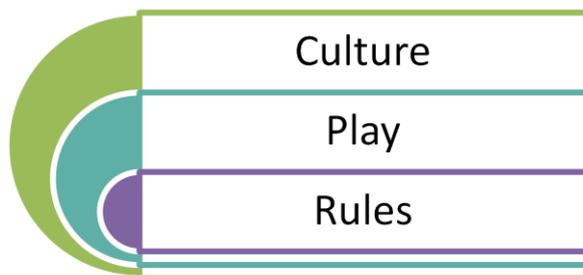


Figure 4: Rules are the structured, formal part of the Play experience, which are part of a game's Culture. (Salen and Zimmerman 2003)

While the book acknowledges that rules, play, and culture are interdependent (see Figure 4) and cannot exist without each other (p. 102). However, the authors focus on identifying specific attributes. This focus requires a tradeoff: holistic understanding cannot be developed through a “bottom up” approach to understanding video games. The authors state that they made this tradeoff deliberately in order to achieve other goals (p. 121).

As my research is aimed at practising designer/researchers, it is useful to compare the practical utility of the MDA text (Hunicke, LeBlanc, and Zubek 2004) and *Rules of Play* (Salen and Zimmerman 2003). The MDA text offers several significant ideas, but perhaps the most useful is its consistent,

broad view of the phenomenon of video games which is still grounded in player experience. The MDA diagram in Figure 3 relates three concepts in multiple, complex relationships, yet is clear. The three elements can be considered as three points on a scale, or the diagram can be seen as an illustration of the dynamics of meaning: the game is a conduit for a bidirectional “flow” of intention. By connecting these ideas, each of which builds on ideas previously found in the literature, MDA represented a significant advance in clarity over prior theory. Suddenly designers were able to discuss a video game holistically, yet also deconstruct the experience from both the player and designer perspective.

Rules of Play also aimed to be comprehensive, but the authors did not aim to achieve the holistic quality of MDA. While they did discuss the relationship between Rules, Play, and Culture as shown in Figure 4, when the text is considered as a whole, this relationship felt more unidirectional and less interconnected than MDA. Perhaps the practising designer might find the most utility from this diagram by regarding the three ideas of Rules, Play, and Culture as three nested lenses through which one might view the phenomenon of video games.

2.4.1 Schell's Conception of Mechanic

In *The Art of Game Design: A Book of Lenses* (2008), Schell discussed game mechanics in some detail, stating that six elements comprise game mechanics:

1. **Space:** Not just a physical or virtual playing field, space is used in the broader sense
2. **Objects, Attributes and States** (Characters, props, tokens, scoreboards)
3. **Actions:** “What can the player do?” These are not just written into the code of the game, but are also part of the social norms around the game.
4. **Rules** (foundational, operational, and other types) make possible all the other mechanics.
5. **Skill** of the player: physical, mental, and social
6. **Chance** adds surprise to the experience

By including space and chance, Schell's definition of mechanic is broader than MDA's, and begins to blur into a larger concept: something more like “the game experience”. This blurring is undesirable in that clarity is lost, but blurring also has merit: it suggests a more holistic conceptualization of “mechanic” might be beneficial.

2.4.2 Systems in Games

In their consideration of “mechanic”, some designers focus on “systems”, defined by Cook as “a curious and intriguing set of rules that interact in unexpected ways” (Cook 2007a). The focus is on complexity, not reality. However, because reality is a rich source for complex systems, designers often use it as a basis for systems-based design, as Cook notes: “Common sources of inspiration include probability, combinatorics, spacial relationships, physics, timing and economic game theory” (Cook 2007a).

Realistic simulations such as flight simulators are different from the reality-inspired systems in games. One key difference between simulations and games, Cook argues, is that “truly deep systems often lay bare their mechanics in order to provide advanced players with absolutely clarity on their inner workings” (Cook 2007a). Note how different this “laying bare” is to simulations which generally keep the complexity “under the hood”: when a trainee flies an aircraft, the goal is to keep the plane in the air. When a stall buzzer sounds, the trainee is expected to blindly trust, not comprehend, the mathematical model of the turbulence simulation. In comparison, most designers of systems-based mechanics expect, or even hope, that players will reverse-engineer the mathematical model in their game. Costikyan and others argue this, as the next section shows.

There is tension in the literature of game design around the value of simulation and system-based design. System-based designs can result in games that are fascinating to the designer, but not visible or meaningful to the player (Costikyan 2006). Costikyan made this point to caution designers, but also noted that simulation “can be quite powerful,” because

[simulations] offer insight into a situation that mere narrative cannot. [Simulation] allows players to explore different outcomes—in the fashion of a software toy—and thereby come to a gut understanding of the simulation's subject. Having played at least a dozen different games on Waterloo, I understand the battle, and why things happened the way they did, and the nature of Napoleonic warfare, far better than if I had merely read a dozen books on the subject. (Costikyan 2006, 204)

The player is not interested in solving the puzzle of the system. However, when it works, systems games can be surprisingly successful. Cook stated “the industry’s most original forms of gameplay were conceived by people inspired by systems,” and cited Tetris for simple systems, and SimCity or Populous for complex ones (Cook 2009b).

In summary, a common position in the game design literature is that commercial games should not aim to simulate reality: games should be fun for the user, not the designer: “In game design, player

experience should always trump so-called 'realism'" (Salen and Zimmerman 2003, 353). This position is not universally accepted by all designers (Ququasar 2011) and is disputed by some (Cook 2006a). Later in this chapter I look at how the field of simulation has taken a firmly pro-realism position, and established a related term: "fidelity."

2.5 Hierarchies of Player Experience

In activity-goal alignment theory, the concept of "activity" is clearly important to understand. This section reviews the literature on classification of player experience, including activity in games. In chapters 5 and 7 I propose tools that require their users to identify a particular activity within the game play experience. To make this choice well, it is helpful to understand how the literature represents different types of player activities.

What is a game's activity? There are many reasonable-sounding choices:

- A holistic activity ("winning the game")
- A strategic activity (e.g. "beating the demon boss")
- Fine-grained events (e.g. "pushing the left arrow button")

In the video game literature, player experience is typically divided according to player personalities and/or types of pleasure (Lazzaro 2012; Bartle 1996; Caillois 2006). These theories are too broad to apply in the context of activity-goal alignment, as they aim to deconstruct the player experience from the top down. Similarly, LeBlanc et al suggests the term "fun" is too broad to be meaningful, and identified types of fun in video games (Hunicke, LeBlanc, and Zubek 2004). Lazzaro used audience research methods to further identify four specific types of fun: Curiosity, Fiero, Amiero, Desire (Lazzaro 2012). Lazzaro's theory is commonly cited by practising designers today (Schell 2008). These theories do not focus on multiple granularities of player experience; they refer mainly to a broad summary of play experiences.

Other approaches to understanding player experience offer multiple granularities of play experience. Recently, Cook proposed a novel bottom-up approach which describes player activity at multiple levels of detail. Skill Chain theory (Cook 2007b) describes how simple activities contribute to complex ones in a given game. Cook offers a worked example in his Skill Chain diagram of Tetris (Cook 2007b); however, Cook does not explain if his theory could categorize similar player activities across various games.

2.5.1 Skill Chains

Skill Chains (Cook 2007b) combines game mechanics with simulation and user experience in a hierarchical, holistic model. Cook applied an idea from cognitive psychology: people master complex skills by first learning component skills (Wood, Bruner, and Ross 1976). Absurdly simple skills lead to mastery of one extremely challenging skill: winning the game, as Cook illustrated in Figure 5.

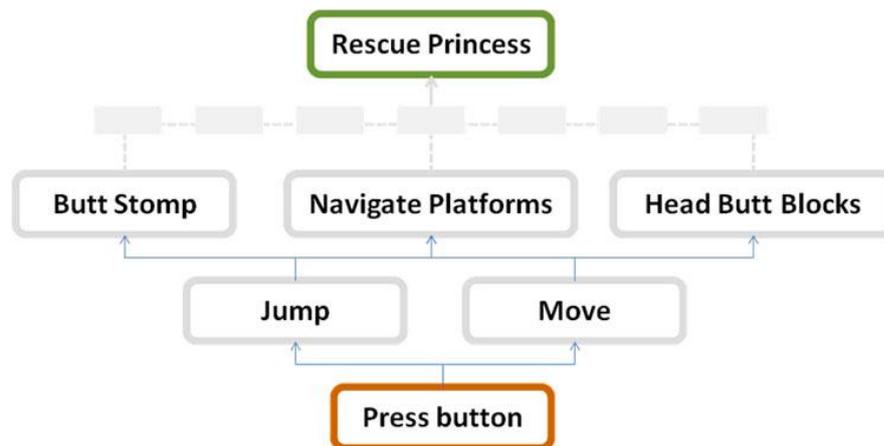


Figure 5: Cook's Skill Chain diagram shows how complex skills ("Rescue Princess") are built from previously mastered simple skills ("Jump"). (Cook 2007b)

The shape of a Skill Chain node-graph is unique to any given game. The nodes are called Skill Atoms. Cook gave the example Skill Atom in Figure 6, which uses button, gear, eye, and lightbulb icons for each of the four key elements:

1. Action (red button): Player does something. E.g. "press a button" or "navigate a complex maze"
2. Simulation (gear): Game calculates a result of player action in its simulation of the game world. E.g. A door might open; a monster might be damaged
3. Feedback (eye): Lets player know how the simulation has changed state. E.g. "green door is unlocked!"
4. Modeling (light bulb) occurs entirely in the player's mind: "the player absorbs the feedback and updates their mental models on the success of their action. If they feel that they have made progress, they feel pleasure. If they feel that their action has been in vain, they feel boredom or frustration." (Cook 2007, p3)

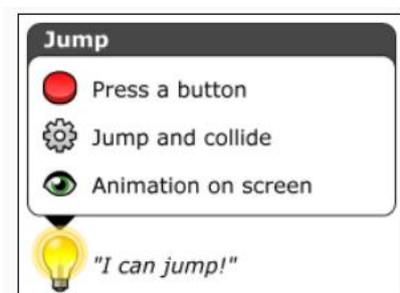


Figure 6: Cook's example Skill Atom is Jump, from the *Mario* game series. (Cook 2007b)

Skill Chains are relevant to the idea of scoring activity-goal alignment for several reasons. First, Cook set a precedent for using a hierarchy to explain player progression in a single diagram. Secondly, Skill Atoms are a holistic, applicable conceptualization of “mechanics.” Cook avoids losing the essence of “mechanic” through taxonomic division, while still explicitly giving insight into components that comprise mechanic, and preserving a clear path from this idea to practice. Third, Skill Atoms are a useful tool to practising game designers refining a playable game or prototype. The broader audience (students, academics) and the conceptual stage of development are not addressed (Cook 2009a). This limited scope and audience might sound like a disadvantage, but this limitation also has a strength: the tool can be made more specific to fit the needs of its niche audience, in a way that a general-purpose theory cannot be.

Skill Atoms set an important precedent on how to remix established theory for practising designers. Cook states he took “tokens, verbs, rules, aesthetics, etc and remixed them...” into Skill Atoms (Cook 2007, p3). Remixing of theory can be done many ways, but Cook’s approach, depth and scope is a clear precedent for the way I combine theory to form useful design tools in this thesis.

2.5.2 Sharma et al’s 3-Level Hierarchy of Mechanic

Cook’s concept of hierarchy of mechanics, ranging from tiny actions to grand, game-encompassing mechanics, seemed to me like a fundamental idea that should have a lot of exploration. However, I did not see such ideas in the game literature. Specifically, I could not find any existing theory where user activity (or “core mechanic”) was broken into hierarchical categories. I asked two different expert individuals and posted a query on one online community of game studies experts, and no one stated they had heard of such a system.

However, I did find that researchers from other fields have similar relevant theory. Sharma et al (2007) cited how DARPA (2005) classified learning-related activities using an eleven level hierarchy from Memorisation to Differing using multiple criteria (speed, quality, reorganization, and scale). Sharma et al found it useful to simplify this into three levels: strategy, tactics, and motor (p. 1046). For example, if one used Sharma et al’s hierarchy to describe a soccer game, it might be summarized in this way:

1. "Strategy" (e.g. in soccer, switched from zone defence to man-on-man). A player might change this decision every 3 minutes or so.
2. "Tactics" - (e.g. "move closer, so opponent's options for passing are reduced.") I make these every 5 seconds or so.
3. "Motor" is tacit movement and placement of feet, made tens of times per second.

Sharma et al's work is in the field of artificial intelligence research, which is reasonably relevant to game studies as they were developing a computer opponent for a real-time strategy game. I discuss Sharma's work in my considerations of future research topics in the final chapter.

2.6 Fidelity in Simulators

This section reviews the literature around the term "fidelity" from the field of training simulation, setting out key points regarding how and why simulations work, how simulations are different from learning games, and what ideas from its literature relate to this thesis.

The term "simulation" can be confusing, as it has three separate meanings in the literature.

Simulation can mean:

1. A serious game that simulates a real-world activity (Aldrich 2009). e.g. *Microsoft Flight Simulator* (Williams 2006). This research uses the term "simulator" (not "simulation") for this meaning.
2. A procedural representation of a real-world phenomenon, often in a software program e.g. physics simulation (Hecker 2000).
3. A mental process in a human (Chi 2008) e.g. "in her mind, she is simulating our reaction."

The real-world value of the first meaning of simulators is beyond question. It is now "unthinkable" to graduate pilots, navigators, emergency responders, and a few other professions without completing training on a simulator (Crookall 2011). Survey studies show simulators are effective, although publication bias and other weaknesses were noted (Sitzmann 2011).

There are many kinds of simulators. "Lightweight" or computer-based simulators which require standard computer hardware, usually based on game technology (A. L. Alexander et al. 2005), are most relevant to this research. Learning goals for such simulators can be divided into two types. One type is problem-specific (Halpern 1998). Problem-specific learning transfers well: "learning to fly a particular flight pattern in a Cessna should transfer to a Saab aircraft, given that the pattern shares the same logical and procedural structure in both aircraft" (A. L. Alexander et al. 2005, 3). Note that the underlying theory, "identical elements," is valid only for short-term tasks that require only low-order thinking, such as assembling machines and pairing adjectives (Schendel and Hagman 1982; Underwood 1951, cited by Baldwin and Ford 1988). The value of simulators in higher-order thinking is less clear, as seen in the discussion around "fidelity."

2.6.1 Fidelity

The idea of “fidelity” in simulations is important to this thesis because the idea of fidelity is remarkably similar to activity-goal alignment, and is one of the most important ideas in simulation literature. However, fidelity is not as similar to activity-goal alignment as it might appear.

In the context of simulators, “fidelity” is the extent to which the virtual environment emulates the real world. Many subcategories of fidelity exist (e.g., Allen, Hays, & Buffardi, 1986; Hays & Singer, 1989; Lane & Alluisi, 1992; Lintern, Roscoe, Koonce, & Segal, 1990; Rehmann, Mitman, & Reynolds, 1995, as cited in A. L. Alexander et al. 2005), although the term is reasonably consistent in meaning¹¹. Fidelity is often split into physical and functional. With physical fidelity, the representation “looks, sounds, and feels like the operational environment, but includes the physics models driving each of these variables” (Baum, Riedel, Hays, & Mirabella, 1982). In relation to video game design literature, physical fidelity maps to the Aesthetic component of the MDA model, and to the visual component of the Skill Chain. Functional fidelity is more closely related to the idea of mechanics from the game literature: it refers to the hidden systems and equations that drive the operational equipment of the simulator.

Isolating a feature to impact only one type of fidelity is not easy. For example, Alexander et al consider a design decision: should one make in-game chat text-based, or voice-based, in a multi-player game? In-game text chat is thought to increase interest, realism, motivation to perform, and attention to details (A. L. Alexander et al. 2005), which improves functional fidelity, but text chat windows requires a the user interface change, reducing the rendering window size, which reduces physical fidelity. The correct design decision is not clear.

Some theorists consider 3D graphics a crucial part of good learning game design (Rice 2007), but others have shown that visuals are not required for good learning. Specifically, visual fidelity is remarkably unrelated to the concept of transfer (Moroney and Moroney 1999). Intentionally reducing visual fidelity can actually improve transfer: beginning pilots using a simple wooden airplane with crude visual representations learned well, possibly due to reduced stimulation that allowed them to focus on basic tasks (Caro 1989). Similarly, a game-like firefighting simulation that

¹¹ “Fidelity” is notable partly because of the reliability of the term itself in the literature. Unlike “mechanic”, simulation researchers enjoy a shared understanding of this term even though it embodies a complex collection of ideas. The success of the term “fidelity” thus forms a useful precedent for learning game research, which struggles with varying meanings used for the same term, a problem Aldrich (2009) called the “Tower of Bable.”

emphasized strategy, without any visual representation of the “on-the-ground” situation, was found effective (Toups, Kerne, and Hamilton 2011)¹².

Others have argued that fidelity, in general, is sometimes overemphasized, and that this overemphasis can impair learning efficiency. Consider impossible events such as a pause button, teleporting or “avatar flying.” Some argue that these should be removed from simulators because they are unrealistic and hurt fidelity (Turner et al. 2000) while others argue that impossible events bring benefits (reduced time in simulator, superior understanding) that outweigh the cost of reduced fidelity, and thus can increase overall efficacy of a training session (Darken and Peterson 2002).

2.6.2 Fidelity and Reality in Commercial Games

Fidelity is not a common term in commercial game design literature, as commercial game designers do not usually seek to represent reality (Salen and Zimmerman 2003, 446) or teach. However, there is some interconnection between fidelity and ideas in video game design literature. Most games contain some simplified simulation of reality, although it is modified so much that it has no meaningful connection to reality. For example, in platformer games like *Mario* (Becker 2007), characters bounce off walls in a way that simulates the conservation of energy: the harder you hit, the farther you bounce. Particle behaviour roughly simulates motion based on wind, friction, and other physical properties. However, as discussed previously in this chapter, designers are careful not to feel constrained by the laws of reality, if breaking the law is more fun for the player. Platformer games usually allow the player to steer the player character while airborne, which breaks both physical law and requires overriding physics simulators (spek 2011), but most designers agree that one should break any law that lets the player have more fun. This is consistent with literature around play: as discussed in Section 2.2.2, player agency is essential to play.

Simulation of non-player characters behaviour in games is a notorious design challenge (Chris Crawford 2003). Most games use a “bottom-up” approach, preferring to build simple behaviours that appear complex, rather than correctly simulate the complexity (Mark 2008). However, leading developers believe this will change: “Think of how much aviation changed in 100 years from the Wright brothers to today's jet engines and stealth bombers. Convincing AI is our jet engine. It's certainly possible, but I think it will take decades to get there” (Mark 2008).

¹² The paper's title “Zero Fidelity” can be confusing because the simulation in question had high functional fidelity even though there was almost no visual representation.

2.6.3 Fidelity in Learning Games

Fidelity is part of learning game literature but is not always discussed in a positive light, possibly because it can constrain learning goals. Play theorists argue that fidelity is always related to the representation of reality, whereas play and games can be used to teach a variety of things, not all of which are real (Singer, Golinkoff, and Hirsh-Pasek 2006). Thus, functional fidelity is not necessarily valuable in learning games. Persuasive games, for example, seek to convince the player of a certain position, not teach objective reality (Bogost 2007a). For example, Squire noted that the designers of *Operation: Resilient Planet* originally correctly simulated the ecology of predators and prey, but later replaced it with a “fake” ecosystem by stripping out the simulation and simply scripting events which enabled them to focus instead on the player experience (Squire and Patterson 2009).

2.7 Learning Game Design: What Makes a Good Learning Game?

Learning game design theory is clearly important to this research. There is some agreement, in both theory and practice, on certain key elements of a good learning game; however, every survey of learning game design I reviewed has at least minor variations on these elements (Kirriemuir and McFarlane 2004; K. A. Wilson et al. 2008; A. L. Alexander et al. 2005; Egenfeldt-Nielsen 2006). Relevant design elements from recent survey (Crookall 2011) include:

- Engagement
- Trust
- Challenge
- Balance between simplicity and realism (according to learning objectives)
- Integration of affective and cognitive dimensions
- Clear learning and game goals

Gee has also written on this topic, and his prominence in the field of learning game design merits some discussion. Coming from a background in semiotics, Gee’s “What Video Games Have to Teach Us about Learning and Literacy” proposed novel relationships between play, identity, and learning (Gee 2004). His outsider’s perspective allowed him to skirt certain issues which had long engaged researchers (e.g. narratology/ludology). His fresh, informal theorizing from simple autoethnographic observations captured a “common sense” middle ground of learning game literature. Gee gives 36 properties that relate game design to learning (2004). He later refined these for an academic audience (Gee 2009a). Gee’s ideas often align with accepted theory. For example, Gee states that a game’s user interface for weapon selection is “a **model of the game’s weapon system**, an abstract

picture of it made for planning, strategizing, predicting, and problem-solving.” Gee’s idea of how the aesthetics, dynamics, and mechanics interrelate matches well to the MDA model, which he cites only in later work (Gee 2009b), suggesting that he independently developed many of the same ideas.

While broad ideas such as “trust” can help inform a designer’s thinking, they are difficult to apply in production. For that phase of development, other researchers offer heuristics for learning game design. These three examples are from MIT’s Education Arcade (Klopfer, Osterweil, and Salen 2009):

- **“Find the fun in the learning”**; focus on and enhance that fun as a core game dynamic
- **Go with the right tech.** Not enough time and effort is spent on simple technologies
- **Don’t over focus on “soft skill”** aka digital literacy skills. Include it only when it fits.

While Klopfer et al give broad, overarching advice, Shelton and Scoresby gave three specific approaches to learning game design. The first approach is “Classic”: Entertainment games are modified for instructional purposes: in-game rewards (points, achievements), which are extrinsic to educational goals. These are good for less complex learning goals such as multiplication tables. The second approach focuses on the use of mainstream games (“Commercial Off-The-Shelf” or COTS¹³) in education. For example, the hit game *Civilization* has been used in history classes (Squire et al. 2003). COTS is used for a variety of learning goals, including literacies, critical thinking, and formation of learning communities (Kirriemuir and McFarlane 2004). Disadvantages are known: COTs-based classroom exercises place a heavy burden on the instructor and it can be difficult to assess students’ progress (Kirriemuir & McFarlane, 2003). Off-task activity is another challenge, as progress through any video game is inherently unpredictable (Squire et al. 2004). The third approach uses custom-built learning games. Compared to COTS, learning games have built-in progression, scaffolding, and assessments. This focuses the students on complex learning issues aligned with meaningful goals. Custom games are criticized for having weak engagement (Squire et al. 2004); however, some learning situations do not require strong engagement. For example, corporate training simulations can work because audience has extrinsic motivations, or participants don’t like playing games anyway. Others have noted custom learning games are prone to design failure and risky to build (Rice 2007).

¹³ In the context of learning games, the acronym “COTS” or “COTs” stands for “Commercial Off-The-Shelf” video games. COTS games are primarily intended to entertain, not teach.

Each of these approaches is suitable to certain types of learning goals. Clearly, identifying types of learning goals is important to learning game designers. The next section explains how learning goals might be distinguished from each other.

2.8 Deep Learning and Higher Order Thinking

Aside from activity-goal alignment theory, this research relies on the concept of “higher order thinking” and Gee’s related concept of Deep Learning (Gee 2009a). I explore Gee’s Deep Learning theory in chapter 6, so it is not presented in depth here. I give a broad definition of the term in chapter 1, and here provide a discussion of the relevant literature, including Gee’s key references, showing how the concepts of higher order thinking relate to deep learning.

2.8.1 Deep Learning

The idea of deep learning is common in a branch of cognitive science known as “learning science” or “the learning sciences” (R. K. Sawyer 2005). Although I was unable to find a clear definition of the term “deep learning,” its meaning is reasonably clear from the learning science literature. Sawyer used the phrase “the notion of deep learning” and set out a table that clearly distinguishes “Deep Learning” from traditional classroom practice, shown in Figure 7.

Table 1.1. Deep Learning Versus Traditional Classroom Practices

<i>Learning Knowledge Deeply (Findings from Cognitive Science)</i>	<i>Traditional Classroom Practices (Instructionism)</i>
Deep learning requires that learners relate new ideas and concepts to previous knowledge and experience.	Learners treat course material as unrelated to what they already know.
Deep learning requires that learners integrate their knowledge into interrelated conceptual systems.	Learners treat course material as disconnected bits of knowledge.
Deep learning requires that learners look for patterns and underlying principles.	Learners memorize facts and carry out procedures without understanding how or why.
Deep learning requires that learners evaluate new ideas, and relate them to conclusions.	Learners have difficulty making sense of new ideas that are different from what they encountered in the textbook.
Deep learning requires that learners understand the process of dialogue through which knowledge is created, and they examine the logic of an argument critically.	Learners treat facts and procedures as static knowledge, handed down from an all-knowing authority.
Deep learning requires that learners reflect on their own understanding and their own process of learning.	Learners memorize without reflecting on the purpose or on their own learning strategies.

Figure 7: Sawyer compared deep learning to traditional classroom practice (R.K. Sawyer 2005, 4) .

Sawyer's table might serve as a definition of deep learning by example, as it makes the idea comparatively clear. Additional insight into the meaning of deep learning in cognitive science can be found in the writings of Bransford, an eminent learning science scholar. Bransford used the term "deep" in one of his three key findings about learning sciences:

To develop competence in an area of inquiry, students must: (a) have a deep foundation of factual knowledge, (b) understand facts and ideas in the context of a conceptual framework, and (c) organize knowledge in ways that facilitate retrieval and application. (Bransford 2000, 15)

Bransford explained his intent behind advocating a deeper conceptual understanding in learning:

Deep understanding of subject matter transforms factual information into usable knowledge. A pronounced difference between experts and novices is that experts' command of concepts shapes their understanding of new information: it allows them to see patterns, relationships, or discrepancies that are not apparent to novices. (Bransford 2000, 16)

In Sawyer's book that he claimed was an introduction to "the new field of learning sciences" (R. K. Sawyer 2005), Sawyer described a consensus of learning scientists on several key ideas including the importance of deeper conceptual understanding. In short, Sawyer and Bransford largely agree that the concept of deep learning is well established and largely accepted in the literature of learning science.

There are other key principles around which learning scientists have achieved consensus: focusing on learning in addition to teaching, creating learning environments, the importance of building on a learner's prior knowledge, and the importance of student reflection (R. K. Sawyer 2005, 3). These principles are presented as equally important but less closely related to Gee's conception of deep learning, so I do not discuss them further.

How does the concept of deep learning relate to prior and other research around similar ideas? Many scholars discuss the importance of engaging students with activity that is not simply rule-following (J. Herrington and Oliver 1999), employing the term "higher order thinking." Baker (1990) set a low bar for the term: "all intellectual tasks that call for more than information retrieval". In contrast, Lewis and Smith were more specific in their definition, requiring that the student take new information and interrelate it with stored information to address perplexing situations (Lewis and Smith 1993).

In some ways, both deep learning and higher order thinking suggest a hierarchy of intellectual behaviour. Indeed, the canonical hierarchy of intellect, Bloom's taxonomy (Bloom 1956), is sometimes specifically cited in learning game research (Rice 2007). However, not all agree on this single-dimensional approach: Newmann (1990) rejected categorization by particular types of thinking (e.g. critical, inductive, moral), and instead suggested the term can be used to identify educational interventions that are "non-routine challenges" that "promote thoughtfulness" (p.44). However, Newmann still distinguished types of learning by higher and lower order.

"Higher order" versus "lower order" thinking might appear oversimplified to the reader. Is it valid to divide types of learning goals into only two categories, given that there are so many kinds of learning? For example, is it valid to cluster activity-goal alignment scores as "high" or "low," as I do in the final chapter of this dissertation? Philosophers have questioned the validity of categorizing indeterminate phenomena (Alfino 1991). While a detailed review of this debate is beyond the scope of this research, one might summarize it this way: One side of the argument, which I call "deconstructionist" here, asserted that a divisive approach to ill-defined multivariate domains is a valid, if questionable, form of design-based inquiry, and the other side (positivist) did not. This deconstructionist position might seem reasonable and appropriate in the context of this research; however, there are reasons for concern. For example, one might use higher-order thinking to create the logical fallacy known as a false dichotomy (e.g. "you are either with us or against us"). In the literature around higher-order thinking cited above, such problems were not evident. For example, Herrington and Oliver did not seek merely to divide all learning goals as being either authentic or not. Instead, they considered how authenticity might help us understand how learning goals vary. Herrington et al's approach, I argue, seems reasonable as a researcher as well as useful as a designer. Of course I acknowledge that any binary division is a simplified view of a scalar value, which in turn is a simplified portrayal of one aspect of a learning goal.

Is it possible to teach higher-order thinking? Newmann stated the answer is complex, but that research shows some success in teaching deductive and inductive reasoning, moral reasoning, informal reasoning, creative thinking, and metacognitive strategies (Newmann 1990).

2.8.2 Rice's Cognitive Scale

Rice identified the idea of higher order thinking in the context of learning game design, and proposed a cognitive scale, complete with scoring rubric, to assess learning games. Rice stated that complex interaction is required for higher order thinking (Rice 2007, 91). Rice argued "higher order thinking" is a worthy endeavour for learning games. This argument fits the literature. Many veteran designer/researchers have specifically advised designers to target higher order learning goals (Devlin

2012d; Klopfer, Osterweil, and Salen 2009). Gee also valued higher order thinking as an appropriate target for learning games (Gee 2006).

Rice did not develop a detailed justification for the theoretical foundations he uses in his work, and I observed some confusing contrasts. For example, Rice cites the upper levels of Bloom's taxonomy in explaining what "higher order thinking" is, states the research aims to identify "gaming elements leading to higher cognitive processing" (p. 91) and focuses on cognitive science in his discussion of Gee (p. 92). These statements suggested his work should be considered in a cognitive science context, but he does not state this, and his methods do not resemble most cognitive science research methods. Rice cited *Math Blaster* as an important example of "lower order thinking" in part due to its commercial success, noting that *Math Blaster* tests simple mathematical calculations and memorisation, using simple narrative and action gameplay as a reward.

Rice cites Squire's (2004) ongoing work as a good example of how video games can be used to teach higher-order thinking. Squire has been studying how social studies students "developed various unique and meaningful understandings in world history, geography, and politics through *Civilization III* gameplay," developing solutions to challenges of using commercial games in the classroom for many years (Squire 2010). Another example of higher-order thinking in learning game is *Operation: Resilient Planet* (Chmiel 2010), which teaches logical reasoning through a scaffolded mechanic of drag-n-drop argument construction. It is cited by experts as an exemplary project (Squire and Patterson 2009).

Rice proposed a "cognitive viability" scoring system whereby the designer assigned a point for each item on a 12-item checklist, and the total. For example, the rubric scored one point for meaningful interaction with non-player characters (e.g. computer-controlled sidekick) "because it offers additional opportunities for thinking." Rice warned against virtual environments of highly realistic empty streets, in which users could roam around but not be involved in additional interactions such as chatting with NPCs or other players. Pre-scripted dialog was deemed sufficient "provided the interaction is meaningful somehow" (Rice 2007, 95)

Rice coined the term "virtual interactive environments" ("VIEs") to refer to a type of game specifically designed to teach higher-order thinking. Rice notes that "Virtual implies that complex three-dimensional graphics are used to create a form of Virtual Reality (VR)" and specifically excludes 2D games and text-based games from VIEs. Then Rice states Cognitive VIEs are required, not just helpful, to teach higher-order thinking. This statement apparently conflicts with *Civilization*, an example of higher-order thinking he provides later in the paper, and other literature.

Despite certain reservations noted here and discussed further in chapter 6, Rice's ideas and approach are useful and I use Rice's work in several places in this research. Rice's work partially informs the AGA-based Assessment Tool developed in chapter 6. The relation between Rice's work and this research is discussed in the final chapter.

2.8.3 Lower Order Thinking Games

The literature is somewhat confused on the value of games that target lower order thinking, but most theorists state that not all such games are entirely useless. "Drill and kill has some value on lower learning levels, and this may help explain the high sales ranking of *Math Blaster* titles. Students weak in math indeed do need time to practice their math skills"(Rice 2007, 90). Devlin states that mastery of multiplication requires "sufficient repetition, and I know of no better way to achieve that than with an entertaining video game" (Devlin 2012c).

There is some reason to believe deep learning and higher-order thinking could be taught with games. Highly regarded designer-researcher Katie Salen led a well-funded team development team to create an award-winning learning game *GameStar Mechanic* that teaches higher-order thinking. Studies found it effective for teaching a particular type of higher-order thinking (Games 2008). This game and its learning goals are discussed in chapter 5.

2.9 Conclusion

This chapter has aimed to provide an understanding of the ideas related to activity-goal alignment (e.g. fidelity, mechanics, alignment), as well as discuss concepts used in this thesis to extend Shelton's activity-goal alignment (e.g. Gee's "deep learning"), as well as present fundamental theories of learning upon which this research is based. In doing so, it has laid the groundwork necessary to understand learning games through the lens of activity-goal alignment.

3 Method

3.1 Introduction

In this chapter I present my methodology. I begin by describing my practical activities within this research, and in the rest of the chapter I discuss the method behind this practice. I position this research in relation to all design research, and discuss differences between my approach and conventional approaches. I introduce analytic autoethnography, discuss its weaknesses, present my qualifications as an autoethnographer, and argue that it fits the constraints and goals of the research. Finally, I discuss ethical considerations.

3.1.1 Practical View of this Research

Before discussing the merits of my method, I present a “hands-on” view of the activities I have undertaken in this research. From a practical point of view, I have created and demonstrated two design tools based on activity-goal alignment. The first is the “AGA Scoring Tool,” discussed in chapters 4 and 5. The second is the “AGA-Based Assessment Tool” discussed in chapters 6 and 7. For the AGA Scoring Tool, I took the following steps:

1. Document the discovery of the problem through analytic autoethnography

In chapter 4, I give an autoethnographic narrative that documents a problem related to activity-goal alignment in a recent learning game development project *Battlefood*. I then analyse and amend my narrative to discuss my various modes (diagrams, meetings, written documents) in my attempts to discuss activity-goal alignment with my co-workers.

2. Design and create an AGA Scoring Tool

In chapter 4 I explain how I developed requirements for the AGA Scoring Tool, and document the important steps in my iterative self-reflective design method to create the tool itself. I then present the final version of the AGA Scoring Tool, and discuss the purpose of its components, explaining how each part responds to a need I identified previously.

3. Examine activity-goal alignment of existing games

In chapter 5 I applied the AGA Scoring Tool in discussions of existing learning games. I compared games with high and low activity-goal alignment to show how scoring activity-goal alignment can be applied to clarify a discussion of video game design.

In developing the AGA-Based Assessment Tool, I varied my approach. Instead of autoethnographic narrative, I critically reviewed existing assessment tools, as the literature forms a useful foundation for a new tool. Using a similar procedure as with the previous tool, I identified criteria for a tool, develop a tool against those criteria, and exercise the tool in discussions of several existing learning games.

In summary, this research uses various methods of autoethnographic investigation: narrative of game development, discussion of existing games, critical reviews of existing assessment tools, and recounting the logic behind the process of tool design. The next section explains the reasoning behind this approach.

3.2 Position within Design Research

This thesis is located in the field of design, and as such is strongly informed by ideas in the design research literature. This research is practiced-based, in the sense of Roth's statement "Design research is tied to practice and driven by its needs" (Roth 1999, 18). More specifically, this research takes a constructionist position which holds that designing (making things) in itself is not research without scholarly reflection upon the process of making (Cross 2001; Dorst 2008, cited in Feast and Melles 2010). This position manifests throughout this thesis. For example, I critique learning game assessments in chapter 6. I break from academic practice such as seen in Dondlinger (2007) by criticising certain peer-reviewed journal articles as being too removed from practice, and instead suggest commercial taxonomies of learning games could be just as valid a data source. Section 3.4 of this chapter expands on this reasoning.

This example shows my practice-based approach to design research is essentially constructionist: although I prioritise practical concerns over scholarly ranking in my practice, I do not reject the scholarly approach. I test theory by building things, and then test them, in an iterative process. As I do, I reflect on the existing theory in action to develop further theory.

3.2.1 Learning Game Design as a "Wicked Problem"

Cross argued that a designerly approach is sufficiently different from science and humanities that it represents a "third way of knowing" (Cross 1982). I do not go that far, although Cross's argument is remarkably convincing, and I acknowledge that Cross's position is not dissimilar to my own.

Buchanan (1992) suggests design research is useful for what Churchman termed "wicked problems" (Churchman 1967): a category of problems that possess the following properties:

- *Wicked problems have no definitive formulation, but every formulation of a wicked problem corresponds to the formulation of a solution*
- *Solutions to wicked problems cannot be true or false, only good or bad*
- *For every wicked problem there is always more than one possible explanation, with explanations depending on the designer*
- *No formulation and solution of a wicked problem has a definitive test*
- *Every wicked problem is unique (Rittel and Webber 1973, 163)*

Both instructional design and game design present problems that meet all of the above criteria, and have been named wicked problems (Eladhari and Ollila 2012). If the practice of instructional design and game design are both wicked problems, and learning game design problems combine instructional design and game design problems, then it follows that learning game design problems are wicked problems. Of course, it is theoretically possible to constrain a learning game design until it is simple enough not to meet the standards described here, but I have never observed encountered such a simple design problem in my practice. I argue it is reasonable to generally consider the challenge of designing a learning game a wicked problem.

Wicked problems are distinct from "tame problems" which Rittel and Webber argue are appropriate for scientific methods (1973). The research questions I ask are not suitable for precise measurement. The problems are so complex and poorly understood that not only can I not measure them precisely, but the idea of measurement is inappropriately constraining. Thus I looked outside of the scientific approach in choosing a methodology to conduct this research.

3.2.2 Design Research for Learning Games

What does design research mean, in the context of learning game research? John Christopher Jones' seminal book "Design Methods" gives four important properties of design methods:

- (1) the user should inform design
- (2) practice is informed by theory while theory should derive from practice
- (3) good design practice combines intuitive and rational approaches
- (4) good design practice is iterative (Jones 1992)

As a form of design research, this thesis seeks to implement all of these principles in its method. Specifically, point (4) is addressed because the development of the AGA Scoring Tool undergoes

multiple iterations. The choice of analytic autoethnography as a method reflects the importance of principle (3), as it combines intuitive and rational approaches. The groundedness of the AGA Scoring Tool in usefulness to practising designers reflects principle (2).

My interpretation of points 2 and 3 suggest that, as a design researcher, I must analyse activity-goal alignment in an authentic context. Barab and Squire (2004) stated "...design-based research focuses on understanding the messiness of real-world practice, with context being a core part of the story and not an extraneous variable to be trivialized." Accordingly, in chapter 6 I analyse the AGA Scoring Tool not by theorizing about it but by employing in a real-world practice of design analysis: I discuss certain learning games that teach three topics (math, human cellular biology, and game design), and suggest that activity-goal alignment is a tool that reveals an important truth in a complex design. This discussion is meant to exercise the AGA Scoring Tool, not suggest a new approach to a specific topic (e.g. teaching math via video games).

The first principle, user-centred design, is particularly important for this research. The game design literature has convincingly built a case for user-centred design. Video games, along with most types of software design, adopted the philosophy of user-centred design in the 1990s (Grudin and Pruitt 2002). By 1992 user-centred design was common practice (Vredenburg et al. 2002) and was codified as a standard in 1999 (ISO 1999). In modern design practice, game designers are expected to constantly ensure that the user experience is coherent and powerful in every key decision made in game development (Schell 2008a). For example, user-centred design is an essential principle of the Agile development methodology (Martin 2003), which is commonly used in game development (Keith 2007). My choice of method, such as autoethnography and comparative analysis of existing assessments, responds to this need for a user-centred approach to design research.

3.2.3 The Case against the "List of Attributes" Approach to Understanding Games

In chapter 6, I conclude that many existing assessments of learning games use lists of attributes, and criticise this overall approach. Given how many papers use the "list of attributes" approach in design literature, a sceptical reader might wonder if it is justified to switch away from such a common approach. In this section I lay groundwork for this proposal by generally considering the methodological implications of the deconstructionist approach.

Learning game research often aims to explicitly identify and discuss isolated attributes of learning games. An important precedent-setting example was Malone's "Key Characteristics of a Learning Game," which listed four key attributes: Challenge, Curiosity, Control, and Fantasy (1987). Malone's work is often cited in the literature, and numerous researchers have proposed additions to his list

(Schaller 2005; Pavlas and Bedwell 2009; E. Z. F. Liu and Lin 2009). While not building on Malone's particular work, many recent scholars (e.g. Dondlinger 2007; Schaller 2005; Consalvo and Dutton 2006) continue to study isolated attributes to make meaning. However, Shelton rejected this approach in developing and applying activity-goal alignment (Shelton 2007) as I reviewed in chapter 2.

While acknowledging that this attribute-isolating approach has value, I argue that the following reasons justify alternate approaches:

1. Building on early works carries the risk of repeating early assumptions. Malone's work uses scientific methods for a wicked problem, which I find is a poor fit. If my research built on Malone's work, it would suffer from that problem as well.
2. The list of attributes of learning games is endless. Learning game designs vary widely (Klopfer, Osterweil, and Salen 2009), and prior research has found that attributes vary among games (Dondlinger 2007). If learning games vary widely, and there is no list of attributes that the literature agrees comprises a learning game, then not all attributes are relevant to all learning games. Thus it stands to reason that a search for a definitive list of attributes will never end. I do not argue that endless lists are completely useless, but I see it as a drawback in that there is no hope of mastering the literature.
3. This "list of attributes" approach has been extensively researched already. The history and volume of research using this approach can be seen in the number of citations in this section and the literature review. Compared to the humble volume of alternate approaches, I suggest alternate approaches are worth considering.

How should practising designers use lists of attributes? While attributes are often cited in scholarly articles generally, it is unclear how practising designers should apply them. Commercial game designer Perry cautioned designers against misusing his "book of lists" (Perry and DeMaria 2009, 1), arguing for more holistic design reasoning. I argue against a naïve checklist approach to assessment of learning games in chapter 6.

To restate my position, I acknowledge that a great deal of learning game research has focused on identifying lists of attributes, and this research has great value. I aim for my research to supplement, rather than contribute directly to, the attribute-isolating approach. I suggest that it is worth trying alternative approaches to the attribute-based method typically employed, in order to understand activity-goal alignment.

3.2.4 Practical Implications

This position affects many decisions within this research. For example, in chapter 6 I review commercial learning game assessment tools. These tools typically seek to identify attributes of learning games that distinguish one type of game from another, in the eyes of the taxonomy users (which may be academics, players or designers). These assessment tools often use some kind of method. This method of classification is intended to be useful to the users, by the authors. Thus the assessment authors reveal what they believe is important about learning games.

Here I survey existing classifications of learning games, but not in the scientific sense. Classification often aims to objectively identify shared qualities and characteristics without bias, but some argue that classification is often subjective: "Identifying a genre taxonomy is a subjective process and people may disagree about what constitutes a genre, or the criteria for membership of a particular genre" (Finn and Kushmerick 2006).

Within the designerly approach, there are three reasons why taxonomies, not other forms of game analysis, are studied in this thesis.

Firstly, it is difficult to 'play favourites' with a particular learning game attribute when reviewing many games. Designers may be tempted to focus on attributes that are apparent in the particular interventions they design even if that attribute is rare or unimportant, as Reeve noted (1993). Secondly, taxonomists revisit their criteria over and over as they update their catalogue. Buckleitner noted this repeated reflection can develop a deeper understanding (2012). Finally, compared to other authors of attribute lists, taxonomy authors must populate their taxonomies. This means, they must be familiar with many types of learning games, a type of literacy that is valued in this context (Klopfer, Osterweil, and Salen 2009).

To create these categories, the author must decide what is important in learning games, according to their users. Thus the categories can be seen as a list of attributes of learning games. For these reasons, I review only taxonomies that are intended to have practical value to their readers.

3.3 Autoethnography

Analytic autoethnography is the most important method in this research. Although the narrative in chapter 4 is most obviously autoethnographic, all methods in this research are somewhat autoethnographic. In the introductory chapter, I explained how the need for this research has come from my own practice. From the literature review to the final conclusion, I have allowed my own

practice as a game designer and research to influence and guide my research. The risk of subjectivity is apparent in the preceding statement, yet below I argue that my method addresses this risk.

3.3.1 Ethnography in Design Research

Recent design researchers suggest ethnography offers solutions to known weaknesses with existing methods: “We need a better way to respond to the challenge of communicating design findings, and ethnographic writing methods provide at least a start in that direction.” (Frankel 2009, 3514) Autoethnography of games, used with other methods, has been used to understand game design (Itu 2010), although it is not a dominant method. What is autoethnography and why is it an appropriate choice of method for this research?

Autoethnography is a form of ethnography. Ethnography has been defined as “a process in which the researcher participates, overtly or covertly, in people’s everyday lives for a sustained period, watching what happens, listening to what is said, asking questions and collecting any other relevant information” (Griffin and Bengry-Howell 2008, 2), p.2). While this definition hints at ethnography’s anthropological roots, it is now used in a variety of disciplines (p.2).

I have emphasized the most important ideas in Tedlock’s discussion of several important characteristics of ethnographic methods:

- Ethnography seeks an **insider's perspective**, aiming to understand a phenomenon from the points of view of those involved—which can be diverse, complex and contradictory.
- **Natural setting**: i.e. not in the research laboratory or in an artificially manipulated or controlled environment.
- **Loose structure**: Research questions, hypotheses and data collection procedures may develop as the study proceeds, although most ethnographic studies do begin with a set of research questions, an overall research design and a strategy for data collection and analysis.
- **Reflexivity**: Unlike many positivists or experimental researchers, ethnographers do not set out to deny their impact on the world they are studying, rather their aim is to understand this in a systematic and rigorous way as an inevitable part of the research process (Tedlock 2005)

Reflexivity is perhaps the most difficult aspect of ethnography for traditional researchers, as it is negatively associated with subjectivity among traditional “hard science” researchers (Tedlock 2005). However, reflexivity can be the foundation of a fundamentally different approach to research. Steier (1991) regards research itself as a bridging process, connecting understanding between communities that speak different language. In this context, he regards reflexivity as essential to good research. In the case of design research, renowned industrial designer David Kelley uses ethnographic methods as a way to bridge the user experience and the developer’s perspective. Further, Kelley has shifted his entire design consulting practice to “design thinking,” a process that is based on building reflexivity and creativity for the clients (Kelley and Littman 2001).

As Tedlock noted, the strength of ethnographic research like this comes from highly personal involvement: “By entering into close and relatively prolonged interaction with people (one’s own or other) in their everyday lives, ethnographers can better understand the beliefs, motivations, and behaviours of their subject than they can by using any other approach” (Tedlock 2005, 155). Like any method, ethnography has limitations. One fundamental limitation is particularly important here. Ethnography traditionally suggested researchers attempt “participant observation”, an oxymoronic term that requires ethnographers to somehow be both dispassionate observers and close to the subjects they study (Tedlock 2005). This limitation of ethnography can be resolved by choosing autoethnography as a method because autoethnography allows the researcher to be self-reflexive: the observer and the subject are identical (L. Anderson 2006).

3.3.2 Autoethnography and Its Benefits

Autoethnographies are “case studies that follow the tradition of ethnographic research” that value inner knowing. This “Value of Inner Knowing” is the important difference from ethnography (Duncan 2008, 30). It provides access to “insider meanings” (L. Anderson 2006) which are essential to truth-finding in subtle, complex problem that are difficult to articulate. In seeking an answer to her research question “How can I improve my practice of hypermedia design?” Duncan’s research requirements resemble mine, so her choice of method is particularly relevant to this research. Duncan explained why she chose autoethnography:

I needed a method in which the lifeworld [sic] and internal decision making of the researcher were considered valid and noteworthy. I needed methods that encouraged systematic reflection and ensured a scholarly account. I needed a means of analysing evidence that not only organized a record but also enabled discovery. What I needed was autoethnography. (Duncan 2008, 28)

As a method of grounded research, Duncan found autoethnography prevented theory from becoming disengaged from practice. This is important to learning game design research, to avoid repeating past failings (Reeves 1993), and is advocated as best practice by both academic (Klopfer, Osterweil, and Salen 2009) and practice-based (Tate, Haritatos, and Cole 2009) thought leaders.

For example, autoethnography brings a unique strength to my procedure in chapter 6, where I play each game once but analyse that experience twice. By drawing two analyses from a single play experience, I control for contextual factors (e.g. time of day, attitude, and/or in-game choices) that plague other methods. It is common knowledge that context is crucial to fairly assessing the experience of playing any game (Chris Crawford 2003), so I argue this advantage is significant enough to outweigh the risk of bias.

3.3.2.1 Autoethnography Addresses Key Weakness of Design-Based Ph.D. Research

One important drawback for design-based research is the size of effort required. While design-based research must "address complex problems in real contexts in close collaboration with practitioners," and use iteration and refinement, many design methods require more time and effort than one student can provide (J. Herrington et al. 2007). I argue that the autoethnographic method is appropriate to address the problem of practicality for Ph.D. design students. Autoethnography allows efficient iterative study required of design-based research. As compared to other forms of ethnography, autoethnography allows efficient iteration of data and analysis (Hayler 2011).

3.3.3 Risks and Limitations of Autoethnography

While autoethnography may enjoy many advantages of ethnography while mitigating some of its risks, it carries its own risks. This section lists relevant risks identified in the literature, and explains how this thesis addresses them. Several types of author bias are broken down and discussed specifically.

Reviewing other autoethnographers' methodological findings, Chang named five key risks of autoethnography:

(1) excessive focus on self in isolation of others; (2) overemphasis on narration rather than analysis and cultural interpretation; (3) exclusive reliance on personal memory and recalling as a data source; (4) negligence of ethical standards regarding others in self-narratives; and (5) inappropriate application of the label 'autoethnography'. (Chang 2008, 22)

My choice of Anderson's analytic autoethnography mitigates many of these risks. Specifically, I argue that risks (1) and (2) are low for this research because only one section of chapter 4 in this research uses the narrative approach to autoethnography. The remainder of the research balances my personal experience with others' views and opinions, following Anderson's precedent as explained in Section 3.4 of this chapter. Risk (3) is mitigated by the use of varied data sources, including meeting notes, emails, and sketches supplied in chapter 4. Risk (4) is mitigated through pseudonyms and obtaining consent as discussed in section 3.4.2. Risk (5) requires me to define what I mean by "autoethnography" and distinguish it from other similar forms, such as autobiography or memoir.

Author saturation and solipsism (Hayler 2011) is another known risk of autoethnography. To understand this risk, it is helpful to recognize two types of autoethnography. Anderson coined the term "analytic autoethnography" (L. Anderson 2006) as distinct from "evocative autoethnography" (p.374). Anderson describes analytic autoethnography as "autoethnography that is consistent with traditional symbolic interactionist epistemological assumptions and goals rather than rejecting them" and suggests five attributes:

complete member researcher (CMR) status,

analytic reflexivity,

narrative visibility of the researcher's self,

dialogue with informants beyond the self, and

commitment to theoretical analysis. (p.378, my line breaks)

I claim that this research fulfils all five of these attributes. I believe this fulfilment is largely self-evident. Specifically, points 2-4 are apparent in my autoethnographic discussion of *Battlefood* in chapter 4: my findings are clearly grounded in analytic reflexivity because I have presented the narrative combined with analysis, not just raw recounting. The researcher's self (e.g. design decisions and predispositions) is also apparent in the narrative text. My commitment to theoretical analysis is also apparent in my relation of my findings to Shelton's activity-goal alignment theory, Gee's deep learning theory and other relevant theory from my field.

Even with these mitigations, my choice of autoethnography is not free from risk. I acknowledge the warnings of past autoethnographic design researchers: autoethnographic research must be executed with care to provide useful results.

3.3.4 Author's Background

After noting that autoethnography requires a deep belonging to the practice being studied, Hayler (2011) states that the analytic autoethnographer must be a "Complete Member Researcher (CMR)" (L. Anderson 2006). Autoethnographers must demonstrate that the researcher has the authority to be a valid source of data. I do so in the following discussion of highlights from my career by detailing the relevance of my experience and background to technical and practical work in the game design field.

I have been fortunate enough to be at the forefront of the personal computing revolution. I was 12 years old when my family first bought an Apple II+ personal computer, which makes me part of the first generation of children who grew up using personal computers at home. I have been designing and building video games since 1982, making me part of the first generation of developers of video games for the personal computer. I began using vector based graphics software to create artwork and content on personal computers in 1985, which was also innovative and cutting edge at the time. I began building 3D models on a type of advanced computer known as workstations in 1988, five years before it was practical on personal computers. In 1990, before personal computers commonly had hardware acceleration that allowed 3D graphics to be practical, I was building 3D video games using tools I built myself. Each of these opportunities allowed me to encounter new problems. For example, in building *Car and Driver* (Electronic Arts 1992), we used a method of applying 2D images to 3D surfaces called "texture mapping." This is a common practice for any video game artist today, but in 1990, it was rare outside of research laboratories. In fact, to the best of my knowledge, *Car and Driver* was the first game that ran on personal computers to use texture mapping in a 3D world. I had to find a way to place and edit the position of the image on the 3D surface. This was an entirely new problem. I designed and wrote a tool that placed UV coordinates in space, using AutoLISP, an obscure programming language. To do so, I studied matrix algebra and improved my programming skills sufficiently to make a tool that worked. From this experience, I developed early knowledge of UV coordinates, AutoLISP, texture stretching, and other key concepts that was very rare at the time.

I have been fortunate to have the opportunity to work in design capacities in a variety of video game development projects over 22 years, writing backstory, designing levels, exploring game mechanics, programming tools, and creating content (models, textures, and animations) for around 40 games. I helped design and build early 3D games for the PC, including *F22* (Sega Enterprises 1991), *Ultima Underworld: The Stygian Abyss* (Origin Systems 1992), *Car and Driver* (Electronic Arts 1992), and *Descent* (Parallax Software 1994).

My work has been recognised as innovative by reputable entities. As an independent consultant, I was recruited to work on projects by the U.S. Department of Energy, NVidia, and Disney, and was invited to the Game Developer's Conference Advisory Board, one of the most prestigious boards in the video game industry. This recognition of my work is relevant to my claim for credibility.

In short, my experience with early technical video game production, along with my breadth and depth of design experience, allows me to reasonably claim to be professional game designer.

3.4 Analytic Emphasis

I am following Anderson's approach to autoethnography in this research, which, as discussed above, is analytic autoethnography. I do so in part to address the weaknesses of the autoethnographic method discussed in the preceding section. My interpretation of analytic autoethnography is broadly informed by grounded theory (Glaser and Strauss 1977). Grounded theory has these properties:

- Grounded theory supposes a continuous interaction between the data collection and data analysis,
- Grounded theory fits with a multitude of qualitative methods, such as observations, conversations, interviews, autobiographies and organizational reports, and
- Grounded theory is generated by themes that result out of the data itself (Charmaz and Henwood 2008).

Grounded theory is not a method per se; rather, it is a broad framework for specific methods. Next I discuss methods used in this research that are not typical narrative autoethnographic methods, yet retain qualities of autoethnographies and also ascribe to the ideas of grounded theory.

3.4.1 Alternatives to Narrative in Autoethnography: "Textual Analysis"

After autoethnographic narrative, my primary method of investigation is a form of analytic process and writing called "textual analysis." There are many methods that have been termed textual analysis (McKee 2003), and there is significant confusion particularly in the field of video game studies around this term (Carr 2009). Broadly speaking, McKee defined textual analysis as "a way for researchers to gather information about how other human beings make sense of the world." When I use the term textual analysis, I do not refer to those methods that focus on rigid coding or categorization of written words. Instead, I mean a form of autoethnographic investigation that is informed by design research methodology. In recognition of the weaknesses of autoethnography

discussed above, I aim to complement the subjective narrative by focusing on theory-making. To make this difference clear, I discuss the more specific meaning of the term “analysis”.

I adopt the meaning of “analysis” given by analytic autoethnographer Anderson (2006). Anderson rejected looser definitions that could “cover everything from social science articles to diaries and grocery lists” and then gave his definition:

I use the term analytic to point to a broad set of data-transcending practices that are directed toward theoretical development, refinement, and extension. (L. Anderson 2006, 387)

Anderson claimed this definition is consistent with “Lofland (1970, 1995), as well as with Snow, Morrill, and Anderson (2003)” (L. Anderson 2006, 387).

Duncan’s reflective journal entries are an example of how this definition can guide analytic writing to be strongly personal and practice-based yet still amenable to theoretical discussion. Duncan writes:

[I] realised that I had taken a fork in the road of design and developed a particular attitude towards the task of writing for multimedia. An interest in pointing to the real world of the user began to emerge as a significant theme. (Duncan 2008, 32)

In this quote, I can easily imagine Duncan’s next sentence might reference existing theory around user/player perception, for example. It is also useful to note that this form of analysis can benefit from alternate modes, not just text. In addition to journal entries, Duncan collected computer screen images, storyboard sketches, loose notes, and diagrams during her practice, and discusses how these inform her findings.

How does this definition of “analysis” affect this research? While my own writing in this dissertation is more conventionally academic in style and approach, I follow Anderson’s and Duncan’s precedent to guide my writing in this research, but I also incorporate a designerly approach to analysis, which involves a reliance on my own observations. For example, in the second section of chapter 6, I discuss *GameStar Mechanic* using my own play experience, in-person observation of my son’s play experience, and watching online videos of TheSnivy10, an anonymous player. By using these multiple sources of first-hand observation, I adopt Duncan’s aims to portray the “real world of the user,” not just my personal opinion.

It is important to consider the limitations of textual analysis. It is clearly impossible to fully convey the experience of any video game through text, yet in this method I state I am using textual analysis to document the key elements of an existing learning game. A sceptical reader might question this

method, and I would admit that it could be better; however, I do not see how I can improve on it. While I implore the reader to play each game I discuss, and use the AGA Scoring Tool themselves, rather than just reading my version, I have no ability to ensure this happens. This thesis must stand alone within the limits of its textual mode, even as it signals the greater understanding that might emerge from a more active participation from the reader.

Another important limitation of the textual analysis in this thesis can be seen when I analyse several learning games. I generally use a “side-by-side” comparison: I discuss a game’s activity-goal alignment using textual analysis and then again using the AGA Scoring Tool. It is obviously impossible to fully suppress my knowledge gained from the first approach when I do the second, even though I tried to clear my mind and consider the game anew when I employed the second approach. I consider this an inherent drawback of the method. I aim to mitigate this drawback by reversing the order of approaches among games, but acknowledge that this mitigation is insufficient to completely dismiss the drawback.

3.4.2 Discussion of Design Practice

In the second half of chapter 4, I document the design and development of the AGA Scoring Tool. That section’s method is different from the rest of this dissertation, so I discuss it here. This section does not simply discuss the final version of tool itself. Instead, it documents the development of the tool, revealing my design process. Why do I include this section?

As with any doctoral dissertation, this research must be explicit and thorough in its presentation of evidence of competency (Curtin University, Office of Research and Graduate Studies, Faculty of Humanities 2012). Previously in this chapter I noted that good design practice combines intuitive and rational approaches, and that good design practice is iterative (Jones 1992). In order to demonstrate that this research is a form of good design practice, and given the importance of process to design practice, I believe it is necessary to present a detailed account of the design process. In section 3 of this chapter I argued that autoethnography is an appropriate method for doing so. Additionally, the documentation of the design process may be helpful for readers who might wish to modify or extend this tool, as I discuss in that chapter

3.5 Ethical Considerations

In this section I make two broad claims of ethical compliance, and then discuss the more specific ethical implications of my research.

Broadly, I claim that this research complies with the ethical standards for academic researchers set out by the National Health and Medical Research Council of the Australian Government (2007). In particular, I claim to comply with the requirements listed in the “Responsibilities of Researchers” section. For example, by publishing the AGA Scoring Tool on a website, this research aims to comply with its ethical obligation of dissemination, including “Support communication of findings to the wider public” (section 4.1) and “Disseminate all research findings” (section 4.4).

The remainder of this section discusses the more specific issues regarding ethical concerns in this research. I have two purposes for the discussion: to further defend my broad claims above, and to set out my self-imposed ethical guidelines and requirements.

3.5.1 Additional Considerations Regarding Ethics of This Research

It is appropriate for me to present my experience with ethical compliance in research, in order to defend my broad claims above. I have some experience with ethical compliance, as I was granted ethical approval for in-classroom research with minors in 2011 by Murdoch University’s Research Ethics Office, and was also approved by Western Australia’s Department of Education and Training. In writing this successful application, I was mentored by veteran researcher Doctor Jenny Robinson, who has been successful in over forty ethics applications at multiple universities. Through my employment at Murdoch University I have also enjoyed the informal mentorship of Doctor Ingrid Richardson and Doctor Mark Cypher, both of whom have extensive experience in ethical research. These professional interactions have helped me develop my own ethical compass in academic research.

Ethically, autoethnographic research is obviously different from other ethnographic methods, since there are no participants other than the author. The author is obviously fully informed of the benefits and damage they risk through participation, which eliminates the majority of ethical risks typical in ethnographic methods.

In the more particular consideration of the boundaries of autoethnographic ethics, Bruni (2002) notes that research activity should not physically, socially, or psychologically harm any living being, and that the ethical autoethnographer must not engage in any research practices that harm themselves or others involved in the study. Bruni notes that these concerns “have no definitive answers. They do not suggest a development of a checklist of acceptable practices; rather, they open up the murky swamp of autoethnography to critical scrutiny [...]”. This section addresses these ethical questions in relation to the method.

3.5.2 Harm from Disclosure of Information about Participants

One main ethical consideration for autoethnographers relates to disclosure of information. Autoethnographers must acknowledge that the reflective, self-questioning, self-interrogating, and deconstructive nature of autoethnography requires that researchers carefully consider potential harm to themselves and those involved in the activities being studied. Certain fields carry higher risks than others. In health science, autoethnographic research could identify criminal behaviours such as drug abuse or sexual molestation, risking obvious and major harm to self and others (Bruni 2002). In social studies, the stakes can be lower: Ellis reports having angered participants, harming personal relationships (Ellis 2007). Ellis's position was unenviable: the value of the research was in direct conflict with the harm. A revealing of truthful account was not just necessary for the research, but essential for her self-growth, Ellis states (p.16).

Ellis and Bochner (2000) suggests autoethnographers should be upheld to “[...] a demanding standard of ethical self-consciousness; showing a concern for how other people are portrayed, for the kind of person one becomes in telling the story, to provide a space for the listener’s becoming, and for the moral commitments and convictions that underlie the story.” In acknowledgement of this demanding standard, I have both changed the names and received unqualified permission from my colleagues to discuss their positions, thoughts, attitudes, and actions in the autoethnographic narrative in chapter 4.

3.5.2.1 Workplace Autoethnographies: Liu’s Example

Liu (2009) offers a useful example for ethical autoethnography in the workplace, in her thesis documenting her managing role in a Toyota factory. Liu noted the difficulty of achieving ethical ideals:

[...] these obligations can prove difficult to uphold in some autoethnographic research settings. Obtaining informed consent may sometimes be an unreasonable goal, even irresponsible, or put the researcher in harm’s way, or prevent a project from being undertaken that would encourage healing. (L. Liu 2009, 52)

Liu claimed sensitivity: “My training and education has heightened my awareness of how my own actions impact on those of others and how, in turn, I am personally implicated by these actions of others” (p.59) and acknowledged risk:

Within workplace-based autoethnographies, because my identity is already disclosed in the text, it may be argued that the identity of others could be gauged through their association

with me in the organisation. Complete confidentiality is therefore difficult if not impossible. With this reality in mind, Chang (2008: 69) states that if all the usual devices for ethical protection have been exhausted then “you simply have to use the real identities of others with their consent.” (L. Liu 2009, 59)

Liu cited Chang in setting her ethical target and claims to meet Chang’s criterion that ‘although perfect protection of privacy is not always possible, you should model an honest and conscious effort to adhere to the ethical code of research’” (Chang 2008; cited in L. Liu 2009, 61). Chang’s position is a good model for protection of privacy in this research.

Liu states her ethical position regarding disclosure:

In this thesis, however, I do not name others and have taken steps to protect the identity of all individual parties. Sometimes this has implied that I have had to make small changes to relevant details and/or context to provide this protection. This has not affected the integrity of the analysis. (p.61)

3.5.2.2 Full Disclosure Not Appropriate

HAGAno (1983) argued that it is not reasonable to require that disclosure and consent in all situations in which autoethnographic data is collected:

Scientists may routinely make observations about people without informing them. No disguises are involved; it is just that social scientists constantly observe others’ behaviour. These everyday observations of our family, friends, and self frequently become parts of the theories or data of the researcher. (p.157)

Ellis asks rhetorically: “do I say before any interaction ‘hi, nice to see you, now remember I’m a researcher’ ”(Ellis and Bochner 2000). I agree that this is a challenge faced by any researcher, and address it in this research by obtaining consent: I did not work with anyone in this research who did not agree to allow me to represent and quote them.

3.5.3 Self-Interest

Autoethnography also carries an ethical risk of self-interest. While this risk is not unique to autoethnography, Tedlock gives several vivid examples of autoethnographic researchers confessing self-interested motives in their choice of location and topic (Tedlock 2005, 469). This might be seen as a mere limitation akin to convenience sampling, but also carries ethical risk that goes beyond mere quality of data: did the researchers noted by Tedlock enjoy a material advantage from their

research that influenced their research findings and decisions? For this research, this risk is mitigated by the shared responsibility inherent in Ph.D. study: unlike solo autoethnographers, my advisors share responsibility for my research (Curtin University Council 2009) and are obliged to review my work and point out ethical issues that arise. I claim that this oversight is sufficient to mitigate this risk.

3.5.4 The Ethical Risks in This Method, and Their Mitigation

A review of the method in light of the ethical discussion above suggests that the main ethical risk is of unintentional identification of participants or disclosure of personal or sensitive information. This autoethnography happens in a social or team setting, and the data contains information about on others, not just the author. Following Liu's example, this thesis acknowledges that some risk exists. Design decisions and opinions of colleagues are visible in the data, and these colleagues could conceivably be identified due to the rich context, even though I use pseudonyms in the data. It is conceivable that these positions might be embarrassing at some future date. To mitigate this risk, colleagues have reviewed the published work given written consent and permission to refer to these data, and nonessential facts have been changed to help disguise identities. While it is acknowledged that some non-zero risk of harm remains, this thesis asserts that this risk is low enough to be acceptable within the ethical boundaries that govern this research.

3.6 Contribution to Literature

Here I argue that this research could contribute to the analytic autoethnography methodology literature in these ways:

1. Analytic autoethnography as an explicit method in design research is rare, yet I have found it a good fit. This finding may be of interest to other design researchers, as well as useful as a comparative reference for other autoethnographers.
2. I address unusually specific research questions, compared to most autoethnographic work, finding analytic autoethnography is an effective method. This finding supports Anderson's argument that analytic autoethnography is compatible with a more rigorous approach to research than was previously supposed (L. Anderson 2006).

Is the mere choice of analytic autoethnography as a method really notable enough to be considered a contribution to the literature? I argue yes here: I have found others discussion of analytic autoethnography that is grounded in related fields very useful, and suggest that others would find

this research similarly useful. For example, analytic autoethnography was discussed in design studies (Frankel 2009), and interactive “hypermedia” design (Duncan 2008). My field of learning game design is different from these fields, and these three examples are quite different from most practitioners of analytic autoethnography (L. Anderson 2006). Autoethnography was developed from social science, and has spread to other disciplines (Chang 2008). Today it is a common method in creative writing, psychology, and fine arts. However, based on my searching of the literature, I believe this research is one of a few examples of analytic autoethnography being used in anything similar to software design.

For any research to contribute as I argue above, it must discuss the reasoning behind its method in some detail. Indeed, there have been numerous calls for a more explicit defence of methods in this type of research. Other autoethnographers who state that autoethnographic research should explicitly defend and declare methods (Duncan 2008a). In game studies, Consalvo and Dutton (2006) agree: “[...] the more qualitative studies have been less forthcoming about how games were studied, other than the assumption that they were played and carefully thought about by the author.” Accordingly, in this chapter I have presented my method in sufficient detail to both satisfy these calls and present my reasoning to others who might consider using analytic autoethnography. By contrast, I encountered similar research that aimed to assess learning games (Rice 2007; Buckleitner 2012; Gee 2004). The discussions suggest their researchers may be using analytic autoethnography, but the authors did not describe their methodology as autoethnography, so their research was not useful to my consideration of method.

When compared to other autoethnographic research, my aims are unusually specific. I have not observed other autoethnographers using specific aims in the literature; however, the literature is vast and it is quite possible that I have failed to uncover important texts. My specificity of aims informs my methodological choices, which one might observe are somewhat different from other autoethnographers. Analytic autoethnographers usually systematically investigate something subtle to develop theory, as I do. However, this research has a more applied approach: I investigate if scoring activity-goal alignment is useful. By comparison, Anderson, a pioneer of the method, used analytic autoethnography to find a “better understanding [of] serious leisure pursuits more generally, as well as for answering the personal question, should I continue to jump out of planes?” (L. Anderson 2011). While many others have also used analytic autoethnography in a variety of research questions, most are broader than this research. Accordingly, those who seek to use analytic autoethnography to address research questions similar to mine might find this research useful.

4 Scoring Activity-Goal Alignment

4.1 Introduction

In the introduction to this dissertation, I argued the need for an AGA Scoring Tool. In this chapter I further explain this need, beginning with an autoethnographic account of the development of *Battlefood*, a learning game whose design I led. This narrative relates my exploration of activity-goal alignment in my own work. From this, I develop my reasoning for the features of the AGA Scoring Tool, documenting how I used an iterative and self-reflective designerly approach to address known knowledge and needs. Finally, I present the tool itself and set out a recommended approach and procedure for using the tool.

4.2 The Story of *Battlefood's* Development

The following narrative describes the development of a learning game called *Battlefood*. *Battlefood* is a video game intended to improve players' knowledge of basic nutrition. Important to this narrative is the fact that I am a veteran professional game developer, as I outlined in chapter 3.

For years I had imagined that a food fight game mechanic, using unhealthy foods as dangerous weapons, could teach nutrition. I got my chance to explore this design idea in 2011, when I was lead researcher for an exploratory pilot project for the Young and Well CRC, a \$37 million cooperative research venture between a consortium of Australian universities and the Australian Federal government.

4.2.1 First Concept of Food Weapons System

In an unpublished grant application I described my vision for the core mechanic of *Battlefood*:

Using a multiplayer shooting game mechanic, our game is based on a "food weapon system:" a mathematical model of human energy consumption and expenditure, with simple nutrition model. Using minor variations on proven mechanics, the game offers a progression from fast action to tactical and, later, strategic, excellence via mastery of weapons system.

The learning goal was aimed at young players to "understand the math behind the nutrition," and the game was designed around incentivising players to seek this learning themselves. The game was self-assessing, a feature of good learning games (Gee 2009a): players could not win without understanding the game's model of nutrition.

I hoped that the masterful player would have something like twenty hours of intense interaction with the Food Weapon System, and imagined this interaction as an experiential, self-motivated type of learning that fit the positivist theory perfectly. I imagined that winners of *Battlefood* would walk away with a conceptual understanding of human metabolism of a depth and subtlety that no other form of study could match.

There were others involved in the designing and building of *Battlefood*, but the core team members were a programmer whom I identify with the pseudonym “Sam”, a student artist “Andy”, and a nutrition expert, “James.” Andy, Sam, and I met weekly, while James’ availability was limited to approximately one meeting a month.

4.2.2 Significance of the Food Weapons System

At the start of the project, I regarded the “food weapons system” as the most important part of Battlefield’s design. This design sprang from an observation common among game designers: Players typically exploit models in games (Schell 2008, p180). The more specific case of weapons system mastery is an established, if tacit, heuristic in gamer culture: to win a first-person shooting game, study the weapons (Bowler 2008). Accordingly, I assumed players would study *Battlefood*’s weapon system in great detail, if the game rewarded such mastery to progress.

It is well recognised that weapon systems are extremely important in video games that revolve around shooting as their core mechanic. For example, consider the intensely competitive multi-player team-based video game *Team Fortress 2* (Valve 2007). Each weapon is very different and understanding these differences is crucial to winning, so players carefully study attributes such as range, damage and accuracy and related them to in-game effects. Bowler (2008), a *Team Fortress 2* player and early-career professional game designer, illustrated a typical player’s mental model of *Team Fortress 2*’s weapon system balance (see Figure 8).

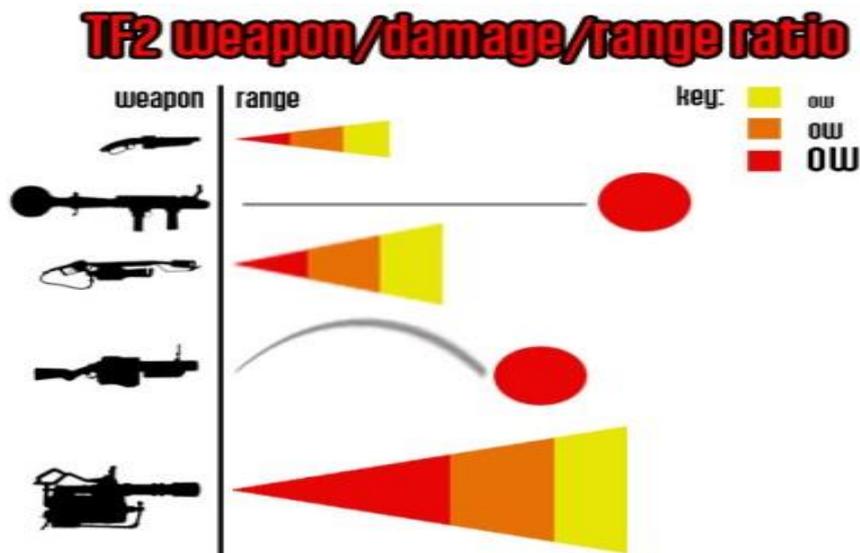


Figure 8: Player's conception of the difference between weapons in the game "Team Fortress 2." (Bowler 2008)

Bowler's chart only shows attributes that directly affect gameplay (range and damage), even though the weapons have other obvious differences such as sound, appearance, and visual effects. For example, the flamethrower is compared directly to the rifle. To a nonplayer, these two weapons appear to be so dissimilar as to be unrelated, but in the player's mind, they are reduced to two variables: range and damage. This suggests that in the player's mental model of the game, the weapon's "winning affordance" is worthy of careful, methodical study.

The wealth of veteran player knowledge about a game's weapons system is substantial. Commenting on the essay accompanying Bowler's chart, *Team Fortress 2* player Chuck compares his gaming experience to academic progression: "Having been first schooled on years of Doom & Doom2 deathmatch, followed by graduate studies in Q [*Quake*] & Q2 [*Quake 2*], with my post-doc work in CS [*Counterstrike*] [...]" While this may be a playful comparison, it illustrates how players build knowledge about games' weapons systems over years. In my own *Team Fortress 2* play experience, I reverse-engineered the game's weapon system by informed experiments while playing the game. I shoot a wall to observe blast radius. I aim near targets and observe if there is any "auto-aim" feature correcting my aiming mistakes. This approach is not subverting the designer's intention, as the training modules in *Team Fortress 2* focus beginning player's attention on these differences as well.

Returning to discussing the Food Weapon System in *Battlefood*, I was designing the game's weapon system to embody the learning goal: understand how nutrition relates to our health. I aimed to bring the typical deep understanding of fantasy weapons to a real-world question. In Shelton's terms, I

was aiming for extremely high alignment between learning goal and activity, although I had not encountered his theory at the time.

In this design, I was following certain theoretical advice that I had encountered. For example, Sid Meier said games can be seen as a series of interesting decisions (cited in A. L. Alexander et al. 2005). To make decisions interesting is a key challenge for game designers. I approached *Battlefood* using a design style that Cook called “system design:” the game design starts with a complex system, often a simulation (Cook 2009b). The complexity defines the possibility space of player decision. As discussed in chapter 2, Meier and Cook’s ideas should be combined: the system’s complexity must be revealed as interesting decisions. In this case, I decided weapons choice was the interesting decision.

In any game, designers can assume players want to win (Koster 2005). To win in a multiplayer combat game, players must somehow distinguish themselves from other players. Players assume that any decent multiplayer game provides some skill-based method of distinction. Normally, this distinction is a player’s skill, especially accuracy of aiming. I steered players away from focusing on accuracy skills by adding “auto-aim,” a feature that helped the player hit opponents even if their aim was off. This reduced the challenge of the “shooter” mechanic, and required them to seek distinction using other elements of the game. My design then focused player attention on choosing the correct weapon. For beginning players, whoever could distinguish between healthy and damaging food would score highest. *Battlefood*’s design de-emphasized other player skills to keep players focused on the learning goal. For example, popular first-person shooters typically require fine motor skill and quick reflexes to win (Bowler 2008). In addition to auto-targeting, the speed of the projectile was designed to make reflexes less important, when compared to other similar shooting games. To win, players would have to find another skill to differentiate them.

I also designed for advanced players, who exploit the food weapons system to their advantage. I intended this player to focus more and more on the learning goal. Advanced players should aim to understand the food weapons system in ever-greater detail. This system was imagined to be sufficiently complex as to provide many avenues to mastery, thanks to the simulation’s mathematical model of nutrition. This in-depth understanding of the mathematical model was the learning goal, but I intentionally did not design a particular path for advanced players; instead, I provided a “sandbox,” the simulation, that would allow clever players to engineer new ways to combine or sequences foods to achieve specific purposes. I also hoped for an “arms race” between defensive and offensive strategies: players could eat certain healthy foods to protect against certain lethal combinations of unhealthy foods. In this vision, the simulation must be factually accurate,

complex, and impartial. I imagined a simulation that used more than two or three type of information from each food weapon, and generated many subtle variations to explore. The mathematical formula at the heart of the simulation would keep the game design tied to the real world knowledge. For example, I imagined the final game would tutor and assist the player in progressing, by rewarding the players with hints and gifts “You won a chicken/chips combo-pak! Their saturated fat is 80g per serving–super-lethal!”

4.2.3 The Fatal, Hidden Assumption: There Is No Equation

While I did a good job of considering the other assumptions underlying the Food Weapons System design, I failed to investigate a crucial assumption. I had assumed there existed a widely accepted mathematical model of human metabolism, detailed enough to predict weight gain/loss from any reasonable combination of calories, protein, fats, and carbohydrates. I assumed the nutrition information on the back of a box of cornflakes could all be plugged into an equation, along with body metrics, and I could know if I would gain or lose weight, healthiness, and other clear, definite answers.

The idea was players would exploit the subtle effects of this equation. For example, any player can roughly assess a burger’s destructive power as similar to fried chicken: both contain a lot of protein, and quite a lot of unhealthy fat. However, a player can discover that a fried chicken is more damaging when used in combination with chips, because both the type of fat used in both the chicken and chips is the same, so the body gets overloaded on that type.

As the project progressed, I contacted several experts in nutrition, and eventually recruited James, who had both research and clinical intervention expertise on obesity, exercise, and nutrition. He agreed to participate and in our first meeting, I explained the vision and presented him with a spreadsheet that listed seven key attributes of any food (calories, protein, carbohydrates, fat) and any person (age, height, activities). I asked him to create a formula that would predict the weight gain/loss of the person. He was unable to do so easily, to my surprise. James and I worked together for several months, researching and developing the food weapons system, but we could not come up with a formula.

There were two important problems that prevented us from reaching agreement right away: My ignorance of nutrition, and my inability to express the concept of “activity-goal alignment”. James struggled to reach an understanding of the need for a nuanced, mathematical “food weapon system” model, and often suggested simplified models, or models that incorporated training or other unrelated elements. Although I tried several times, both in written form and in meetings, to

convey the “activity-goal alignment” idea, and although James seemed to understand and express the idea back to me in meetings, it was apparent from his suggestions that he didn’t really understand the implications. For example, in an early meeting with Tim, I explained again that we needed a mathematical formula that predicts weight gain/loss from any combination of a few properties of any food. After the meeting, James wrote:

After I left the office I thought a little more about some of the options (sitting in traffic) that we could explore.

I think we may be able to build something into the injury model e.g.

Lack of Vit D and calcium increases risk of bone related injury

Iron and B12 leads to energy depletion

Etc.

So I will have a think about that as I think we could do something with the healthfulness from that perspective. But give me some time and I will see what I can bring up.

This email reveals that he was casting the net wider, trying to extend the food weapons system to include injury and nutrients. He was also proposing that the simulation be simpler cause/effect, not a comprehensive equation that would predict weight gain/loss.

As James patiently disabused me of my “common sense” notions of metabolism and nutrition, I was dismayed to discover how little uncontroversial knowledge exists about human metabolism. James explained that while there were published models, experimental evidence was not strong, and one person’s claims might conflict with other equally reputable claims. Many models did not predict an individuals’ weight gain/loss in a way the food weapons system could use, as they dealt with broad population segments such as age ranges, not individuals.

4.2.4 Major Redesign 1: Add a “Training Mode”

About one-third of the way through development, I became convinced that my initial assumption was wrong: There is no widely accepted mathematical model of human energy consumption in the literature. I believed no one really knows what will happen to a male 18-year-old if he eats 80 grams of saturated fat, versus 60, per week. There are many variables (James said genetics play a huge role, for example) and many of these variables’ roles and function are hotly debated. I perceived this was a big problem for the game design. As designer, I had to make a choice: Choose a mathematical model that was unproven, or change the design so as not to require a mathematical

model. I did not wish the game to be criticized for teaching wrong information, so I changed the design.

James was particularly keen on incorporating exercise and training, partly because the modelling is less controversial. His influence was apparent in the first major revision of the design. After the conversation where I accepted James' position, I wrote an email to the team: "The big idea is: fight mode is more about fun, and training mode delivers more of the learning. So, we can add fun stuff to fight mode that isn't strictly about learning." By adding training, I was seeking an alternative mechanic to that could have high activity-goal alignment, without throwing out the whole game and starting again.

As a result of other demands on my time as a university lecturer, there was need for more rapid decision making than I could provide. I empowered Sam, who was funded up to full-time, to make most design decisions himself, rather than delay the project. His design influence reflected his personal passion for "hard-core" shooter games such as Counterstrike. We built a video demo that showed a **playable** prototype with a very simple caloric in/out model, no food weapons system, and functional core multiplayer shooter mechanic. There was little progress on the training component. Sam, the programmer, and I explored why he was resisting this element in meetings.

From these meetings, I realised that this "fight/train" split had two important weaknesses. First, it weakened the clarity and power of the original vision, as the core mechanic was now split. Second, it divided the learning from the play: The "fun" part was mostly in the multiplayer fighting, while the "learning part" was mostly in the training. As I mentioned in the introduction, this felt like a move toward "chocolate-covered broccoli", a term describing a schism between learning and play known to be problematic in past learning games (Klopfer, Osterweil, and Salen 2009). While the fight/train system was not nearly as split as games like *Math Blaster*, it still felt like the wrong design direction. Perhaps if James had access to activity-goal alignment theory he could see that "training mode" lowered the game's activity-goal alignment and we could have addressed this weakness earlier. I discuss this later in this chapter.

A third reason we did not pursue training mode was related to the project budget. The project that provided funding to build *Battlefood* was studying cocreation with years 5-7 students at a local primary school. They had co-designed an unrelated game which we wanted to build for them, so we reduced the budget for *Battlefood* to build it. This reduction increased our concern that it would not be possible to finish the game with the training features required.

4.2.5 Food Pyramid Redesign

For these reasons, we began looking for a new vision that was simpler to build, and retained the “core” play mechanic and high activity-goal alignment of the original vision. To find this, we began from our improved understanding of the subject matter. By this time, thanks to ongoing conversations with James, I had a much clearer idea of what is well understood and important for our players to learn: The accepted knowledge (“eat a lot of veggies”) is perhaps best quantified in a simple diagram like the food pyramid. These diagrams, James explained, essentially quantify “rules of thumb” about portion size and proportions of “greens” vs. “sweets.” It seemed to me that even these simple-sounding claims were contested by various authorities, but James felt that the science behind most of the food pyramid was sufficiently uncontroversial to endorse.

When we were half-way through development and playtesting weekly, we decided to return to the original vision but replace the Food Weapons System with a “food pyramid” mechanic. In this version, if player collected the quantities of foods in the categories recommended by an Australian government-approved set of healthy eating guidelines (fats, sweets, grains, dairy, etc.), that player would “power up” and become invulnerable to attack for a limited period of time.

I believed this change would reduce the learning potential of the game. Without a food weapons system, the possibility space for the player to explore was greatly reduced. The “killer weapons” were easy to identify after an hour or less of gameplay by experienced gamers. Two months later James reviewed the design and noted that the Food Pyramid failed to account for age and gender, and suggested we instead use the “Recommended Daily Servings Chart” (Department of Health and Ageing, Australian Government 2012). I developed another user interface based on this chart which James approved in mid-May. This redesign showed the categories of food, and recommended portions of each category. Sam didn’t like the aesthetics of this redesign, and was concerned it would cover too much of the 3D view, so redesigned it. This redesign removed the representation of portions. We revised this to include names of categories, to slightly improve the learning potential of the game.

4.2.6 Final Design

In the final stages of the project we had playtested the game enough times with a broad enough audience to discover the game was fun for a wider variety of ages than anticipated. We considered revisions for future versions of the game that would make both the activity and the learning goal suitable to younger audiences, but decided against making these revisions because the game was

“feature complete,” which means that further development efforts are meant to focus on finishing the features already in the game, not exploring major design decisions or adding new features.

4.3 Modes of Communication around Activity-Goal Alignment

In the narrative above, I have aimed to explain the context that created the general need for activity-goal alignment theory in my design practice. Next, I explain how the theory illuminated my pre-existing notions. During the development of *Battlefood*, I took notes and captured our attempts to communicate about activity-goal alignment prior to my discovery of Shelton’s activity-goal alignment. In my attempts to explain, I experimented with alternate modes including verbal conversation, whiteboard sketches, and emails. Below I present and discuss selected attempts.

4.3.1 Model Complexity as Poor Substitute for Activity-Goal Alignment

In this section I describe how I developed original ideas that were not as effective as activity-goal alignment in explaining activity-goal alignment. During the *Battlefood* project, James and I discussed how the simulation model relates to the learning goals. The discussion began when James suggested that the game could “add on” to an oversimplified model of metabolism: if the learning scope were expanded to include exercise, diabetes, or injury recovery, then useful learning goals might be met, even without the good simulation. I resisted this suggestion because these extensions could not be integrated into a single simulation; instead, multiple small simulations would be needed. This is technically possible, but the interactions between these small simulations would be difficult to build, and could reduce activity-goal alignment.

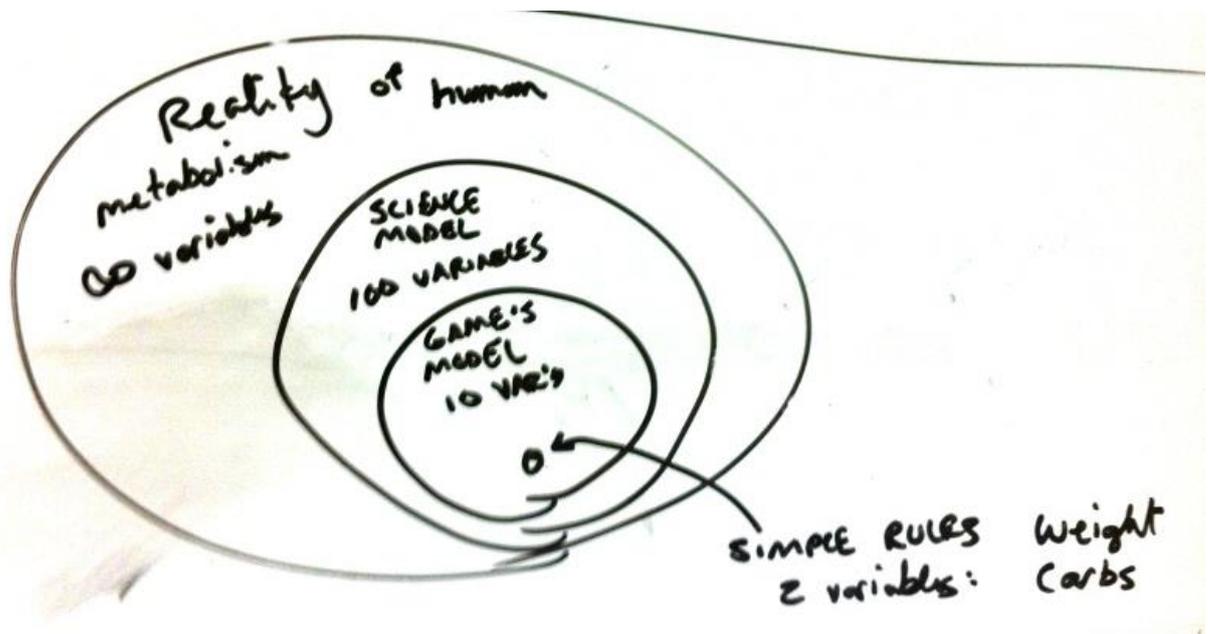


Figure 9: My whiteboard sketch of the relative complexity of four nutritional models: the original proposed in-game model, science model, all within "reality:" true human metabolism. Note that I used number of variables as an approximate measure of complexity.

In that context, I created the drawing in Figure 9, aiming to compare the various types of models we might consider. The simplest model, the innermost tiny circle, used only 2 variables - weight and calories (incorrectly labelled "carbs" in the drawing). The next circle, labelled "Game's model," showed my imagined ideal level of complexity for a game, in which ten variables are supposed. The "Science model" circle with 100 variables was supposed to be too subtle and complex for the game's purposes. Finally, reality, in all its infinite complexity was shown in the largest circle to show that the scientific model, while complex, is still not perfect.

The drawing implicitly suggested that the game should have one comprehensive simulation of human metabolism, rather than a series of connected components such as stomach capacity limiting system, energy level, health level. This approach had an important weakness: it suggested that simulation accuracy, rather than effectiveness of the learning game, was important. Simple models are just as useful as complex models (Chwif, Barretto, and Paul 2000), and while I and my team knew that theoretically, but this drawing confused that knowledge. Accuracy had no place in the discussion, but there was a need to create a linear relationship between learning game effectiveness and simulation. "Accuracy" was the closest idea I could think of on the fly.

In review, I observe that the lack of terminology may have misdirected James toward accuracy when I wanted to talk about activity-goal alignment. A term like endogenous, as discussed in chapter 2, was too theoretical and broad; it would have been too easy to simply say "yes the game is

endogenous because the core mechanic is the learning goal.” However, I can imagine that if James and I had discussed and agreed upon the meaning of a designer’s term like activity-goal alignment, especially if we could have scored it using the tool I later developed and discussed our differences, it could have made the discussion more fruitful and substantially reduced confusion between the team members.

4.3.2 Design Document

Design documents are one of the oldest and most common tools for communicating about game design ideas (Adams 2010). There are many forms of design documents, from concept, to “bibles” to small component documents. *Battlefood* used a large initial concept document, which was replaced by multiple small documents and task lists, in an informal agile development process.

In the initial concept document, the alignment between activity and goal was an important idea to convey. However, it was difficult to express. In an early draft of the grant application for the project, I wrote:

Using a multiplayer shooting game mechanic, our game is based on a “food weapon system:” a mathematical model of human energy consumption and expenditure, with simple nutrition model. Using minor variations on proven mechanics, the game offers a progression from fast action to tactical and, later, strategic, excellence via mastery of weapons system.

In retrospect, I see that my phrase “excellence via mastery” meant “high activity-goal alignment” but the terminology was unclear. Before the project was approved, in notes I wrote while developing the grant application, I discussed objectivity in the game design:

Will an objective simulation (of the human body diet) have important benefits to a game that teaches healthy eating? We hypothesize yes, because “healthy eating” is subjective, and as such [sic]. Human body simulation is objective.

In this incomplete sentence, I relate simulation to learning outcomes, by considering subjectivity. Subjectivity is an essential challenge of teaching nutrition. What is “healthy eating”? Food culture experts state that subjectivity and cultural context is an important component of any answer (Murcott 1993).

I needed a better approach than simulation to relate activity and game mechanic. My attempt to fit this game within the simulation literature is somewhat confusing because the literature of simulation has very little to say about subjectivity, as presented in the literature review in chapter 2. In most typical simulations, the goal is to be objective about dealing with real-world problems (A. L.

Alexander et al. 2005). Games, on the other hand, present a designer's view of a story world, which is highly subjective (Schell 2008). Activity-goal alignment is similar to the concept of "fidelity" in simulation, as I discuss in chapter 2, but allows for subjectivity because it is designed for games, not simulations. Shelton developed activity-goal alignment and used it in the design of a poetry-based game, which is a good example of subjectivity in game design (Shelton 2007). Rereading the quoted paragraph above, and recalling my intent when writing it, I believe I was really trying to discuss activity-goal alignment.

4.3.3 Playtests

Battlefood prototypes were playtested eight times with ages 6 to 40, men and women, in classroom and non-classroom settings. The playtests were informal observation in structure, mainly involving the developers observing and recording players' reactions. Overall the game appeared to be an engaging activity, from the earliest prototypes to the final version. The relationship between food types was believed to be clear and apparent to a variety of ages with no prompting or assistance outside of the game itself, in the final version. Specifically, I observed children aged six to fifteen correctly identifying the types and quantities of foods necessary to make a healthy diet within five minutes of playing the game for the first time. While my observations were not methodical enough to be considered valid or reliable, and this research makes no claims about the efficacy of the game overall, these preliminary playtest results suggested that the game was linking the play activity with the learning goal.

Unfortunately, the prototypes were not set up to explore the relationship between activity-goal alignment and efficacy. Playtest sessions were too short to observe mastery, and having only one version of the game, whose improvements from playtest to playtest made it near-impossible to relate observed differences in activity to design choices around activity-goal alignment.

4.3.4 Subject Matter Expert and Activity-Goal Alignment

Generally speaking, I believe there is great value in activity-goal alignment as a tool for helping non-designers communicate with designers about activity-goal alignment in the game. As noted in the autoethnography above, James and I revised the game design to include a "training mode," which would have lowered the game's activity-goal alignment. I eventually realised this and we removed the feature, but only after some time. I believe this problem could have been spotted earlier by using activity-goal alignment. Specifically, if James had understood activity-goal alignment theory, I speculate that he and I could have realised that "training mode" lowered the game's activity-goal alignment as we discussed it, and saved the team time. I can easily imagine how a Venn diagram of

activity-goal alignment, such as the one in Figure 12, would have greatly improved James and my ability to communicate about the core problems in this game.

4.3.5 Programmer and Activity-Goal Alignment

Sam's job was to codify the Food Weapons System. As such, he drove the need to be specific and mathematical. This need often led to discussions that I now believe could have benefitted from activity-goal alignment theory. For example, Sam would ask "how many calories should be consumed when the player is running?" A time-correct simulation suggested a very low rate, compared to eating. As in the real world, it would take weeks of playing to lose weight. This obviously conflicted with the "fun" element of the game design: we felt it was obvious that players would expect meaningful changes during a match time of 3 minutes. This prompted a discussion about how much to accelerate time in the game world. At the core of this conversation was the question of simulation. Overtly discussing the concept of activity-goal alignment would have been helpful in this discussion. For example, we could have discussed the question: if we accelerate time, does the player activity become less aligned with our learning goal? This question keeps the design focus on the learning goal. Without it, we focused on the fun, and allowed the learning goal to be de-prioritized. This kind of trade-off is not always wrong, but in retrospect I believe activity-goal alignment would have helped us keep a broader perspective: instead of worrying about factual accuracy, we would have focused more on the overall learning goal of the project in this particular discussion.

I also think that a visual diagram, not just a text discussion, could have helped us focus on the learning goal. I can imagine a significant increase in communication clarity if we had scored *Battlefood's* activity-goal alignment as we worked, and kept that score visible (i.e., printing out the Venn diagram, and putting it on the wall near our working spaces, updating it as we changed designs). When I recall my mental state and my attempts to share the holistic perspective with the programmer and artist, I feel it is likely that an activity-goal alignment scoring diagram would have helped us make more balanced design decisions.

4.4 Development of AGA Scoring Tool

The preceding sections narrated the development of the learning game *Battlefood*, and documented efforts to align the activity and the learning goal of the game. In this project, certain learning game design questions were raised but not answered. Answering these questions requires an improved method of communication in the design process. For example, a common design question was: how

do we best balance the teaching goal, the in-game activity afforded by our choice of simulation complexity? This question might be answered better if I had some better way to communicate about the relationship between player activity and learning goal during design decision-making. I describe how I developed a possible solution in this section.

Specifically, I explain my reasoning for the features of the AGA Scoring Tool, documenting how I used an iterative, self-reflective designerly approach to communicate known knowledge and needs.

4.4.1 Early Efforts

Before I had discovered Shelton's activity-goal alignment theory, I had already been thinking along similar lines to Shelton. For example, I had illustrated the idea of alignment from educational theory using Venn diagrams in a lunchtime presentation to peers at Murdoch University in 2010. I showed the slide in Figure 10 to discuss an example: what might someone learn from playing soccer?

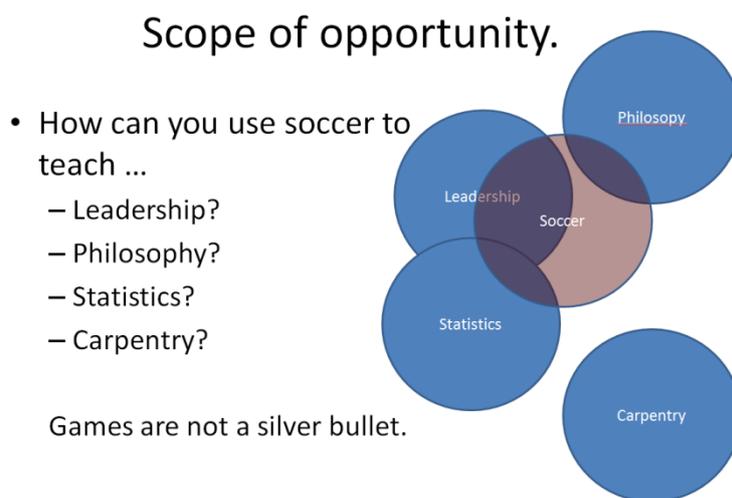


Figure 10: Venn diagram showing different skills one could learn from soccer. Note how many topics align with the core activity of soccer, while some (carpentry) do not.

In the following discussion I was pleased to note that the diagram was useful and clear to a wide audience of academics, including those who did not know the game design literature. In the following slide, I defined this overlap between activity and goal as "alignment" as shown in Figure 11.

Good learning games are **ALIGNED**.

A good curriculum designer would **UNDERSTAND** the game
 read, study, watch other players
 play it herself

IDENTIFY “core mechanics” –

what is at the heart of the game, that players do, moment to moment?
 What’s going on in their heads? What separates beginners from experts?

And then **ALIGN** to educational goals

What do the ‘core mechanics’ teach

What do I need taught?

Is there a match?

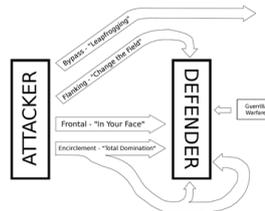


Figure 11: Alignment in Learning Games, a slide from my “brown bag” faculty presentation.

In discussing this slide, I said soccer might be a good tool to teach strategy because to soccer players are usually pursuing and exploring strategies as they play. The incomplete overlap on the Venn diagram was intended to suggest that soccer can, but does not always, teach strategy, and some parts of strategy are not a good easily learned from playing soccer. I go on to lay out a three-step process for a teacher to achieve activity-goal alignment (referring back to Figure 11):

1. Understand the game
2. Identify the “Core Mechanics”
3. Align the game with educational goals

In this process, I presume a great deal, including that there exists a game for the learning goal in question. In the third “core” slide, I give an example of a game with low activity-goal alignment (Figure 12), in that the activity (choosing from a list) has little relation to the learning goal (ecology

and hunting practice). The Venn diagram illustrates the small overlap.

Bad learning games don't align.

The gameplay (matching) is unrelated to the subject matter (habitats).

“fancy flash cards”

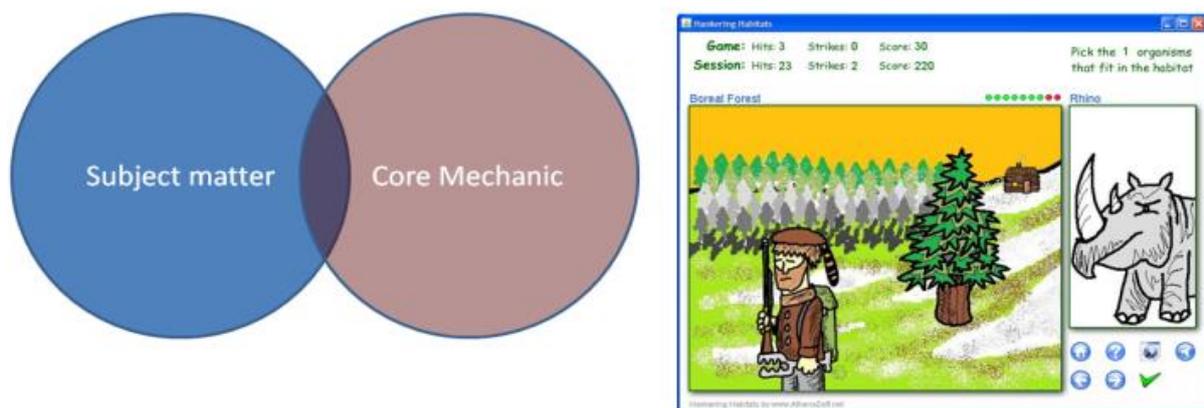


Figure 12: An example of low activity-goal alignment. This game's activity closely resembles the multiple choice activity possible with a paper worksheet.

Shelton's paper describes activity-goal alignment in a very similar way to my explanation. His definition of game play activity expands somewhat beyond the strict definition of the term I used (core mechanic), which is a useful expansion. For example, without Shelton's theory, I might have stated the game *3rd World Farmer* (Hermund et al. 2008) had core mechanics that did not meaningfully describe the real activity or learning moment in the game, but as I discussed in chapter 2, “mechanic” is too narrow a term to consider player's attitude, context, and other important factors. “Activity” is a superior term in that it allows a broader view if needed, but also allows the game designer to discuss specific actions, as was shown in my discussion of granularity of activity in chapter 2. I further discuss types of player activity later in this chapter.

After discovering Shelton's activity-goal alignment theory, I began to apply it in my game design practice, and immediately found that I wanted something concrete when I communicated about it. I worried that learning game designers would incorrectly equate activity-goal alignment with effectiveness: like fidelity, they assume that a low-scoring activity-goal alignment game will not teach effectively. To avoid this, I initially felt the need to linearize the discussion. I felt learning game designers needed to follow a chain of logic in their thinking, using questions to prompt their

reflection. To that end, I began building a prototype scoring tool named the “AGA Scoring Tool”, following iterative design and reflective practice guidelines discussed in chapter 3.

4.4.2 AGA Scoring Tool - Prototype

The AGA Scoring Tool began as a prototype: a form consisting of list of questions. It was a first-pass synthesis of the observations of data gathered. I tried this prototype form on one game and improved the wording, reviewed the key ideas and processes noted in autoethnographic section, and revised it to “AGA Scoring Form Prototype,” which simply required me to rate each item 0-10, as shown Figure 13.

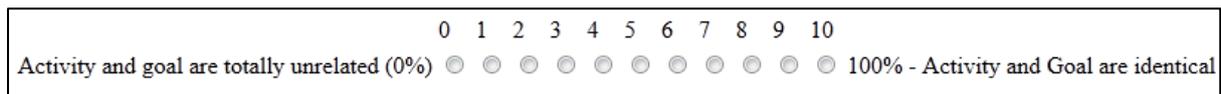


Figure 13: Numeric rating of activity-goal alignment used in the AGA scoring form prototype.

I used this first version to fill out several games, and immediately noticed several problems.

- **Single mechanic, multiple goals.** Many learning games aim simultaneously for various learning goals (spelling, typing, arithmetic, history, systems thinking). For example, *GameStar Mechanic* uses a common mechanic known as “platformer” (jumping over gaps between floating platforms in a 2D side view) to teach systems thinking, level design, and other learning goals. Some of these learning goals are closely aligned with the activity, while others are not. When I needed to be able to connect multiple learning goals to the same activity, I realised this prototype lacked that ability.
- **Not simple enough.** The prototype was too long and linear to be practical in a production setting. While it might have been useful for scaffolding a first-time designer through broad, holistic pondering of the game design, it was restrictive for the designer who understands the game overall, and will synthesize secondary concerns into scoring activity-goal alignment without prompting.
- **Less scholarly, more arty.** As I scored, I realized that an activity-goal alignment score is essentially a judgement call, not a measurement. Reducing my judgement to a numeric rating of 1-10 felt wrong, as if I were pretending to measure data, and I wondered if the nature of activity-goal alignment might be better expressed as a visual, interactive tool rather than a form. I wondered if a combination of graphics and text might be more effective.

4.5 AGA Scoring Tool V1

To amend the problems with the prototype, I rebuilt the tool to use an interactive web-based Venn diagram, shown in Figure 14.



Figure 14: The tool's Venn diagram creator, with four learning goals overlapping one core mechanic.

I chose a Venn diagram because I recalled my early presentations using Venn diagrams, as shown in Figure 12, and recalled how effective they were. I also recalled that Shelton characterized the degree of alignment using simple, clear labels: “high activity-goal alignment” or “low activity-goal alignment.” I felt that the relationship might be useful if expressed as clearly and powerfully as “high/low” but with more subtlety than a binary division allows, and felt a Venn diagram would be able to so.

I also felt a Venn diagram addressed the weaknesses of the numeric rating in the prototype discussed above, as I aimed to approximate, not measure, the degree of alignment. Venn diagrams can do either (Ruskey and Weston 1997), but I observed that Venn Diagrams are often used to visually represent relationships between abstract concepts in learning game design literature, which supported my intuition that a Venn diagram might suggest to game designers approximation, not measurement.

4.5.1 Critique

After creating the interactive diagram shown in Figure 14, I tested it by using it to critique a number of games. These trials showed that the diagram focus was a big improvement, when compared to clicking a number in a row. Moving and re-moving symbols around a visual space is an artistic act of composition, where, in a design context, precision can do more harm than good (Fish and Scrivener 1990). The imprecision of the sphere positioning encouraged me to spend time pondering and consideration, an activity that I found useful to making a good decision. I would place a sphere, examine the overlap, reflect on the play experience, and move the sphere, in seconds.

It might seem like one can re-score a row of buttons as one ponders, but I found this was not the case: clicking felt too declarative; it felt like I had given a final answer, and did not encourage continued reflection. With the movement of spheres, I was able to easily both create, and reflect on, the essential elements of activity-goal alignment for the games as I worked. However, this version of the diagram left too much unsaid. For example, there was a reason why I made “interdisciplinary thinking” overlap less than “systems thinking” but I could not express it in the diagram alone. It needed some of the textual expression that the first form provided.

I reflected on the critiques and repeated several more iterations of this nature, which are not documented here as they offer no additional insight in the design process. Finally, I combined the essential parts of the survey form and the Venn diagram tool to create a final version: “AGA Scoring Tool v7”.

4.5.2 Final Version: AGA Scoring Tool v7

The final version of the AGA Scoring Tool is publicly available at <http://www.whitkin.com/aga>. It has four key elements: context, Venn diagram, discussion, and score. These sections are identified in Figure 15 and explained below.

This is a tool for you, a learning game designer, to understand the activity-goal alignment of an existing game. (What is AGA? Read [this](#))

This is not a test or survey; instead, the form guides your thinking as you learn how AGA might be useful.

What game are you scoring?

Envision a typical play experience (e.g. classroom/home, alone or with parents/friends, hardware and controls, player persona / age motivation etc). Describe it:

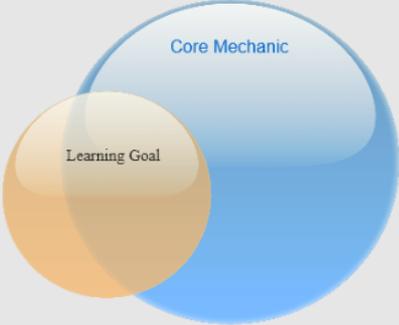
CONTEXT

In the window below, you will name the activity and goal(s), and relate them in a Venn diagram. (see example)

This diagram tool allows only one activity, so choose the most important one (the [core mechanic](#)). You may wish to do this again if your game has more than one core mechanic.

Enter an activity name and goal name, separated by a comma

VENN DIAGRAM



(to save this diagram, [capture the screen](#))

Often, activity-goal relationships are too complex for a diagram to capture. Discuss the relationships in your diagram. Note missing important factors (e.g. where, in the player experience, does learning really occur?)

DISCUSSION

You are ready to assign an activity-goal alignment score. For each learning goal, note how closely it aligns to the activity / activities. Use a percentage score or just "low/high" (see example here).

SCORE

Figure 15: The four elements of AGA Scoring Tool v7.

4.6 How to Use the AGA Scoring Tool

In this section I set out how a designer might go about using this tool. I describe options for using the tool, based on my own experience using the tool.

4.6.1 Basic Procedure

The basic procedure for using the tool is set out in the form itself. There are four main steps as indicated in Figure 15. First, the tool asks the designer to identify a particular game to score. This question intrinsically tells the player that this tool is not a general-purpose framework: it requires one specific game. For example, one might type “*GameStar Mechanic*.”

The second step further grounds the designer’s consideration, by asking them to “Envision a typical play experience (e.g. classroom/home, alone or with parents/friends, hardware and controls, player persona / age motivation etc.). Describe it:” This task combines several typical approaches to understanding play experience, as discussed in Section 3 of chapter 2. This open-ended, multi-line text box is intentionally non-prescriptive, allowing designers to reflect on what they want to.

The third step is a diagram creation tool. It has two substeps. First, the designer decides an activity to consider. This decision is not as simple as it might appear, and so bears some further consideration.

4.6.2 Activity: How to Choose

This tool requires the designer to identify a particular activity within the overall game play experience. In chapter 2, I discussed how Cook and others have identified multiple granularities in any single game activity. If there are many possible choices, ranging from a holistic activity “playing level five,” or fine-grained activities such as “pushing the left arrow button,” how should a designer choose which one to name in this form?

I decide which activity to name by noting during gameplay what seems to be the most important activity. This is usually obvious when I identify the core mechanic, but not always. For example, the exponential growth of enemy cells in the learning game *Re-Mission* is a minor component of the core mechanic, but may have been important to a powerful learning outcome. I discuss this example in the following chapter.

The tool then asks the designer to explain the meaning of their diagram in text form, stating that activity-goal relationships are often too complex for a diagram to capture. It reminds the designer to consider missing important factors: where, in the player experience, does learning really occur? This

discussion is intentionally free-form: it allows discussion in text form the most essential points. The goal here is to gain the benefits of the mode of textual discussion, in a brief, compact form. It is not intended that the designer write a full-blown academic essay here.

Finally the designer is asked to score each activity. The designer can give a qualitative or quantitative score and be as specific as they desire. For example, for *Re-Mission*, I wrote “Cancer knowledge: 25% Treatment Adherence: 25% Self-efficacy: 50%” which were simply my estimate of the overlaps. Instead, I could have written something nonspecific such as “Self-efficacy: high.” The design responds to weaknesses in the prototype, where I found that more strict approaches to scoring were worse, as discussed above.

4.7 Discussion of AGA Scoring Tool Design

. In this section I discuss some possible concerns and implications of the design of the final version of the AGA Scoring Tool shown in Figure 15.

The reader might wonder: why is there only one activity portrayed in the diagram, given that learning games can have multiple, diverse activities? While I acknowledge that a few learning games do have more than one activity, I do not allow multiple activities for these reasons:

First, I am convinced it would reduce the fundamental value of the tool, which is its clarity. I believe it would be more appropriate to use some other kind of diagram, not a Venn diagram, to portray such complex relationships. I acknowledge that other types of diagrams might be more accurate and complete picture, but imagine they would not be as simple to read. I briefly experimented with combining multiple goals and activities into a single Venn diagram, and felt that while it might be useful to veteran designers, especially for their own use, the number of overlapping circles is high enough to likely confuse communication with fellow designers, other team members, and/or other broader audiences. In developing this tool, I chose to prioritise simplicity over accuracy.

Second, and more importantly, my experience designing learning games leads me to believe that designers typically consider one primary play activity at a time, even when a game employs multiple activities. The interrelation of activities is then considered, but in a more holistic way. I do think the relation between activities is important, but I regard it as a different design issue, related to but separate from activity-goal alignment theory, and thus beyond the scope of this research.

It is my hope that designers will modify this tool to suit their needs. This particular tool is designed to meet specific needs, and although I have tried to keep the design useful across a variety of genres, it is likely to reflect my own needs and ways of thinking.

As with any tool I have used in video game production, I do not view the current version to be in a “final” state. I intend to continue modifying it throughout my design career, as I encounter new problems. Further, I do not claim it is complete. I discussed in the introduction chapter that this research does not consider certain types of games. I can easily imagine that a designer from a substantially different subculture (e.g. gambling, sports) or building a very different kind of learning game is likely to modify this tool to suit their needs. For example, a Massively Multiplayer Online game designer might find it more useful to survey, rather than dissect, their activity-goal alignment, since such games often have many different, equally important activities. Perhaps they might cluster many small activity-goal Venn diagrams in a single illustration that relates each player character class to subsets of activities and goals within the broader game. In this case, I would expect them to override my finding that multiple mechanics would damage the clarity of the tool’s Venn diagram, since their situation is obviously different.

While acknowledging that it is not perfect, I have intended the AGA Scoring Tool to balance simplicity and depth, for the specific audience of practising learning game designers. To discover whether it has succeeded, some experimental application is necessary. The following chapters document further testing and application of this tool.

5 Using Activity-Goal Alignment Scores in Analysis

5.1 Introduction

This chapter addresses the first research question of this thesis: how might practising learning game designers and researchers find it useful to score activity-goal alignment theory of a learning game? In this chapter I argue that scoring activity-goal alignment clarifies important design elements of learning games. To provide evidence, I compare using the AGA Scoring Tool to textual discussion by evaluating several learning games using both methods. First I compare *Re-Mission's* activity-goal alignment to two different games on the same topic: *Immune Attack* lacks *Re-Mission's* sense of urgency, but resembles the game in many other ways, so serves as a baseline. *Infection! The Game* concisely demonstrates how exponential growth can create a sense of urgency, which suggests that scoring activity-goal alignment is useful but insufficient to explain a user experience. In the second discussion, I investigate two approaches to teaching game design through game play: one with high activity-goal alignment and one with low activity-goal alignment. This investigation revealed that scoring activity-goal alignment is very flexible: it can compare two extremely different games that address the same learning topic.

In this chapter, I discuss the following games:

1. *Re-Mission* (Tate, Haritatos, and Cole 2009), a prize-winning learning game about cancer
2. *Immune Attack!* (Stegman 2010b) about the immune system, resembles *Re-Mission*
3. *Infection! The Game* (Anton 2009), a 2D game featuring exponential enemy generation
4. *GameStar Mechanic* (E-Line Media 2010) teaches kids how to design videogames
5. *Understanding Games* (Zecher and Straka 2007) teaches game design

I selected the games by considering four main criteria: likelihood of familiarity to the reader, usefulness in as a comparative example in relation to the other games of the same topic, expected range of activity-goal alignment scores, and range of topics. Because the play experience is important to understanding games, I selected two well-known learning games in hopes that the designer-reader may have already played the game. I pair the well-known games with lesser known games on the same topic. I also wanted to achieve a spread of low, medium, and high activity-goal alignment scores, so I selected games I hoped would cross the activity-goal alignment spectrum. For example, from my prior play experience I recalled that *Understanding Games* has activity that clearly did not align with the learning goal. The AGA Scoring Tool should yield a low score in this case, so I

chose this game. Finally, I chose games with dissimilar topics (immune system and game design), to ensure a variety of types of learning goals in the games.

Aside from cited sources, I gathered data for this analysis from the following sources: the games' consumer-facing websites (e.g. E-Line Media 2012), my own play experience, and my casual observations of play experiences of those around me¹⁴.

I intentionally use two styles of writing in this chapter, so that I can exercise scoring activity-goal alignment in a variety of contexts. I start with a formal, comparative writing style: first I describe the game factually, then I gather my notes from my own playing experience and secondary data (e.g. observe others playing; read discussions) and describe these in a written section titled "My play experience". Then I use the AGA Scoring Tool to generate a brief critique of the game, expressed in a scoring section. In discussion section, I compare the AGA score outputs to the written analyses for each game, finding interesting advantages and drawbacks when scoring AGA. I then switch to a more flowing, less structured writing style for the second discussion, to show how AGA can be useful in the critical discourse common in game studies literature.

5.2 Human Cellular Biology

In this section, I discuss the design of several games that aim to teach about human cellular biology, especially regarding the immune system. For the first game, I divide my discussion into three sections:

1. Introduction, where I set out the basic features and learning goals claimed by the developers of the game
2. My Play Experience, where I present a textual analysis of the activity-goal alignment in the game, comparing my own play experience to others' experience, in light of the discussion in the introduction
3. Scoring activity-goal alignment, where I use the AGA Scoring Tool to study the game's activity-goal alignment.

I set out my thought process in some detail, including documentation of multiple iterations of scoring, so that the reader can appreciate how the AGA Scoring Tool interacts with, aids and

¹⁴ The sceptical reader is encouraged to form a basis for criticism of my findings by scoring the games in this chapter for themselves. All of the games discussed can be accessed via the links in the references. The tool itself and instructions are available at <http://whitkin.com/aga>. The scoring process might take five to fifteen minutes, excluding any time required to play the learning game in question.

constrains analytical thinking. I provide this detail only for this first game because there is insufficient value in repeating it for the other games.

5.2.1 *Re-Mission* - Introduction

Re-Mission is a prize-winning learning game about cancer. Designed for young people diagnosed with cancer, the game lets the player control Roxxi, a humanoid “nanobot” designed to stop the cancer in the bloodstream (Figure 16). The primary activity is familiar to most gamers, typically known as a “3D shooter” mechanic: aiming at and shooting enemies, 3D navigation. The enemies are various types of cancerous cells. There is arguably a secondary activity that is more specific to the learning goal: the player is mitigating side effects and treatment-related problems such as bacterial infections, nausea, and constipation.

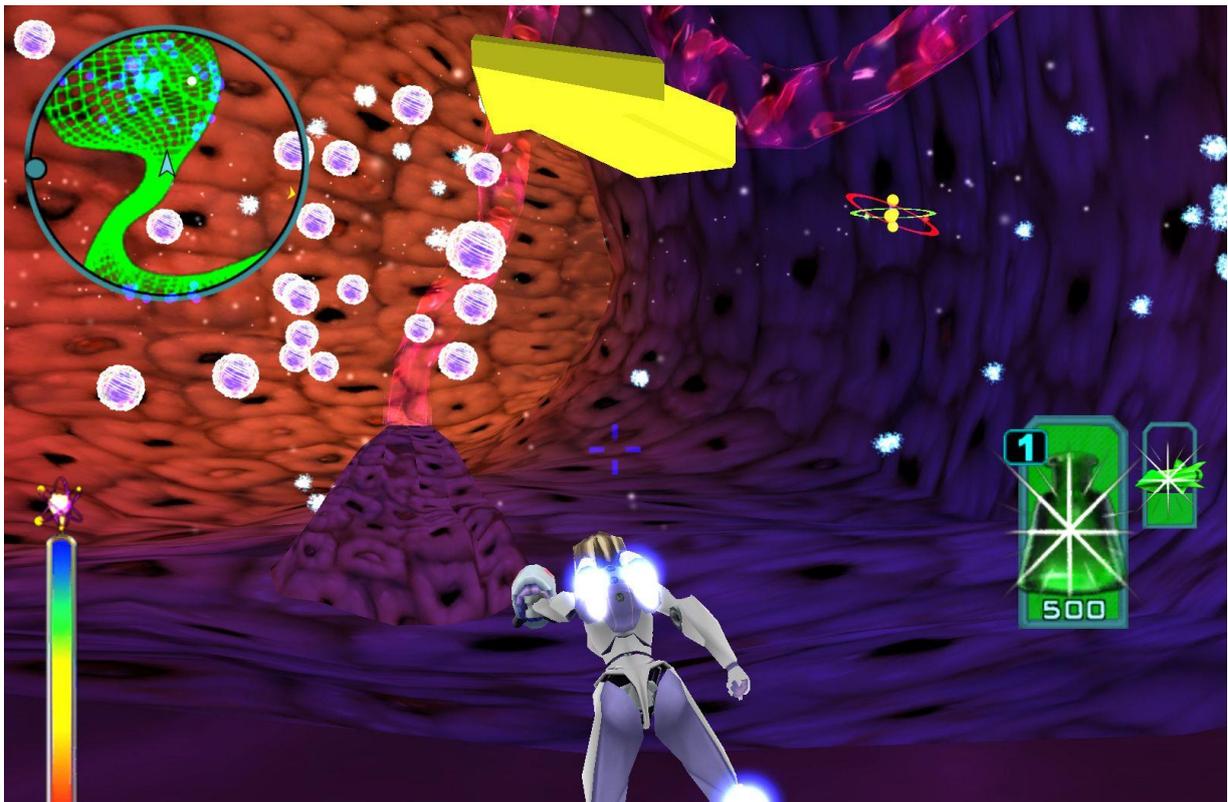


Figure 16: Hopelab's *Re-Mission* is a 3D shooting game where the player fights cancer cells inside a human body. (Tate, Haritatos, and Cole 2009)

The developers stated (Tate, Haritatos, and Cole 2009) the game aimed to improve players' knowledge of cancer, treatment adherence, and self-efficacy¹⁵. To address the learning goal of increasing players' knowledge of cancer, the game uses consumer medical terminology and concepts

¹⁵ Self-efficacy is a psychological measure of one's belief in one's own abilities to improve. Self-efficacy does not distinguish between false confidence and reasonable hope. Self-efficacy is not exactly the same as self-confidence, but shares some of its meaning (E. S. Anderson, Winett, and Wojcik 2007).

integrated into gameplay: for example, weapons shoot chemotherapy, antibiotics, antiemetics, and stool softeners. Kato et al (2008) studied young cancer patients who played *Re-Mission* and found the game was effective in improving compliance with pill-taking regimens, which is an important outcome as compliance can literally save lives. The high rigor of this study may explain why this game is often cited as exemplary in the learning game literature (Klopfer, Osterweil, and Salen 2009).

In addition to these three specific learning goals, *Re-Mission* developers elsewhere indicated the game aimed to address additional related goals:

- *Compliance with oral chemotherapy regimens and other prescribed medications*
- *Prompt reporting of symptoms and side effects*
- *Proper nutrition to increase the body's ability to fight cancer*
- *Anxiety, nausea, and pain management through breathing and muscle relaxation exercises (HopeLab 2012)*

Although pill-taking was the primary measured outcome of their published study, the designers did not set pill-taking as an explicit learning goal of the game. In other words, they did not seek to teach players how to take pills, but instead aimed to teach players how important pill-taking is.

5.2.1.1 My Play Experience

I played *Re-Mission* for several hours, until I felt I had progressed from new player to intermediate, grasping the bulk of the experience comprising the core mechanic. After installing and launching the game, and after a short non-interactive video that introduced my mission through a comedic cut-scene, I was able to experience the core mechanic.

The core mechanic was clear and familiar, as it resembled that of most 3D shooting games of its era. My play activity consisted of developing precise 3D spatial navigation and shooting skills. Success required aiming accurately with a mouse, conserving ammunition, learning the behaviour patterns of enemies, and trading off resources such as health to achieve objectives. I felt intense pressure, caused by overwhelming enemies. A pounding soundtrack and constant threats kept my self-perceived stress level consistently high. Unlike the game they used in their control cohort, *Indiana Jones*, *Re-Mission* gave me no time to reflect, and no puzzles to solve. It was a highly engaging shooting game that was neither too easy nor too difficult.

I noted a difference between this game and other 3D shooting games: my internalization of the importance of the in-game missions. In most shooter games, my enemy is clearly a fantasy opponent (e.g. zombie, alien). In war-based shooter games, my personal political views on the worthiness of war make it difficult for me to deeply empathize with my avatar's motives. Although I do immerse in the role-play of a soldier at a tactical level (e.g. I feel frustrated when I do not complete my missions), at some level I am unable to accept war games' portrayal of reality.

However, *Re-Mission* was different. Here, I was fighting a real-world enemy (cancer cells) that I personally hate. My father has battled lymphoma, and I could easily empathize with him, and the metaphor of battle, of combat, was a good fit for how I felt about it. *Re-Mission* allowed me to role-play a fantasy role I actually desired: I would have happily accepted a real-world mission of controlling a nanobot through my father's body, attacking cancerous cells. In this way, *Re-Mission* combined fantasy and reality in an emotionally powerful way. Developers cited a youth with cancer: "Many times we as survivors feel helpless battling our disease, being Roxxi in the game makes you feel like you are battling the disease itself" (Mel Majoros 2012). This quote fitted with my own play experience.

My play experience with *Re-Mission* suggested the game may achieve a crucial learning outcome that its designers did not declare: Humans are notoriously bad at understanding the implications of nonlinear functions, despite the simple mathematics required to explain them (Bartlett 1976). While classroom exercises aim "to give [students] experiences where they discover the exponential function on their own using arithmetic that they understand," I found *Re-Mission* to be a far more powerful method of understanding this. After *Re-Mission*, I found that I not only was able to visualise the exponential growth of cancerous cells, I was able to relate this to my imagined adherence to any self-administered treatment

A certain aspect of the game design enabled this, in my view. The mathematical model of *Re-Mission's* malignant cell reproduction is exponential: I could watch them divide, again and again. This was crucial to my play experience, in that it drove my strategic decisions. If I didn't shoot the cells quickly, they were able to overwhelm me. After a certain point in their exponential growth, I would be virtually certain to lose. My play strategy (exponential growth of enemies requires early, intense attack) was very closely aligned with the real-world learning goal (exponential growth of cancer requires early, intense treatment). In the following discussion, I ascribe my realization to an unstated learning goal: the "tacit implication of exponential growth."

It was reported that although only 28% of participants fully adhered to the requested 1 hour of game play per week, "significant positive effects, even for participants who played the game <50% of the requested duration" were found (Kato et al. 2008, e314). The researchers were surprised by this finding, and asked: If playing the game was effective, why wasn't playing longer more effective? In considering this finding, I developed a theory that explained why efficacy of the game was not as linearly related to play time as I imagined. I had personally experienced the "tacit implication of exponential growth" in the first fifteen minutes of playing. Additional playing time delivered a little more factual information: for example, I learned about different kinds of treatments and how they affected cancerous cells differently, but none of this had the emotional impact of that first realization. To this day, I understand the implication of exponential growth in a way no parabolic chart has ever achieved for me. I wondered if other players had a similar experience. Did a single realization occur, which could happen after only a brief play experience? Or was there a gradual building of a general understanding which would improve gradually over lengthier play experience? I did not collect the data needed to answer this question.

Re-Mission may be helping cancer patients accept the side effects of their pills as necessary, a well-documented problem (Levine et al. 1987). The theory of delayed gratification has been used to explain that all humans unreasonably overvalue short-term gain in decision-making (Chapman 2005).

I believe that part of *Re-Mission's* efficacy is related to players' improved understanding of the tacit implication of exponential growth. Tacit knowledge is related to intuition, a huge, poorly understood, part of decision-making (Khatri and Ng 2000). I found *Re-Mission* had changed my intuitive risk/reward analysis of cancer treatment. I already knew, in theory, why cancer treatments such as radiation and chemotherapy are usually so intense, but I also can easily imagine being a teen that chose not to take my pills if they caused vomiting or similar side effects. *Re-Mission* illustrated the benefits of intense, early treatment. If I were diagnosed with cancer and having treatment options explained to me, I believe the game experience would cause me to favour "hit it hard, early" strategies. I investigate this finding further in following sections.

To restate, I found the game's explicit design aims were very well aligned with the core mechanic. In particular, the core mechanic was nearly perfectly aligned with an understanding of the exponential growth of cancer that is so unforgettable, deep and powerful that it was easy for me to accept that this game does cause patients to improve their treatment adherence.

Was this high alignment accidental or planned? The designers of *Re-Mission* published some of their reasoning and process of design decision making (Tate, Haritatos, and Cole 2009). In this paper, I

can observe that some of their ideas are consistent with Shelton's theory. They give a core principle of their design process, which they describe as: "live out contingencies in the virtual world instead of real life." This is closely related to, and possibly another phrase for, activity-goal alignment. This quote from *Re-Mission's* designers further shows how they grasp the essential idea in activity-goal alignment:

Live out contingencies in the virtual world instead of real life (e.g., IF the patient skips chemotherapy doses, THEN Roxxi's chemoconcentrating blaster misfires every third shot— cancer cells survive and become drug resistant) (Kato et al. 2008, 310)

This connection between misfiring weapon and real-world behaviour is a near-perfect example of high activity-goal alignment: the learning goal (teaching kids not to miss treatments) and the in-game activity (misfiring when treatments are missed) are closely aligned.

When I played *Re-Mission*, I did not experience some of the specific features mentioned in the developers' descriptions, including the misfiring mentioned above. The vast majority of game play experience was focused on shooting cancerous cells. The activity of taking pills on time was not a part of the core mechanic I experienced. The game did allow me to experience the catastrophic consequence of missing even a single cancerous cell, but it did not link the real-world activity of pill-taking to that consequence. It is puzzling that the developer's ideas related to activity-goal alignment were not more commonly observable in my personal game experience, or the videos of gameplay, or player testimony (Mel Majoros 2012).

The game was most effective teaching skills related to its core mechanic's activity: I was coached, assessed, encouraged, and ultimately required to improve my aiming and flying skills to win. Through the lens of activity-goal alignment, the primary activity of shooting cancer cells has little relation to the learning goal of pill-taking, and the secondary activity, treatment of side effects, did not occur very often during gameplay. It is interesting to think that, if another game with higher activity-goal alignment had been built with the same treatment-adherence goal, it might have been difficult to find a way to make the game as engaging as shooting cancer cells, and thus perhaps it would be less effective than *Re-Mission*. This suggests that activity-goal alignment, although important, can be overemphasised in design decisions.

In summary, I felt that, the designers did not cite Shelton's activity-goal theory, they did understand similar ideas, and this awareness may have helped them select the activity-goal alignment observable in the game's design.

5.2.1.2 Activity-Goal Alignment Score for *Re-Mission*

Here I describe my activity-goal alignment scoring for *Re-Mission*, using the AGA Scoring Tool v7. I provide not just the score but my thought process in some detail, including multiple iterations, so that the reader can notice how the AGA Scoring Tool both aids and constrains analytical thinking. As explained in the previous chapter, the AGA Scoring Tool has four elements: context, Venn diagram, discussion, and score. Each of these elements is addressed explicitly below.

One important distinction between Shelton's activity-goal alignment and previous similar theories is Shelton's emphasis on player context. Accordingly, I did not use my own context of play because the game is clearly targeted at cancer patients, not me. Based on my observations of my and others' play experience, and the stated intended audience, I wrote a persona for the context: "Middle-class Australian teenager, diagnosed with cancer, living at home. He plays on the family PC in the living room. His sessions are not social: they are single-person self-challenges. He engages intensely (1-4 hours per day) for several weeks, until he feels he has beaten (or failed to beat) the game. He plays occasionally after that."

In identifying a Core Mechanic, or primary player activity (the blue circle in the Venn Diagram in Figure 17), I chose the phrase "shooting cancer cells," after rejecting broader genre phrases like "3D shooter" and deciding that the role-playing elements, while significant, were not central enough to constitute a true role-playing game. I used the three learning goals that were studied, although there are others named by the developer, as these three seemed most important to the developers. I used their phrasing (HopeLab 2012). I used lower case in my text to remind me that this assessment is a "rule of thumb." I wanted the tool to remind its viewers that activity-goal alignment score is informal, speculative and flexible, as opposed to a formal, rigid assessment, which I associate with proper capitalization.

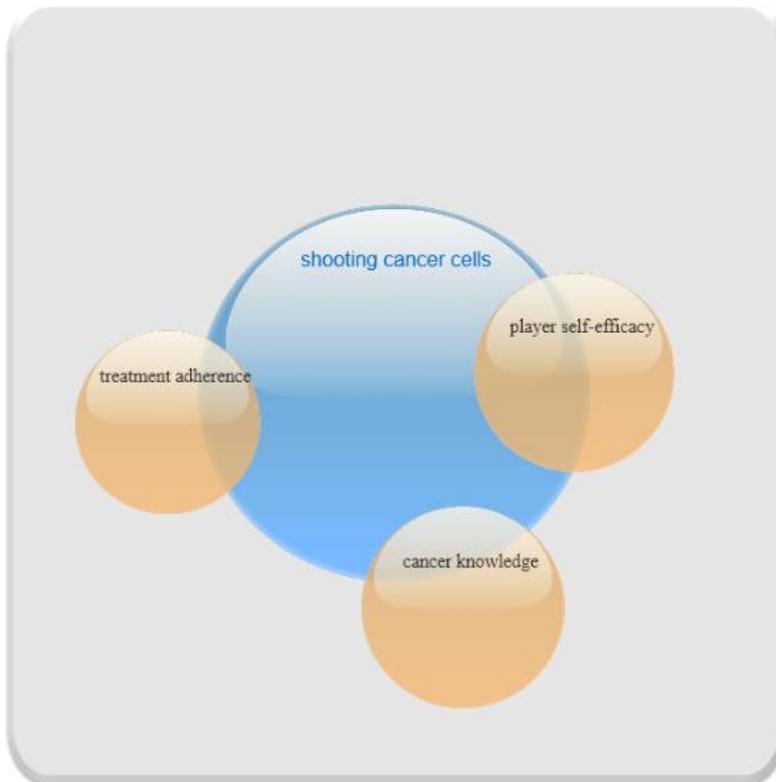


Figure 17: My initial positioning of Re-Mission's activity-goal alignments was quick and intuitive.

Regarding the placement of overlap, I first created the three learning goals (the orange circles in Figure 17), and then considered each one as I placed it on the blue activity circle. I found this a very useful tool to enable reflection. I was careful to replay the actual game experience in my mind as I placed the circles. I adjusted the placement of each circle many times as I reflected. My first iteration had three goals placed quickly and intuitively, without much rational justification, as shown in Figure 17.

Once I set up an initial placement, I imagined explaining my diagram to various audiences, including the *Re-Mission* developers, other game developers, and academic researchers. For each audience, I imagined how I might explain or defend the Venn diagram's degree of overlap. Then, I reviewed the notable moments in my gameplay experience, and recalled my observations of other players' experiences. I asked myself if this overlap was a fair summary of my perception of a typical play experience and kept moving them until their position balanced the tensions and reflected my feeling. Instead of considering the angular position or size of the learning goals, I changed only their radial distance at first.

The description above may make my process seem more linear than it actually was. I did not sequentially place the circles, recall my experience, look at the circles, move them, and then recall

again. Instead, I entered a mental state that bore hallmarks of a Flow state (Csikszentmihalyi 1990): focused and highly present, I worked in rapid iterations: recalling, moving circles, and reflecting on the impression the diagram gave, all at once. I discuss the significance of this state and this particular observation in the final section of this chapter.

I did not feel that the size of the learning goals was of primary importance and initially left them all the same size. After some consideration I thought it might help to make size reflect importance of treatments, so I increased the size of the “treatment adherence” goal, and shrank the other two, and decided a fourth learning goal, exponential reproduction, was justified. This diagram can be seen in Figure 18.

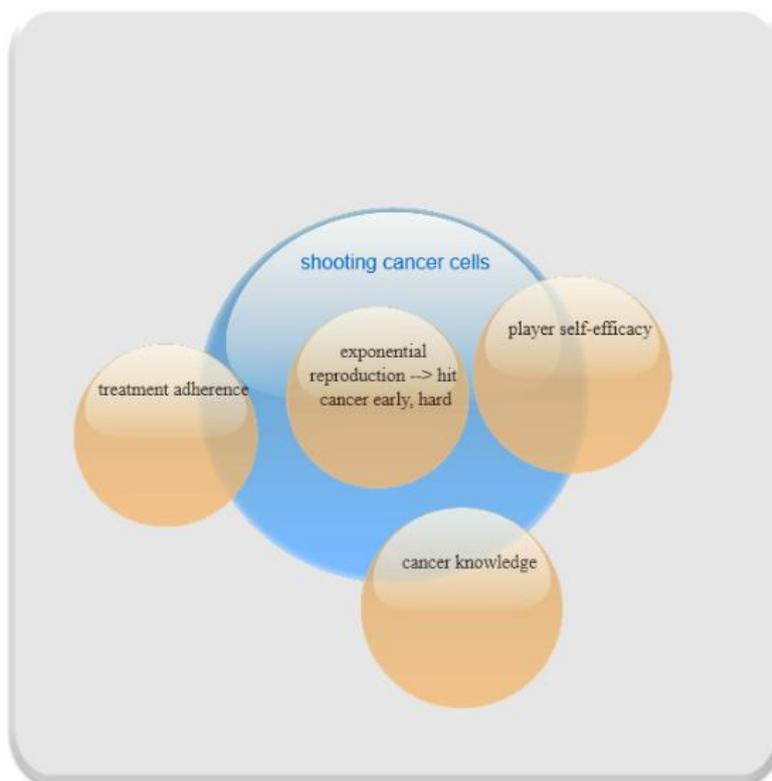


Figure 18: An interim version of the Re-Mission activity-goal alignment shows varying sized learning goals.

After reflecting on this diagram, I briefly played the game again to refresh my recollection of the play experience.

I had initially decided to that the learning goals' size would relate to their importance. Upon reflection, I decided to change the meaning of the size for two reasons. Firstly, I felt that importance of goals was too high-level a consideration for an activity-goal alignment analysis, as it is usually set at the beginning of the project, and is not negotiable by game designers. Secondly, I felt that the

importance did not sufficiently relate to my observations in the game, so I was portraying a fixed decision made by the designers, which was available without my help. For these reasons I decided to change the meaning of the circle size to “observable impact” in the game play experience and resized them as shown in the final diagram (Figure 19).

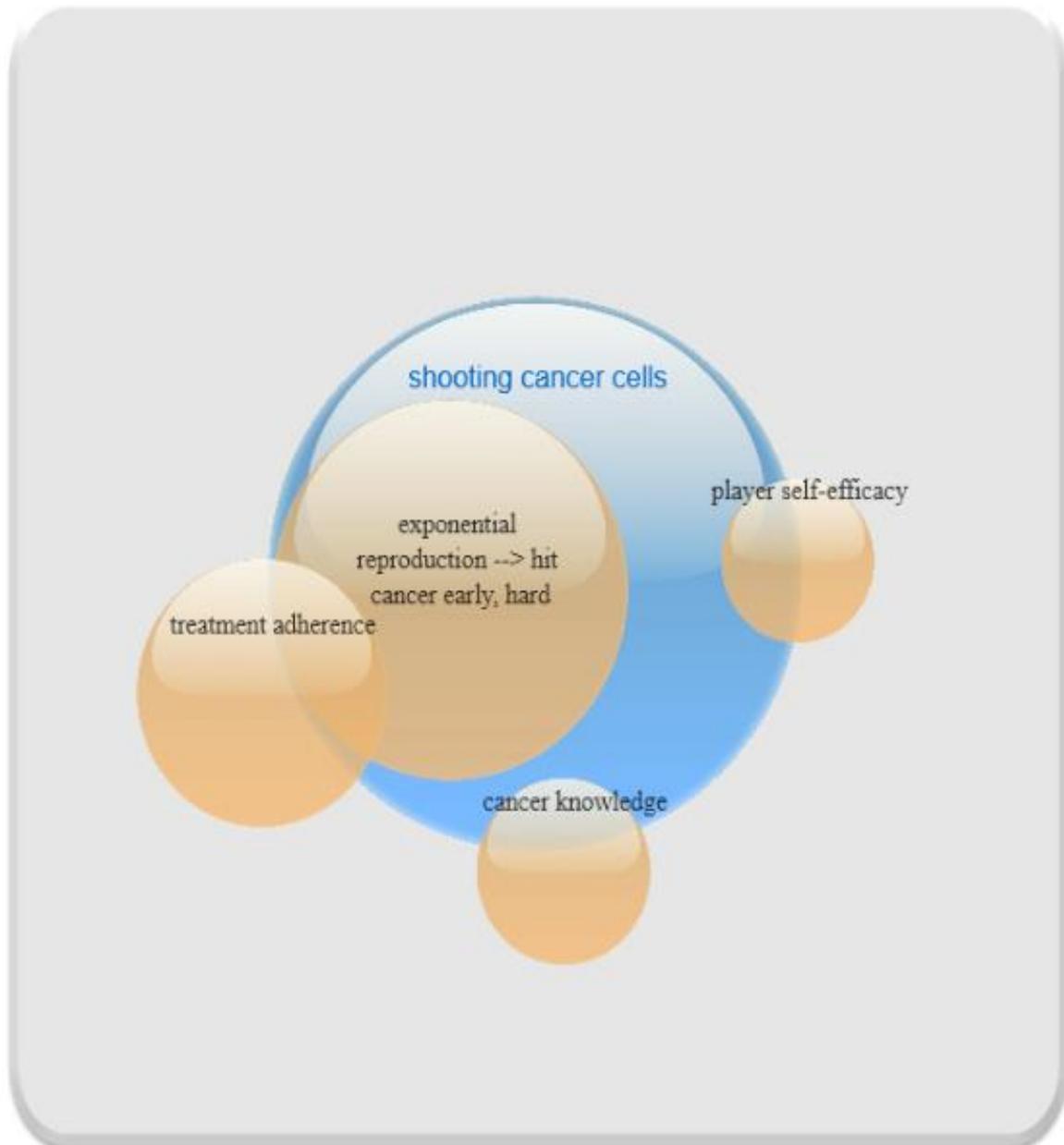


Figure 19: My final activity-goal alignment diagram for Re-Mission.

The form then asked me to explain the diagram. I wrote:

I show weak overlap with cancer knowledge because of the nonstop action: it was too intense for significant factual learning. I was focused on winning, and winning did not require

comprehension of cancer knowledge. I felt the shooting mechanic would build my self-efficacy, though could also imagine self-efficacy being further improved through social connections. The weak overlap with treatment adherence because the pill-taking behaviour was rarely referred to, and there was no visual component. The game could have more directly portrayed treatment adherence. For example, by seeing the actual pills destroy the cancer cells, rather than the abstract shots in the game, Re-Mission might have more solidly built the connection between pill-taking and cancer-killing. However, this idea is only loosely based on behavioural theory, and I cannot be sure that such a game would be more effective than the existing game. The game was perfectly aligned for an undeclared, specific learning goal: that exponential reproduction of cell growth means it is vital to treat cancer aggressively. I felt this at a tacit "belief" level, a feeling that was far stronger than mere knowledge, after playing.

The form then asked me to score the game in text form. I looked at the circles and wrote down my estimate of the percentage overlap, rounding to nearest 25% increments. I wanted to be less accurate and considered using text descriptions, but "low" and "high" did not fairly represent what I felt, and other words felt too vague. I wrote "Cancer knowledge: 25% Treatment Adherence: 25% Self-efficacy: 50% Exponential: 100%".

The core mechanic (shooting) and the real-world activity of pill-taking are highly dissimilar, which might erroneously suggest that the activity-goal alignment is low. However, the designer's stated learning goals do not specifically name pill-taking, so perhaps pill-taking is not appropriate to consider in activity-goal alignment theory. Instead, pill-taking is a real-world behaviour that serves as a measure of efficacy for the game.

In scoring *Re-Mission's* learning goal that related to exponential reproduction, I posed the following question to myself: What was special about *Re-Mission's* approach to portraying exponential growth of enemies? Could this realization have been achieved in any game, or is there something important about this particular implementation? To investigate, I tried to isolate the specific element of exponential multiplication by playing other games that had exponential growth of enemies.

The rate of enemy creation is an important design decision that must be made for any shooting game. To set this rate, designers must make at least two types of decisions: the first is a mathematical model (linear or exponential), and the second is a phase known as "tuning," where the designer modifies behaviour of the core mechanic by changing the values of many specific,

interdependent values, often through playtest feedback (Schell 2008, 392). For example, how many enemies per second are created? How many spawn points? How fast can enemies move?

As these decisions are crucial to the play experience, the math behind enemy spawning is sometimes studied and critiqued by reviewers and players, not just developers. For example, one player of a casual multiplayer game *Battlenations* (Z2Live 2011) posted “Please increase the rate at which weaker enemies spawn so I am not continually wait[ing] on troops to get out of the hospital and have a hope of levelling them up at a reasonable rate” (dangerB13 2012). Exponential models are not common in modern games, and have been criticized as a disingenuous way to end games quickly in 1980s arcade game designs (Dugan 2010). At least one player agreed, critiquing the exponential spawn rate in a modified version of the 3D shooting game *Caster* (Elecorn 2010): “The multiplying enemies multiply too much, way too much” (CDB531 2012).

Re-Mission is not the only learning game that uses exponential growth of hostile organisms in the bloodstream to portray the impact of delayed treatment. The Immune System Defender (Nobel Media AB 2004) and *Infection! The Game* (Anton 2009) both do so as well, and I compare them to *Re-Mission* below.

5.2.2 *Infection! The Game*

Infection! The Game (Anton 2009), which I will henceforth call *Infection*, is a simple 2D online video game. This game is useful in this discussion because it isolates exponential regeneration as a pure mechanic.

I could not find any particular learning goals stated by the developer for this game. The credits link to a web site that does not discuss the game but seeks to sell an “Apprentice Doctor Foundation Course” for \$19.95, making me believe this is an advertising game, not a learning game. I feel the game’s design is simple enough to clearly suggest a learning goal: teach the perils of untreated exponential growth of enemies. Note that this learning goal is independent from cancer, or even human health: it could apply to alien invasion, or weeds invading a suburban lawn.

The mechanic of the game is apparent in Figure 20: the player moves an attacking macrophage around a circular field, destroying bacteria, which move around and divide (exponential regeneration). Each round can end three ways: the player clears all bacteria; the round timer runs down, or the level of infection (shown in a bar graph at the bottom of the screen) exceeds a certain limit. To kill bacteria, the player mouses over the bacteria, and the macrophage moves toward it. When they overlap the bacteria is engulfed and killed. Armoured bacteria take longer, as the macrophage must bump them several times before they revert to simple bacteria which can then be

engulfed. Clicking the mouse on a bacterium launches a small arrow in the direction of the mouse pointer. If the arrow hits a bacterium, the bacterium is tagged. The macrophage moves toward any tagged bacteria and attacks them without player guidance. There are four arrows which cannot be lost, but the macrophage must run over the spent arrows to collect them for reuse.

There are two modes. In Beginner mode, shown in Figure 20, the levels are really simple. As levels progress, bacteria become more armoured, move faster, and reproduce more quickly. In Expert mode, the game is tuned to be more difficult (more bacteria spawning, heavier armour), and a new element is added (a poisonous object that freezes the macrophage).

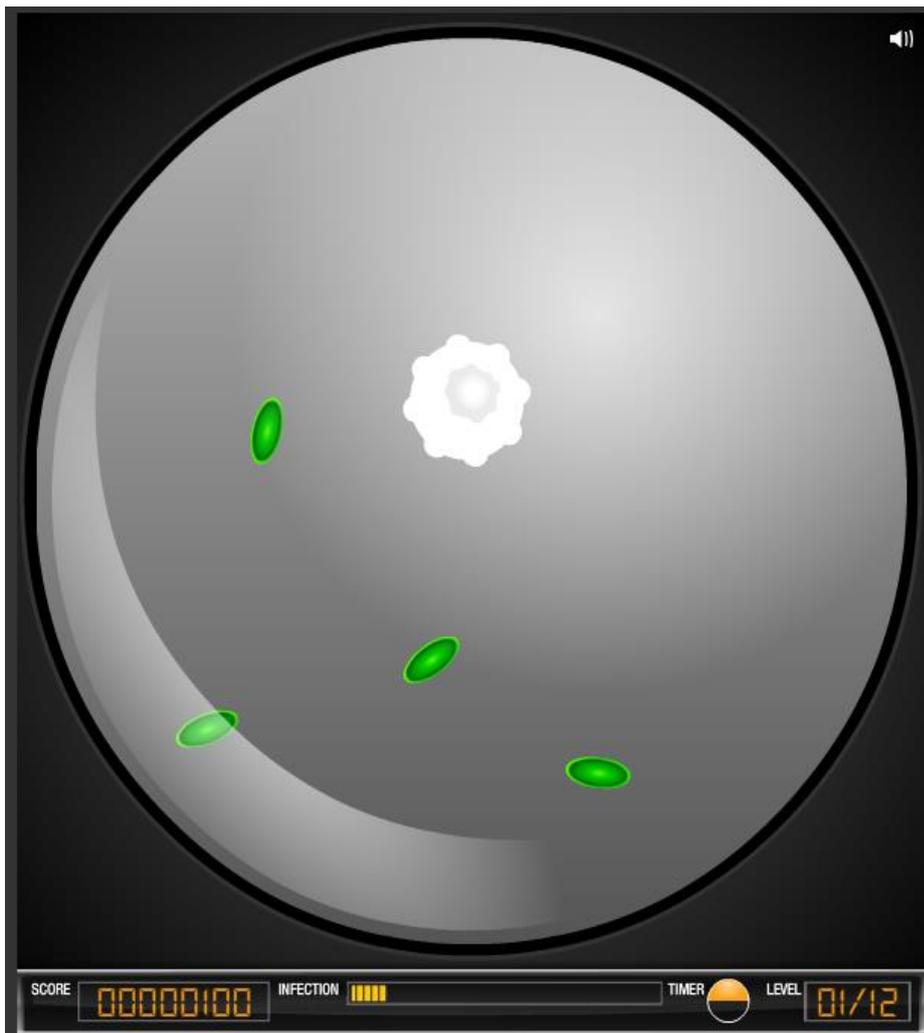


Figure 20: *Infection: The Game* is a simple 2D game that lets the player attack viral cells as they multiply. (Anton 2009)

5.2.2.1 My Play Experience

I played the game for 30 minutes and finished it on both modes. Overall I did feel a similar sense of urgency and challenge in *Infection* that I had felt in *Re-Mission*, despite the relaxing music and

abstract aesthetic. The exponential growth of the bacteria was similarly challenging and engaging in both games. I lost rounds many times, allowing me to experiment with strategies and tactics, repeating levels multiple times until I found an optimal solution. Expert mode added a freeze mechanic (the red bugs shown in Figure 21) which required me to approach the problem afresh; however, it was not difficult for me to develop a new strategy.

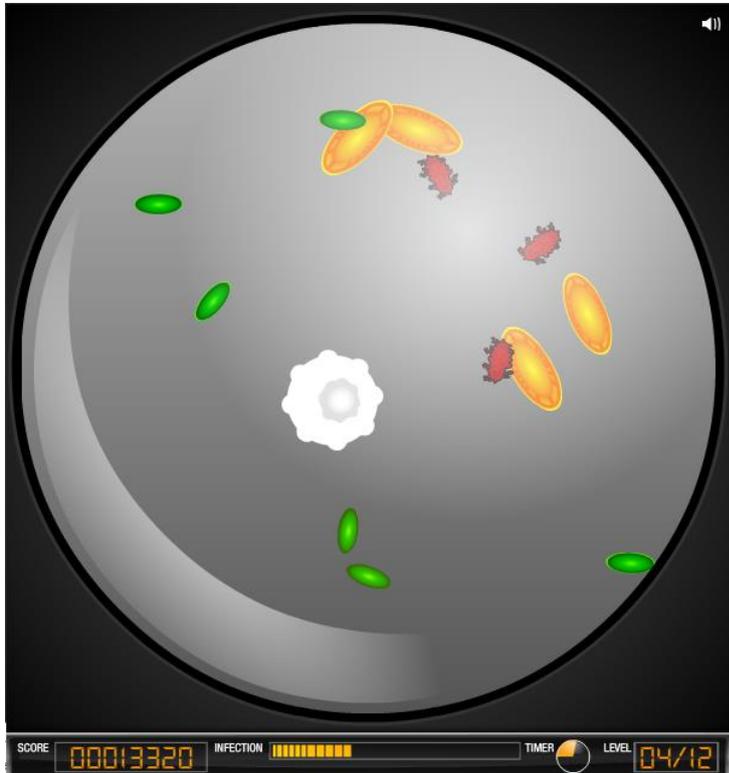


Figure 21: In Expert Mode of Infection, red bugs cause the player's avatar to freeze. This image also shows the indication of bacteria division (orange colour). In this image, all the orange bacteria happen to be armoured (indicated by the small circles near edges). (Anton 2009)

The rate of regeneration is intense enough at higher levels to cause the player to discover the same "hit early and hard" strategy, but the limited playing field makes it much less threatening: I could see all my enemies, and assess how urgent the situation was. *Re-Mission*, by comparison, had the limited field of view that is part of any 3D game. Since areas were always off-camera, and the bacteria moved rapidly, I found it difficult to visually verify that I had killed all the bacteria, so I was never sure the situation was under control¹⁶.

¹⁶ Although *Re-Mission* had a mini-map that offered a top-down 2D view, I found it too visually complex and difficult to read at a glance, and did not use it much.

5.2.2.2 Activity-Goal Alignment Score for *Infection*

I have abbreviated the discussion for this score¹⁷, presenting only the four key outputs from the AGA Scoring Tool:

- Context: I wrote: “middle-class 16-year-old assigned to play in school.”
- Diagram: The diagram is shown in Figure 22.

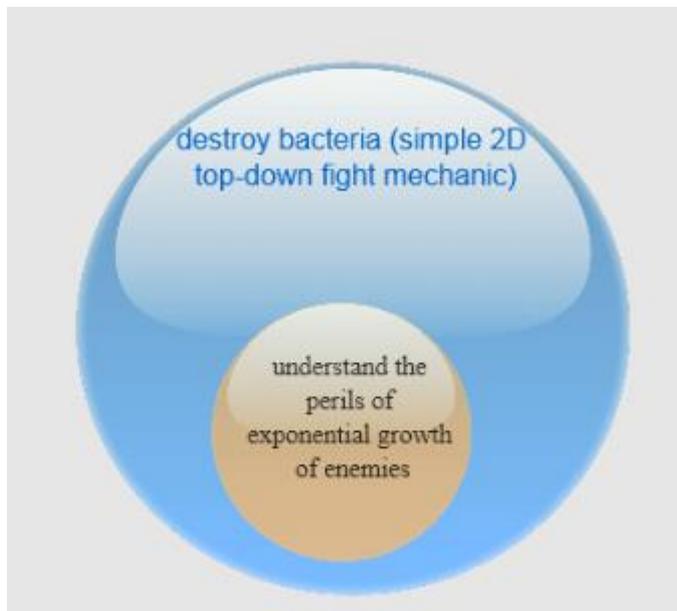


Figure 22: Activity-goal alignment of *Infection*.

- Discussion: I wrote: “While acknowledging that I defined the learning goal, not the designer, it is not surprising that this game has such high alignment. If the goal had included bacterial behaviours, this game would not be sufficient. While the game does not cover all kinds of exponential growth, it imparts a solid tacit understanding of the perils of exponential reproduction of an enemy.”
- Score: I wrote: “100%. The learning goal of exponential growth is fully aligned with *Infection*’s core mechanic.”

I found scoring this game was very quick and easy with the AGA Scoring Tool. At first the tool felt too structured for such a simple game, since it seemed apparent that the activity-goal alignment was very high, but I found that the structure was helpful in getting me to reflect on various aspects of the game.

¹⁷ I provided a level of detail for Re-Mission, but my process did not vary enough, and my changes were too insignificant to justify the reader’s attention through such detail for several more games.

Infection was interesting in terms of activity-goal alignment because it exhibited high activity-goal alignment for the particular problem of exponential growth, but despite this I did not achieve the depth of tacit realization I got in *Re-Mission* around the particular problem of exponential growth. This divergence between high activity-goal alignment and efficacy is interesting. With *Infection*, I believe the primary barrier to deeper tacit learning was its lack of context in the game's storyworld. When I played *Infection*, I played like a pure gamer. I only weakly connected the game's feeling of urgency to the threat to real-life infection. It was not a purely abstract game like Tetris: I did consider the game objects "bacteria" and felt some kind of satisfaction in killing them all. In *Re-Mission*, I felt the importance of my mission strongly, and I believe this was because *Re-Mission's* storyworld situated me in the harsh reality of cancer. The 3D environment constantly reminded me that I was fighting inside a human body, trying to save a life. *Infection* did not offer this fictional context. Before *Infection*, I felt that *Re-Mission's* story and context was nice but it was unclear how important it is. Is the story even necessary? After *Infection*, it was clear that the story element in *Re-Mission* is important, maybe even necessary, to its efficacy, especially as the targeted players have such a strong emotional connection to the story.

5.2.3 *Immune Attack*

Immune Attack (Stegman 2010a) appears similar to *Re-Mission*. The player pilots a nanobot through a 3D environment (shown in Figure 23) of blood vessels, completing missions that retrain broken immune cells. However, *Immune Attack* differs from *Re-Mission* in some important ways.



Figure 23: Immune Attack superficially resembles Re-Mission, but the game play experience is quite different. (Stegman 2010a)

Immune Attack has different learning goals from *Re-Mission*'s goals. Although not stated as clearly and specifically as *Re-Mission* did, *Immune Attack*'s designers gave some indication of their intended learning goals: "We expect that the presentation of cell biology in *Immune Attack* will leave students with an understanding of interactions between different cell types, how these interactions are mediated by proteins, and how cells modify which proteins they express in response to their environment" (Stegman 2010b). The caption under the demo video on the website for the game suggests the game also seeks to teach more specifically as well: "...you [the player] will learn about the biological processes that enable macrophages and neutrophils—white blood cells—to detect and fight infections" (FAS 2011). In reflecting on that and the designers' statements in other publications (Stegman 2010a), I believe it is fair to define *Immune Attack*'s learning goals by stating that the game has these three learning goals:

1. Factual knowledge about the immune system (e.g. names of cells in cellular biology),
2. A generally accurate mental model of the immune system (e.g. a player would realise that macrophages are an important element of the immune system, because they destroy foreign invaders, even if the fact wasn't stated explicitly), and
3. Experiential understanding of specific immune system dynamics (e.g. players have seen "seen" macrophages follow scents through intercellular space to find tagged invaders).

Note that *Immune Attack* aims to educate but does not seek to change any player behaviour, whereas *Re-Mission* is designed to both educate and change a specific behaviour (treatment adherence) among its players.

5.2.3.1 My Play Experience

When I played *Immune Attack*, I completed most of the game. After reading the walkthrough notes which described the game from start to end, I decided that the last two levels were not sufficiently different from the first five levels to justify the time spent playing. *Immune Attack* felt like a character-based adventure role-playing game: carefully crafted objectives and missions that were intended to challenge me to identify objects in 3D and remember information, rather than the twitch and precise accuracy required in *Re-Mission*. This gentle pace suited the learning goals well, allowing me to explore my environment in a self-paced way.

On the particular issue of exponential increase of foreign organisms, *Immune Attack* was very different from *Re-Mission*. The most comparable level in *Immune Attack* was level four, "Eat the Pseudomonas!", where the player navigates through 3D space, tagging enemy cells which are then dispatched by NPCs (macrophages). The enemy cells did not move or attack as they did in *Re-Mission*; instead they floated serenely in intercellular space. Their numbers did increase, but there was no visual representation of their reproduction, other than a counter in the upper right corner of the screen that would sometimes increase by one. The regeneration rate felt linear, not exponential. The primary challenge was finding them, these tiny green pill shapes in the large 3D space, crowded with numerous other large or entangled cells. The core mechanic felt more like a "spot the hidden object" than "fight for your life". There was a small amount of time pressure but it was nothing like the intensity of *Re-Mission*. As a result, I cannot imagine learning the implication of exponential growth by playing *Immune Attack*.

Regarding the more general play experience, I found myself alternating between frustration and pleasure while playing *Immune Attack*. I enjoyed how the game kept the science rigorous and technical. It made a good effort to engage me in the truth, allowing the player to play with the fundamental actions that really do occur in the immune system. I felt that I was likely to retain an intuitive understanding of the interplay between the elements of the human immune system as a result of my play experience. I did often wish the game was harder. The designers did a good job of making the game fool-proof: clear instructions were provided at every turn. However, this clarity came at a cost: I rarely felt a sense of personal achievement. It felt like I was working through the tutorial mode, but there was no interesting application of my newly-mastered vocabulary, mental model, and control skills. I kept hoping the next level would say "Great work! You have proven you

can handle a macrophage in isolation...but that's not a real-world situation. In this level, you will attempt to beat pneumonia in an AIDS patient. If you ever get stuck, click help. Go!"

As I played, I found myself questioning the designers' knowledge of commercial game design literature. When compared to *Re-Mission*, *Immune Attack* felt like it had been built outside of video game culture. Notably it does not use standard 3D game navigation controls (Gkikas, Nathanael, and Marmaras 2007). For example, in *Immune Attack*, the S key is used to stop the ship's motion, whereas in most games it would move the ship backwards. Moving the mouse moved a targeting reticule, where it normally rotates the avatar. To rotate the avatar, the player must hold down the right mouse button. In this mode, using an elastic controller which made positioning nonstandard, the view could be freely modified. These controls were not terrible: stop in particular was useful in level four, but I did not perceive a sufficient benefit over well-established standards to justify this change. The game also did not allow the user to modify the controls (i.e., change key mappings). Other well-established commercial game conventions were reinvented or ignored: repetitive audio hints were not suppressed, and designs of mini-map and targeting icons were needlessly unfamiliar.

Immune Attack's representation of real-world information, including facts, behaviours, and aesthetics, felt trustworthy and rigorous. I did not sense any inconsistency or vagueness in the facts underlying the supporting material, the in-game text, the spoken dialog, the 2D user interface, or the visual 3D representations.

5.2.3.2 Comparison of *Immune Attack* and *Re-Mission*

I felt the designers of *Immune Attack* prioritized factual accuracy over poetic license in almost all design decisions except for the player's character (the nanobot). *Re-Mission* designers, on the other hand, stated that fun came first in their design decisions (Tate, Haritatos, and Cole 2009), which matched with my *Re-Mission* play experience.

While *Re-Mission's* production quality does not match a \$10 million budget commercial game, it does have a sufficient degree of polish¹⁸ to be credible as a commercial entertainment product.

Immune Attack did not feel to me like it had achieved that level of polish. There were many problems that iterative debugging and beta testing usually flushes out. For example, I noted a bewildering mismatch between 3D in-game visuals and the 3D cut-scene visuals for the macrophage transmigration. The designers state they plan to fix this a future version (Stegman 2010a).

¹⁸ This word is commonly used in video game production, in line with Merriam-Webster's 3rd meaning: "to bring to a highly developed, finished, or refined state" (<http://www.merriam-webster.com/dictionary/polish>)

There was also a lack of sophistication in the instructional design. The biology facts on every transition page were unrelated to their context, which made them distracting and added to my cognitive load. The scriptwriting was painfully slow-paced and lengthy. The character believability (as experienced through voice acting quality, dialog, and storyworld) was not sufficient to enable the character empathy or storyworld immersion they seemed to be intended for. I found myself skipping the briefings and stumbling through the mission through trial and error. This hampered my learning because there was some learning material contained in the briefings.

After struggling through the control system and noting various minor problems, such as the lack of clear feedback when I damaged my ship, the overall construction of the game's "story arc" and core mechanic felt coherent to me. I worried that a teenager might not agree, and wondered if the designers had their target audience clearly in mind. In other words, I would not expect a student to play this game instead a commercial game. The flaws prevent sufficient engagement, although it is enjoyable at times. The quality of the content is very high.

5.2.3.3 Scoring Activity-Goal Alignment of *Immune Attack*

In this section I note the salient output from my application of the AGA Scoring Tool to *Immune Attack*.

- Context: High school student, middle class and moderately motivated, playing in a classroom as directed by their biology teacher, somewhat distracted by desire to chat with a friend playing separately nearby.

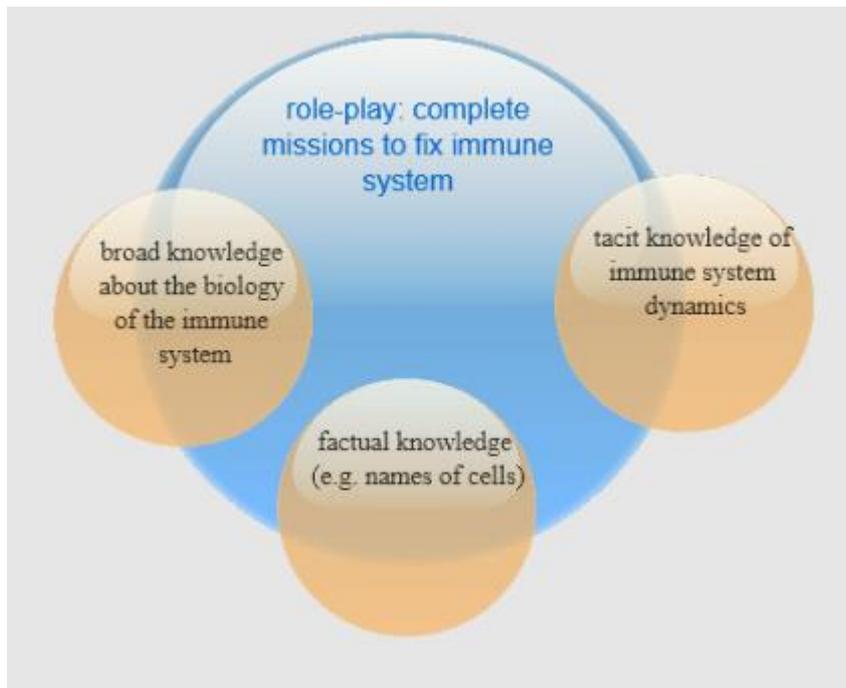


Figure 24: *Immune Attack*'s activity-goal alignment diagram.

Here I discuss the meaning of the activity-goal alignment scoring diagram in Figure 24. The tacit knowledge that came from completing missions has stayed in my memory as a result of my game-playing, even though the missions themselves did not feel particularly compelling. I can recall the large, slow-moving macrophage, rolling down the blood vessel and being retained by the markers on the walls. I can recall the macrophage roaming through the intercellular spaces, following scents to hunt down tagged intruders. It all seems real to me in a way it didn't before playing the game. The factual knowledge has stuck less well, but again I find that it fits in my mental model nicely. The scientific, long names can feel awkward and pompous when I use them in a doctor's office discussion (partly because I am not quite sure if I'm using them correctly, and the doctor is sure). These scientific names felt compatible with the game's fiction. I found myself attempting to grasp the proper names so I could better role-play, and this effort was at a nearly subliminal level. The game clearly did not intend to portray the entire immune system, so there was much about the system that was not fully presented, but the broad knowledge supplied was also compatible with all the game activity. I would have ranked it higher, but the game just wasn't much fun—it felt very earnest. In this way, I felt the activity itself was too unrefined to allow for a really high activity-goal alignment score.

- activity-goal alignment score :

Factual knowledge (e.g. names of cells): 70%

Tacit knowledge of immune system dynamics: 70%

Broad knowledge about the biology of the immune system: 60%

In scoring this game, I was surprised by how well the activity and goal actually aligned. This score focus lost some of the point of the textual discussion, but it also guided my thinking to realize the game had high activity-goal alignment. Without the tool, my score would have been lower because of the game's lack of polish. This is a reason I could not have defended, and the tool helped me see that.

5.2.4 Conclusion

I conclude the first of two discussions in this chapter by summarizing the most important observations. For additional clarity, I boldface the crucial points from each observation.

Most significantly, I found **the AGA Scoring Tool captured the gist of my observation of activity-goal alignment**. In all three games, the diagram was especially useful in enabling me to reflect nonverbally yet specifically. In addition, I thought the resulting image was useful as a visual illustration of activity-goal alignment. I found the tool able consider multiple learning goals against one activity without losing significant clarity.

Also significantly, I found that the **activity-goal alignment score was more comparable and consistent than a textual discussion** when comparing various learning games. The scoring tool implies a process. Following this process caused me to be more consistent and objective in that I was able to capture a more accurate recollection of the activity-goal alignment. My text discussions explored unique aspects of each game, revealing what was most important for each game. However, the freedom also reduced the quality of my analysis: without the guiding force of a tool, some of my scoring discussion varied in quality and criteria.

There are drawbacks to the tool's conciseness. I found **the AGA Scoring Tool does not allow the designer to distinguish between types of learning goals**. Learning games can aim to change a player's belief, behaviour, factual knowledge, and/or attitude. I had originally decided not to attempt to represent these differences in the interest of clarity. While using the tool to assess *Re-Mission*, I felt some desire to add detail to the diagram. However, upon reflection, I remain convinced that adding such detail would hurt the diagram more than it would help. Clarity is the diagram's key advantage over textual discussion. Also, as Shelton does not specifically explain how activity-goal alignment theory suggests different considerations for each type of learning goal, I

believe textual discussion is more effective for this and other more complex considerations of activity-goal alignment theory.

I observed that the activity-goal alignment score in *Infection* was higher than *Re-Mission*'s score, but *Re-Mission* is a more impactful game to me. From this, I find that **scoring activity-goal alignment is useful but insufficient to fully explain a user experience**. This is not surprising, but bears repetition to ensure that the limits of activity-goal alignment are clear: learning game design is too complex to summarize in a single theory such as activity-goal alignment.

5.3 Game Design

In the second half of this chapter, I apply a similar analytic and scoring process to a different genre of game for comparative purposes. I discuss a nearly self-referencing topic: the design of video games that teach video game design. At least two such games exist, and I compare their very different approaches by scoring their activity-goal alignment.

5.3.1 *GameStar Mechanic*

GameStar Mechanic is a learning game whose objective is to help children adopt a designer mindset and develop literacy around game design. It is a highly praised modern learning game, having won numerous prestigious awards. The lead designer, Katie Salen, has been a celebrated leader in the learning game design field for many years. There are three core mechanics in *GameStar Mechanic*: playing games, repairing existing but broken games and designing games from scratch.



Figure 25: *GameStar Mechanic* both teaches game design, and lets players build their own game. Note the "edit/play" rocker switch in the upper left. (E-Line Media 2012)

The game incorporates a game engine. Players can build simple 2D side-scrolling games using “tiles” or “sprites” which are resources needed to build a game. The player character, enemy characters, points, walls, exit points, or power-ups are all examples of tiles. A few example tiles are shown in the grid panel on the above the “Publish” button in Figure 25: the blue figure in the upper left, for example, can be dragged and dropped onto the map on the right side of the screen to set the starting position of the player character. A few properties for each tile such as colour, speed, jump strength, or lives, can be set. A few general properties such as gravity, wraparound screen can be set as well. Compared to game engines designed for children such as *GameMaker*, *GameStar Mechanic* is very limited in features and scope.

Salen said *GameStar Mechanic* was meant to fill a particular need: most game building software doesn’t “explicitly embed the practice or thinking of game design within the experience and often focus on game programming as the primary pedagogy” (Salen 2007). Salen’s focus on procedure,

not skills or factual knowledge, is consistent with Gee's position on the potential of learning games¹⁹. Games²⁰ further explains how *GameStar Mechanic* is an example of Gee's "Discourse" of game design.

By Discourse, Gee (2005) refers to the ways of doing, talking, thinking and using tools that game designers enact in professional practice. Participants can then appropriate this Discourse in the context of a familiar game-play activity, which Gee (2003) has argued fosters a safe space for self-expression and early learning success. More sophisticated and abstract practices like programming may be pursued once specific goals emerge from the players themselves. (Games 2008, 136)

In other words, players first learn a culture of professional game design practice, and then learn skills later.

GameStar Mechanic aims for two basic types of learning: high-order reasoning and analysis, and the craft/skill of video game design (E-Line Media 2012). The first learning goal is broken down into several subtypes: "Systems Thinking" is how the developers describe a high-level design reasoning that includes comprehension and creation of dynamic systems, "a characteristic activity in both the media and in science today". "Interdisciplinary Thinking" focuses on ability to synthesize knowledge from different domains in meaningful ways. "User-Centred Design" spans the higher-order thinking and the design skills: "Students act as sociotechnical engineers, thinking about how people interact with systems and how systems shape both competitive and collaborative social interaction." Ancillary skills are also claimed, including "Specialist Language" (technical terminology and semiotics) and socially mediated analytical reflection: "Students learn to explicate and defend their ideas, describe issues and interactions at a meta-level, create and test hypotheses, and reflect on the impact of their solutions on others" (E-Line Media 2012).

The second learning goal, design skill, is clearer. The lesson plans list rules, game space, core mechanics, goal creation, balance, playtesting and other ideas commonly seen in professional game design texts. For example, developers provide a detailed curriculum for in-class teacher-led exercises. Lesson 6.1 does a good job of exploring the true nature of rules. In class, teachers invite students to play, then modify "musical chairs," an effective method commonly used in leading professional game design workshops (Fullerton, Swain, and Hoffman 2008). By contrast, the in-

¹⁹ This is perhaps unsurprising because Gee was named as an advisor on the project

²⁰ The surname of this author, Ivan A. Games is, potentially confusing in this context. This footnote reassures the reader that, yes, Games is the last name of a researcher of games.

game tutorials (called “quests”) seem intended to be student-led, standalone teaching tools, intended to both teach and encourage game players to be designers.

5.3.1.1 My Play Experience

In addition to playing the game myself, completing the tutorials and building a few games, I observed my six-year-old son progress through the first half of the tutorials and build two games. I also watched self-narrated walkthrough videos by "TheSnivy10," a YouTube username for someone I would guess was an English boy about ten years old (TheSnivy10 2012), as well as several other similar videos (*How to Use GameStar Mechanic (from a 7th Grader's Perspective)* 2012; ARimeth 2011). Finally I reviewed the content on the game's website, and several academic publications that discuss the game (Games 2008; E-Line Media 2012).

A large part of *GameStar Mechanic* is what I call “tutorial mode.” A series of brief games are embedded in linear progression through a fictional storyline that aims to teach principles of level design. The designers also provide classroom lesson plans. Neither my son nor TheSnivy10 showed any real interest in the learning moments within the quest; they both clicked past the instructional screens as fast as possible to play the games. They both displayed strong moments of excitement and engagement around elements typically seen in entertaining video games, such as player skill progression and violent power fantasy (C. Crawford 1984). For example, at one point TheSnivy10 vocalized: “I have a gunnn. I have a gun so I have one thing to say to you: you want to kill me? Well, DIE IDIOT!” 2:00). While this quote was not as shocking in context as it appears in text, and was not typical of the rest of his dialog, it is consistent with the broader point: the pleasure players found in *GameStar Mechanic's* quests was often, even usually, unrelated to the learning purpose of the quest. Given that the apparent intent of the tutorials is teaching game design, is the lack of engagement with the learning component a problem? I believe it is not a huge problem in the context of the long-term learning goal, especially if one considers the possibility that an intervention may use both extrinsic and intrinsic motivations, and these are not always in opposition or conflict with each other (Vallerand et al. 1993). A common problem in learning games is insufficient player engagement, so any sign of player engagement during a tutorial is a good sign. Also, observation is not infallible: the player may not indicate, or even realize, what they are really learning (Ambinder 2009). It may be that unobservable learning is occurring, and it certainly means the game has engaged the player, which may be sufficient for later learning to occur.

I observed that the tutorial's content was humble in its learning goals, especially when compared to the lesson plans. For example, Quest 3 introduced the idea of game rules. The supportive text suggests rules are complex and important, but the game did not afford the freedom necessary to

grasp the meaning of rules. It only allowed the player to change one number (the time limit). The classroom exercise, by contrast, had the student add a new rule of their own creation, which is a far freer task. I observed players completing the tutorial and realizing that they can change a game, but failing to grasp the broader point: that "rules" are a primary form of creative choice that game designers employ. I have sympathy for the designers, as I cannot conceive of how the broader point might have been incorporated into the game. Was this learning goal too ambitious? Or perhaps the learning goals of the game design and the in-classroom experience were actually different?

The tutorials are challenging in that they require significant game-playing skill to succeed, and there is no way to skip ahead. This requirement has two drawbacks. It requires a level of game playing skill which prevents some players (e.g. people with muscular dystrophy) from progressing. Second it increases the time necessary to experience the tutorial, creating a crucial barrier for teachers. Elsewhere the designers argue that it is important for teachers to actually play through the quests, so they know what they are assigning (Klopfer, Osterweil, and Salen 2009). While one self-described teacher welcomed the challenge of beating the *GameStar Mechanic* quests because it gave them both an improvement in game playing skills, and the motivation to continue (Anon. 2011,), I have not observed this enthusiasm and willingness to spend time playing games to be a common attitude among the teachers I personally know, and conflicts with my reading of the literature.

By requiring player skill to progress, *GameStar Mechanic's* design sends a message to their audience: you must have player skill to be a designer. Is this true? My reading of the literature leads me to believe that it is helpful for both teachers and designers to have sufficient player skill to playtest and gather useful insight into play experience nuance; however, mastery is certainly not required, high player skill is not as important as other design skills, and the skill of designing games is very different from that of playing games.

Another important element of the game is "sandbox" mode. I use the term "sandbox"²¹ to mean that the game does not impose win/lose conditions but enables players to create their own experience, usually through assembling supplied elements. I observed players assuming designer identities in their relation to this mode. At 2:05 TheSnivy10 invites the viewer to play the games he built: "If you are looking for an easy game, play my game LeQuest Part 1." TheSnivy10 then states they were "fun if you want a really easy experience" and "perfect for beginners" (2:22) which is a mature sign of any designer: able and willing to both imagine what would be fun for other types of

²¹ This definition of "sandbox" fits with commonplace usage in commercial game literature (Schell 2008) but learning game researchers use the term somewhat differently (Squire 2007).

players beside himself, and to actually build it. This is a clear example of user-centred design, one of the four stated learning goals, and also seems to fit Gee's goal of "situated knowledge" (Gee 2009b).

GameStar Mechanic enables players to form an online community around the games they build. This is a socially mediated design feature, which has been described as beneficial by many theorists, including Salen. While it can provide tremendous value, social communities can also have negative consequences for young members. I observed one example of the negative influence of peer review in *GameStar Mechanic*. TheSnivy10 said he never made Part 2 of his game *LeQuest* because of "mean comments" about his first game (TheSnivy10 2012, 2:14). This statement conflicted with my observation of the community, as a whole. I read twenty comments, chosen semi-randomly as a convenience sample and got the impression of a supportive community. Typical comments for a top-ranking game are shown in Figure 26.

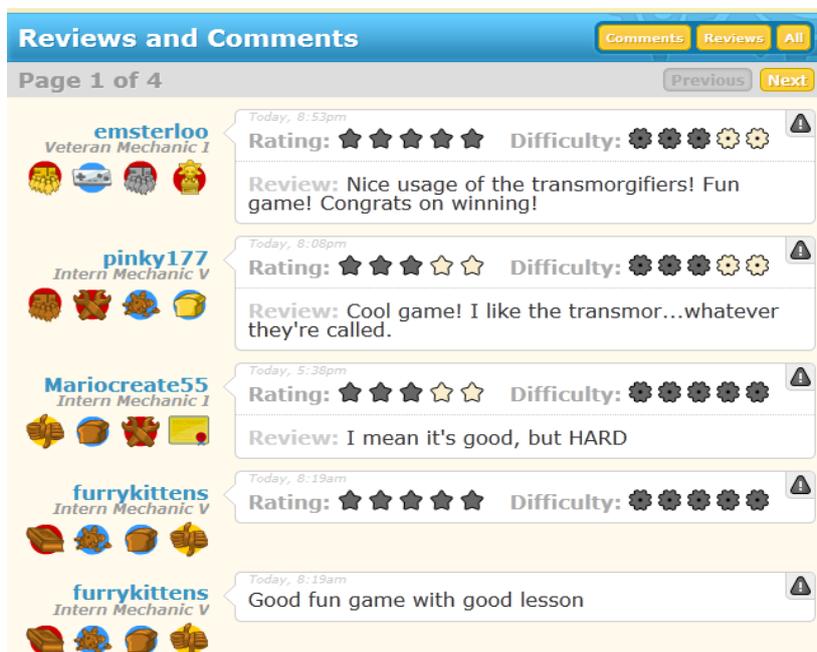


Figure 26: example player feedback on user-created *GameStar Mechanic* game (<https://GameStarmechanic.com/game/player/1022261>).

I observed designerly thinking in some of the comments. While many comments were bragging about player skill "I beat this!" "I'm so proud this level was sooo tough," some comments addressed design issues. The design critiques were very brief: "great game" "nice" "too easy".

Players must earn tools to build levels through the tutorial. I observed this requirement frustrated and limited sandbox mode users who want full access to a variety of sprites to build the most compelling game possible (TheSnivy10 2012). I believe the designers knew this but decided the benefits of this approach outweighed the frustration. Should sandbox games require progression to

access sandbox tools? There is no literature that addresses this question that I am aware of. Some commercial games do so, while others do not. *LittleBigPlanet*, for example, restricts the player to "the tools, stickers and decorations collected from the Story mode, to build your own levels, toys or gadgets" (Sony Computer Entertainment Europe 2012). *Minecraft's* Creative Mode allows players to access every asset in the game, regardless of what they have discovered in Survival Mode (Minecraft Wiki 2012). Having experienced both approaches myself, and observed my son and six other children play *Minecraft*, I personally believe sandbox games should provide unfettered access to all tools, but I acknowledge that other approaches can work as well.

Another example of the lockdown approach in *GameStar Mechanic* can be seen partway through the tutorial mode. After the edit mode is introduced, the *controls* that would allow a player to modify a level to be easy are disabled. This seemed to be intended to enforce a single path of progression through the quest. To me it felt as if the designer lost sight of the larger learning objective²².

5.3.1.2 Activity-Goal Alignment Score for *GameStar Mechanic*

Overall, I struggled to score *GameStar Mechanic*, as both the game's activities and the range of learning goals are very broad. I changed each goal's position several times before I was satisfied. In this section I present both my final score and my reasoning.

For the Context section, I wrote: "A 9-year-old boy with typical middle-class gamer motivations including power fantasy, challenge, and status. He enjoys creative activities outside of gaming - Lego, building toy models, and is attracted by the "build your own game" message in the marketing—that's why he started playing. He plays alone, with parent in the house but not nearby, on a Windows 7 laptop. His sessions are 30 to 120 minutes long, once a day."

The diagram requires a single "core mechanic" or primary activity to be identified. *GameStar Mechanic* has many such components and none are clearly "core." To address this problem, I considered several approaches. First, I created a series of diagrams that represent each mechanic in the game. There were so many diagrams that the score was more confusing than clarifying. Secondly, I tried picking the most important mechanic but found this excluded too much of the player activity. Finally, I attempted to merge all mechanics into a super-category, but found the

²² Should games about game design prevent cheating? I and many industry-leading professional game designers I know personally started by building "cheats" for games they loved. They quickly grew bored of this and began to improve, not cheat, games. Preventing cheating makes sense in multiplayer games where cheaters ruin the experience for others but this is a single player game. If a player understands the tools well enough to cheat, hasn't the game achieved its learning goal? Perhaps, rather than reducing player choice, the game could detect and shunt clever cheaters towards tutorials that are designed for subversive players? Although fascinating, answering these questions is beyond the scope of this research.

resulting “mechanic” was too vague to assign to a learning goal. In the end, I felt the game’s activity-goal alignment is best expressed with two diagrams: “tutorial” and “sandbox.” These terms were discussed in the preceding section, and summarized in the diagrams. I grouped the community aspect into “sandbox,” since the primary focus of conversation seemed to be player-created games.

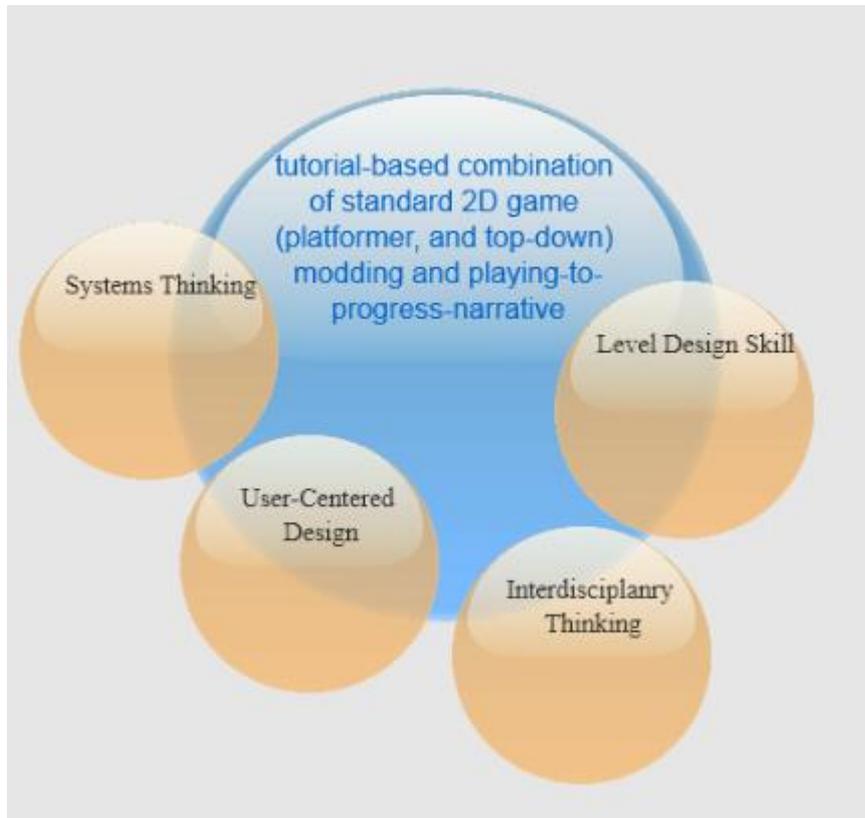


Figure 27: Activity-goal alignment diagram for "tutorial" play mode of *GameStar Mechanic*.

Referring to the “tutorial” play mode shown in Figure 27, I placed the “Systems Thinking” goal to weakly overlap the “tutorial” activity for the following reasons. While *GameStar Mechanic* does teach some system thinking, it is my view that the limited options for modification in the narrative-driven tutorials prevent a complete or thorough understanding of systems thinking. However, the game does introduce a few key concepts that I would regard as “systems thinking” in the tutorials, so I rated the “Systems Thinking” goal as partially aligned. I placed “Interdisciplinary Thinking” partially aligned with the tutorial mechanic for similar reasons. I could not find a place where the designers clearly define how Interdisciplinary Thinking occurs. The closest I found is a statement that students “become intelligent and resourceful as they learn how to find and use information in meaningful ways” (E-Line Media 2007), which does not exactly match my understanding of the term Interdisciplinary Thinking. *GameStar Mechanic* does require persistence, but that is not really interdisciplinary thinking. Players cannot be resourceful in creating their own tiles, which enable

free-form design, because tiles must be unlocked by completion of designer-imposed tasks.

GameStar Mechanic's tutorials are too heavy-handed, single-purpose and linear in structure to align perfectly with my understanding of Interdisciplinary Thinking.

I set Level Design Skill as mostly aligned, since the tutorials are clearly designed to teach the player to build good levels. It assesses players in their skill quite cleverly. However, many aspects of level design (e.g. sound, multi-player, camera movement) are not addressed, and the limitations of the underlying engine greatly limit the types and range of level design skill the game can teach. I partially aligned User-Centred Design because game design is almost inherently a user-centred design activity, and *GameStar Mechanic's* implementation allows for a very rich, in-depth experience with user-centred design. One such example was clearly documented in Games' dissertation (Games 2009, 162), when a player explains how he changed the game to improve the play experience. However, when considering only the tutorial mode as this diagram does, there is not a lot of teaching or assessing of User-Centred Design principles or practice.

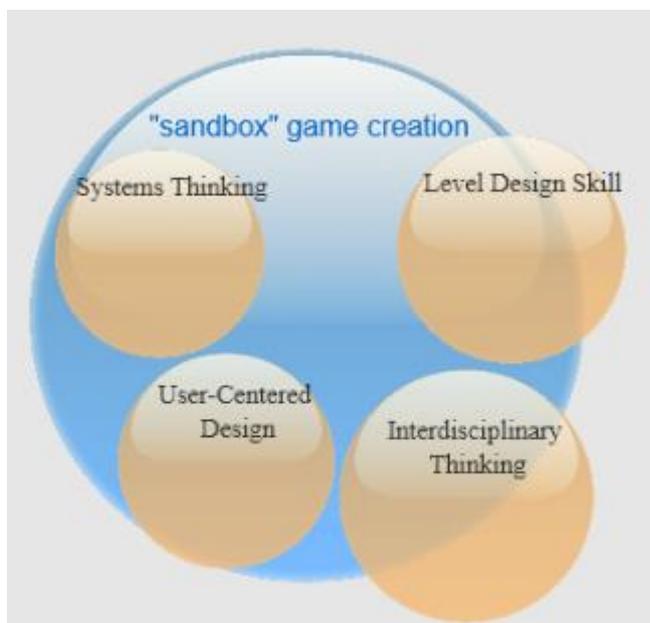


Figure 28: Sandbox mode of *GameStar Mechanic*. This diagram shows its activity-goal alignment.

Figure 28 shows sandbox mode of *GameStar Mechanic*, where players create their own games with no guidance, using tiles they earned by progressing through the tutorial mode.

I think the Systems Thinking goal is very well aligned to the activity of free-form game creation. The act of building a game strongly relates to considering the game experience as holistic.

The Level Design Skill goal is strongly, but not completely, aligned with sandbox creation because while one obviously must build levels to build these games, the skill of level layout is not always the focus of the design activity. Building a game is more than learning level design. I felt interdisciplinary thinking is also mostly, but not fully, aligned, for similar reasons.

Of all the games scored in this chapter, this score feels the least solid to me. The combination of the extremely high-order learning goals, the diverse range of activities, and the broad, complex game design make it difficult to confidently assess the activity-goal alignment even using two diagrams of primary play activity. Nevertheless, I feel this activity-goal alignment score is a reasonably correct assessment of my view of *GameStar Mechanic's* activity-goal alignment.

5.3.2 "Understanding Games"

The second game assessed in this section is "*Understanding Games*," an award winning learning game that teaches game design. A screenshot is shown in Figure 29. Its creator describes it as

a series of four games explaining the basic concepts of video games. The tutorial-style episodes deal with rules, motivation, learning and identification in video games. The player is guided through each episode by the narrators Bob and Bub, who explain core concepts of games to the player. The player can experience these concepts directly while playing the integrated games. (Zecher 2008)



Figure 29: Players learn how to design by playing simple games, then learning what that play experience means from text. (Zecher 2008)

5.3.2.1 My Play Experience

This game felt more like a lecture than a game to me. I thought of the interactive elements as “mini-games” - illustrations of mechanics, not complete games²³. The learning occurred in the discussion, which was presented as text on screen, and was not interactive (Figure 30).



Figure 30: Most of the player’s time is spent reading explanatory text in *Understanding Games*.

While the overall application was reasonably effective as a teaching tool, I felt it could have been nearly as effective (but not nearly as engaging) if it was redone as a PowerPoint slide show.

However, I would never have looked at it if it was another slide show, so I recognize the novelty of the format has value. The designers’ key contribution was embedding interactive experiences within an online lecture. The experience of playing a familiar game mechanic within a lecture was a useful convenience, allowing me to refresh my memory and allowing the designers to stop and start the game experience to illustrate various points.

5.3.2.2 Aga Scoring Tool Analysis

Here is the result from AGA Scoring Tool v7:

For Context, I wrote: “An 18-year-old male high school student, middle class, studying game design in a multimedia class at a public high school, is required to play this game for homework. He plays at home on a laptop. His parents are in the house, not involved. He starts and completes the game in fifteen minutes. It is not an engaging game experience. He reflects on rules and interaction more deeply than if he had read them in a book.”

²³ A “mini-game” (Chow, Susilo, and Zhou 2010) is a short game contained within a larger game. This short game has a single, clear mechanic and extremely simple controls. “click on the dancing rabbit” is a mini-game

Diagram: The scoring diagram is shown in Figure 31.



Figure 31: Activity-goal alignment diagram for *Understanding Games*.

Description: I wrote: “The core mechanic is progression through the tutorial, reading text and pressing spacebar. The various mini-games intentionally use simple mechanics: Pong, Tag, etc. (Is this really a game? There is no win/lose; it feels more like an e-learning app. But, since its author calls it game, and it has embedded mini-games, and it has won awards as a serious game, I’ll consider it a game.)”

This description suggests that a “game about game design” might sound highly aligned, but activity-goal alignment theory would suggest an alternate view: the learning goal is designing games, but there is no opportunity to design games.

5.3.3 Conclusion from Game Design Discussion

In this section I examined two games that have very different core mechanics to achieve the same learning goal, to see if the AGA Scoring Tool would be useful in this type of comparative analysis.

I was surprised by my own activity-goal alignment score for *GameStar Mechanic*. I had chosen it because I imagined it would be a clear example of high activity-goal alignment score. On average the game did score highly, but by applying the tool I was able to penetrate the complexity of the game to find that several learning goals were only partially aligned with certain activities.

For the second game, "*Understanding Games*," I was unsurprised that I scored it low, since prior to completing the tool, I had realised its activity and goal were not aligned. I was interested to find that it was easy to apply the AGA Scoring Tool to such an outlier in the spectrum of interactivity in video games, especially since Shearer discussed how the lack of interaction in similar pseudo-games made it difficult to fairly assess them (Shearer 2011).

5.4 Discussion

How did the AGA Scoring Tool perform, compared to Shelton's textual approach? I discuss this crucial point with particular reference to my earlier discussion of *Re-Mission* in section 2 of this chapter.

In this chapter I found a clear answer to one of central aims of this thesis: **the AGA Scoring Tool was useful to me as a learning game designer/researcher**. Scoring activity-goal alignment focused my investigation, yet I was able to reflect on the game experience holistically and in detail freely, but my attention stayed squarely on topic. I believe this was partly due to the script-like nature of a form which kept me in a reactive analytical mode: rather than wandering in my attentions, exploring multiple aspects of the entire game experience, I had a process to follow.

As a designer, I am suspicious of processes and procedures, but I found that this tool was not so prescriptive that it suppressed a holistic designerly approach. For example, the diagram allowed me to resize the learning goals, but did not define what the size meant. This lack of definition was useful to me, as it caused me to consider what attributes the learning goals shared. This consideration led me to realise that I felt the learning goals were not equal in importance. Thus the tool helped me realise something important that was beyond what the process suggested. Since I had failed to articulate this in my freeform analysis, I found this tool useful indeed.

The diagram also let me overlap one learning goal over another. Although the tool does not dictate the meaning of overlap here, I believe viewers of Venn diagrams assume that overlap has meaning. I overlapped learning goals to suggest some shared element between them. For example, I overlapped "treatment adherence" and "exponential reproduction" as way of visually suggesting that these two goals were interrelated. I felt it was necessary to state my intention of the meaning of the goal overlap in the description text part of the tool.

In my previous experience using this tool during development, I noticed it did not fairly summarize every aspect of the game. In comparing *Re-Mission's* activity-goal alignment score to the textual

discussion I wrote, I observed this lack of comprehensive coverage again. The textual discussion is more thorough and custom-built than the tool's score, capturing important insights unique to the game and the experience that no semi-automated diagram can capture. In that way, it is clear that the AGA Scoring Tool cannot completely replace Shelton's text-based approach. However, although I lost comprehensive coverage, I gained wonderful clarity.

I also felt the **AGA Scoring Tool helped me stay objective** in my analyses. The diagram forced me to summarize my own thoughts, and I found I could not justify any more than moderate activity-goal alignment score for *Re-Mission*. This surprised me because after writing the textual analysis, I was left with a favourable impression of the game, so I had imagined I would rate the game much more highly than I actually did. The tool helped me reflect on all the drawbacks as well as the strengths of the game's activity-goal alignment, which reduced the bias that might have come from my feeling toward a game at a particular moment in time.

I found a limitation of the AGA Scoring Tool in scoring *Re-Mission*. I scored one learning goal, "perils of exponential reproduction," at 100%. While the tool did clearly illustrate the perfect alignment I observed, and I was able to show my feeling about the relation between treatment adherence and exponential reproduction, I was unable to illustrate that this goal was a belief, not a thought. The tool does not allow the designer to distinguish between types of outcomes (belief, behaviour, factual knowledge, and attitudes). In designing the tool I decided not to attempt to represent these differences, in the interest of clarity, but in applying the tool in this case, I felt the lack as a disadvantage. I discuss this issue further in the final chapter.

5.5 Summary of Findings

The two conclusion sections in this chapter, as well as the preceding section, provide the basis for the findings discussed here.

5.5.1 The AGA Scoring Tool improves the designer/researcher's ability to quickly, holistically and accurately perceive a learning game's activity-goal alignment.

I found that the AGA Scoring Tool was effective in focusing my investigation on the specific question: what is the observable activity-goal alignment in a learning game? I was able to reflect on the game experience holistically and in detail freely, but my attention stayed squarely on topic. In addition to the speed benefits discussed below, I found the tool was flexible enough to allow me to analyse a variety of games, some simple and some complex. When I compared the tool's output (the diagram

and text) to my textual discussion, the tool's conclusion reasonably matched my intent in the textual discussion. From this, I conclude that the AGA Scoring Tool is generally useful to a learning game designer/researcher, either as a stand-alone tool or in conjunction with a more detailed textual analysis. This finding addresses the first research question of this thesis, but it is an insufficient answer without the caveats and explanations in the following findings.

5.5.2 Activity-goal alignment scores are not as comprehensive as textual analysis.

I found the limitless, flexible nature of textual discussion is something the activity-goal alignment score method lacks. Textual discussion was clearly more thorough and custom-built than the tool's score, capturing important insights unique to the game and the experience that no semi-automated diagram could capture. There was no need to limit the discussion to the confines of activity-goal alignment theory. For example, the AGA Scoring Tool does not allow the designer to distinguish between types of outcomes (belief, behaviour, factual knowledge, and attitudes). I discussed how one might be able to redesign the tool to address this specific point, but I feel the problem is more fundamental. The idea of scoring activity-goal alignment necessitates a lot of simplification, and something will always be lost in that process. In this way, I feel it is clear that scoring activity-goal alignment is not general-purpose: it cannot completely replace Shelton's format, or any text-based approach, for every type of discussion of activity-goal alignment, and certainly does not substitute for a broad consideration of learning game design.

This negative-sounding finding does not mean scoring activity-goal alignment has no value. I see it as one useful tool among many. I hope it is especially useful in the production context, where the production team already has a good understanding of the unique qualities of the game, but is in danger of missing the forest for the trees during the design process. This danger was noted also by Shelton, as discussed in chapter 2, so I hope this research usefully contributes to Shelton's work.

5.5.3 The activity-goal alignment score is a qualitative measure: a "rule of thumb"

It is apparent that this tool is a "rule of thumb" for practising designers. It would likely be a waste of time to improve the precision of an activity-goal alignment score. Consider a well-known rule of thumb is the "rule of thirds," which states that picture composition is often more interesting if the subject is approximately one-third of the way from the edge of the composition. "Approximately" is the key word in the previous sentence. This "rule" is intended only as a rough guide, to help amateur composers break the instinctual habit of centring the primary subject in every composition. It is unimportant if the subject is 30% or 34% across the horizontal axis of an image. However, it is a "rule" that should be broken: other factors are usually far more important, so the

rule is only meaningful as a rough estimate. However, I have observed that amateur designers are sometimes tempted to use measure precisely when placing their subject according the “rule of thirds.” To measure for the rule of thirds is to miss its point.

Similarly, the scoring of activity-goal alignment is a rule of thumb. Each game’s needs, constraints, and play context is too unique to make anything more than a rough estimate of alignment meaningful. Shelton’s discussion clearly suggests activity-goal alignment theory is a structured lens for a design decision, not a quantity to be precisely measured. The tool’s modality is confusing, I fear. Unfortunately for this research, Venn diagrams are often used for measuring factual data in other fields where precision is very important, so their very nature suggests that precise measurement is not only possible but might be a reasonable aim. In design language, the Venn diagrams erroneously afford the possibility of measurement. I fear the mere suggestion of scoring activity-goal alignment tempts designers try to measure it more accurately, just as an amateur photographer might erroneously employ tape measure when first trying to follow the rule of thirds. I feel such an effort would be unlikely to generate meaningful results.

5.5.4 Scoring Activity-Goal Alignment with a Tool is both Faster and has Quality Advantages over a Textual Discussion

In chapter 2 I discussed how existing methods of understanding activity-goal alignment employ textual discussion. In scoring activity-goal alignment via textual discussion, I found the mechanics of capturing my reflection as textual discussion while trying to recollect and reflect was an important barrier to quality in my design practice. I had to write some of my thoughts down hastily, reflect partially, write down new thoughts, reread the initial thoughts and rewrite in light of the following thoughts, then try to merge and integrate. This activity was reasonably effective in that I did emerge with a discussion that captured important parts of the activity-goal relationship. However, I noted room for improvement: the mechanics of writing required time and mental effort sufficient to interfere with my design activity. A key quality of the state of Flow is a seamless, continual blend between reflection and action (Csikszentmihalyi 1990). Unlike my text writing, I entered something like a mild Flow state while using the tool to score *Re-Mission*’s activity-goal alignment, and I believe this state was enabled by the tool. Unlike a textual discussion or a hand-drawn Venn diagram, the AGA Scoring Tool enables simultaneous reflection and revision because it is so quick and easy to revise. It takes a single click-and-drag, which I could do in less than one second, to restate my position on the game’s activity-goal alignment. As a result, I was able to move circles as fast as I recalled and reflected. By comparison, it took tens of minutes for me to convert a thought into a revision to my textual analysis. In short, this tool made me realise that writing is too slow for an ideal

process of recollection and reflection necessary for comprehensive perception of a game's activity-goal alignment, and reasonable alternatives, such as this tool, exist.

5.5.5 Activity-goal alignment is a phenomenon that emerges during the act of playing a game.

Activity-goal alignment is based on player activity. As it depends on the player's decisions during play, similarly to many other theories in game design such as Flow (Csikszentmihalyi 1990) or Hard Fun (Lazzaro 2005), activity-goal alignment is at least a little different every time the game is played. I feel it would be difficult to judge a game's activity-goal alignment using only its artefacts (the code, artwork, instructions) because there is too little basis for understanding player experience. Recordings of the game play experience (text description, gameplay video, or in-person recounting) may be sufficient sometimes. Playing a game oneself is especially powerful approach to scoring activity-goal alignment because it is only then that a complete understanding of the play context, which is so essential to understanding activity, is possible. Since autoethnography enables such an understanding, as I discussed in chapter 3, autoethnography is well suited to scoring activity-goal alignment. However, since the designer is not always the target audience, a combined approach such as I have used here, seems appropriate. More well-resourced researchers might employ more elaborate approaches to scoring, such as recruiting larger samples of target audiences for observations of play, to further improve their scoring of activity-goal alignment. In any case, one must approximate activity with a succinct statement to give a single score of activity-goal alignment for a given game. This must be done with the understanding that activity is at least a little different for each and every play session.

In other words, because activity-goal alignment is unique to every play session and player, one must carefully consider audience and context when scoring activity-goal alignment.

5.5.6 A Game May Need Multiple Activity-Goal Alignment Scores.

It may not be meaningful to assign a single activity-goal alignment score for a learning game, because some learning games are too complex to reduce to a single primary activity. One example discussed in this chapter, *GameStar Mechanic*, has at least two different main activities: building games and following tutorials. The AGA Scoring Tool can only assess one activity at a time. Thus, to score multiple activities, multiple activity-goal alignment scores are needed.

Moreover, summarizing or averaging the activity-goal alignment scores across multiple learning goals can easily fail to yield a meaningful score. For example, consider a hypothetical video game named *SoccerGame*, designed to teach real-world soccer skills. It might have high alignment for strategy (player reasoning for field position is largely the same between video game and real-world

soccer) but very low alignment with motor skills (it is hard to practice dribbling by clicking a button). Is it meaningful to say anything overall about *SoccerGame's* activity-goal alignment? I argue that it would be relatively meaningless to say "In total, *SoccerGame* has 50% activity-goal alignment" because the learning goals are too unrelated: combining the learning goals of strategy and motor skills doesn't make sense. Instead, I would suggest designers consider each learning goal separately in this case. However, I found in this chapter that sometimes the learning goals are related enough to make a combined score have some meaning, as with *Re-Mission*, even though important differences remain.

5.6 Conclusion

In conclusion, through discussion of five activity-goal alignment scores, this chapter has addressed one of the research questions in this thesis by concluding that scoring activity-goal alignment is generally useful to the learning game designer/researcher.

6 Review of Learning Game Design Assessment

6.1 Introduction

The following two chapters aim to extend activity-goal alignment theory to address assessment of learning games. In chapter 5, when I assigned an activity-goal alignment score for several games, I noted that some learning games are widely agreed to be excellent yet had low activity-goal alignment scores. Clearly, scoring activity-goal alignment alone is insufficient to assess a learning game. This chapter addresses the need for improved methods of game assessment I argued for in chapter 1, exploring how scoring activity-goal alignment might be used to improve assessment of learning games.

This chapter begins with a conundrum: For twenty years, top scholars have argued for improvement in learning games, yet some modern learning games do not follow this advice. I regard this situation as a problem, and argue that improved methods of assessment might help. I criticise existing approaches to assessment, finding that the typical modality (checklists or essays) might be improved upon. In particular I discuss Gee's principles of deep learning: I find these excellent in a text discussion, but these principles do not clearly differentiate deep learning games from other types of games when used as an assessment checklist. Finally, I develop criteria for a new assessment tool that might help learning game designers avoid reinventing the wheel.

6.2 Prevailing Advice in Learning Games: Aim Higher

In chapter 2 there was some discussion of prevailing advice from the game design literature. Here I argue that experienced learning game academics often exhort researcher/designers to aim higher in their choice of learning goals. Regarding math games²⁴, for example, Devlin (2012c) has argued three key points in his series of blog articles:

1. Most math games today provide only repetitive practice of basic computational skills,
2. Such low-level learning goals are the “low hanging fruit” of math games, and
3. There are already many such games, so there is no need for more.

²⁴ I use the term “math games” in this chapter to refer to learning games that aim to teach any aspect of mathematics, from basic arithmetic to advanced concepts

Devlin expands this last point into strongly worded advice: “What distresses me is that the medium [learning games] offers so much promise for good mathematics learning, it is **a waste of time, effort, and money to focus on** the lowest level — **repetitive practice** of the basic, procedural, computational skills. We’ve done that. Let’s move on”²⁵ (my emphasis). Devlin then focused attention on higher learning goals:

The good news is, since repetitive practice of basic computational skills is a tiny part of learning mathematics— albeit an important part, in my view (some educators disagree)—most of the fruit in the math [education] orchard is still waiting to be picked. The bad news is, that fruit is a lot higher up, and thus more difficult to reach. (Devlin 2012c)

Devlin believed that math is more art than science. He argued that we need to shift emphasis in math education away from calculation, just as mathematics scholars shifted emphasis from repetitive calculation to conceptual understanding in the late 1800s.

While Devlin is focused on mathematics, similar advice can be seen throughout other areas of the education literature, even within “strong science” areas such as chemistry. For example, the American Chemical Society called for a shift away from the traditional procedural approach towards the Constructivist approach²⁶ to chemistry education (ACS 2009). The branch of cognitive science known as “learning science” commonly uses the term “deep learning” to describe this general position (R. K. Sawyer 2005), as I discussed in chapter 2.

Gee noted a general lack of deep learning games: “The creation of deep serious games for such deep learning remains today more a hope for the future than a realised possibility, although there are intriguing beginnings here and there” (Gee 2009a, 65). Similarly, Squire and Jenkins implied that game designers should aim higher than digital worksheets: “The wave of educational games released in the 1980s and early 1990s largely ran counter to prevailing educational concerns” and were criticized for “dumbing down” for educational audiences (Squire and Jenkins 2003). This statement fits with Robertson's critique of gamification (Robertson 2010).

Where might designers turn for an authoritative guide to deep learning? I suggest Gee’s six principles of deep learning is a good choice. Games that “really use the power of gaming” possess

²⁵ Devlin’s phrase “lowest level” reinforces the concept of putting learning goals on a spectrum from lower to higher order, as I discussed in the literature review chapter. I apply this idea to assessment of learning games in Chapter 7.

²⁶ the Constructivist approach to education rejects the “vessel” idea and instead posits that students construct their own knowledge, as I discussed in chapter 2.

certain qualities, argued Gee (2009a). Gee gave six required properties of deep learning, in the form of questions:

- *Property 1: Does game play allow and encourage the player to “psych out” and take advantage of an underlying rule system to accomplish personally held goals to which the player is emotionally attached?*
- *Property 2: Does the game allow the player microcontrol that creates either a sense of embodied intimacy or a feeling of reach in power and vision?*
- *Property 3: Does the game offer the player experiences that meet the conditions for good learning (discussed above)?*
- *Property 4: Does the game allow, encourage, and help players find and use effectivity–affordance matches between smart bodies or tools and worlds?*
- *Property 5: Does the game use modelling or models to make learning from experience more general and abstract?*
- *Property 6: Does the game allow and encourage the player to enact his or own unique trajectory through the game, thereby creating his or her own story? (Gee 2009a, 78)*

Gee discussed these same six principles in some detail.

Gee’s position fits with activity-goal alignment theory. Although he does not use the term “activity-goal alignment,” Gee’s definition of effectivity-affordance match suggests compatibility with Shelton’s approach. Gee refers to “models to make learning from experience.” The term “model” refers to small simulations within the game. Unlike simple learning games which present facts without any kind of simulation, Gee argued that models provide experience, not memorisation of facts, and that this type of game design is necessary for “deep learning”²⁷.

²⁷ It is crucial to grasp the primary purpose of any learning game when considering how models or simulations relate to the idea “deep learning”. There are many ways to incorporate simulations or models into a learning game design, and sometimes some simulations are not the primary activity. For example, a chemistry game might have the player role-play a chemical engineer, and interact with a simulation of the physics chemical bonding only occasionally, such as during a workbench experiment. Does such a game have high-activity goal alignment and/or incorporate “deep learning”? The answer depends on the learning goal: If the game’s purpose is to advocate a career, one might make the case that the game simulates the daily activities of the chemical engineer. Thus the game might still possess Gee’s “deep learning” even though the physics model is not used often in game play. It can be difficult to identify the primary learning goal of the game, as there are often many unrelated goals. I provide several worked examples of identifying the primary learning goal of complex learning games in chapters 5 and 7.

Gee's implicit advocacy of high activity-goal alignment is perhaps most apparent in his third property, which he says lies at the foundation of how games recruit good learning: "Games put players in worlds where they experience things." Gee argues that experiential learning is a superior theory to traditional cognitive science models that suggest the brain is like a digital computer. As I have argued in chapter 4, experiential learning is the crux of high activity-goal alignment.

In chapter 2 I noted Rice also argued for experiential learning, but unlike Gee, Rice specifically focused on virtual presence. Gee's difference in position is apparent in his third principle, which presented the concept of "microcontrol." Gee said microcontrol means that "the player can affect the movements and actions of that element or elements at a fine-grained, detailed level." Gee explains that microcontrol is related to the 3D virtual worlds, which Rice argued are essential (Rice 2007), but flips the tables: Instead of building the computer experience around the physicality of the human body, Gee said in commercial games "the player feels that his or her body has extended into and is intimately involved with the virtual world" (p. 68). This subtle but crucial difference can be understood by example. Gee uses the internet-controlled watering can: "if a person is using a web cam to manipulate a watering can in a fine grained way to water plants in another country, the person feels that his or her body has, in a sense, extended to that other country." Gee's point matched my own experience designing *Battlefood*: the alignment between the player's action in the game and the learning goal intended is high in games that afford microcontrol. This correlation also fits with the broader prevailing view of intrinsic motivation in the learning game design and psychology literature, as I discussed in chapter 2.

It is not just microcontrol that fits with activity-goal alignment. The last three of Gee's principles of deep learning suggest that a game with a high activity-goal alignment score is positively related to Gee's concept of "deep learning." In addition to Walker and Shelton's argument to that effect (2008), my personal experience in learning game research leads me to conclude that games with high activity-goal alignment embody Gee's deep learning principles more than games with low activity-goal alignment. Gee suggests modelling allows the player to experience objectives in their own way, and narrative overlays objectives with many kinds of meaning to the player. In both of these statements, I understand him to be assuming that a game's features should be designed to relate the mechanic with the outcome. In essence, I see Gee working from a foundation of activity-goal alignment.

The debate around gamification is relevant to this discussion. Although both sides have cited Gee, perhaps due to his credibility, I feel that the arguments that oppose gamification typically fit Gee's position best. For example, Bogost cited Robertson in his earnest statement that "gamification is

bullshit” (Bogost 2011b). Gee’s approach fits with Bogost’s and Robertson’s position much better than those who advocate gamification (Bunchball Inc 2011). Robertson contrasted games with gamification by emphasising meaningful choice:

Games give their players meaningful choices that meaningfully impact on the world of the game. Deciding to run two miles today rather than one, or drink two liters of Coke instead of four are just choices of quantity. Deciding to dump my sniper rifle for an energy sword is a meaningful choice. It’s going to change how I move, who I fight, when I run. It’s literally going to change whether I live or die. (Robertson 2010)

In other words, for games to rise above gamification, they must have a meaning-making activity that is closely related to the game mechanic. Likewise, activity-goal alignment can be seen in Bogost’s work: Bogost stated gamification “fails to embrace the complex responsiveness of ‘real’ games, games that make solutions seem interestingly hard rather than tediously so” (Bogost 2011a). In summary, gamification opponents generally argue that both high activity-goal alignment and deep learning are important to excellent learning game design²⁸.

In conclusion, I generally find that it is the prevailing view in the learning game literature that researcher/designers should not design more of the same “digital worksheet” games that exist today. Instead, they should aim to build deep learning games by setting high-order learning goals and aiming for deep learning. I specifically cite the term “deep learning” because I generally find Gee’s position convincing. His logic is clear and his citations are well defended from a variety of fields, as one might expect from a scholar of his stature.

How might Gee’s ideas be best applied in practice? Gee is clearly a masterful writer: he hits a useful compromise between detail and breadth of coverage in the length and nature of his attributes. However, in the following chapter, I argue that amateur learning game designers can easily misinterpret any academic writing, and suggest the problem might be modality, not the content or quality. There is one important variation on modality commonly seen in the learning game literature: Assessment tools are often provided as part of an academic paper but, unlike textual discussion, offer the practising designer a more specific method of analysis. Next I discuss several examples of assessment tools.

²⁸ I relate gamification to both activity-goal alignment and deep learning further in chapter 7.

6.3 Existing Assessment Tools

One crucial tool for any designer is a method of assessment: how might a designer know if their design is good or not? For learning games, there are many such methods, yet I am proposing a new one. Why are none of the existing methods sufficient for my purposes? This chapter can be seen as an argument for the need for a new method of assessment of learning games. In this section, I review existing methods of assessing learning games from both the commercial and academic literature. This review defines an important limitation of existing methods, especially in light of the need for “deep learning” as discussed in the preceding section.

6.3.1 Scope

There are many types of assessments of learning games. Certain types are not useful in the context of this research. Here I name the types of assessments I have excluded, and explain why I excluded them.

In a sense, almost any discussion of learning game design could be used as the basis of assessment. For example, the designers of *Re-Mission* discussed the key principles they used in their design (Tate, Haritatos, and Cole 2009), listing certain properties. However, they did not seem to intend this list to be used in such a manner. Given that assessment is not the same as advice, I feel it is useful to work from literature whose authors intended to develop assessment criteria. I make one arguable exception to this: Gee’s Deep Learning principles (Gee 2009a), discussed in this chapter.

I leave aside assessments based on theory that I found insufficiently supported or justified, however interesting the idea. For example, Becker proposed the “Magic Bullet” (Becker 2008, iv), an unusual approach to assessment that aimed to allow designers to visualize the relative proportions of required versus optional or incidental learning in a game. Becker’s highly flexible, individual approach has strengths, but I felt Becker’s design diverged too far from its foundations in the literature and I believe that a clearly repeatable method or process is important, which Becker’s method lacks. For these reasons, I do not include such theories here. I also do not include assessments that do not exercise their theory. For example, Liu and Lin (2009) examined 196 games using expert opinion and content analysis to identify 43 indicators in five categories. However, the authors did not use these indicators to assess any games, nor has anyone else that I could find²⁹. Finally, I only discuss literature that overtly assesses learning games. For example, (S. De Freitas and

²⁹ Specifically, I did not find any discussion or application of Liu and Lin’s 43 indicators in any subsequent paper written by either of them, or any text that cited this paper, in a Google Scholar search on 1 June 2012.

Oliver 2006) offered a four-factor framework for tutors to embed games in their lesson plans that merely acts as a guide to discussion: it leaves judgement to the reader.

In short, I exclude assessment tools lacking worked examples, clear reasoning, any evidence of efficacy, and/or case studies.

6.3.2 Shearer's Basic Assessment

In his Master's thesis, Shearer (2011) developed and tested an assessment method for learning games, and applied it to a type of learning game called "distracted driving games." These games aim to educate young drivers on the dangers of being distracted while driving. Shearer developed a "best practices checklist" to assess the quality of distracted driving games. Shear cited ten scholars including Gee, Prensky, Salen, Beck, Quinn, and Aldrich, and identified ten commonly named desirable attributes from the literature. Shearer stated his criteria for a game metric to be included in the best practices checklist: it "had to be mentioned by no less than three theorists, and it could not have anything to do with the aesthetics of the game." I found this statement reasonable and generally agree with his ideas, but wished for more discussion of his reasoning: why were aesthetics excluded? He said he dropped the less commonly cited attributes and finally identified six attributes: Story Line, Challenges, Rewards, Motivation, Interactivity, and Feedback. For each of these, he named more specific sub-criteria. For example, Challenges had three criteria:

- a. Reachable, but not too easy or too hard to obtain*
- b. Risk and uncertainty*
- c. Appropriate skill level or difficulty depending on the audience (p. 30)*

Shearer's assessment tool records only the presence or absences of these six attributes. Shearer briefly discussed some of the attributes and their meaning for seven specific learning games.

I selected Shearer's work from among similar literature for two reasons. First, Shearer limits his investigation to a particular type of learning game. This selection allows him to develop theory that is a useful position between real-world practice (Tate, Haritatos, and Cole 2009) and generalized theory (Gee 2003). Specifically, Shearer evaluated several games he did not build, which removes the risk of positive bias yet keeps his first-person play experience firmly in the foreground. Second, I was able to follow Shearer's reasoning for his assessment criteria.

Shearer stated "some game metrics were not as important as others, because very few theorists included them." By valuing the frequency of mention as a proxy for significance, Shearer's approach

could contain the “danger of indiscriminate, unqualified use of quantitative citation” (Garfield 1963) of the criteria. With this potential limitation in mind, I suggest that Shearer’s work might be useful as a “lowest common denominator” of learning game assessment. I used Shearer’s work to assess the well-known learning game *Math Blaster* in this chapter.

6.3.3 Rice’s Cognitive Viability Index

Rice developed the Video Game Cognitive Viability Index shown in Figure 32 to help teachers assess learning games for higher-order thinking with a rubric (Rice 2007).

Characteristics	Y/N 1/0
Requires users to assume a role in the game, rather than simply play.	
Offers meaningful interaction such as dialogue with NPCs.	
Has a storyline.	
Has a complex storyline with characters users care about.	
Offers simple puzzles.	
Has complex puzzles requiring effort to solve.	
Uses three-dimensional graphics.	
Allows multiple views or camera pans and the ability to zoom in and out.	
Allows different ways to complete the game.	
Simulates complex processes requiring adjustment of variables by users to obtain desired results or adjusting variables leads to different results.	
Allows interaction through use of avatars.	
Avatars are lifelike.	
Requires interaction with virtual elements within the game.	
Requires knowledge of game elements beyond mouse prompts, number entry (e.g., combining elements to create new tools, understanding complex jargon).	
Requires gathering of information in order to complete.	
Requires synthesis of knowledge in order to complete or successfully engage elements in the game.	
Environment effectively replicates real world.	
NPCs display AI characteristics.	
NPCs display effective use of AI resulting in dynamic experiences for the user.	
Offers replay ability with varying results.	
Total score: (Indicating placement on the Video Game Cognitive Viability Index)	

Figure 32: Rice (2007) provided this scoring rubric for teacher evaluation of higher order thinking in a learning game.

Rice's basic reasoning and approach made sense to me, but I could not fit his specific criteria to Gee's conception of deep learning. Gee's idea of "microcontrol" might seem similar, but I found crucial differences which I discuss later in the following chapter. Additionally, some of Rice's specific items did not fit with my understanding of the literature. For example, Rice's criterion "Uses three-dimensional graphics" conflicts with studies showing that while three-dimensional graphics are beneficial for three-dimensional problems, other forms of higher-order thinking are just as well addressed with other modalities (W. Winn et al. 2002) and quality of learning outcome is independent of visual fidelity (Toups, Kerne, and Hamilton 2011). While there is literature that supports the relation between fidelity and learning outcomes, as I discussed in chapter 2, it does not argue specifically that good learning games should use three-dimensional graphics, as Rice did. I found one aspect of Rice's work very compelling: Rice chose a specific yet broad theory from the field of educational technology: higher order learning theory. As I discussed in chapter 2, Rice made an important insight with this choice, as it captures a crucial distinction that I think learning game designers often fail to make.

6.3.4 Commercial Assessments of Learning Games

It is also useful to look at existing commercial assessments of learning games, as their user-based approach offers a useful contrast to the academic literature. Each assessment was developed by one or more persons, whom I collectively call "the author." The author's audience, whom I call "users," typically include journalists, learning game players, and players' parents and guardians. As might be expected on a commercial site, learning game designers and academic researchers are not always the primary audience.

The first assessment I discuss is named "Evaluation Videogames" (Pixel 2012). It is a "collection of reviews of commercial videogames which have been evaluated in order to assess their educational potential." Its assessment of interaction uses the seven criteria shown in Figure 33 and discussed below.

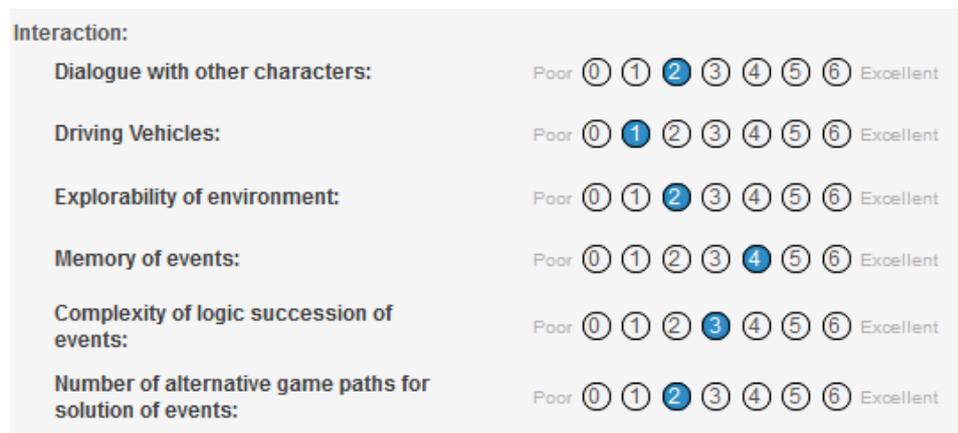


Figure 33: “Evaluation Videogames” uses seven criteria of interaction. (Pixel 2012)

I found the individual ratings sometimes did not match the overall rating. For example, consider the ratings of the particular game shown in Figure 33. The seven interaction criteria were scored between 1/6 and 4/6, which suggests an unimpressive overall rating of 3/6. However, Figure 34 shows the final score for that same game, where the suitability to education was rated a perfect 6/6. Why was the game rated so highly if the interaction was so weak? The text explanation (“You can learn in an active attractive, humorous way”) was not sufficient for me to understand this disparity. I can imagine two reasons for the disparity between individual ratings and summary ratings. First, the site is open and allows anyone to rate a game, which when combined with the lack of clear direction on rating methods, obviously presents the potential for bias problems. I recognise that there are good reasons to allow open sites, and that the site has named the evaluator, perhaps in an effort to reduce bias that anonymous posting would allow. However, a similar mismatch between interaction rating and final evaluation score was noted on other reviews on the site, which makes me think this problem was not unique to this case. However, this possibility has little relevance to my argument so I leave it aside. Secondly, perhaps the reviewers did not perceive a connection between the interaction rating and the educational suitability rating. This relationship is essentially the premise of activity-goal alignment, and I discuss the implications of this possibility further below.

EVALUATION
Usability potential of this game in an educational context:
 Poor ① ② ③ ④ ⑤ ⑥ Excellent

Explain how the videogame can be used in an educational context:
 You can learn in an active attractive, humorous way.

Evaluator:
Name of Evaluator:
 Lisa Fluck

Country:
 Germany

Date of Evaluation:
 19/04/2008

Figure 34: Overall rating of educational potential from “Evaluation Videogames.” (Pixel 2012)

Regarding the overall design of the assessment, I found general validation of my findings that drove the design of the AGA Scoring Tool, as I discussed in chapter 4. Specifically, this tool provides supporting evidence for two important positions I take throughout this thesis:

- 1) an single, important idea can usefully combine a variety of related ideas, and
- 2) it is both possible and useful to summarize this idea using a combination of a scalar rating and a textual explanation.

Specifically, I observed that the simplicity of a 7 point scalar rating had explanatory power that purely text-based assessments lacked, and while a single rating was prone to misinterpretation and/or oversimplification of something as complex as a design principle, the accompanying brief textual explanation gave sufficient context to understand the rating.

However, I observed that this tool has important drawbacks. I also observed that the user of the site is not advised on the meaning of the seven criteria, which I found confusing. For example, “memory of events” could mean the player’s memory or the game’s ability to store player achievements, but there is no way to be sure. From this, I conclude that it is useful to provide more explanation embedded within the assessment tool than was provided in this example.

While this rating system seemed similar to scoring activity-goal alignment, it did not have a particular methodology that would allow a practising designer to rely on the scales. Further, these ratings seemed to vary considerably between listings. Indeed, some listings did not display the rating scales at all. For example, “The River City Project” contained no educational suitability information except a block of text:

[Site Question:] Explain how the videogame can be used in an educational context:

[Reviewer Answer:] The game is a multi-user environment with the purpose to help the learning of scientific enquiry. The game takes place in 1800, the story is about a city that is River City dealing with health problems. The students have to make hypothesis on the reasons of the illness and can organize test and experiments. (Pixel 2012)

This text reviews but does not assess the game. Without a rating or an assessment, I feel the site largely fails to deliver its intended purpose: assessment of learning games. While most listings did have assessments, but this lack of consistency may make it difficult to use this review in a comparative way, as the site seems to be intended. To allow comparisons, I suggest that ideal assessment tools should generate assessments that are more consistent than this tool does.

A published methodology would have helped me understand the reasoning behind the assessment criteria. The tool did not specify give justification for their choice of criteria, although I believe these seven criteria were meant to represent a broad sampling of the game literature. The first three criteria (dialogue with other characters; driving vehicles; environment explorability) appear more specific than the others, and seem related to the concept of fidelity in simulation. The next two (Memory of events; Complexity of logic) are more abstract and might relate to cognitive psychology's mental models. The last one (Number of alternative game paths for solutions) seems to assess "affordance", which is the idea that an artefact's form can alone guide user understanding and behaviour (Gee 2003). A final rating seems to summarize the game's educational potential (Figure 34), as I previously discussed.

This form is designed specifically for learning games, but in contrast to prevailing views discussed earlier in this chapter, this tool does not score anything that clearly relates the player's activity and the learning goal. The site states vaguer criteria: "Usability potential of this game in an educational context" which is a 0-6 rating and a paragraph of text. This vagueness has drawbacks. For example, the video game 3rd World Farmer (Hermund et al. 2008) was rated 6/6 and described this way:

From the pedagogical point of view, and expressly stated by its creators, the ultimate goal of this simulation is to provoke student reflection on the difficulties still faced by many families in the Third World. The social connotations of the game is also supplemented by the inclusion of basic economic principles, namely, scarcity, choice and opportunity cost. (Ivars Vendrell 2012)

While I value flexibility in scoring, I found too much flexibility to form the basis for a specific critique in this example.

To summarize my findings of “Evaluation Videogames,” I found the combination of 7-point rating and a textual explanation (shown in the top two lines of Figure 34) succinct, clear and precise enough to be useful, which validates the ideas I used throughout this thesis. However, I found that the absence of anything resembling an activity-goal discussion, the lack of clear relation between total ranking and its components, and the unclear relation of criteria to theory were important drawbacks to the design of this assessment tool. These drawbacks partially form the basis of the criteria I develop at the end of this chapter.

6.3.5 Commercial Assessment without Scoring

Next I discuss an assessment tool that does not rate games using numbers or scales. Instead, it describes the games in text. These descriptions contained interesting insights into how activity-goal alignment was perceived by the reviewer.

In 2010, the European Network for Growing Activity in Game-based Learning in Education (ENGAGE) released “*Guide to Games for the Classroom*”, a book that contained reviews of 33 learning games and an online, user-editable catalogue of learning games (Engage Learning 2012). Rather than rating games, it offered a 1-sentence summary with no rating (e.g. for a game named “Gears”: “Could be used to train motoric skills, patience and a nice gimmick for discussions in physics.”) The full review followed a form that included a variety of categories, including cultural questions (“Does the game specifically target cultural issues and if so, which ones?”) Notably, the mechanics were not addressed.

Another example of this approach is given by Prensky. He reviewed learning and serious games (Prensky 2005) with a one-line description. Health Games Research gave a similar rating-free summary with basic categorization (Figure 35).



Publisher
SuperBetter Labs

Developer
SuperBetter Labs

Web site(s):
<https://www.superbetter.com/>

Game

2 people like this

"SuperBetter helps you achieve your health goals — or recover from an illness or injury — by increasing your personal resilience. Resilience means staying curious, optimistic and motivated even in the face of the toughest challenges.

SuperBetter is a tool created by game designers and backed by science to help build personal resilience: the ability to stay strong, motivated and optimistic even in the face of difficulty challenges. Resilience has a powerful effect on health — by boosting physical and emotional well-being. Resilience also helps you achieve your life goals — by strengthening your social support and increasing your stamina, willpower and focus. Every aspect of the game is designed to harness the power of positive emotions and social connection for live, feel, and act better."

Topic : Behavior and Personality, Health Care and Medicine, Social Issues and Skills, Stress and Coping

Genre : Educational, Entertainment

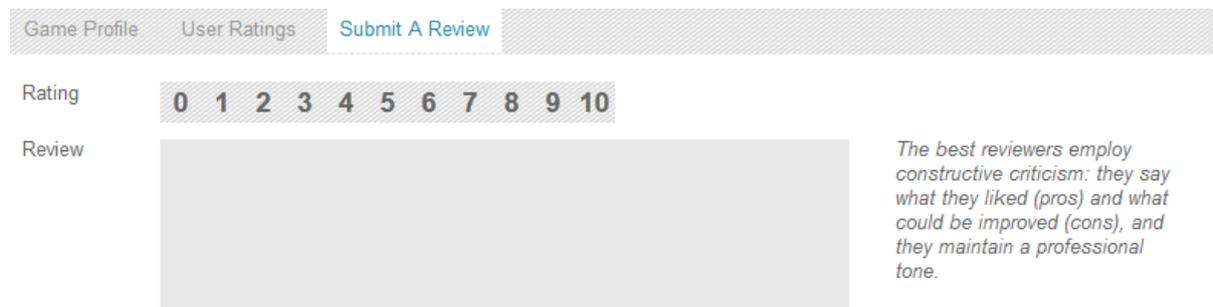
Game platform : Internet

Target population : General Audience

How to obtain : Free, Online

Figure 35: Health Games Research Database entry for *SuperBetter*, showing review criteria: topic, target population, genre, platform, and price/distribution. (<http://www.healthgamesresearch.org/games/superbetter>)

Figure 36 shows that this assessment has two parts: a rating 0-10, and a 2000-character description.



Game Profile User Ratings **Submit A Review**

Rating **0 1 2 3 4 5 6 7 8 9 10**

Review

The best reviewers employ constructive criticism: they say what they liked (pros) and what could be improved (cons), and they maintain a professional tone.

Figure 36: User review form on the Games for Change website gives a text box with some broad guidance and a single rating scale. (Games for Change 2012)

Users were advised: “The best reviewers employ constructive criticism: they say what they liked (pros) and what could be improved (cons), and they maintain a professional tone”. From this description we can expect a highly flexible and free-form review, similar to what one might see on any product reviews for sites like Amazon.com. Notably, users are not asked to consider anything particular about learning games here. I believe this is an important weakness in the criteria. In my review of other text-based assessment sites such as *The Educational Games Database* (TEGD 2012), I observed similar strengths and weaknesses.

The drawback I observed in all these examples of text-based assessments stems from a certain lack of structure. By offering total flexibility, I found the Health Games Research assessment tool allowed too much subjectivity and variability to make the assessment useful across various games. Bias also seems an important risk with these text-based assessments. However, I appreciate the power of free-form text to allow the reviewer freedom to capture something as complex, unique, and dynamic as the quality of a learning game. The Games for Change review form (Games for Change 2012) employs a 1-10 rating scale shown in Figure 36. This rating alone could not hope to match the richness of text, and the form includes a review text box as well, possibly for this reason. This finding further supports the position, taken throughout this thesis, that any learning game rating or scoring system is best accompanied by a text description, and perhaps a brief one is best. However, I do not suggest that this idea should be required of any valid learning game assessment tool, as it might be possible to develop a tool that lacks a text description but is still useful. For this reason, I do not argue that textual descriptions should be in every assessment tool, as seen in the design criteria I provide at the end of this chapter.

The flexibility of text-based assessment can be structured. One common approach to structuring text is the highly structured classification system. Here, I discuss an example. “Serious Game Classification” (Ludoscience 2012a) lists over 2,600 serious games using multiple criteria: gameplay, purpose, market, and target audience. Each of these criteria is explained to the user in clear language:

Gameplay: Does this title features [sic] stated goals to reach like any "game"? Or is the player totally free to make his own choices in a "play-based" way? Besides these two overall gameplay types, the core rules of each title are analysed and represented as GamePlay bricks.

Purpose: Besides its play value, does this title feature other purposes? For example, is it designed to train you? To broadcast a message? To tell you a story? (Ludoscience 2012b)

In explaining their conception of the idea of "Gameplay", the authors state that all video games can be divided by type of gameplay: "Game-based" means "...goals are used as reference to evaluate the performance of the player, in order to judge if he has 'won' or 'lost', or to give him an objective mark called 'score'" (Ludoscience 2012a). The alternative to "Game-based" was called "Play-based" which has no stated goals to reach. This type of game is colloquially known as "sandbox" or "digital toy" (Starr 1994).

Games were classified using multiple criteria, including "gameplay." The authors have published numerous papers (Djaouti et al. 2008) explaining their cognitive science-based justification for the use of a novel set of icons named "gameplay bricks" shown in Figure 37.



Figure 37: Ludoscience's "Gameplay Bricks." Orange indicates rules stating goals; blue indicates rules defining constraints. (Djaouti et al. 2008)

The brick's definition was quite specific, as shown in the example in Figure 38.

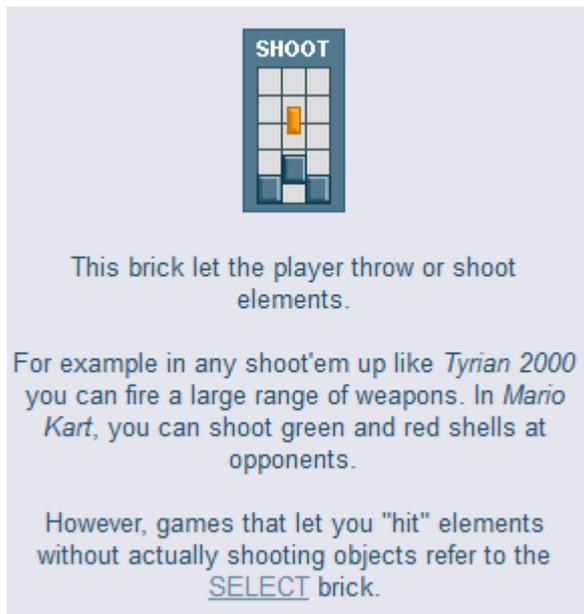


Figure 38: Ludoscience's "Shoot" gameplay brick. (Djaouti et al. 2008)

I found the visual icons of the bricks appealing and the ideas behind them logical; however, in practice, I found them insufficient to serve as a classification of all games. Specifically, I found the bricks were too generic to understand the core mechanic of some common games. For example, the review of the serious game *3rd World Farmer* (Hermund et al. 2008) identified typical farming game mechanics, suggesting that *3rd World Farmer's* mechanic is similar to most farming games; however, the key point of the game was not represented. The designer of *3rd World Farmer* stated that in other farming games "...it's always possible to prosper by playing cleverly and making the right game choices. It's not always like that in *3rd World Farmer*." The player can easily die even when playing optimally in *3rd World Farmer*. I tried to convey this crucial point using Ludoscience's structured classification system, but found I was unable to. Perhaps the problem is inherent in any modular, predetermined library of game mechanics. If this collection of mechanics is unable to represent a simple serious game like *3rd World Farmer* fairly, it does not bode well for more ambitious efforts.

Ludoscience's novel conception of "bricks" is distinct from more common explanations of game mechanics such as Cook's "Game Atoms" (Cook 2007b) or the MDA framework (Hunicke, LeBlanc, and Zubek 2004). The difference is that bricks are intended to classify existing games by their mechanic, whereas Game Atoms are intended to help designers build novel mechanics. In other words, bricks compare, but do not explore, game mechanics. I felt that Djaouti et al's overall approach was primarily based on cognitive science. For example, they begin their investigation by advocating the Input/Process/Output model commonly seen in both early computer science and

educational literature. This science-based approach contrasts with the design-based approach common among theorists such as Gee. Gee typically begins with his observations of the users' experience of game play, not abstract principles, and then develops theories based on how those observations inform a variety of various existing theories (Gee 2003). By contrast, Djaouti et al aimed to define a strict set of bricks that fit both their idealised input/process/output model and their observations of existing game mechanics.

In chapter 2, I discussed the term "mechanic," which is commonly used throughout game design literature. I find it curious that Djaouti et al do not use the term "mechanic" in their description of bricks, or relate their coined term "brick" to the existing concept of a game's mechanic. Although it is possible that the terminology did not translate fairly into English, the terminology reinvention added to my general impression that Ludoscience's theory was developed in relative isolation from the most commonly accepted ideas in learning game design literature. In sum, my experience with Ludoscience's *Gameplay Bricks* led me to conclude that a novel diagram or graphic is compelling and engaging, but if it is not sufficient to explain the key points of a game, other free-form modes of expression are needed as well. I also observed ontological limitations that supported my argument opposing overly scientific approaches to design problems, given in chapter 3.

6.3.6 Summary of Existing Assessment Tools

To summarize this section, I reviewed selected methods of assessments and found that commercial approaches often do not state or defend their basis for assessment clearly, while academic assessments are often based on particular ideas or literature that makes them unsuitable as general-purpose methods of assessment. Gee's Deep Learning principles are a notable exception: they seem to hold promise as a sound basis for a method of assessment.

One point in common in the preceding review of assessment tools is their modality: many use a "list of attributes" approach. In chapter 3, I argued that, historically, this approach stems from a powerful precedent set in the learning game literature. Next, I provide evidence for my argument that a new approach might be justified, by showing how several attribute-based approaches can all fail to be effective.

6.4 Application of Selected Existing Assessment Tools on *Math Blaster* series

Since its release in 1983, the *Math Blaster* series of learning games has sold several million copies (Lucas Allen 2012) and ranked 112th overall in Amazon software sales in 2007 (Rice 2007, 89). The

series consisted of twelve products separated by age-based learning goals in 2012 (Knowledge Adventure 2012a).



Figure 39: In *Math Blaster* series, the exciting shooting activity freezes until player answers an arithmetic problem. (Knowledge Adventure 2012b)

1.1 My play experience

Starting in 1983, I have played various versions of *Math Blaster* many times. Knowledge Adventure has recently launched "*HyperBlast*", an iOS application in the *Math Blaster* series (Knowledge Adventure 2012b). In 2012 I observed my 6 year old son playing the game and played it myself, and found the activity-goal alignment similar to past versions.

It is immediately apparent from a short play session that *HyperBlast's* player activity and learning goal are unrelated. The core mechanic is "tunnel running:" the player zooms through three-dimensional tubular tunnels on a motorcycle, dodging obstacles by tilting the device physically, and/or shooting them with two weapons. The action is fast-paced and highly engaging, with intense music, copious sound effects and satisfying visual explosions. Suddenly, a large evil-looking creature blocks the way. The player controls are disabled, the motorcycle freezes, and a disembodied math problem appears in text on screen, similar to Figure 39. The player is expected to answer by touching one of six numbers on the creature's six arms. The player is rewarded with points proportional to the number of correct answers. The creature does not speak, explode, smile, or otherwise interact with the player, nor do any of the game controls have any bearing on the math problem. There is no explanation of the meaning of the standard math symbols, nor is there any exploration of the idea of addition: how to add, why it is useful, or what its limitations are. The earned points have some benefit in the three-dimensional tunnel running portion of the game, but that is the limit of the connection between in-game activity and learning goal.

6.4.1 Criticisms of *Math Blaster* Series

Many notable theorists have criticised *Math Blaster* as not meeting the full potential of learning games. *Math Blaster* “has received high ratings from educators, but not from games designers” (Becker 2007). Osterweil described *Math Blaster* as a canonical example of “chocolate covered broccoli” (Osterweil 2006). This phrase is a negatively framed version of the idea of “sugar-coating:” In this metaphor, learning is the broccoli, and the game is the chocolate, and the resulting combination is neither more palatable nor healthy than the two taken separately. Osterweil said the problem occurs when designers consider learning as incompatible with fun. Rice also essentially argued this point (Rice 2007) as discussed in chapter 2. Klopfer, Osterweil, and Salen cited *Math Blaster* as the example for their statement “designers of educational games that try to inject content learning into a game where it doesn't fit may create experiences that are somewhat entertaining, but their educational value is highly suspect” (Klopfer, Osterweil, and Salen 2009, 25). Rice (2007) argued that *Math Blaster* teaches “lower order thinking:” simple calculations and memorisation using simple narrative and action gameplay as a reward.

Squire and Jenkins (2003) found university students were largely critical of their childhood learning games’ quality, and quote one student citing *Math Blaster* as a memorably bad example:

I played ‘educational’ type games for the computer a long time ago in middle school (remember Math Blaster?) and they sucked. (p. 12)

Many respected scholars thus consistently use *Math Blaster*’s core mechanic as a canonical example of what not to do. It might seem like any modern learning game designer who has read the literature could not avoid understanding the problem of *Math Blaster* style mechanics. However, as I argue below, it is not only possible, but easy, to misunderstand the problem.

6.4.2 Developer’s Defence of *Math Blaster*

How do *Math Blaster*’s designers conceive of *Math Blaster*’s learning goals? What is their justification or defence against all this criticism? Knowledge Adventure’s white paper titled “The Blaster Method” (Knowledge Adventure 2005) related the game’s design to learning theory. It started with a broad claim: *Math Blaster* builds “mathematical ‘pre-skills’ by emphasizing visualization, pattern recognition, memory, logical deduction and many other important thinking skills.” The authors also claimed that the *Math Blaster* series follows these game design principles: motivation, modelling, practice, evaluation, and active learning.

I could not see how these claims were supported in the provided discussion. For example, in their explanation of active learning, the authors cite the National Council of Teachers of Mathematics: “good mathematics instruction must be active in order for students to learn appropriately.” To me, NCTM’s principles seemed to criticize *Math Blaster*’s fundamental approach: “Students must learn mathematics with understanding, actively building new knowledge from experience and prior knowledge” (NCTM 2012). I did not see how *Math Blaster* provided either experience or new knowledge, as players must only memorize facts to win. I have not seen where logical reasoning, procedural thinking, and deeper understanding of mathematics are useful, let alone scaffolded or required, in the challenges presented in *Math Blaster* series.

Math Blaster’s learning goals are not met with their player activity. My experience with *Math Blaster* agree with Rice’s statement that *Math Blaster* is essentially a “computerized worksheet,” enabling only practice of arithmetic (Rice 2007), and does not aim to teach how arithmetic works. Other respected scholars agreed (Devlin 2012b; Klopfer, Osterweil, and Salen 2009). Becker was more generous in her assessment: “Much of what needs to be learned in order to complete the game lies in the psychomotor domain, and to a lesser extent in the first half of Bloom’s cognitive domain (knowledge, comprehension, and application)” (Becker 2007).

The *HyperBlast* white paper authors also claim “Children enjoy what they are learning and are better able to retain and use the skills they have learned from their time in the Blaster world” (Knowledge Adventure 2012a, 34). I find this problematic for two reasons: the authors do not provide any evidence, nor do they explain how this claim relates to the discussion of flow or active learning. I could imagine a reasonably convincing argument that supported active learning theory in *Math Blaster*, built on observations of kids in classrooms taking control of their own educational experience by playing games instead of listening to lectures, but I would suggest that this argument also feels anaemic as the active learning has little to do with the learning goal of mathematics.

In both papers discussed above, the developers of *Math Blaster* argued that the game had learning value related to theory. I generally found their arguments unconvincing: the authors made broad theoretical claims were insufficiently supported by the citations in their argument, and do not provide sufficient detail to understand how the game’s design embodies the theoretical position they argued. This might be simply a problem of language and culture: if the developers are designers, not academic scholars, they may have used theory correctly but only failed to articulate it. Many academic critics have discussed *Math Blaster*, and although most have criticised its design in relation to learning theory, at least one did not. Next I examine one academic discussion that supported *Math Blaster*’s design.

6.4.3 Becker's Defence of *Math Blaster*

In discussing the design of the *Math Blaster* series, Becker stated:

Contrary to popular complaints (Gee, J. P., 2005; Prensky, M., 2006), for Math Blaster the lack of integration of the learning and game elements as defined by Clark Aldrich (2004) are not detractors. This is further backed up by a study involving middle school students who were asked to critique educational games created by other students, Rieber, et. al. found that important game characteristics for the students included story, challenge, and competition, but did not include integration of storyline with educational content or production values. (Rieber, L. P., Davis, J., Matzko, M., & Grant, M., 2001). (Becker 2007, 7)

Becker essentially states that children enjoy low activity-goal alignment games. I agree. However, Becker posits that low activity-goal alignment is not a detractor, and I do not agree. Children enjoy many poorly designed games, especially in a classroom setting. To say children enjoy *Math Blaster* does not defend *Math Blaster's* design. Instead I ask: Is *Math Blaster* a good example of how games can teach math to children? I would argue not.

Becker cited a study where children assessed fourteen simple child-designed video games and one simple adult-designed learning game. In interviews with the teacher, Becker found that the children valued attributes such as competition, storyline and resemblance to commercial games. Others have also found that "children were perfectly content" with "sugar coated" games like *Math Blaster* (L. Rieber et al. 2001). I question how this finding relates to the quality of learning gamed design. Rieber et al studied middle school children in a classroom. It is well established that classroom game-playing behaviour is fundamentally different from deeply engaged gaming sessions because children do not have the freedom to choose (Gee 2003; Klopfer, Osterweil, and Salen 2009). Children use a different standard for playing games in a classroom than they do in the situations that Gee observed (Gee 2003).

Gee's findings fit my own experience. I have spent months meeting weekly with middle school children in a classroom, where I regularly observed the children employing dual standards in their gameplay. In the classroom, during several sessions I allowed the children to play any game they wanted and observed them. I found the children enthusiastically played almost any game I suggested, even if it was a clearly broken prototype, if it was made available to them. They played many simple, crude 2D Flash-based games such as *Tank Wars* (2 Flash Games LLC 2012) happily. However, when recess time came they usually abandoned these crude games to play outside. Further, when we discussed gaming in general, they described a much higher standard in their play

sessions at home. Many played sophisticated three-dimensional games such as *Minecraft* and *Portal* on consoles and powerful personal computers, and many complained about lack of engagement, unrealistic storylines and interface control problems. One constantly asked me to help him to mod *Minecraft* to enable his dream scenarios. These were not “anything goes” players outside of the classroom. I was convinced by these observations that the most important reason children play games at school is to avoid harder deskwork and extract whatever pleasure might be had. Their classroom gaming did not seem to relate to the game’s quality or efficacy.

I feel that Becker’s argument makes sense in the classroom context. However, we cannot generally assume that if children want to play a video game in a classroom, that the game is well designed. In this way, I continue to find the arguments that largely oppose the basic design of *Math Blaster* convincing.

In the next section, I show how various assessment tools fail to reveal *Math Blaster*’s design problems.

6.4.4 Attribute-Based Evaluation Methods

In my review of existing assessment methods, I noted that the majority used a checklist-style list of attributes or principles. Other designers have specifically found that the attribute-based checklist approach has important weaknesses (Ciampa-Brewer 2011; McDougall and Squires 1995), and I believe these findings are directly relevant to the field of learning game design. In this section I aim to demonstrate the weaknesses of the attribute-based approach to game assessment.

I begin by using Shearer’s assessment method to discuss *Math Blaster*, as shown in Table 1.

Table 1: *HyperBlast*, a *Math Blaster* series game, rates well using Shearer’s Best Practices Checklist (Shearer 2011).

1. Story Line	x
Specific problem	x
Rules	x
Clear objectives	x
2. Challenges	x
Reachable, but not too easy or too hard to obtain	x
Risk and uncertainty	x
Appropriate skill level or difficulty depending on the audience	x

3. Rewards	x
Emotional, need/like	x
Positive rewards foster motivation	x
4. Motivation	x
Experience bar or tracking device	x
Experience points	x
Measures progress through story	x
5. Interactivity and Engagement	x
Action choices	x
Multiple aims both long and short	x
Engaging theme	x
Clear goals	x
6. Feedback	x
Positive and negative	x
Leads to engagement	x
Choices and consequences	x
Learn a lesson	x

To complete Shearer's checklist for *HyperBlast*, I used a combination of three types of observations: I recalled my own play experiences past and present, as well as my recent observations of my children's play experiences. For example, I asked myself: does *HyperBlast* have an engaging theme? I recalled the vivid graphics and the compelling music, and the fast-action shoot/dodge mechanic, and I found that the answer is clearly "yes". Does *HyperBlast* give action choices? Yes: I am free to shoot enemies, or move to avoid them. Are there multiple aims, both long and short? Yes: I must complete the problems in front of me to progress through the game.

I found that every box on Shearer's list needed ticking. Shearer developed this checklist to assess simple learning games which lacked very basic features such as points or rewards. For that purpose, his list could be useful. However, I find that Shearer's bar is too low: The "lowest common denominator" approach is not useful for learning games like *Math Blaster*.

Next, I explore two checklists that aim higher. Gee and Morgridge (2008) established 15 criteria and used them to critically assess the learning game *SMGL*, stating: "we can use these features—and others like them—as a checklist." I followed their precedent, to assess *HyperBlast* as shown in Table 2.

Table 2: *HyperBlast* scored against Gee and Morgridge's 15 assessment criteria for learning games.

Game Attributes	
long, complex, difficult	x
strong identities	some
think like scientists	
players are producers, not consumers	
lower consequences of failure	x
user-selected difficulty	x
Learning Attributes	
problems are well ordered	x
mastery and challenge cycle	x
regime of competences - doable, but challenging	x
think about relationships, not facts	
explore, think laterally	x
has 'smart tools'	
cross-functional team (raiding party)	
performance before competence	x
just-in-time information	x

Gee and Morgridge's principles felt more meaty and defensible to me: I felt this checklist set a reasonable standard. The half-filled checklist generally suggests that the game is not necessarily well designed, as compared to Shearer's checklist. However, I was surprised at how many of the criteria were observable in *HyperBlast*. For example, I observed kids playing *HyperBlast* who showed "performance before competence": the player can guess at answers and progress some ways through the game. However, my interpretation of the author's intent seems to conflict with what I observed. A sophisticated, contextualized understanding of Gee and Morgridge's intent is necessary to correctly tick the box.

Of course, it is the responsibility of the assessor to understand the tool they are using, but in recognition of the idea of affordance, I feel the assessment tool should be designed to be difficult to misuse. A design checklist is a tool that seems easy to use. Would an amateur designer realise that they do not possess a sufficiently sophisticated understanding of the literature when using a checklist such as this? I feel it is likely that many amateur designers will simply read the brief

sentence and assume that their intuitive grasp of its meaning is sufficient, reflect on their observations of player behaviour, and tick the box or not.

The results of such a tool could be harmful, not just useless, because others cannot tell how sophisticated an understanding is behind a single tick. Consider a learning game assessment tool Y, an amateur designer A who uses that tool to create an assessment of a learning game, and a veteran designer B, who reads A's assessment without knowing A. If A ticks "performance before competence," and B sees that tick, how is B to know if A understands the idea well enough? I feel a tick, alone, is insufficient information. One of the merits of free-form text is the assessor must express themselves. If A had to write a few words to justify their tick, B might be able to guess A's competency. Thus I suggest a well-designed assessment tool will both help the assessor gain a reasonable understanding of the ideas implicit in the assessment, and give the reader some basis for deciding if the assessor is competent to assess.

I also found the completed checklist difficult to interpret as a holistic assessment. What does a half-filled list mean? Are some of these items more important than others? Is the game bad because half the boxes are empty, or is the game pretty good but could be even better? If *Math Blaster*, a game that I argued has design problems, can tick half of these boxes, how could a designer hope to assess a more subtle or complex game's design? For me, any half-filled checklist is difficult to interpret holistically. From this, I argue that checklists are a modality of learning game assessment that have significant room for improvement.

However, I acknowledge that my discussion so far is not yet a solid argument against checklists as assessment tools in general. There are other variables still to consider. Perhaps Gee and Morgridge's focus was overly broad, or there were simply too many items on their list to achieve clarity. Perhaps a "top 10" list that was more focused would give clearer results. Fortunately Gee also developed the Deep Learning principles (Gee 2009a), as discussed previously in this chapter. I adapted these six principles into a checklist, and scored *HyperBlast* against the resulting checklist, as shown in Table 3. I enumerated the five subconditions within the third principle because Gee's discussion made it clear that these five subconditions were sufficiently distinct that they needed to be considered separately, not combined.

Table 3: *HyperBlast* scored with Gee's Six Deep Learning Properties.

1	Psych out rule system (exploit emergent properties)	
2	microcontrol → embodied intimacy	x
3	game experience enable five subconditions:	

3.1	game has goals	x
3.2	goals require reasoning	
3.3	immediate feedback	x
3.4	interpret knowledge to progress	
3.5	enables gamer communities	x
4	effectivity-affordance matches bodies/tools <—> world	
5	has modelling to generalize learning from experience	
6	gameplay lets player enact unique story	x

As with the preceding checklists, I reflected on my play experience in light of Gee's accompanying text description as I decided whether or not to tick each item in Table 3. This is a highly personal, intuitive process, embedded in my own play experience. I did this evaluation with my autoethnographic method's warnings in mind. Specifically, I attempted to consider confirmation bias as I completed this list by reflecting on my observations of others' play experiences, not just my own, as discussed in my presentation of my *HyperBlast* play experience. For example, the second item asks if the game has microcontrol that leads to embodied intimacy: "a type of melding of one's self and one's character." I observed that the fun part of *HyperBlast* clearly has some microcontrol. Although not at the level of a highly polished video game like *Tomb Raider*, I felt a moderate connection to my ship when dodging and shooting through the tunnels. Doing the math problems, however, had essentially no microcontrol or embodied intimacy. How should I tick this? I chose to use my high-level, intuitive-emotional reaction to the game as a whole: When I consider *HyperBlast*, do I recall any sense of control/embodiment connection? I found that what came to mind was the tunnel running mechanic. The math part did not register, emotionally. For this reason, I ticked item 2. However, I am not satisfied: both ticked and unticked fail to capture the gist of my experience. To me, it is clear that the design is weak, although it is unclear if there is a single underlying problem. Is the checklist format, as opposed to the level of detail given, the phrasing of the question, or some other cause, preventing me from expressing my experience? In considering these possible barriers, I felt that the checklist format was an important factor.

Additionally, I found that Table 3 fails to suggest direction toward a solution the problem I and others have observed with *HyperBlast*. This finding is not surprising because Gee noted that it is insufficient to consider each of these six properties on their own: "Deep games—entertainment or serious—not only have these properties, but implement them powerfully" (Gee 2009a, 78). Gee also stated that all games can still teach something, stating that the commercial game *Sam and Max*

lacked all of these six properties, yet “probably helps people with mental alertness and general problem solving skills” (Gee 2009a). Thus an attribute-based method seems like a poor choice in assessing a learning game against Gee’s conception of deep learning. Though he refers to only his own list, I believe Gee’s point criticises any attribute-based method of assessing learning games when there is a broad, general idea behind the attributes.

I do not argue that lists of attributes are useless to game designers while they are practising game design. When I was designing *Battlefood*, I reviewed several lists of attributes, notably Schell’s lenses and Gee’s deep learning principles. I found that checklists were somewhat useful summaries, reminding me to broadly consider a number of particular aspects of my design. However, this consideration was always subverted in relation to a few high-level goals. I did not actually tick off items, or count how many items I ticked.

Of course, authors of lists of attributes do not necessarily intend designers to use them as methods of assessment. Indeed, they are often intended to be tools for reflection and discussion. For example, Schell created a deck of 100 lenses of game design, and advised designers how to use them (Schell 2008). He suggests the designer focus on a few lenses, rather than attempting to consider them all at once. For example, he suggested randomly drawing a few cards from the deck and carrying them around, or selecting a card and discussing it in relation to a game with co-workers. Notably, he did not suggest designers check every card one by one, in an attempt to decide if their game is good or bad.

A key problem with checklist-based assessments above is that they seem to suggest that all items should be weighted equally. Unless it is stated otherwise, I feel that checklist-based assessments generally imply that all items should be weighted equally. In reflecting on my experience as a learning game designer, I believe a few key ideas are essential to good learning game design, rather than a long list of various ideas. In particular, I agree with Shelton, Malone, and many prior theorists that activity-goal alignment and its related predecessors should be considered not as one of many items on a checklist, but as a general, guiding principle. For example, consider the criticism of *Math Blaster* I presented above. *Math Blaster* clearly possesses many enviable attributes such as engaging gameplay; however, it has one crucial problem, which is that its core mechanic has no connection to its subject matter. Rice and many others have argued that problem outweighs all the other positive attributes. Thus I feel that a different approach to assessment could be useful.

6.4.5 Summary

In summary, I find the literature around *Math Blaster* largely supports Devlin's argument: New math games should aim to teach mathematics in a deeper and more meaningful way than *Math Blaster*.

I find many existing assessment tools do not sufficiently identify *Math Blaster*'s design problems, and argue against the usefulness of the checklist approach to learning game assessment. This position is consistent with other designer/researchers' criticisms of checklists as a form of assessment (McDougall and Squires 1995). More broadly, the attribute-based method of assessment is, I argue, insufficient for amateur designers because significant learning game design experience is required to correctly interpret these assessment tools.

6.5 Criteria for Assessment Tools for Amateur Designers

I have argued the need for a new approach to learning game assessment. Specifically, I believe there is need for a tool that helps designers of learning games focus on the prevailing advice in the literature. In that context, and based on my discussion of weaknesses in existing methods of assessment in the preceding sections, I give three types of criteria for learning game assessments.

First, I conclude that any learning game assessment must:

1. Define a clear, reliable theoretical basis for assessment. In other words, it must set out reasons for claiming that certain games are “good” or “bad,” and defend them in the light of past work.
2. Judge clearly and consistently enough to avoid fruitless argument about the judgement. There must be enough flexibility to allow the wisdom of the assessor to prevail over blind rules, with enough structure (e.g. specific enough rule-based procedure or guidelines) to ensure the tool is not just a foil for personal preference.

Additionally, assessments aimed at the practising designer should be:

3. Easy to use: The designer must find the assessment's outcome easy to interpret in practice
4. Prescriptive: a designer would be likely to understand how to modify their game to improve the assessment.
5. Simple and Quick: Practicing designers and reviewers should find it easy to use on the fly.
6. Versatile: useful for the broadest possible range of learning games.

Note that I leave as much as possible, such as the modality and theoretical foundations, to the discretion of the tool designer. I do this in order to make these criteria as general as possible: learning game design is a broad, diverse field. Although I have argued that activity-goal alignment is important, here I aim to create a “superset” of learning game assessment criteria. If I specified any particular theory in the criteria, it could limit the usefulness of these findings in the research of others.

In my review of existing tools in preceding chapters, I did not observe any existing assessment tools that meet these criteria. The next chapter charts my attempt to develop such a tool.

6.6 Conclusion

In this chapter I have found that the literature, while diverse, consistently offers a prevailing view that learning game designers should be more ambitious in their aims than past efforts by advocating both deep learning and high activity-goal alignment. Specifically, I discuss *Math Blaster* in light of Devlin's argument: New math games should aim to teach mathematics in a deeper and more meaningful way than *Math Blaster*. Having identified a clear problem through discussion, I seek tools that amateur designers can focus on. In failing to find such a tool, I generally find the attributed-based approach to assessment misses the forest for the trees: significant experience designing learning games is required to correctly interpret these assessment tools. Finally, I have set out criteria for a learning game assessment tool that might do the job.

7 Activity-Goal Alignment-Based Assessment of Learning Games

7.1 Introduction

In this chapter, I address the second question of this thesis: How might combining the ideas of scoring activity-goal alignment and deep learning offer useful new approaches to learning game assessment?

In the introduction (chapter 1) to this dissertation, I argued the need for an improved assessment tool for learning games. In the previous chapter I argued that learning game designers need to aim high, and set out criteria for any assessment of learning games. This chapter responds to those challenges: I combine activity-goal alignment scoring with the idea of higher order thinking to create an AGA-Based Assessment Tool, using Gee's conception of deep learning. I then exercise the tool in a comparative discussion of recent math games: *HyperBlast* (Knowledge Adventure 2012b), *Equatia* (Luckybird Games 2012) and *DragonBox* (WeWantToKnow 2012). I find that the AGA-Based Assessment Tool, while less comprehensive, is simpler and clearer than other assessment methods. I argue that there is need for clarity over comprehensiveness, and suggest this tool might help amateur designers aim higher because it emphasizes two important ideas in a form that is more difficult to misinterpret than its predecessors.

7.2 Development of the AGA-Based Assessment Tool

In this section I discuss my development of a new AGA-Based Assessment Tool, extending the concept of scoring activity-goal alignment to propose a new way of assessing learning games.

The AGA-Based Assessment Tool aims to focus designers on considering whether or not their game meets Gee's deep learning properties of good digital games (Gee 2009). Although Gee's writing is excellent and many designer/researchers are likely to understand generally what Gee meant, Gee has not provided a particular tool or process to aid a developer in evaluating the deep learning in a particular learning game. Accordingly, I aim to develop the "AGA-based Assessment Tool," which is more applicable than Gee's paper by itself. This new tool addresses the problems I noted in the last chapter when I adapting Gee's paper into a simple checklist-based assessment. I further justify this aim in the Discussion section of this chapter.

The AGA-Based Assessment Tool proposed in this chapter relates three theories: Gee's conception of deep learning in games, higher order learning, and activity-goal alignment. Here I discuss how these three ideas relate. In the introductory chapter I defined lower order learning tasks to include memorisation, simple calculation, and tacit, skill-oriented forms of knowledge such as driving, hammering or sewing, whereas higher order learning is more abstract and complex, such as understanding the interdependence between creatures in ecology. In the literature review chapter I reviewed theories around the concept of higher order thinking, which relate to Gee's discussion of the concept of deep learning from cognitive science. In the previous chapter I stated that Gee's implicit advocacy of high activity-goal alignment in his deep learning theory is perhaps most apparent in his third property, which he says lies at the foundation of how games recruit good learning: "Games put players in worlds where they experience things." As I have argued in chapter 4, experiential learning is the crux of high activity-goal alignment.

In this way, I have already related two of the theories to each other. It remains to specifically show how higher order thinking relates to Gee's ideas around deep learning.

As discussed in the preceding chapter, Gee identified two types of learning games: simple ones, and those that aim for "deeper conceptual understanding" (Gee 2009a):

Games can be used for different types of learning. For example, we could, and do, use games for skill-and-drill, for a sort of Trivial Pursuit that takes knowledge to be memorizing and repeating 'facts.' Or we could seek to use games for the creation of deeper conceptual understandings and for problem-solving abilities that go beyond being able to pass paper-and-pencil tests. (p. 65)

In his rejection of skill-and-drill type designs as appropriate for deeper conceptual understanding, Gee clearly connects his deep learning theories to higher order learning goals. Later Gee gave examples of deep learning games, including *Civilization*, that are clearly aiming at higher-order thinking through high activity-goal alignment: "[...] urban planning, social activism, some type of science or business, or the exemplification of a particular perspective connected to ways of understanding and changing the world." (p. 65)

If Gee's deep learning is related to high activity-goal alignment, showing how high activity-goal alignment is apparent in Gee's discussions and principles, and if a game has both higher order learning goal(s) and high activity-goal alignment, as I have argued, then it is more likely that Gee's principles of deep learning will be observable in the learning game in question. The tool I propose aims to allow designers to indirectly assess the likelihood of compliance with Gee's deep learning

principles for any given game, by using activity-goal alignment and higher order learning theories as proxies. It is apparent that this method will not comprehensively assess a game for Gee's deep learning, and that is not its aim. The tool's aim is to help the designer realise whether their game is similar enough to Gee's theory to attempt a more comprehensive assessment, or not.

While Rice used higher order and lower order as a spectrum, Gee's discussion of examples suggests a more binary distinction: all learning games belong to one of two groups: "deep learning" or not deep learning (which Gee called skill-n-drill). I also divide all games into two categories, but use Rice's criteria for division: lower or higher order thinking. I do not use Gee's distinction, even though it resembles Rice's, because I believe most designers are more capable of correctly assessing whether or not a game uses lower or higher order thinking than they are of assessing deep learning. Deep learning theory is novel, complex and holistic. Rice lightly adapted the age-old learning science theory, which is widely accepted and familiar to many as I discussed in chapter 2. Though it lacks the explanatory power of Gee's theory, it is simple and clear, which is crucial for this tool's aims.

I additionally use the activity-goal alignment score to split all games into two groups: low activity-goal alignment and high activity-goal alignment, using the process I outline below. These two bifurcations can be combined in a 2x2 matrix which yields four categories of learning games. This matrix, shown in Figure 40, is the AGA-Based Assessment Tool.

	Low AGA	High AGA
Lower Order Thinking	1. Digital Worksheet	2. Learn by Doing
Higher Order Thinking	3. Pointsification	4. Deep Learning

Figure 40: The AGA-Based Assessment Tool is a matrix diagram relating lower/higher order thinking to low/high AGA. Each cell is named, coloured, and numbered for ease of reference in the discussion. Cell 1 is light red, cell 2 is light green, cell 3 is bright red, and cell 4 is bright green.

I intend this tool to be used by learning game designers with a wide range of experience and skill. I have aimed the following procedure to be suitable for an early-career designer who has read some literature and has a specific learning game, or design, in mind. To use the AGA-Based Assessment Tool, the designer answers two questions about a learning game:

1. Name the game's **primary learning goal** and decide: is it higher order or lower order? If the game primarily aims to teach mechanical calculation, memorisation or other forms of rote

learning, or motor skills (e.g. accurate aiming and clicking with a mouse), it has a lower order learning goal, and so belongs in either category 1 or 2. For other types of learning goals, and especially if the primary learning goal is to master a system that requires higher order reasoning or thinking such as economics or military strategy, it belongs in category 3 or 4.

2. Name the player's **primary activity**, and decide: is there high activity-goal alignment, or low between this activity and the learning goal named in step 1? In chapters 4 and 5 I explained how to identify a game's primary activity. While the AGA Scoring Tool can be employed to help make this assessment, for some games it is not necessary. A shorthand method may be sufficient, such as this: One reflects on their play experience, or imagines it if the game is not built yet: What does the player typically do in the game? Are they running around shooting things? Are they primarily exploring an abandoned factory? Are they calculating odds, predicting behaviours of opponents? The game's primary activity is often its core mechanic. If the primary activity is not similar to its learning goal, it has low activity-goal alignment³⁰.

If this procedure seems insufficiently structured or defined, or is otherwise unclear for a specific game design, I recommend first scoring the activity-goal alignment of the game using the AGA Scoring Tool. As discussed in chapter 4, that tool's process is more specific and aims to clarify related ideas, e.g. what constitutes a primary activity, and how the learning goal relates to it. After scoring the game, the 2-step procedure given here may then be sufficient.

If it is unclear how to use the tool in more complex situations (e.g. learning games with multiple learning goals, split activities, or lacking win/lose states), see the Discussion section of this chapter.

7.2.1 The Four Categories of Activity-Goal Alignment-Based Assessment

The answers to the two questions in the procedure above, in combination, assign a game to one of the tool's four categories. Here I state the meaning and reasoning behind each category, briefly discussing example games and relevant theory.

Games in the **first category, "digital worksheets,"** for learning game design research.

³⁰ For additional information and guidance on deciding if a game is high or low AGA, see the discussion and worked examples in Chapter 6. For additional information and guidance on deciding if a game has high or low order learning goals, see the discussion of high order learning theory in Chapter 2.

	Low AGA	High AGA
Lower Order Thinking	1. Digital Worksheet	
Higher Order Thinking		

Figure 41: The first category of the AGA-Based Assessment Tool is named “Digital Worksheet.” The background of this category is intended to be coloured light red.

“Digital worksheet” games have both lower order thinking and low activity-goal alignment, as shown in Figure 41. The term “digital worksheet” is commonly used in the literature to describe games that aim to teach rote memorisation or simple calculations, much like classroom drill worksheets. *Math Blaster* fits this category, as I discuss later in this chapter. This category is not always a bad choice, but designer/researchers should defend this choice carefully, as it runs counter to the prevailing advice to “aim high”. Digital worksheet games are simple to design and offer clear measurement of efficacy. Digital worksheet games provide some extrinsic motivation, which might be some improvement over paper worksheets, even if not comparable to the extremely high motivation levels seen in commercial video games.

“Digital worksheet” games typically require motivation from outside the game itself. For example, construction workers are required to memorize hazardous material symbols. A simple matching video game can make this tedious task more fun and improve training efficacy; however, this game would not aim to be as intrinsically motivating as commercial games. So while designers of hazardous material training are justified in building a “digital worksheet” game, researchers should aim higher. This type of learning game has been built and studied for over twenty years now. We should not mistake this weak motivation for the powerful intrinsic motivation that commercial video games typically provide. Generally speaking, these games are the “low hanging fruit” of Devlin’s argument and Devlin opposed building more such worksheet math games.

“Digital worksheet” game designs have been argued to do harm to players (Rice 2007; Osterweil 2006; Bogost 2011b; Robertson 2010). I discuss two types of potential harm here. First, they reinforce the idea that learning games do not deeply engage in their subject matter, which set player expectations low. This in turn makes students less willing to regard learning games as serious teaching tools. Second, they imply that learning is merely a tedious chore that gets in the way of fun. This attitude conflicts with theorists who have argued that powerful learning is fundamentally playful (Osterweil 2006).

In colouring the background light red, I suggest caution is warranted. This tool essentially advises designer/researchers who build digital worksheet games to defend this choice carefully, given the prevailing advice to aim high in learning game research. These games are relatively easy to design, and Bogost criticized all designers who aim to profit from the hype of games without expending effort (Bogost 2011b). I do not go that far, but argue that most research featuring “digital worksheet” learning games is not sufficiently novel to warrant research resources. In short, I agree with Devlin (2012c): “Let’s move on.”

Games in the **second category** in the AGA-Based Assessment Tool are often good examples of learning game design. They combine lower order thinking with high activity-goal alignment score, as shown in Figure 42.

	Low AGA	High AGA
Lower Order Thinking		2. Learn by Doing
Higher Order Thinking		

Figure 42: The second category of the AGA-Based Assessment Tool is named “Learn by Doing.” Its background is coloured light green.

These games teach lower order types of knowledge via their core mechanic. For example, vehicle simulator learning games aim to develop tacit understanding of vehicle dynamics and controls. Trainees must memorize control locations and gain experience with the controls’ effects on the vehicle in practical situations.

These games provide skills and/or tacit knowledge through practice. After playing, trainees might know the following:

- It’s easier to balance a bike if you go faster.
- Don’t drive tractors along steep slopes; point it up or down the slope.
- This jet fighter doesn’t stall in vertical climb.
- This car requires 10 meters per 10km/h to stop.

The four sentences above merely summarize a key point from an experience. Experiential knowledge cannot be obtained solely by reading a book. Games offer opportunities beyond traditional training simulations because they enable the human to “map” their experience to domains outside their

physical body's scope, as Gee argued in his discussion of "microcontrols" (Gee 2009a), which I discussed in chapter 2. The core mechanic of games in this category is often a complex simulation of a physical real-world event or phenomenon. Flight simulators, driving instruction, and heavy vehicle trainers all use complex physics simulations to teach lower-order skills. However, compared to higher order learning goals like economics, this category of game aims to teach simple forms of knowledge through its core mechanic.

This category's learning goals are constrained to lower order, which means that many of Gee's principles of "deep learning" are not able to be fulfilled, so this category is distinct from "deep learning" category. For example, the first deep learning principle asked:

Property 1: Does game play allow and encourage the player to "psych out" and take advantage of an underlying rule system to accomplish personally held goals to which the player is emotionally attached?

Clearly, a lower order learning goal does not fit with idea of "psyching out" a rule system. One does not "psych out" a flight simulator in order to learn to fly a real plane.

	Low AGA	High AGA
Lower Order Thinking		
Higher Order Thinking	3. Pointisification	

Figure 43: The AGA-Based Assessment Tool's third category is named "pointisification."

The **third category, "pointisification"**, is coloured bright red because it runs counter to prevailing advice in the learning game design literature, especially including Gee's "deep learning" theory. Pointisification games have higher order learning goals being addressed by unrelated activities. Often these activities are simplistic extrinsic mechanisms used by video games to measure progress, reward, or incentivize, as Robertson (2010) described clearly.

I use Robertson's term "pointisification" for this category over the more common term "gamification" because the debate about gamification has broadened the term's meaning to include too wide a variety of games (Mullany 2012). However, this category could be named "gamification"

if one used Bogost's definition of gamification in his "Gamification is Bullshit" argument (Bogost 2011b). Bogost's term "exploitationware" (Bogost 2011a) also fits this category. "Chocolate-covered broccoli" (Osterweil 2006) is a closely related idea that is in common usage in the learning game literature. None of these existing terms are an exact fit, as they do not distinguish using only high / low order thinking and high / low activity-goal alignment, but they substantially agree on the overall meaning.

Games in this category feature shallow mechanics that do not meaningfully embody the learning goal. One example is *Multiplayer Classroom*, a so-called learning game which adopts game design principles to improve motivation without substantially changing the learning activities in the classroom (Sheldon 2011). To understand why Sheldon's design fits this category, start with Gee's observation that games have unique properties that suggest a new approach to education (Gee 2003). *Multiplayer Classroom* does not adopt Gee's ideas, and for good reason: a game that truly embodied the potential of video games would conflict with standard classroom procedure. Since the designer cannot typically change the rules in the classroom, the designer works within these constraints, adopting game-style points system in the classroom, and found that it increased motivation. This same dynamic occurs in many corporate gamification projects, argued Bogost (Bogost 2011b). Sheldon found that this design works: students' motivation is increased. Similar findings are commonly employed in defence of gamification (Bunchball Inc 2011), and are typically not disputed by its opponents of gamification.

If this type of design works, what is wrong with it? Robertson explained the problem: by using game language, pointsification designers promise an experience that contains the psychological, emotional and social power of a great game (Robertson 2010). Robertson stated this will ultimately lead to disappointment when the experience doesn't deliver on that promise: "the only rational thing for them [the designers] to do is to turn round and say 'Games don't work! We gamified the dickens out of this thing, and it still didn't make as much money/reach as many users/generate as much social heat as World of WarCraft/ Farmville/ Minecraft'."

In other words, and as I argued in the previous chapter, pointsification games spend the social capital, or player goodwill, that other games earn. While some small parasitic activities probably will be tolerated, such applications cannot scale too much. It is possible to exhaust the reputation of games, which might explain the rise and fall of the term "edutainment" (Klopfer, Osterweil, and Salen 2009). Robertson's argument that bad gamification parasitically damages the brand of games (Robertson 2010) largely supports this position.

The **fourth category**, “Deep Learning³¹,” signifies good examples of learning game design.

	Low AGA	High AGA
Lower Order Thinking		
Higher Order Thinking		4. Deep Learning

Figure 44: The fourth category is named “Deep Learning.” The background of this category is bright green.

As shown in Figure 44, this category identifies games that teach higher order learning goals using activities that are well aligned to those goals. In the previous chapter I found that *GameStar Mechanic* teaches system thinking and other higher-order thinking goals through a reasonably well aligned player activity, so it fits this category. It is very difficult to design a deep learning game, as Bogost discussed (Bogost 2011a). Gee stated games that are both fun and teach certain specific higher-order learning goals are rare, despite many years of attempts, which indicates that it must be difficult to design and/or build them (Gee 2009a). I imagine an amateur designer might aim to build a deep learning game by creating a game mechanic that comprises a higher-order learning goal. I have found that this simple-sounding task is much harder than it sounds, and is likely to be completed only after a long design process. During this process, the designer is likely to use tools more complex than either the AGA Scoring Tool or AGA-Based Assessment Tool. It is here that more subtle, refined approaches to game design, such the six properties Gee proposed (Gee 2009a) are most useful.

I argue that the categorization in the AGA-Based Assessment Tool implicitly advocates for excellence in learning game design, as identified by many leading game designers. While I do not go so far as to map specific attributes from prior research to these categories, I do argue that common advice about learning game design maps well to these categories. For example, Halverson (2005) clarified previous taxonomies of learning games (Malone and Lepper 1987) by identifying games with generic mechanics suitable to any low order learning goal as “exogenous.” Since exogenous games have low order learning goals and low activity-goal alignment, they belong in category 1. Halverson continued this work, using the term “endogenous” to describe games that target more complex learning goals

³¹ I have used Gee’s term “deep learning” (Gee 2009a), even though Gee did not exclude games with lower order learning goals, as this category does. Gee’s deep learning can occur in a flight simulator, for example.

beyond memorisation, in part by integrating learning content into the structure of the game. I related endogenous games to high activity-goal alignment in chapter 2. Since they have high order learning goals, and high order alignment, endogenous games clearly belong to category 4. Earlier in this chapter I argued that Gee's deep learning principles steer designers toward building games that fit in category 4. Rice's Cognitive Viability Index (Rice 2007), broadly interpreted, suggests designers aim to build category 4 games, which is unsurprising since Rice and I both employ higher order learning theory. Finally, Klopfer et al (2009) offered several principles to learning game designers. Some of their advice falls outside the scope this research, but the relevant points fit well. For example, they advise designers to "put learning and game play first:" one cannot prioritize either learning or game play over each other. This is essentially an argument for high activity-goal alignment. They also argue to "find the game in the content," stating: "In designing any learning game, it is all too tempting to simply graft academic content onto existing forms [of video game mechanics]...Dimension M grafts algebra onto a first-person shooter" (p. 94). This is clearly an argument condemning low activity-goal alignment.

In all of this literature one can see more than impartial differentiation. As I argue in this chapter, prevailing advice advocates for both deep learning and high activity-goal alignment. By making category 4 bright green, I intend to reflect the literature's prevailing position: advocacy of deep learning, and advocacy of high activity-goal alignment.

7.3 Activity-Goal Alignment-Based Assessment Of Three Math Games

To show this tool in action, I discuss several games that aim to teach mathematics to children. I investigate how the designers see their games, and compare their view to this categorization's critique of the games. I use the tool to clarify the relationship between *Equatia*, *HyperBlast*, and *DragonBox*. Specifically, I argue that one recent math learning game, *Equatia*, shares *Math Blaster's* design problems, while another, *DragonBox*, does not.

7.3.1 *HyperBlast*

As discussed in the preceding chapter, I find that *HyperBlast* (Knowledge Adventure 2012b) has lower order learning goals and very low activity-goal alignment, as shown in Figure 45. Thus I assign it to the first category: *HyperBlast* is a "digital worksheet" game.

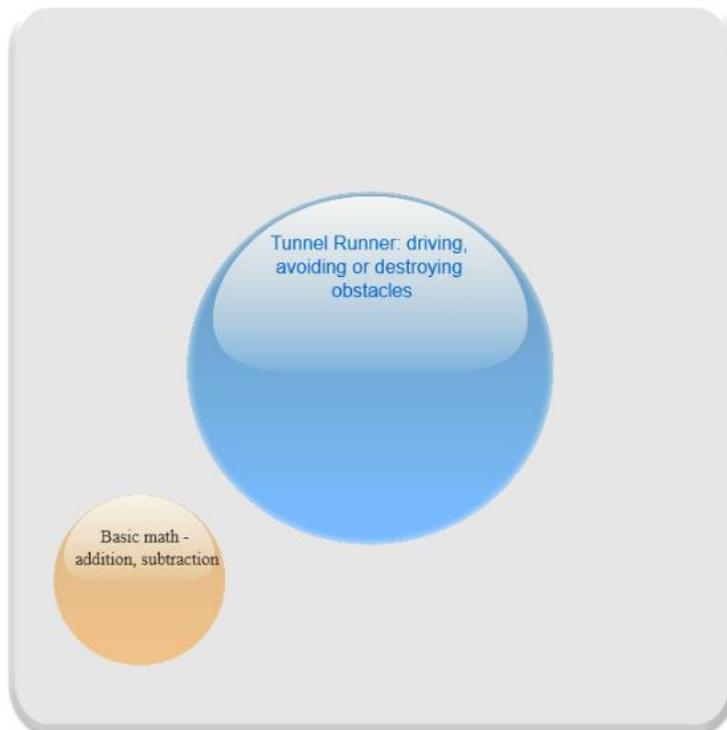


Figure 45: *HyperBlast*, a Math Blaster branded learning game, does not relate its tunnel running mechanic to its learning goal of arithmetic practice

In my explanation of the first category of the AGA-Based Assessment Tool, I stated that digital worksheets can range from good to bad, depending on context. Thus, to pass a negative judgement on a game, it is insufficient to assign it to the first category. I must take a position and defend it for the specific game in question, in light of the guidelines provided. My analysis led me to conclude that *HyperBlast* has two fundamental design limitations: weak learning goals and the divorce of the learning goal from the gameplay. I recall that these two weaknesses were observable in the original 1983 version of *Math Blaster*.

Hutchison (2008) argued that games should be understood in the context of the history of the development of their underlying technology (the computer), which afforded their developers different options at different times. His term “techno-historical” refers to this combination of technology and history in video game criticism. I believe it is clear that the techno-historical landscape must play an important part in assessing any long-lived game series such as *Math Blaster*. Few games have survived over 25 years, as *Math Blaster* has. In 1983, *Math Blaster*’s design was arguably constrained by the limitations of the player’s computer’s performance. Without any real competitors and much less clear literature on math game design, *Math Blaster* was not a terrible design for its time. However, computers in 2012 are far less constraining for math game designers than they were in 1983, yet *Math Blaster*’s design has not changed substantially.

In 2012, Devlin stated that most math games use the same mechanic as *Math Blaster* (Devlin 2012b). I have also observed that there are many math games with almost exactly the same learning goal. The designers of the *Math Blaster* sequel *HyperBlast* did not address these weaknesses, despite the numerous critiques in the literature. Instead, they only made it more visually compelling. Devlin (2012b) noted that *Math Blaster* long ago plucked the low-hanging fruit of arithmetic practice. Today that branch is now bare, yet many developers continue to seek fruit there. Given *Math Blaster's* unique position in the market, it is disappointing that 25 years later, the developers do not aim higher than refreshing the lipstick on an old game.

HyperBlast does enable children to practice arithmetic, and in that limited sense it does help players learn some small part of mathematics. However, I argue that this small good is overshadowed by the long-term harm that shallow learning game experiences like *Math Blaster* do, by damaging the reputation of learning games, as I discussed earlier in this section. Specifically, I believe that every version of *Math Blaster*, which is perhaps the most popular learning game in the world, damages the brand of all math learning games by reinforcing the public's low expectations that math learning games just practice rote memorizations. I previously quoted adults who cited their early experience with *Math Blaster* when criticising learning game design from a consumer perspective. If the public today generally considers *Math Blaster* the definitive math game, and recognises it as the shallow experience it is, then *Math Blaster* has done harm indeed.

7.3.2 *Equatia*

Equatia is a recently published math game whose design is discussed in a Master's thesis³² from the School of Education at Stanford University (Fukumoto 2011). A screenshot of the game during a typical play session is shown in Figure 46.

³² Fukumoto published a "condensed version" of his thesis. My comments here are based only on this version as I was unable to obtain the full text of Fukumoto's thesis.

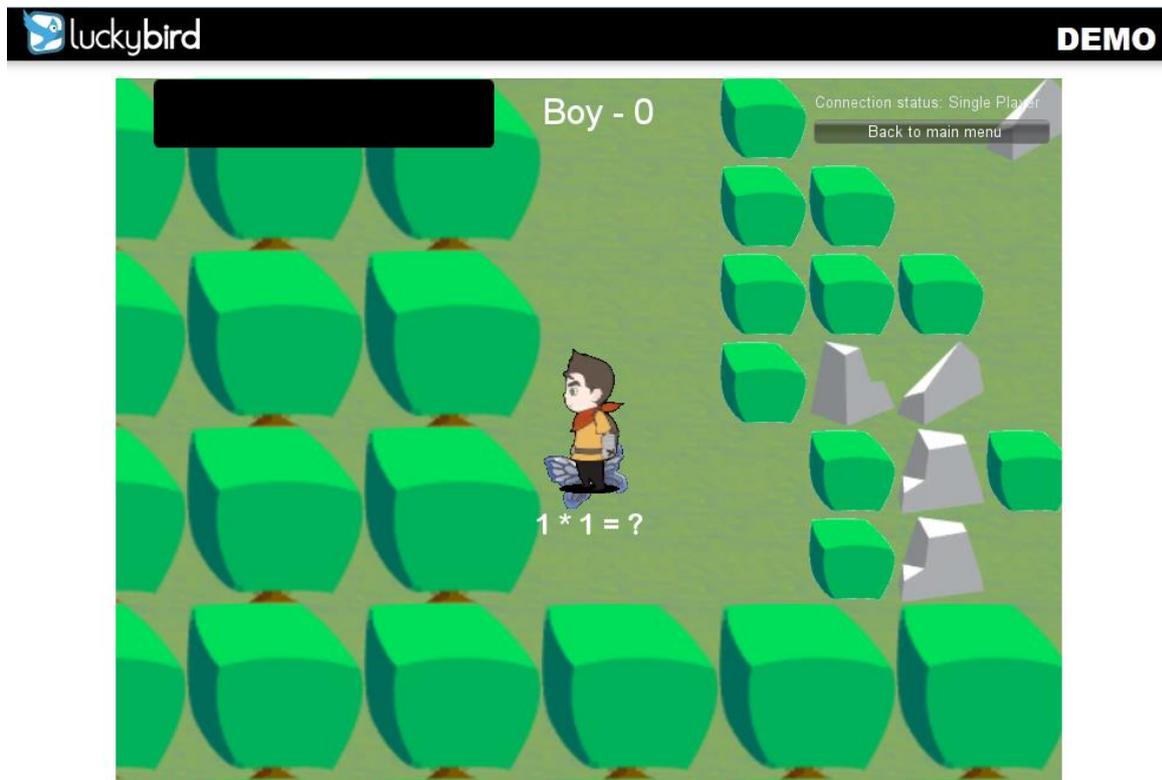


Figure 46: Screen capture of Equatia's game play, showing a 2D scrolling adventure game with a disembodied floating arithmetic question. (Fukumoto 2011)

I downloaded and played through the *Equatia* demo on a PC, and observed two different children playing it briefly. I also watched users' demo videos (*Equatia Demo, v.1 Run-Through* 2010), the development website (Luckybird Games 2012), read the online dissertation summary provided by the developer (Fukumoto 2011), and read all online discussions I could find. The following is a summary of my play experience.

The instructions for *Equatia*'s controls serve as a summary of player activity: "Move with arrow keys or ASDW. When questions appear, type in [the] correct answer and hit Enter to attack." In these instructions, the vast distance between gameplay activity and goal are clearly observable. Gameplay and learning are divorced, just as in *Math Blaster*. Correct answers activate the player's sword, which cuts down enemies and trees, creating pathways to access special items. The game activities are familiar to almost all gamers: explore, destroy enemies, collect items, and attack monsters. However, none of these activities are available unless you type the answer to a math problem. Sometimes there is no time pressure, and other times the player's character is being hit by an enemy while she tries to type the answer to twelve times nine. In short, *Equatia* offers fun game activity interrupted by factual questions that are obviously unrelated to the game's story and mechanic.

Fukumoto argued that there were several differences between *Equatia* and “edutainment” games. I observed most of these differences in my play experience: In *Equatia*, the player answers math questions constantly, rather than having the action freeze during questions, as *HyperBlast* does. In *Equatia*, the player must type answers instead of clicking multiple choice. *Equatia*’s core mechanic is slow-paced exploring instead of a fast-action tunnel game, so the contrast between answering a math problem and exploring in the game is less jarring than in *Math Blaster*. However, are these improvements significant enough to justify research resources? I did not find the argument convincing.

Unlike *Math Blaster*, *Equatia* is accompanied by a coherent presentation and defence of its learning goals. *Equatia* does not aim to teach active learning or higher-order thinking. In his thesis, *Equatia*’s designer Fukumoto (2011) scopes his goal tightly: “[...] that all students have the foundational arithmetic skills needed to succeed.”

Fukumoto’s clear statement of the low-order learning goal and the obvious lack of alignment between that goal and the clarity of the primary play activity (2D scrolling adventure) made *Equatia* easy to categorize using the AGA-Based Assessment Tool I developed in the previous section. I used the AGA-Based Assessment Tool to identify *Equatia* as a digital worksheet, which is category 1 shown in Figure 40.

As with *HyperBlast*, it is not sufficient to critique *Equatia* by merely assigning it a category. Like *HyperBlast*, I found that *Equatia* has two fundamental design limitations: weak learning goals, and the divorce of the learning goal from the gameplay. However, unlike *HyperBlast*, *Equatia* is a modern design by an academic who presumably had the incentive, means and opportunity to consider recent literature.

Devlin clearly stated that there was no need for more math games that practice basic computation skills. To me it is clear that *Equatia* did not follow this advice. Unlike the authors of the *Math Blaster* series, Fukumoto states that he is aware that this approach is not recommended practice in the literature: “While exogenous mechanics are generally criticized in educational games, I believe in this case it is appropriate.” His reasoning is unclear to me. Why does he believe it is appropriate? I offer some possible reasons in the Discussion section.

7.3.3 *DragonBox*

DragonBox (WeWantToKnow 2012) is a game that aims to teach algebra, and is being hailed as a new leap in math learning games (J. Liu 2012). I played *DragonBox* on an iPad 1, and observed an

adult, a 12-year-old, and a 6-year-old play it, as well as observing publicly accessible video recordings of others' play experience.

The game starts as an abstract card-like single player game. The first screen has two trays, each containing a number of cards, as shown in Figure 47. One of the trays contains a box with a star on it, sparkling and glowing. The player is instructed to isolate the box by tapping green cards, which vanish when tapped. This activity alone is sufficient to beat the first three levels. Then the game adds a new element to the mechanic: "night" and "day" versions of cards are introduced. When overlaid they become green swirl cards, which the player taps to erase.

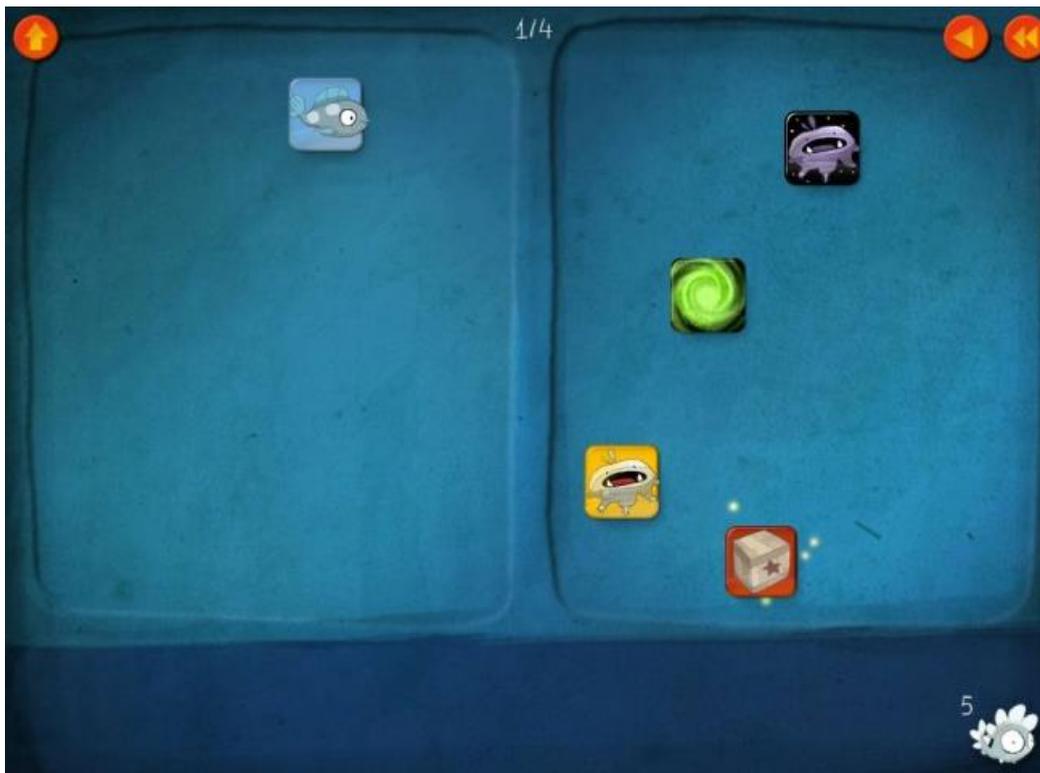


Figure 47: DragonBox begins as a "math-free" zone: the play area and symbols are not obviously math-related. (WeWantToKnow 2012)

After several levels of moving cards around and tapping on swirls, the game adds a few cards at the bottom. The player drags these onto the trays — but whenever she drags a card onto one side, she also must drag a copy to the other side as well. This simulates adding the same number to both sides of an equation. Later, the player can flip these extra cards from day to night (and vice versa) before dragging them onto the trays.



Figure 48: *DragonBox* mid-game uses a linear layout. (WeWantToKnow

As the game progresses, it gradually transitions to traditional algebraic symbols (Figure 48, Figure 49). For example, cards are positioned above and below each other, with a bar in the middle. These, of course, are fractions — multiplication and division — but the player doesn't need to know that to play the game. Later the sparkling box is replaced by a little white sparkling square with an "x" on it. So the player is now isolating the "x." Sometimes instead of pictures there are letters on the cards like "c" or "b" and sometimes there are things that look like dice, white squares with some number of dots on them.

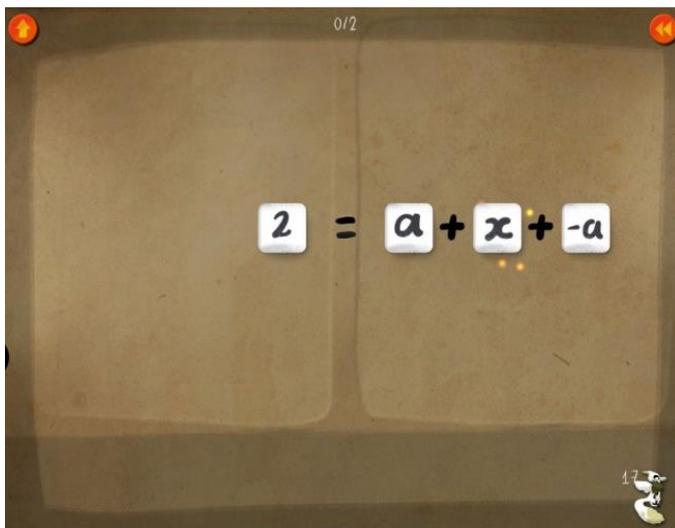


Figure 49: *DragonBox* ends up presenting algebra using standard symbols. (WeWantToKnow 2012)

Regarding its learning goals, *DragonBox*'s designer clearly stands opposed to Fukumoto's position: "Teaching algebra from arithmetic, concrete to abstract is to my mind crazy [...]" (jbhkb1 2012). In my play experience, the game successfully tutored me through the fundamental activity of algebra:

the balancing, the finding of equality between two separated but mysteriously related sides. In a lengthy online discussion of *DragonBox* among users and educators on a reputable online forum ("YCombinator Discussion of *DragonBox*" 2012), a range of learning outcomes were given:

- *The game highlights the malleability of numbers. For me, this was the deep revelation about algebra, and it took me a long time to get there. The way the game is designed using touch, lets you fling numbers around and place them on top of each other and such, and doesn't let players get hung up on numbers or placement. We get to the understanding that algebra requires very quickly.*
- *One of the things my eldest daughter struggled with early on in Algebra was the variables. She kept insisting on having a 'value' for the variable up front because the abstraction bothered her...This program elegantly sidesteps that issue by starting off with boxes. Boxes are the real world equivalent of variables and they aren't numbers so they don't trigger that association per-maturely [sic].*
- *[in DragonBox] the rules of algebra have been mapped, transparently and isomorphically, to the game rules. Playing the game doesn't just help you memorise the rules of algebra, it actually makes you think about how you use those rules to manipulate and transform expressions, which is the fundamental skill underlying "real" algebra*
- *[...] initially ignoring "this is math" which lets players avoid the mental barriers they may have erected about that particular subject. Being able to "do" first makes explaining the "why" later much easier and more interesting.*

The activity-goal alignment for *DragonBox* is very high, as illustrated in Figure 50. One might criticise this activity-goal alignment score because of the scope of the game's learning goal: teach algebraic thinking. Gee and others have argued that some authentic activity, or at least an alternate framework, is needed for deep learning. Because the game completely focuses on the balancing of icons, and the representation of variables beyond conventional symbols, there is always a danger of users progressing in a game without really understanding algebra itself. Clearly this game cannot teach the entirety of an algebra class. The designer acknowledged the general difficulty of setting the scope of a learning game but did not justify the particular scope he set in relation to *DragonBox*'s design (Huyhn 2012).

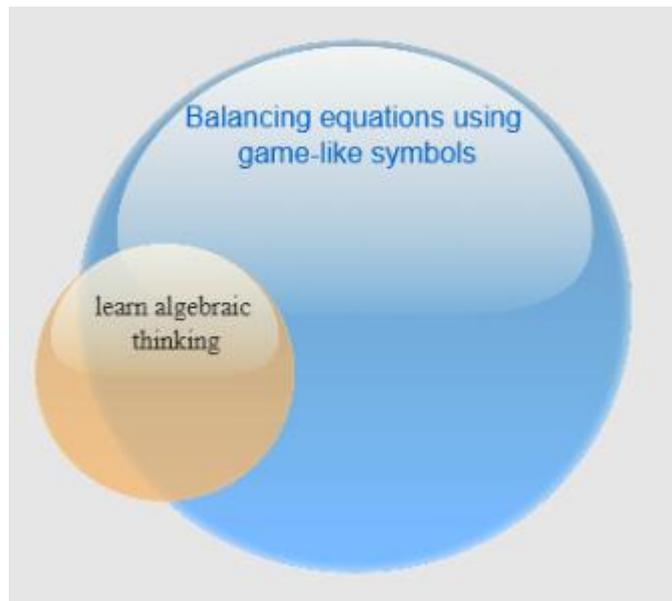


Figure 50: *DragonBox* has strong activity-goal alignment.

Nonetheless I assess *DragonBox* as a Deep Learning game because I believe learning to balance algebraic equations is a sufficiently high order learning goal to qualify. As I show in Figure 50, I find it has reasonably high activity-goal alignment. This puts it in category 4 of the AGA-Based Assessment Tool: deep learning.

	Low AGA	High AGA
Lower Order Thinking	1. Digital worksheets (<i>Equatia</i> , <i>Math Blaster</i>)	2. Learn by Doing (flight simulators)
Higher Order Thinking	3. Pointsification (<i>Gamified Classroom</i>)	4. Deep Learning (<i>DragonBox</i>)

Figure 51: Output of the AGA-based Assessment Tool. The three games in boldface are discussed in this chapter.

I critiqued *Equatia's* design because its learning goal and activity are insufficiently novel to generate significant advances in our understanding of learning games. specifically, arithmetic practice is well addressed by existing games (Devlin 2012a). I find design of *DragonBox* is more appropriately ambitious and focused on enabling autonomy: he wants the user to play with an important concept in mathematics. Ironically, given which one was funded research, *DragonBox* has been an unusually successful commercial product. The developers claim it outsold *Angry Birds*, one of the top-selling video games on the Apple App Store, in Norway for a brief period of time (J. Liu 2012). This is very

impressive performance for any game, let alone a learning game that is experimenting with new approaches to teaching math.

7.4 Discussion

In most of this section, I discuss the implications of the findings above, in relation to a broad consideration of the learning game design literature and my dissertation's overall purpose. First, only in the following subsection, I identify common points of confusion and recommend approaches for amateur designers to use this tool.

7.4.1 Discussion of the Practical Application of the AGA-Based Assessment Tool

- **Multiple learning goals:** Many learning games have more than one learning goal, yet this tool allows for only one to be considered. The designer should choose the most important learning goal. They may also choose to analyse the game multiple times, if the game has several equally important learning goals. I gave an example of this in my discussion of *GameStar Mechanic* in chapter 5.
- **Multiple or temporally-split activity games.** Some learning games have more than one core activity. In this case, the designer should choose the primary, most memorable activity that comes to mind when they reflect on their play experience. The process of unprompted recollection of an experience removes the irrelevant or least meaningful elements (Holzman and Gardner 1960). This filtering can be useful in this context. For example, in my discussion of *HyperBlast* in this chapter, I noted that the game has two activities: answering math questions, and tunnel running. By simply recalling my play experience, I found it was clear that the primary activity is the tunnel running mechanic. I spent the most time in that part of the game, and it was the most memorable and engaging. The other activities felt like supporting, or of secondary importance, to my play experience.
- Some large-scale games have **multiple, unrelated play modes**. For example, *GameStar Mechanic* lets the player either progress through a storyline, or build their own games using a construction set. This assessment tool is only relevant to one activity at a time. Consequently, the designer may wish to assess multiple areas of such a game separately.
- Note that this tool is intended **primarily for learning games**. If you are considering an e-learning application, persuasive game, simulation, or any software that does not have typical

learning game features such as learning goal and a win/lose state³³, this tool is unlikely to be helpful.

From practical considerations, I now turn to the broader question: what is the value of this tool?

7.4.2 Value of AGA-Based Assessment Tool

Consider the output of the AGA-Based Assessment Tool shown in Figure 51. In comparison to the preceding discussion, Figure 51 makes three points apparent:

1. One can immediately see a big difference between *Equatia* and *DragonBox*.
2. It is apparent that *Equatia* is not as ambitious as *DragonBox*, as it aims to enable practice only.
3. The underlying theory and reasons are very clear.

It might seem like this tool is very simple, perhaps even crude. It does not discuss subtleties; it simply forces games into quadrants. This crudeness might seem distasteful when compared to Gee's elegant deep learning properties, but particularly in consideration of the shortcomings of existing assessment tools discussed in the previous chapter, I argue it is useful.

In short, I argue that the combination of simplicity, clarity and theoretical depth shown in Figure 51 is difficult to achieve in an equivalent text discussion, or in a checklist-style format.

7.4.3 Scope

I claim a vast scope for this tool: it can assess any learning game. A sceptical reader might ask: Given the wide variety of video games, is this scope too broad? Specifically, can any tool fairly summarize all of the game design literature? Of course, no single assessment method can represent the entire body of game design literature in detail; however, I have not observed any instance where this tool conflicts with a major position within the game design literature.

7.4.4 Deep Learning in Serious Games is a Major Challenge

The sceptical reader might wonder why, if deep learning is so valuable, why the most popular learning games are “digital worksheets.” In the typical classroom, one rarely finds deep learning games (Corbett 2010), with a few notable exceptions (World Peace Game Foundation 2012). Why not? Specifically, if *Math Blaster* is bad, why is it so popular? I noted previously that the series has sold millions of copies. How could it be bad?

³³ Because there is not agreement in the literature on exactly what a game is, some software labelled “game” may not be within the scope of this research. I discuss my definition of game in the introductory chapter of this dissertation.

Deep learning games are not commonly used in education, despite their apparent advantages. Here I present my summary of discussions found elsewhere (e.g. Klopfer, Osterweil, and Salen 2009), using classrooms as the context.

Most of *Math Blaster's* buyers are teachers and parents, not the players themselves (Knowledge Adventure 2012a), and for reasons remarkably unrelated to game quality: teachers have been found to prefer games with easy-to-understand learning goals, sophisticated marketing campaigns, modest technology requirements, and believe they motivate students to work through problems (Lee et al 2004, cited in Rice 2007). Klopfer et al corroborated these findings, discussing the superficiality of analyses of learning claims among buyers, explaining how the fears and hopes in parents are translated into brand recognition, endorsements, and claims that need have little relation for the actual effect of the product. I note a wide disparity between *Math Blaster's* claimed and actual learning outcomes. This disparity, distressingly, seems not to be noticed by or important to the game's buyers. In other words, there is reason to believe *Math Blaster's* success is despite, not because of its design quality. *DragonBox's* success suggests a possibility for hope in relation to these reasons: perhaps modern buyers are better able to recognise design quality in learning games.

However, there is another important barrier to deep learning games. Halverson stated deep learning games are not appealing to school administrators because they "provide inefficient and unpredictable environments for learning school-based material." He called it a "difficult challenge" because they don't fit classroom-based, assessment-driven pedagogy. He said the games teach through a protracted and indirect process "with a steep learning curve when compared with standard curriculum units on mathematical fractions, Egyptian history, or European expansion." Rice, Halverson and Klopfer et al share a similar dim view of the road ahead for deep learning games. They all noted that educators often buy learning games to achieve low-order learning goals because they are easier to measure and their fine-grained progress integrates more easily into standard school schedules (Halverson 2005; Klopfer, Osterweil, and Salen 2009; Rice 2007). This problem is not limited to educational contexts.

Bogost noted that deeply embracing a gamerly approach is difficult because it threatens the status quo (Bogost 2011a). He discussed how a gamerly approach might require reinventing customer interaction, not merely scoring points for closing sales deals. Gee asked if it is possible to build deep learning games: "I would argue that if we are to make deep serious games that really use the power of gaming, then these [deep learning] features will have to be present and implemented well. In the end, I am not sure this can always be the case when we leave the domains (content) usually covered in entertainment games"(Gee 2009a). However, Gee offered some hope: "Games do not have to be

long and complicated to meet the conditions—*Diner Dash* meets them and is neither.” Gee noted that *Sam and Max* series of games met none of these conditions and still may teach something, but comparing *Sam and Max* to *Diner Dash*, he stated:

[...] but Diner Dash does more—in addition, it makes the player embody and empathize with a set of connected problems (a problem space) connected to a certain identity or way of being in the world. (p.78)

Here we can see Gee suggesting that these games work because of a multivariate alignment of various elements of game design. In this case he advocated for greater overlap between player empathy, the game's affordance of identity, and a set of connected problems. One might see activity-goal alignment as a simplified version of this way of thinking.

To restate, achieving deep learning in a serious game is a worthy challenge for designer/researchers, and many theorists before Gee have suggested that great rewards will come to those who master it. Gee has shown how multiple fields are moving closer toward developing a sufficiently clear understanding of this challenge that we can hope to conquer it in the future. While acknowledging that this has been argued for some time, I still believe it is true. I also believe that the alternatives, such as building less ambitious games, are worse, as I have argued in this chapter.

7.4.5 Assessment as Advocacy

Throughout this thesis, I have made the advocacy of activity-goal alignment apparent, and defended it as the prevailing view of the literature. In addition, the design of the AGA-Based Assessment Tool implicitly advocates Gee's position about the importance of deep learning in game design. Many existing assessment tools are not so clearly aligned with any particular theory and it might seem counterintuitive to advocate a position using assessment, when compared with a legal judge deciding the fate of a criminal. The judge is meant to objectively identify the truth in the various arguments presented. Similarly, it might seem like an assessment tool should discover which learning game is objectively better than another. Is advocacy appropriate in an assessment tool?

I argue here that the model of the legal judge, is not appropriate in learning game assessment. One cannot pretend assessment is merely a question of factual accuracy: even in the context of traditional assessments such as classroom tests, “Assessment tasks and standards consist of series of value judgements which are never independent of particular political and social contexts” (Tuck 1995). In other fields, advocacy is discussed explicitly, as part of assessment design (De Leon 1990), but not all learning game assessments state their basis or position, as the preceding chapter revealed. I argue that learning game assessment criteria should not be unknown, unstated, or

hidden behind an all-purpose curtain of supposed “objectivity.” In other words, I feel that learning game assessments should take a position in the literature and advocate a particular theory.

Especially for designers and researchers, assessments that argue a theoretical position are not only acceptable, but significantly superior. In chapter 6 I argued that existing assessment tools that explained their reasoning were better than those that didn’t. As a designer conversant in the literature, I found only assessments with solid theoretical or logical foundations felt sufficiently trustworthy and convincing to bother using.

Of course, not everyone will agree that assessments should advocate, but I do not feel this inevitability merits much discussion. However, I briefly discuss the views of a few of the most likely dissenters here. Some proponents of gamification may take issue with the negative position implied by the assessment tools in this research, even though some proponents have stated that high alignment is valuable (Mullany 2012). Of course, they may disagree with the priority of activity-goal alignment theory in comparison to other design ideas. However, I believe many defenders of gamification would support the foundational theory embedded in this tool, even if they might not like the implications of the AGA-Based Assessment Tool.

Designers and researchers of “digital worksheet” games might take issue with my suggestion that such games are poor examples of learning game design. Indeed, after advising designers not to repeat known problems, e.g. “Dimension M grafts algebra onto a first-person shooter,” Klopfer et al stop short of attacking such games: “While these projects may work, we prefer a different approach” (Klopfer, Osterweil, and Salen 2009). Even Devlin, who fiercely advocates against such designs (Devlin 2012a) acknowledges that they can be effective. In response, I repeat that I do not condemn all digital worksheet games as ineffective. Instead this thesis argues that digital worksheets do not have the potential for important advances in learning when compared to deep learning games. There are many perfectly commendable commercial applications of digital worksheet game designs. I do suggest that the justification arguments for additional research on digital worksheet games should be particularly rigorous in light of the existing work.

7.4.6 For True Innovation, Beware Commercial Customers

I have argued that learning game research should aim higher: designers should strive for major innovation, not just minor increments. How does this fit with market-based innovation? If this innovation is not demanded by customers, should it be pursued anyway? I argue “yes,” using the following example. In 2011 the game designer of *Equatia*, Fukumoto, collaborated with Learning Lab, a technology-based component of a private charter school (Rocketship Education 2012) that stated

they focus on increasing test scores by improving skills against specific "micro-standards." They stated: "[...] our 2-4th graders use *Equatia* for the first 10 minutes of their math lab block for fact fluency practice (a whole block is typically 35 minutes)." Fukumoto stated that *Equatia*'s design was substantially affected by Rocketship partnership, but also that his original overarching design remained intact.

By taking "significant design input" (Fukumoto 2011) from his education partner in his game's design, Fukumoto took a customer-focused approach. There is evidence that commercial ventures are more successful if they heed their customers (Jaworski and Kohli 1993). However, this evidence primarily focuses on successful versus failed commercial ventures, not research. Slater and Narver (1998) state that focusing on customers can mislead designers of innovative or disruptive products. They note a "chorus" of literature advising such designers to lead, not follow, their customers. However, the literature is complex: cocreation experts find that including sophisticated customers in design process can lead to innovation (Candy, Amitani, and Bilda 2006). However, given that Rocketship has clearly stated that their focus is on "fact fluency," perhaps overreliance on customer input is relevant to understanding why *Equatia*'s design conflicts with prevailing advice.

Few educators will welcome the disruption that deep learning games require in the classroom because of the systemic barriers Klopfer, Osterweil, and Salen (2009) discussed. It makes sense that traditional educators continue to demand minor variations on "digital worksheet" learning games; however, I feel this approach will continue to generate the lacklustre results these games have shown for the last twenty years. To address flaws in today's education system, learning game designers should be accepting design guidance from those who are not vested in this system and/or in a position to challenge its foundations.

7.4.7 Is the Modality, Not the Message, the Problem?

In this chapter I addressed the need for alternative modalities found in the preceding chapter by developing the AGA-Based Assessment Tool. Here, in light of the example provided, I more broadly discuss the problems with modality of most learning game literature.

As an author of academic learning game research, I sincerely hope practising learning game designers read the existing literature. I also have sympathy for practising designers who fail to notice a clear position from the literature, or fail to see how to apply the reasonable-sounding advice. It takes significant time and effort it takes to both read and assimilate a prevailing view of the literature. Further, I have found it is easy to bend general ideas to one's existing worldview, and how difficult it is to get guiding principles clear enough in one's head to hone one's design to the

principle, rather than bending the principle to the design. For this audience, I believe the merits of this research are clear.

However, the implications for my second audience, designer/researchers, are not as clear. In the previous section I discussed how Fukumoto, an educated, experienced designer/researcher, acknowledged two prevailing views of educational and game design literature, scaffolding and endogenous game design (i.e., high activity-goal alignment), and then explicitly made design choices that opposed them. How it is possible for a well-educated designer to apparently understand the prevailing view in the literature and not follow its advice? I envisage these possible reasons:

- He was not convinced that deep learning was necessary to justify a contribution to game design research. This is the view most clearly expressed in his writing. He regards his improvements as innovative enough.
- His industry partner didn't want real innovation. Rocketship measures learning success using standardized testing of skills (Rocketship Education 2012).

Thus I have sympathy for Fukumoto's design choices, even though I criticise them, because deep learning games are risky: the design and development path is very difficult, and it is quite possible to fail. Even worse, one may not know one is failing until partway through development. This represents a tremendous risk to the scarce resources available to typical learning game researcher/designers. As Klopfer et al stated, "In designing any learning game, it is all too tempting to simply graft academic content onto existing forms" (Klopfer, Osterweil, and Salen 2009, 31). How can we get future designers to take up the challenge of true innovation?

I suggest that one barrier preventing designers from accepting the challenge is dissemination; specifically, the modality of the literature. For many years, deep learning game research researcher/designers have used traditional modalities such as books, journal articles, and lectures to disseminate their theories. However, given that practice of learning game design has not embraced the advice offered, perhaps the problem is related to the mode of dissemination, not the content, of the literature. Words are powerful but slippery and vague. Perhaps designers need simplicity and clarity more than power in their theory.

Where could we look for alternative approaches? We might look at the extreme of pictograms, one of the most ancient forms of written communication. For example, consider the humble smiley face icon. This pictogram is an extremely simplified drawing of a human head: two eyes, a smiling mouth, and a circle for the head itself. It conveys some of the power and impact of a human emotion without being grounded in race, culture, or gender context. By analogy, I suggest that the AGA-

Based Assessment Tool, like the smiley pictogram, excludes all considerations except for two key concepts in game design: "activity-goal alignment" and "higher order thinking". The meaning of its four-cell matrix is relatively difficult to misinterpret, yet it is not grounded in any particular learning game, audience, or play context. This combination of specificity, simplicity, and ungroundedness aims to achieve simplicity, even at the expense of precision, that makes the theory useful to a practising designer.

To the sceptical reader who is not convinced that words' vagueness is problematic for practising designers, here is an example of the weakness of how the word, however carefully written, can fail when applied in learning game design practice. Education scholar McGuinness (1999) discussed the role of technology in developing thinking. He began by clearly identifying "digital worksheets" as an out-dated approach:

In early examples of computer-assisted learning, the pedagogy was largely drill and practice of skills such as arithmetic facts, word recognition and spelling. Recent innovations are more consistent with a thinking skills approach. (p. 3)

This statement would clearly exclude *Equatia*, which aims to teach through drill. However, the next sentence is more ambiguous.

Interactive exploratory environments or microworlds allow students to direct their own learning through discovery and guided discovery processes. (p. 3)

To me it is apparent that the isolated phrase "interactive exploratory environment" could describe *Equatia*'s design. *Equatia* clearly offers interactive exploration of an environment. From there I can easily imagine claiming *Equatia* is an improved learning game. However, I believe McGuinness intended to exclude games like *Equatia* because all of his cited examples have high activity-goal alignment and higher order learning goals: "[...] Geometry Sketchpad for exploring geometric relationships, STELLA for systems thinking, ThinkerTools for teaching the laws of force and motion" (p. 22).

Overall I found McGuinness a skilled wordsmith. Even though I believe that McGuinness's choice of words allows enthusiastic designers to justify designs that conflict with his advice, I do not criticise his writing. Many theorists have stated essentially the same point clearly over the years. Veteran learning game designers Klopfer et al gave "principles and best practices for moving the field [of learning games] forward in a positive direction" (Klopfer, Osterweil, and Salen 2009). They use simple language and make clear value judgements:

The resemblance to a game is meaningless when the activity is nothing more than answering multiple-choice questions and when success is measured solely as the percentage of correct answers given expressed as a "score" and presented with a fun animation. (p. 4)

However, many researcher/designers read, acknowledge, and ignore this advice. Could anyone have written clearer language to prevent another author from misinterpreting their words? I doubt it. I ask if the modality of written text might be the root problem.

Additionally, I argue that the learning game design literature has another problem: attribute-based forms of knowledge, which I argued in chapter 3 is common in the learning game design literature, do not fit how veteran designers think as they design. Cross stated experienced designers synthesize solutions in a holistic way whereas novices tend to rely on more specific rules in their practice (Cross 2004). Schön stated that designers do not deconstruct their decisions as they make them (Schön 1983), a finding that agrees with my position: I argue that it is impractical to skim Gee's "deep properties paper" or ask myself Gee's fifteen questions on six topics for every little game design decision I make. I hope the assessment tool presented here suggests some directions for future research that employ alternate modalities to attribute-based analyses presented in academic textual form.

Finally, I hope it is apparent that the modality of this tool is separate from the theories it advocates. I recognise that there are many other valid theories that might be formed into similar tools, and hope that future researchers do so. I believe there are many alternates to, or improved versions of, the AGA-Based Assessment Tool which could be valuable contributions.

7.4.8 Implications of Key Findings

One key finding of this chapter is that today's learning game researcher/designers should avoid building "digital worksheet" learning games. Klopfer stated that one big barrier to learning game progress is "Limited Ambition: Funded projects are often not ambitious enough" (Klopfer, Osterweil, and Salen 2009, 19). I agree, and specifically argue that "digital worksheet" games are too inherently limited in reach to justify consuming any more research resources. Fukumoto's success getting *Equatia* adopted by Rocketship is evidence that there is market demand among educators for low-activity-goal alignment games; however, I argue that commercial entities, not academic learning game researchers, should meet this demand. In other words, I am confident that the profit-motivated commercial entities will find new forms of superficial improvements to *Math Blaster* for many years to come. This type of incremental improvement is not an appropriate challenge for learning game researchers. They should be aiming higher.

7.5 Conclusion

In this chapter I developed the AGA-Based Assessment Tool. I aimed to empower learning game designers to assess their own designs, aiming to achieve a pictographic-like clarity that I argue is a useful alternative to text-based modes of assessment. I showed how any learning game can be described in one of four categories, and discussed the implications of these categories for both practising designers and researchers of learning games. I used the AGA-Based Assessment Tool to differentiate three math games, as shown in Figure 51, and argued it has important advantages over other methods of assessment.

I discussed how *Equatia*'s innovations are not different enough from *Math Blaster* to represent a change in category, while *DragonBox* is quite different. Through the preceding discussions, I address my second research question, finding in essence that combining the ideas of scoring activity-goal alignment and deep learning offers useful new approaches to learning game assessment.

8 Discussion and Conclusion

8.1 Introduction

In this final chapter, I reflect on observations made throughout the thesis and discuss their implications for learning game designers. I argue that this research fills a gap in the literature: scoring activity-goal alignment usefully extends Shelton's conception of activity-goal alignment, and is useful when assessing Gee's conception of deep learning. I answer my research questions by first finding that scoring activity-goal alignment highlights the key points in a designer's necessarily complex understanding of the relation between activity and goal in a learning game. I then find that combining the ideas of scoring activity-goal alignment and deep learning offers useful new approaches to learning game assessment. Finally, I suggest directions for further research, such as: might scoring activity-goal alignment be applicable to commercial games?

8.1.1 Review

In this thesis I have proposed, developed and examined the applications of an "activity-goal alignment score," which gives a reasonable approximation of activity-goal alignment in a game.

In the first three chapters, I provided context and background. In the literature review, I found that Shelton's theory and Rice's theory both had interesting ideas that I could combine in my research. The third chapter explained my methodology, which I argued fulfils the requirements of design research. I discussed autoethnography, finding it, despite its limitations, a good choice for this research.

In the fourth chapter I explained how, during the development of the learning game *Battlefood*, the theory of activity-goal alignment might have enabled a better design solution. From this need, I then discussed the development of the AGA Scoring Tool, showing how I evolved it iteratively from a simple diagram to the interactive tool shown in Figure 15. By exposing my process and explaining design decisions, I offered insights into my design process and reasoning which might aid future researchers who wish to amend or modify this tool.

In the fifth chapter I addressed the first research question, applying activity-goal alignment to six learning games. I exercised the tool in a variety of contexts and applications. In order to reveal the usefulness of the tool in context, I first discussed activity-goal alignment in textual form, then used the AGA Scoring Tool in a second discussion, and finally compared the two discussions. The chapter

concluded that scoring activity-goal alignment is useful to a practising designer/researcher, as well as discussing potential problems and limitations relating to the tool.

The sixth chapter began with a notable finding from the preceding chapter: a high activity-goal alignment game is not always better than a low activity-goal alignment game, and asked: is activity-goal alignment theory useful for assessment of learning games? I related activity-goal alignment to a prevailing view: designers and researchers must be ambitious in their designs of learning games. I questioned how we might assess a game design against this prevailing view. In this context I exercised and critiqued four existing methods of assessment from academia: Shearer, Rice, and two from Gee. I also reviewed several commercial assessment tools. I concluded that none met my needs, and all were easy to misinterpret. Finding the attributed-based approach to assessment inadequate, I developed criteria for a new assessment tool that might help designers identify when a learning game's design conflicts with prevailing advice.

chapter 7 presents the development of a new activity-goal alignment-based assessment method to answer the second research question. I argued that the gist of prevailing advice is found in Gee's six properties of "deep learning," which collectively embody two key ideas: low/high order learning, and activity-goal alignment. I combined these two ideas to create an AGA-Based Assessment Tool, and exercised the tool in a discussion of recent math games: *HyperBlast* (Knowledge Adventure 2012b), *Equatia* (Luckybird Games 2012) and *DragonBox* (WeWantToKnow 2012). I concluded that the AGA-Based Assessment Tool might help amateur designers avoid reinventing the wheel because it emphasizes important ideas in a form that is more difficult to misinterpret.

8.2 Significance

In the introduction I argued for the broad significance of this research. My argument had three major points:

- 1) Successful learning games are needed.
- 2) Learning game design research can improve the success of learning games.
- 3) The problem of activity-goal alignment is important to learning game design research.

From these three points, I claimed that any research that adds to the activity-goal alignment literature is significant. In this section, I furthered this argument by showing how this research contributes to activity-goal alignment. I also discussed the significance of other findings in this thesis, in relation to the existing learning game literature.

8.2.1 Scoring Activity-Goal Alignment

Here I presented my findings regarding scoring activity-goal alignment, grounded in the practice of learning game design. In keeping with my autoethnographic approach, it is relevant to discuss how the research findings affect one's own practice, so I began there. I then discussed the theoretical implications of these practical findings.

Generally, I found that **scoring activity-goal alignment was a useful improvement on activity-goal alignment theory in my own practice**. In chapter 4, I found that scoring activity-goal alignment was needed for improving my understanding of activity-goal alignment in my own game production process. In chapter 5, I found scoring activity-goal alignment was useful to developing my position on the activity-goal alignment of a variety of learning games.

I found that the AGA Scoring Tool can **improve both speed and clarity** of considering the activity-goal alignment of a learning game. When compared to the sequential nature of writing text-based discussions, the tool enables **simultaneous** reflection, revision, and communication. The tool makes it quick and easy to revise the essential relationships in a game. Specifically, it takes a single click-and-drag to subtly adjust my position on a game's activity-goal alignment. By comparison, it took tens of minutes for me to convert a thought into a revision to my textual analysis, using Shelton's text-based approach. While the speed was nice, the ability to stay in a **mild flow state** while recording my score was a notable advantage. I discussed the implications at the end of chapter 5.

I also found that scoring activity-goal alignment was especially **useful for preventing errors in textual discussion** (chapter 5), and that scoring activity-goal alignment was **useful for summarizing activity-goal alignment in a variety** of games and discussions. It worked well for scoring a known low activity-goal alignment game (*Understanding Games*) and high activity-goal alignment games (*GameStar Mechanic*). It worked well in a flowing discussion as well in a comparative analysis.

In chapter 5 I noted that activity-goal alignment score **does not always correlate with overall game quality**. It is tempting to assume that all great learning games have high activity-goal alignment Scores, but this is not always true. For example, in chapter 5 I found that, for a particular learning goal and activity, a simple learning game *Infection* had a higher activity-goal alignment score than the acclaimed *Re-Mission*, but was, in my view, a worse game overall. It is possible that a reader could glance at those two diagrams and assume that more overlap means an overall better game. While I argued in chapter 7 that more overlap is desirable in research, the activity-goal alignment score alone is not a measure of learning game efficacy.

In other words, the practising designer must be careful to understand both the learning goal and the activity being considered when reading an activity-goal alignment score. Additionally, I argued that authors citing activity-goal alignment scores should bear responsibility for preventing this problem. I advised presenting activity-goal alignment scoring diagrams in such a way to prevent such a mistake³⁴.

Leaving practice-based discussion to discuss contributions to theory, I found generally that **scoring activity-goal alignment usefully extends Shelton's conception of activity-goal alignment**.

Specifically discussing Shelton's work in chapter 2, I noted that Shelton found designers could use activity-goal alignment to broadly consider their own game design decisions, but he did not assess or score a game's activity-goal alignment. This research does score activity-goal alignment, but I do not argue it supersedes Shelton's approach. By reducing activity-goal alignment to a single axis, the scoring method is more tightly scoped in its consideration of activity-goal alignment than a free-form discussion such as Shelton provides. This is a necessary limitation to achieve its simplicity and clarity, as discussed in the "limits" section of introduction and chapter 7. In this way, this research complements Shelton's work.

8.2.2 Criteria for Evaluation of Assessment of Learning Games

As I discussed in chapter 1, researchers have called for improved methods of evaluation of learning games. In chapter 6, I reviewed several existing commercial and academic methods of assessing learning games (Games for Change 2012; Ludoscience 2012a; Engage Learning 2012; TEGD 2012). In discussing their limitations, I developed broad criteria for any assessment. Restated here, I argued that any learning game assessment must:

1. Define a clear, reliable theoretical basis for assessment. In other words, it must set out reasons for claiming that certain games are "good" or "bad," and defend them in the light of past work.
2. Judge clearly and consistently enough to avoid fruitless argument about the judgement. There must be enough flexibility to allow the wisdom of the assessor to prevail over blind rules, with enough structure (e.g. specific enough rule-based procedure or guidelines) to ensure the tool is not just a foil for personal preference.

Additionally, assessments aimed at the practising designer should be:

³⁴ Specifically, I found caption text under the AGA diagram was a useful place to prevent misinterpretation of the AGA score.

3. Easy to use: The designer must find the assessment's outcome easy to interpret in practice
4. Prescriptive: a designer would be likely to understand how to modify their game to improve the assessment.
5. Simple and Quick: Practicing designers and reviewers should find it easy to use on the fly.
6. Versatile: useful for the broadest possible range of learning games.

These criteria are significant because they consider a particular audience, the practising designer/researcher, and do not require any particular theoretical foundation (as long as there is one), modality, or method. They are designed to guide design of assessments: they measure rigor, usefulness, and other qualities. I found these criteria helpful in creating the AGA-Based Assessment Tool, and hope that other researchers will find them similarly useful.

8.2.3 Contributions from the AGA-Based Assessment Tool

In chapter 7, I argued that it is useful to basically, quickly assess learning games against Gee's principles of deep learning, which I argued is a good choice for foundational learning game design literature, but that no such procedure, tool, or method exists. Informed by the AGA Scoring Tool, I built a new tool and demonstrated its use in understanding hidden design decisions in three modern learning games: *HyperBlast*, *DragonBox*, and *Equatia*.

In summarizing that work, I argued that this work comprises a meaningful contribution to the established body of learning game scoring methods for these reasons:

1. The AGA-Based Assessment Tool has a clear audience (learning game designer/researchers, especially amateurs) whose needs, I argue in chapter 6, are not being met by existing tools.
2. The theoretical foundations (Shelton's activity-goal alignment, higher order learning, and Gee's conception of deep learning) are well developed and defended throughout this thesis.
3. In contrast to the typical approaches (attribute-based checklists, or textual discussion), this assessment aims for unusual clarity in mode and format. I argue that this improvement addresses a crucial shortcoming in the existing literature.
4. It is designed to promote the prevailing position among leaders in the field regarding the core design principles of learning game design

8.2.4 Contribution to Scientific/Designerly Debate in Design Research

Broadly speaking, this research rejects the "list of attributes" approach to understanding video games. In chapter 3 I advocated a design-based approach to design problems, discussing the history of weak outcomes that have resulted in part, I argued, from an overly reductionist approach. In

chapter 6 I further supported this position by showing how damaging the reductionist approach can be. Using a specific case study, I demonstrated how a checklist-based approach to assessment seems simple and clear from the theorist's point of view, but can easily be misleading when amateur designers use it in practice. In chapter 7 I suggested one alternative approach, using a 2 x 2 matrix instead of a checklist. In light of this work, I argue that this research gives useful evidence for the benefits of the designerly approach advocated by many design researchers (Jones 1992).

Jones' position is not universally supported in this field and conflicts with standard approaches in related fields, such as simulation and vocational training. These fields generally follow a more scientific approach. In this dissertation, I have advocated a more designerly approach to understanding learning games. In taking such a position and defending it, this research contributes to the argument against the scientific approach to design research.

8.2.5 Modality of Learning Game Literature

In general, I suggest that the modality of the tools in this research might in itself contribute to the literature. In chapter 7 I argued that the prevailing position of the learning game design literature supports activity-goal alignment, but these key points are being ignored or lost by designer/researchers. In chapter 5 I demonstrated how scoring activity-goal alignment was simpler to understand than conventional textual analysis, and suggested the AGA-Based Assessment Tool offers a similar advantage over existing methods of assessment reviewed in chapter 6. If these tools' modality is indeed important to making prevailing theory more accessible than prior literature, then this research might suggest a way to improve on the dissemination of many types of learning game literature, not just the specific theories used in this thesis.

Dissemination is also important for the topic of assessment of learning games. In chapter 7, I argued that the literature's typical modality (primarily lengthy text-based discussions) is not a good fit for practising designers. In the specific case of learning game assessment, I argued that in contrast to the typical approaches (attribute-based checklists, or textual discussion), the AGA-Based Assessment Tool aims for pictographic-like clarity in mode and format. While acknowledging that this clarity comes at the expense of detail, I argued that in light of failure of designers to implement existing learning game research, and considering the needs of learning game designers, this trade-off is preferable. As J.M. Keynes apparently said³⁵, "it is better to be broadly right than precisely wrong."

³⁵ Edward Tufte attributed this quote to Keynes (Tufte 2012).

8.2.6 Innovation in Method

This research contributes to the analytic autoethnography methodology literature in several ways. I have demonstrated how autoethnography in design research is uncommon enough that this research might be useful as a comparative reference for other analytic autoethnographers (Chang 2008). Unlike the typical narrative-driven approach in analytic autoethnography, I used analytical discussion in self-reflexive ways that are influenced by autoethnography, but are not as descriptive or set along a timeline (L. Anderson 2006). In this work I have shown how analytic autoethnography can be used to address specific aims in design research.

Of these points, I feel that the last might be most interesting. Many designers employ “soft” methods like autoethnography to answer somewhat scientific questions (“is X useful”). By building on others’ use of autoethnography in design studies (Frankel 2009; Duncan 2008), I hope this research makes a small but significant contribution to the literature.

8.3 Questions for Further Research

The subjectivity of the tools generated in this research is an acknowledged limitation. Additionally, this research employed the autoethnographic method, which also has an acknowledged subjectivity limitation. It would be interesting to strengthen this research by having multiple game designers employ the AGA Scoring Tool to score the same educational game and comparing their findings.

In this research I have explored how categorizing learning goals as “low order” and “high order” can be combined with activity-goal alignment to assess any type of learning game. It would be interesting to explore how combining the top prevailing theories into assessment tools might be applied in other, more specific assessment purposes. For example, if assessing mathematical learning games, one might choose Devlin’s advice, which is specific to math games, rather than Gee’s. Thus future research might ask: is a topic-specific theoretical foundation superior to a general theory when considering the activity-goal alignment in a particular type of game?

I noted in my discussion of *GameStar Mechanic* that activity-goal alignment does not have an equivalent to Sharma et al’s three-level hierarchy, and speculated that such a hierarchy might be useful, despite the complexity it might add

Using theory to create a tool or method of assessing learning games is not a new approach. It would be useful to compare other executions to this thesis. For example, TPL-KATS (Hoeft et al. 2003) has been applied to map a player’s mental representations of their teammates (Evans, Harper, and

Jentsch 2004). How might Hoefft et al's approach be productively intersected with Shelton's theory to amend or improve either of the tools presented here?

The idea that modality might be an important consideration in learning game literature suggests many future directions. There are many ways one might adapt the established advice and knowledge might be presented in formats that have more impact on practising designers. This is a large and promising area for further research.

8.4 Conclusion

In chapter 7 I argued for higher ambition in learning game research. Deep learning games are rare, although some do exist. Gee exemplifies how multiple fields are moving closer toward developing a sufficiently clear understanding of this challenge that learning game designer/researchers can hope to conquer it in the future. While acknowledging that this has been argued for some time, I still believe it is true. I also argue that the alternative—building and researching less ambitious games—is worse. I conclude that converting theory like Gee's into practice is a worthy challenge for designer/researchers, and great rewards could come to those who do so.

I conclude on a hopeful note: By doing this research, I have aimed to clarify and underscore the prevailing advice to designer/researchers of learning games: go beyond the "low-hanging fruit" of digital worksheet games and tackle the hard challenges in our field that Gee has set out for us. I hope this research is one small step toward improving the design of "deep learning" video games, for they could be the powerful new tool we need to make formal education relevant in the 21st century.

9 Bibliography

- 2 Flash Games LLC. 2012. "Tank Wars." <http://www.2flashgames.com/f/f-771.htm>.
- Abt, Clark C. 2002. *Serious Games*. University Press of America.
- ACS. 2009. "ACS Guidelines for Chemistry in Two-Year College Programs". American Chemical Society.
http://portal.acs.org/portal/PublicWebSite/education/policies/twoyearcollege/WPCP_012239.
- Adams, Ernest. 2010. *Fundamentals of Game Design*. New Riders.
- Akilli, GK, and K. Cagiltay. 2006. "An Instructional Design/Development Model for the Creation of Game-Like Learning Environments: The FIDGE Model." *Affective and Emotional Aspects of Human-Computer Interaction: Game-Based and Innovative Learning Approaches*: 93.
- Aldrich, Clark. 2009. *The Complete Guide to Simulations and Serious Games: How the Most Valuable Content Will Be Created in the Age Beyond Gutenberg to Google*. John Wiley and Sons.
- Alexander, A.L., T. Brunyé, J. Sidman, and S.A. Weil. 2005. "From Gaming to Training: A Review of Studies on Fidelity, Immersion, Presence, and Buy-in and Their Effects on Transfer in Pc-Based Simulations and Games." In *The Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC), NTSA, Orlando, Florida*.
- Alexander, W. H., and J. W. Brown. 2011. "Medial Prefrontal Cortex as an Action-Outcome Predictor." *Nature Neuroscience* 14 (10): 1338–1344.
- Alfino, M. 1991. "Another Look at the Derrida-Searle Debate." *Philosophy & Rhetoric* 24 (2): 143–152.
- Ambinder, Mike. 2009. "Valve's Approach to Playtesting: The Application of Empiricism" presented at the Game Developer's Conference, San Francisco, California US.
http://www.valvesoftware.com/publications/2009/GDC2009_ValvesApproachToPlaytesting.pdf.
- Anderson, Eileen S., R. A Winett, and J. R Wojcik. 2007. "Self-Regulation, Self-Efficacy, Outcome Expectations, and Social Support: Social Cognitive Theory and Nutrition Behavior." *Annals of Behavioral Medicine* 34 (3): 304–312.
- Anderson, L. 2006. "Analytic Autoethnography." *Journal of Contemporary Ethnography* 35 (4) (August 1): 373–395. doi:10.1177/0891241605280449.
- . 2011. "Time Is of the Essence: An Analytic Autoethnography of Family, Work, and Serious Leisure." *Symbolic Interaction* 34 (2): 133–157.
- Anton. 2009. "Infection! The Game." <http://www.kongregate.com/games/multimediac/infection-the-game>.
- AppData. 2012. "AppData - Facebook Application Leaderboards, Charts, and Metrics." June.
<http://appdata.com/>.
- Archer, Louise, Merryn Hutchings, and Alistair Ross. 2003. *Higher Education and Social Class: Issues of Exclusion and Inclusion*. Psychology Press.
- ARimeth. 2011. "Wuthering Heights in Gamestar Mechanic - YouTube." March 27.
<http://www.youtube.com/watch?v=9xBSRrXoXw0&feature=endscreen>.
- Baldwin, T. T., and J. K. Ford. 1988. "Transfer of Training: A Review and Directions for Future Research." *Personnel Psychology* 41 (1): 63–105.
- Barab, S., and Kurt Squire. 2004. "Design-Based Research: Putting a Stake in the Ground." *Journal of the Learning Sciences* 13 (1): 1–14.
- Barab, S., S. Warren, and A. Ingram-Goble. 2006. "Academic Play Spaces: Designing Games for Education." In *Annual Meeting of the American Educational Research Association, San Francisco, CA*.

- Bartle, Richard. 1996. "Players Who Suit MUDs." <http://mud.co.uk/richard/hcdis.htm>.
- Bartlett, A. A. 1976. "The Exponential function—Part 1." *The Physics Teacher* 14: 393.
- Bateson, G. 1987. "A Theory of Play and Fantasy." In *Steps to an Ecology of Mind: Collected Essays in Anthropology, Psychiatry, Evolution, and Epistemology*. Aronson (Northvale, NJ).
- Becker, K. 2007. "Battle of the Titans: Mario vs. MathBlaster." In *World Conference on Educational Multimedia, Hypermedia and Telecommunications 2007*, edited by Craig Montgomerie and Jane Seale, 2707–2716. Vancouver, Canada: AACE.
- . 2008. "The Invention of Good Games: Understanding Learning Design in Commercial Video Games". UNIVERSITY OF CALGARY.
- Biggs, J. 1996. "Enhancing Teaching through Constructive Alignment." *Higher Education* 32 (3): 347–364.
- Bloch, P. H. 1995. "Seeking the Ideal Form: Product Design and Consumer Response." *The Journal of Marketing*: 16–29.
- Bloom, B. S. 1956. "Taxonomy of Educational Objectives: The Classification of Educational Goals ."
- Blunt, R., and V. A. Annandale. 2007. "Does Game-Based Learning Work? Results from Three Recent Studies." In *Proceedings of the Interservice/Industry Training, Simulation, & Education Conference*, 945–955. NTSA.
- Bogost, Ian. 2007a. *Persuasive Games: The Expressive Power of Videogames*. The MIT Press.
- . 2007b. "The Rhetoric of Video Games." *The John D. and Catherine T. MacArthur Foundation Series on Digital Media and Learning*: 117–139.
- . 2011a. "Persuasive Games: Exploitationware." *Gamasutra*.
http://www.gamasutra.com/view/feature/134735/persuasive_games_exploitationware.php?page=2.
- . 2011b. "Gamification Is Bullshit." *Ian Bogost's Blog*.
http://www.bogost.com/blog/gamification_is_bullshit.shtml.
- Bowler, Steve. 2008. "TF2 vs. Vanilla FPS." April 8. <http://www.game-ism.com/2008/04/08/tf2-vs-vanilla-fps/>.
- Bransford, J. 2000. *How People Learn: Brain, Mind, Experience, and School*. National Academies Press.
- Breuer, J. S., and G. Bente. 2010. "Why so Serious? On the Relation of Serious Games and Learning." *Eludamos. Journal for Computer Game Culture* 4 (1): 7–24.
- Bruni, N. 2002. "The Crisis of Visibility: Ethical Dilemmas of Autoethnographic Research." *Qualitative Research Journal* 2 (1): 24–33.
- Buckleitner, Warren. 2012. "Children's Technology Review » About The Ratings."
<http://childrenstech.com/evaluation-instrument>.
- Bunchball Inc. 2011. "Gamification Works." <http://www.bunchball.com/products/gamification>.
- Burghardt, G. M. 2005. *The Genesis of Animal Play: Testing the Limits*. The MIT Press.
- Burns,, Chris. 2011. "iTunes Rewind 2011 Goes Live: Top Apps, Movies, and Music Listed."
<http://www.slashgear.com/itunes-rewind-2011-goes-live-top-apps-movies-and-music-listed-08201051/>.
- Byrne, John H. 2008. *Concise Learning and Memory: The Editor's Selection*. Academic Press.
- Caillois, R. 2006. "The Definition of Play and the Classification of Games." In *The Game Design Reader: A Rules of Play Anthology*, 122. MIT Press.
- Candy, L., S Amitani, and Z. Bilda. 2006. "Practice-Led Strategies for Interactive Art Research - CoDesign: International Journal of CoCreation in Design and the Arts."
<http://www.informaworld.com/smpp/content~db=all~content=a759323519>.
- Caro, P.W. 1989. "Flight Training and Simulation." In *Human Factors in Aviation*, 229–261. Gulf Professional Publishing.
- Carr, D. 2009. "Textual Analysis, Digital Games, Zombies." In *Proceedings of DiGRA*. Citeseer.
- Cavallari, B., J. Hedberg, and B. Harper. 1992. "Adventure Games in Education: A Review." *Australian Journal of Educational Technology* 8 (2): 172–184.

- CDB531. 2012. "Problems with Trial Mod." *Steam Users' Forums*.
<http://forums.steampowered.com/forums/showthread.php?t=2591952>.
- Chang, H. 2008. *Autoethnography as Method*. Left Coast Press Walnut Creek, CA.
- Chapman, G. B. 2005. "Short-Term Cost for Long-Term Benefit: Time Preference and Cancer Control." *Health Psychology* 24 (4S): S41.
- Charmaz, Kathy, and Karen Henwood. 2008. "Grounded Theory." In *The SAGE Handbook of Qualitative Research in Psychology The SAGE Handbook of Qualitative Research in Psychology*, 240–260. 1 Oliver's Yard, 55 City Road, London EC1Y 1SP United Kingdom: SAGE Publications Ltd. <http://0-srmo.sagepub.com/prospero.murdoch.edu.au/view/the-sage-handbook-of-qualitative-research-in-psychology/n14.xml>.
- Chi, M. T.H. 2008. "Three Types of Conceptual Change: Belief Revision, Mental Model Transformation, and Categorical Shift." *International Handbook of Research on Conceptual Change*: 61–82.
- Chin, A. P, J. M. Marijke, W. M Jacobs, E. P.G Vaessen, S. Titze, and W. van Mechelen. 2008. "The Motivation of Children to Play an Active Video Game." *Journal of Science and Medicine in Sport* 11 (2): 163–166.
- Chmiel, M. 2010. "Game Design towards Scientific Literacy." *International Journal of Cognitive Technology* 15 (1).
- Chou, L., R. McClintock, F. Moretti, and D. H. Nix. 1993. "Technology and Education: New Wine in New Bottles: Choosing Pasts and Imagining Educational Futures." Retrieved August 24: 2000.
- Chow, Y. W., W. Susilo, and H. Y. Zhou. 2010. "CAPTCHA Challenges for Massively Multiplayer Online Games: Mini-Game CAPTCHAs." In *Cyberworlds (CW), 2010 International Conference on*, 254–261. IEEE.
- Chwif, L., M. R. P. Barretto, and R. J. Paul. 2000. "On Simulation Model Complexity." In *Simulation Conference Proceedings, 2000. Winter*, 1:449–455.
http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=899751.
- Ciampa-Brewer, R. J. 2011. "Creative Consciousness: Becoming a Reflective Designer". Office of Graduate Studies, University of Massachusetts Boston.
- Consalvo, Mia, and Nathan Dutton. 2006. "Game Studies - Game Analysis: Developing a Methodological Toolkit for the Qualitative Study of Games." *Game Studies* 6 (1) (December).
http://gamestudies.org/0601/articles/consalvo_dutton.
- Cook, Daniel. 2006a. "What Are Game Mechanics?" *Lost Garden*.
<http://www.lostgarden.com/2006/10/what-are-game-mechanics.html>.
- . 2006b. "Lost Garden: The Wii Help Cat: A Lesson in Interaction Design." *Lost Garden*.
<http://www.lostgarden.com/2006/12/wii-help-cat-lesson-in-interaction.html>.
- . 2007a. "Rockets, Cars and Gardens: Visualizing Waterfall, Agile and Stage Gate." <http://lostgarden.com/2007/02/rockets-cars-and-gardens-visualizing.html>.
- . 2007b. "The Chemistry Of Game Design." *Gamasutra*. July 19.
http://www.gamasutra.com/view/feature/129948/the_chemistry_of_game_design.php?page=2.
- . 2009a. "Project Horseshoe: Multiplayer Game Atoms." *Lost Garden*.
<http://lunar.lostgarden.com/labels/skill%20chains.html>.
- . 2009b. "What Is Your Game Design Style?" *Lost Garden*.
<http://www.lostgarden.com/search?q=simulation>.
- Corbett, Sara. 2010. "Learning by Playing: Video Games in the Classroom." *New York Times*, September 19, sec. The Sunday Magazine.
- Costikyan, Greg. 2006. "I Have No Words & I Must Design." In *The Game Design Reader: A Rules of Play Anthology*. MIT Press.
http://80.250.162.180/2006/lvan.Ignatyev/DiGRA/I%20Have%20No%20Words%20%26%20%20Must%20Design_Toward%20a%20Critical%20Vocabulary%20for%20Games.pdf.

- Cothran, D. J., and C. D. Ennis. 1998. "Curricula of Mutual Worth: Comparisons of Students' and Teachers' Curricular Goals." *Journal of Teaching in Physical Education* 17: 307–326.
- Crawford, C. 1984. *The Art of Computer Game Design*. Osborne/McGraw-Hill Berkley.
- Crawford, Chris. 2003. *Chris Crawford on Game Design*. New Riders.
- Crookall, D. 2011. "Serious Games, Debriefing, and Simulation/Gaming as a Discipline." *Simulation & Gaming* 41 (6) (January): 898–920. doi:10.1177/1046878110390784.
- Cross, Nigel. 1982. "Designerly Ways of Knowing." *Design Studies* 3 (4) (October): 221–227. doi:doi: DOI: 10.1016/0142-694X(82)90040-0.
- Csikszentmihalyi, M. 1990. *Flow: The Psychology of Optimal Experience*. Harper & Row New York.
- Curtin University Council. 2009. *Rule No. 10: Made Pursuant to Statute No. 12 - Enrolment: Degree of Doctor by Research*. http://policies.curtin.edu.au/local/docs/Rule_10_to_Statute12-June2009.pdf.
- D'Angelo, Cynthia, Daisy Rutstein, and Christopher Harris. 2013. "Review of Computer-Based Simulations for STEM Learning in K-12 Education."
- Daastol, E., and C. Jago. 1995. "One Person's Opinion: Should School Be Fun?" *The English Journal* 84 (2): 15–16.
- dangerB13. 2012. "Topic: Please Increase the Low Level Raider/beast Spawn Rate | Battle Nations." <http://battlenations.com/forums/topic/please-increase-the-low-level-raiderbeast-spawn-rate/>.
- Darken, R. P., and B. Peterson. 2002. "Spatial Orientation, Wayfinding, and Representation." *Handbook of Virtual Environments*: 493–518.
- De Freitas, S., and M. Oliver. 2006. "How Can Exploratory Learning with Games and Simulations within the Curriculum Be Most Effectively Evaluated?" *Computers & Education* 46 (3): 249–264.
- De Leon, J. 1990. "A Model for an Advocacy-Oriented Assessment Process in the Psychoeducational Evaluation of Culturally and Linguistically Different Students." *Journal of Educational Issues of Language Minority Students* 7: 53–67.
- Deci, Edward L., and Richard M. Ryan. 1985. *Intrinsic Motivation and Self-Determination in Human Behavior*. Springer.
- Deepend. 2005. *Food Force: The First Humanitarian Video Game*. PC (version 1). United Nations World Food Programme. <http://www.wfp.org/how-to-help/individuals/food-force>.
- Department of Health and Ageing, Australian Government. 2012. "Recommended Daily Servings Chart." <http://www.health.gov.au/internet/healthyactive/publishing.nsf/content/recommended-daily-servings>.
- Devlin, Keith. 2012a. "How to Design Video Games That Support Good Math Learning: Level 1." *Profkeithdevlin*. <http://profkeithdevlin.org/2012/01/29/how-to-design-video-games-that-support-good-math-learning-level-1/>.
- . 2012b. "How to Design Video Games That Support Good Math Learning: Level 2." « Profkeithdevlin." *Profkeithdevlin*. <http://profkeithdevlin.org/2012/01/31/how-to-design-video-games-that-support-good-math-learning-level-2/>.
- . 2012c. "How to Design Video Games That Support Good Math Learning: Level 5." « Profkeithdevlin." *Profkeithdevlin*. <http://profkeithdevlin.org/2012/03/30/how-to-design-video-games-that-support-good-math-learning-level-5/>.
- Djaouti, D., J. Alvarez, J. P. Jessel, G. Methel, and P. Molinier. 2008. "A Gameplay Definition through Videogame Classification." *International Journal of Computer Games Technology* 2008: 4.
- Dondlinger, M. J. 2007. "Educational Video Game Design: A Review of the Literature." *Journal of Applied Educational Technology* 4 (1): 21–31.
- Dugan, Patrick. 2010. "A Brief History of Disingenuous Design Practices." *Play This Thing*. <http://playthisthing.com/brief-history-disingenuous-design-practices>.

- Duncan, M. 2008. "Autoethnography: Critical Appreciation of an Emerging Art." *International Journal of Qualitative Methods* 3 (4): 28–39.
- Egenfeldt-Nielsen, S. 2006. "Overview of Research on the Educational Use of Video Games." *Digital Kompetanse* 1 (3): 184–213.
- Eisen, G. 1990. *Children and Play in the Holocaust: Games among the Shadows*. Univ of Massachusetts Pr.
- Eladhari, M. P., and E. M. I. Ollila. 2012. "Design for Research Results: Experimental Prototyping and Play Testing." *Simulation & Gaming*.
- Electron. 2010. *Caster*. <http://store.steampowered.com/app/29800/>.
- Electronic Arts. 1992. "Car and Driver."
- E-Line Media. 2007. "What Skills Are Learned in Gamestar Mechanic?" *Gamestar Mechanic Learning Guide*.
<https://sites.google.com/a/elinemedia.com/gsmlearningguide/home/introduction/skills-learned-in-gamestar>.
- . 2010. "Gamestar Mechanic." <https://gamestarmechanic.com/>.
- . 2012. "Gamestar Mechanic Learning Guide."
<https://sites.google.com/a/elinemedia.com/gsmlearningguide/home>.
- Ellis, C. S. 2007. "Telling Secrets, Revealing Lives." *Qualitative Inquiry* 13 (1): 3–29.
- Ellis, C. S., and A. Bochner. 2000. "Autoethnography, Personal Narrative, Reflexivity: Researcher as Subject."
- Engage Learning. 2012. "The Portal for Game-Based Learning."
<http://www.engagelearning.eu/teachers/>.
- Entertainment Software Association. 2011. "The Entertainment Software Association - Industry Facts." <http://www.theesa.com/facts/index.asp>.
- Equatia Demo, v.1 Run-Through*. 2010.
http://www.youtube.com/watch?v=6oQXng1JrbQ&feature=youtube_gdata_player.
- Evans, A. W., Michelle E. Harper, and Florian Jentsch. 2004. "I Know What You're Thinking: Eliciting Mental Models about Familiar Teammates." In *Concept Maps: Theory, Methodology, Technology. Proceedings of the First International Conference on Concept Mapping*. Pamplona, Spain. Citeseer.
- FAS. 2011. "Immune Attack." <http://www.fas.org/immuneattack/>.
- Feast, L., and G. Melles. 2010. *Epistemological Positions in Design Research: A Brief Review of the Literature*. Connected.
- Ferenstein, Greg. 2010. "How Social Gaming Is Improving Education." *Mashable*. February 7.
<http://mashable.com/2010/02/07/social-gaming-education/>.
- Finn, A., and N. Kushmerick. 2006. "Learning to Classify Documents according to Genre." *Journal of the American Society for Information Science and Technology* 57 (11): 1506–1518.
- Fish, J., and S. Scrivener. 1990. "Amplifying the Mind's Eye: Sketching and Visual Cognition." *Leonardo*: 117–126.
- Florida, R. L. 2002. *The Rise of the Creative Class: And How It's Transforming Work, Leisure, Community and Everyday Life*. Basic Civitas Books.
- Frankel, L. 2009. "Communicating Design Research Knowledge: A Role for Ethnographic Writing."
<http://www.iasdr2009.org/ap/Papers/Orally%20Presented%20Papers/Design%20Method/Communicating%20Design%20Research%20Knowledge%20-%20A%20Role%20for%20Ethnographic%20Writing.pdf>.
- Freitas, De Sara, and Paul Maharg. 2011. *Digital Games and Learning*. Continuum.
- Fukumoto, Brady. 2011. "Equatia: Stanford School of Education Learning Design and Technology Master's Project". Stanford University.
- Fullerton, Tracy, Christopher Swain, and Steven Hoffman. 2008. *Game Design Workshop: A Playcentric Approach to Creating Innovative Games*. Morgan Kaufmann.

- Games for Change. 2012. "Games for Change: Submit a Review (Way)." <http://www.gamesforchange.org/play/way/>.
- Games, I. A. 2008. "Three Dialogs: A Framework for the Analysis and Assessment of Twenty-First-Century Literacy Practices, and Its Use in the Context of Game Design within Gamestar Mechanic." *E-Learning and Digital Media* 5 (4): 396–417.
- . 2009. "21st Century Language And Literacy In Gamestar Mechanic: Middle School Students' Appropriation Through Play Of The Discourse Of Computer Game Designers". UNIVERSITY OF WISCONSIN.
- Garfield, E. 1963. "Citation Indexes in Sociological and Historical Research." *American Documentation* 14 (4): 289–291.
- Garris, R., R. Ahlers, and J. E. Driskell. 2002. "Games, Motivation, and Learning: A Research and Practice Model." *Simulation and Gaming* 33 (4): 441–469.
- Gaver, B. 2002. "Designing for Homo Ludens, Still." *I3 Magazine*, June.
- Gee, James Paul. 2003. "What Video Games Have to Teach Us about Learning and Literacy." *Comput. Entertain.* 1 (1): 20–20.
- . 2004. *What Video Games Have to Teach Us About Learning and Literacy*. New edition. Palgrave Macmillan.
- . 2006. "Video Games: Do They Have Educational Value?" *CQ Researcher*, November 10.
- . 2008. "Game-Like Learning: An Example of Situated Learning and Implications for Opportunity to Learn | James Paul Gee." January. <http://www.jamespaulgee.com/node/29>.
- . 2009a. "Deep Learning Properties of Good Digital Games." *Serious Games: Mechanisms and Effects*: 67.
- . 2009b. "Cats and Portals: Video Games, Learning, and Play."
- Gee, James Paul, and Tashia Morgridge. 2008. "Being a Lion and Being a Soldier: Learning and Games." In *Coiro J Knobel M Lankshear C Leu DJ Eds*, 1023–036.
- Gentner, Dedre, Keith James Holyoak, and Boicho N. Kokinov. 2001. *The Analogical Mind: Perspectives from Cognitive Science*. MIT Press.
- Gkikas, K., D. Nathanael, and N. Marmaras. 2007. "The Evolution of FPS Games Controllers: How Use Progressively Shaped Their Present Design." In .
- Griffin, Christine, and Andrew Bengry-Howell. 2008. "Ethnography." In *The SAGE Handbook of Qualitative Research in Psychology The SAGE Handbook of Qualitative Research in Psychology*, 15–31. 1 Oliver's Yard, 55 City Road, London EC1Y 1SP United Kingdom: SAGE Publications Ltd. <http://0-srmo.sagepub.com.prospero.murdoch.edu.au/view/the-sage-handbook-of-qualitative-research-in-psychology/n2.xml>.
- Grudin, J., and J. Pruitt. 2002. "Personas, Participatory Design and Product Development: An Infrastructure for Engagement." In *Proc. PDC*, 144–161. Citeseer.
- Halpern, D. F. 1998. "Teaching Critical Thinking for Transfer across Domains: Disposition, Skills, Structure Training, and Metacognitive Monitoring." *American Psychologist* 53 (4): 449.
- Halverson, R. 2005. "What Can K-12 School Leaders Learn from Video Games and Gaming." *Innovate: Journal of Online Education* 1 (6). <http://www.innovateonline.info/index.php?View=article&id=81>.
- Harel, I. E., and S. E. Papert. 1991. *Constructionism*. Ablex Publishing.
- Hayano, D. M. 1983. *Poker Faces: The Life and Work of Professional Card Players*. Univ of California Pr.
- Hayler, Mike. 2011. "Using Autoethnography." *Research in Education* 3 (2) (July): 9–12.
- Hecker, Chris. 2000. "Physics in Computer Games." *Communications of the ACM* 43 (7): 35.
- Hermund, Frederik, Ole Toubro, Jakob Nielsen, and Roman Spycher. 2008. "3rd World Farmer: A Simulation to Make You Think." <http://www.3rdworldfarmer.com/index.html>.
- Herrington, Anthony, and J. Herrington. 2005. *Authentic Learning Environments in Higher Education*. Idea Group Inc (IGI).

- Herrington, J., Susan McKenney, Thomas Reeves, and R. Oliver. 2007. "Design-Based Research and Doctoral Students: Guidelines for Preparing a Dissertation Proposal." In *World Conference on Educational Multimedia, Hypermedia and Telecommunications 2007*, edited by Craig Montgomerie and Jane Seale, 4089–4097. Vancouver, Canada: AACE.
- Herrington, J., and R. Oliver. 1999. "Using Situated Learning and Multimedia to Investigate Higher-Order Thinking." *Journal of Interactive Learning Research* 10 (1): 3–24.
- Herrington, J., T. C. Reeves, R. Oliver, and Y. Woo. 2004. "Designing Authentic Activities in Web-Based Courses." *Journal of Computing in Higher Education* 16 (1): 3–29.
- Hochmitz, Ilanit, and Nirit Yuviler-Gavish. 2011. "Physical Fidelity Versus Cognitive Fidelity Training in Procedural Skills Acquisition." *Human Factors: The Journal of the Human Factors and Ergonomics Society* 53 (5) (October 1): 489–501. doi:10.1177/0018720811412777.
- Hoefl, Raegan M., Florian G. Jentsch, Michelle E. Harper, A. William Evans III, Clint A. Bowers, and Eduardo Salas. 2003. "TPL-KATS—concept Map: A Computerized Knowledge Assessment Tool." *Computers in Human Behavior* 19 (6): 653–657.
- Holzman, P. S., and R. W. Gardner. 1960. "Leveling-Sharpening and Memory Organization." *The Journal of Abnormal and Social Psychology* 61 (2): 176.
- HopeLab. 2012. "For Clinicians « Re-Mission." <http://www.re-mission.net/for-clinicians/>.
- Hopson, John. 2012. "10 Years of Behavioral Game Design with Bungie's Research Boss." *Gamasutra*, June 15.
http://www.gamasutra.com/view/feature/172409/10_years_of_behavioral_game_design_.php?page=3.
- How to Use Gamestar Mechanic (from a 7th Grader's Perspective)*. 2012.
http://www.youtube.com/watch?v=lkCwU0ezUn0&feature=youtube_gdata_player.
- Huizinga, J. 2003. *Homo Ludens: A Study of the Play-Element in Culture*. Vol. 3. Taylor & Francis.
- Huizinga, J. 1939. *Homo Ludens*. Boston: Beacon Press.
- Hunicke, R., M. LeBlanc, and R. Zubek. 2004. "MDA: A Formal Approach to Game Design and Game Research." In *Proceedings of the AAAI Workshop on Challenges in Game AI*, 04–04.
- Hutchison, Andrew. 2008. "Techno-Historical Limits of the Interface: The Performance of Interactive Narrative Experiences."
- Huyhn, Jean-Baptiste. 2012. "How to Design Video Games That Support Good Math Learning? « We Want To Know World." March 29. <http://wewanttoknow.com/2012/03/29/how-to-design-video-games-that-support-good-math-learning/>.
- IGDA LEG SIG. 2012. "Learning, Education and Games Special Interest Group Home Page."
<http://www.igda.org/learning-education-and-games>.
- Institute for Play. 2012. "The Patterns of Play." http://www.nifplay.org/states_play.html.
- ISO. 1999. "ISO 13407:1999 - Human-Centred Design Processes for Interactive Systems". International Organization for Standardization.
http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=21197
- Itu, S. R. 2010. *Serious Games Designed by INGOs as a New Media Tool for Entertainment-Education*. Erasmus University.
- Ivars Vendrell, Rafael. 2012. "Evaluation Videogame: 3rd World Farmer." *Learninggame.org*. February 3. <http://learninggame.org/videogames/scheda.php?id=252>.
- Jacobs, C. D., and L. Heracleous. 2007. "Strategizing through Playful Design." *Journal of Business Strategy* 28 (4): 75–80.
- Jaworski, B. J., and A. K. Kohli. 1993. "Market Orientation: Antecedents and Consequences." *The Journal of Marketing*: 53–70.
- jbhkb1. 2012. "DragonBox Designer jbhkb1 Replies to Y Combinator Discussion." *Y Combinator Hacker News*. <http://news.ycombinator.com/item?id=4105397>.
- Jones, J. Christopher. 1992. *Design Methods*. 2nd ed.

- Kato, P. M, S. W Cole, A. S Bradlyn, and B. H Pollock. 2008. "A Video Game Improves Behavioral Outcomes in Adolescents and Young Adults with Cancer: A Randomized Trial." *Pediatrics* 122 (2): e305–e317.
- Kelley, T., and J. Littman. 2001. *The Art of Innovation: Lessons in Creativity from IDEO, America's Leading Design Firm*. Vol. 10. Crown Business.
- Kerr, S. 1975. "On the Folly of Rewarding A, While Hoping for B." *Academy of Management Journal*: 769–783.
- Khan, Sal. 2011. "Khan Academy, About." *Khan Academy, About*. www.khanacademy.org/about.
- Khatri, N., and H. A. Ng. 2000. "The Role of Intuition in Strategic Decision Making." *Human Relations* 53 (1): 57–86.
- Kirriemuir, J., and A. McFarlane. 2004. "Literature Review in Games and Learning."
- Kissane, B. 2011. "Mathematics Education and the iPod Touch." <http://researchrepository.murdoch.edu.au/6202/>.
- Klopfer, Eric, Scot Osterweil, Jennifer Groff, and Jason Haas. 2009. "Using the Technology of Today, in the Classroom Today: The Instructional Power of Digital Games, Social Networking Simulations, and How Teachers Can Leverage Them". The Education Arcade, MIT. Google Scholar.
- Klopfer, Eric, Scot Osterweil, and Katie Salen. 2009. "Moving Learning Games Forward". MIT: The Education Arcade. http://education.mit.edu/papers/MovingLearningGamesForward_EdArcade.pdf.
- Knowledge Adventure. 2012a. "Math Blaster For Your Classroom." <http://www.knowledgeadventure.com/school/mathblaster/Default.aspx>.
- . 2012b. "Math Blaster HyperBlast for iPhone 3GS, iPhone 4, iPhone 4S, iPod Touch (3rd Generation), iPod Touch (4th Generation) and iPad on the iTunes App Store." <http://itunes.apple.com/au/app/math-blaster-hyperblast/id407597983?mt=8>.
- Koster, R. 2005. *A Theory Of Fun In Game Design*. Pdf. Paraglyph press.
- . 2006. "Response to 'What Are Game Mechanics?'" *Lost Garden*. <http://www.lostgarden.com/2006/10/what-are-game-mechanics.html>.
- Kreutzmann, A., and D. Wolter. 2011. "Physical Puzzles—Challenging Spatio-Temporal Configuration Problems." In *IJCAI-2011 Workshop* 27, 21.
- Kruglanski, A. W. 1975. "The Endogenous-Exogenous Partition in Attribution Theory." *Psychological Review* 82 (6): 387.
- Laurson, S., and H. Thiry. 2008. "What Do We Know about What Works? Review of US Foundations' Programs in Secondary Chemistry Education". University of Colorado at Boulder.
- Lazzaro, Nicole. 2005. "Why We Play Games: Four Keys to More Emotion without Story." *Design* 18: 1–8.
- . 2012. "The 4 Fun Keys." <http://www.xeodesign.com/whyweplaygames.html>.
- Lenhart, A., Pew Internet & American Life Project, and D. John. 2008. *Teens, Video Games, and Civics*. Pew Internet & American Life Project Washington, DC.
- Lepper, M. R, and D. I Cordova. 1992. "A Desire to Be Taught: Instructional Consequences of Intrinsic Motivation." *Motivation and Emotion* 16 (3): 187–208.
- Lepper, M. R., D. Greene, and R. E. Nisbett. 1973. "Undermining Children's Intrinsic Interest with Extrinsic Reward: A Test of the 'Overjustification' Hypothesis." *Journal of Personality and Social Psychology* 28 (1): 129.
- Levine, A. M., J. L. Richardson, G. Marks, K. Chan, J. Graham, J. N. Selsler, C. Kishbaugh, D. R. Shelton, and C. A. Johnson. 1987. "Compliance with Oral Drug Therapy in Patients with Hematologic Malignancy." *Journal of Clinical Oncology* 5 (9) (September 1): 1469–1476.
- Lewis, A., and D. Smith. 1993. "Defining Higher Order Thinking." *Theory into Practice* 32 (3): 131–137.

- Liu, Eric Zhi Feng, and Chun Hung Lin. 2009. "Developing Evaluative Indicators for Educational Computer Games." *British Journal of Educational Technology* 40 (1): 174–178. doi:10.1111/j.1467-8535.2008.00852.x.
- Liu, Johnathan. 2012. "DragonBox: Algebra Beats Angry Birds." *Wired*, June 13. <http://www.wired.com/geekdad/2012/06/dragonbox/>.
- Liu, Lucy. 2009. "No Title (an Autoethnographic Thesis on Toyota Quality Management)."
- Loftus, G. R., and W. W. Nelson. 1985. "Games as Teaching Tools: The Computer Connection." *Humanistic Perspectives on Computers in the Schools* 1 (4): 67.
- Lucas Allen. 2012. "MathBlaster.com Review: This Is Not Your Parents' Math Blaster." *Tech Powered Math*. <http://www.techpoweredmath.com/mathblaster-com-review/>.
- Luckybird Games. 2012. "Equatia." <https://www.luckybirdgames.com/equatia/play/>.
- Ludoscience. 2012a. "The Online Classification of Serious Games." <http://serious.gameclassification.com/>.
- . 2012b. "Serious Game Classification : Classification Guidelines." <http://serious.gameclassification.com/EN/about/bricks.html#GAMEPLAY>.
- Malone, T. W. 1981. "Toward a Theory of Intrinsically Motivating Instruction." *Cognitive Science* 5 (4): 333–369. doi:doi:
- Malone, T. W., and M. R. Lepper. 1987. "Making Learning Fun: A Taxonomy of Intrinsic Motivations for Learning." *Aptitude, Learning, and Instruction* 3: 223–253.
- Mark, Dave. 2008. "Why Not More Simulation in Game AI?" *AIGameDev.com*. <http://aigamedev.com/open/article/more-simulation/>.
- Martens, A., H. Diener, and S. Malo. 2008. "Game-Based Learning with Computers—Learning, Simulations, and Games." *Transactions on Edutainment I*: 172–190.
- McDougall, A., and D. Squires. 1995. "A Critical Examination of the Checklist Approach in Software Selection." *Journal of Educational Computing Research* 12 (3): 263–274.
- McGonigal, J. 2011. *Reality Is Broken: Why Games Make Us Better and How They Can Change the World*. Penguin Pr.
- McGuinness, C. 1999. "From Thinking Skills to Thinking Classrooms." *Research Brief* 115.
- McKee, Alan. 2003. *Textual Analysis: A Beginner's Guide*. SAGE.
- Mel Majoros. 2012. "What We've Heard « Re-Mission." Accessed July 14. <http://www.re-mission.net/what-weve-heard/>.
- Merry, Sally N, Karolina Stasiak, Matthew Shepherd, Chris Frampton, Theresa Fleming, and Mathijs F G Lucassen. 2012. "The Effectiveness of SPARX, a Computerised Self Help Intervention for Adolescents Seeking Help for Depression: Randomised Controlled Non-Inferiority Trial." *BMJ : British Medical Journal* 344. doi:10.1136/bmj.e2598. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3330131/>.
- Milker, MC. 2010. "Are Educational Toys Really Educational? Guest Post from MC Milker." *Parents for Ethical Marketing - Corporate Babysitter*. <http://www.parentsforethicalmarketing.org/blog/2008/01/11/are-educational-toys-really-educational-guest-post-from-mc-milker/>.
- Moen, R., and K. O. Doyle. 1978. "Measures of Academic Motivation: A Conceptual Review." *Research in Higher Education* 8 (1): 1–23.
- Moroney, William, and Brian moroney. 1999. "Flight Simulation." In *Handbook of Aviation Human Factors*, 355. Routledge.
- Mullany, Robert. 2012. "Lipstick on a Super Model – How Uber Could Leverage Gamification | Bunchball." *Bunchball Blog*. <http://www.bunchball.com/blog/post/826/lipstick-supermodel-%E2%80%93-how-uber-could-leverage-gamification>.
- Murcott, Anne. 1993. "Talking of Good Food: An Empirical Study of Women's Conceptualizations." *Food and Foodways* 5 (3) (April): 305–318. doi:10.1080/07409710.1993.9962010.

- National Health and Medical Research Council, Australian Government. 2007. "Australian Code for the Responsible Conduct of Research". NHMRC.
<http://www.nhmrc.gov.au/guidelines/publications/r39>.
- NCTM. 2012. "Principles for School Mathematics."
<http://www.nctm.org/standards/content.aspx?id=26802>.
- Newmann, F. M. 1990. "Higher Order Thinking in Teaching Social Studies: A Rationale for the Assessment of Classroom Thoughtfulness." *Journal of Curriculum Studies* 22 (1): 41–56.
- Nobel Media AB. 2004. *The Immune System Defender*.
<http://www.nobelprize.org/educational/medicine/immunity/>.
- Oke, Femi, Mallika Dutt, and Emily Treat. 2012. "Games That Improve Education and Bring Social Change." *Beyond School Books*. http://www.unicef.org/education/bege_62540.html.
- Osterweil, Scot. 2006. "Rethinking Learning Games: Why They Must Evolve in Order to Survive." In *Proceedings of the 2006 International Conference on Game Research and Development*, 15–16. Perth, Australia: Murdoch University.
- Özdemir, G., and D. B. Clark. 2007. "An Overview of Conceptual Change Theories." *Eurasia Journal of Mathematics, Science & Technology Education* 3 (4): 351–361.
- Page, G. P., J. W. Edwards, S. Barnes, R. Weindruch, and D. B. Allison. 2003. "A Design and Statistical Perspective on Microarray Gene Expression Studies in Nutrition: The Need for Playful Creativity and Scientific Hard-Mindedness." *Nutrition* 19 (11): 997–1000.
- Papert, S. 1993. *Mindstorms: Children, Computers, and Powerful Ideas*. Da Capo Press.
- Paras, B., and J. Bizzocchi. 2005. "Game, Motivation, and Effective Learning: An Integrated Model for Educational Game Design." In *Proceedings of DIGRA*.
- Parker, S. C., and A. Temple. 1923. "Unified Kindergarten and First-Grade Teaching. I." *The Elementary School Journal* 24 (1): 13–27.
- Pavlas, D., and W. Bedwell. 2009. "Ingentaconnect Investigating The Attributes in Serious Games That Contribute to Learning."
<http://www.ingentaconnect.com/content/hfes/hfproc/2009/00000053/00000027/art00005>
- Perkins, D. N. 1992. "What Constructivism Demands of the Learner." *Constructivism and the Technology of Instruction: A Conversation*: 161–165.
- . 1999. "The Many Faces of Constructivism." *Educational Leadership* 57: n3.
- Perry, David, and Rusel DeMaria. 2009. *David Perry on Game Design: A Brainstorming Toolbox*. Charles River Media.
- Piaget, Jean. 1932. *The Moral Judgment Of The Child*. Routledge & Kegan Paul.
http://www.archive.org/stream/moraljudgmentoft005613mbp/moraljudgmentoft005613mbp_djvu.txt.
- . 1970. "Science of Education and the Psychology of the Child. Trans. D. Coltman."
- Piaget, Jean, and M. Cook. 1955. *The Child's Construction of Reality*. Routledge & Kegan Paul London.
- Pinker, S., and W. W. Norton. 1998. "How the Mind Works." *Trends in Cognitive Sciences* 2 (1).
- Pixel. 2012. "Evaluation Videogames." <http://learninggame.org/videogames/>.
- Premsky, Marc. 2004. *Digital Game-Based Learning*. McGraw-Hill.
- . 2005. "Social Impact Games." *games2train.com - Entertaining Games with Non-Entertainment Goals*. <http://www.socialimpactgames.com/index.php>.
- . 2006. *Don't Bother Me Mom -- I'm Learning!* Paragon House Publishers.
- . 2009. "Should a 4-Year-Old Have an iPhone?" *Handheld Learning*.
<http://www.handheldlearning.co.uk/content/view/61/>.
- Proto, Andrew. 2011. "The Gamified Classroom - Gamification Co." *Gamification Co*.
<http://www.gamification.co/2011/09/28/the-gamified-classroom/>.
- Ququasar. 2011. "Goals for the Goal-Less." *Species Development Blog*.
<http://speciesdevblog.wordpress.com/tag/design/>.

- Radoff, Jon. 2011. "Game Player Motivations." *Jon Radoff's Internet Wonderland*.
<http://radoff.com/blog/2011/05/19/game-player-motivations/>.
- Reeves, T. C. 1993. "Pseudoscience in Computer-Based Instruction: The Case of Learner Control Research." *Journal of Computer-Based Instruction* 20 (2): 39–46.
- Resnick, M. 2004. "Edutainment? No Thanks. I Prefer Playful Learning." *Associazione Civita Report on Edutainment* 14. http://www.roboludens.net/Edut_Articoli/Playful_Learning.pdf.
- Rice, J. W. 2007. "Assessing Higher Order Thinking in Video Games." *Journal of Technology and Teacher Education* 15 (1): 87.
- Richardson, Virginia. 1997. *Constructivist Teacher Education: Building New Understandings*. Routledge.
- Rieber, L., J. Davis, M. Matzko, and M. Grant. 2001. "Children as Multimedia Critics: Middle School Students' Motivation for and Critical Analysis of Educational Computer Games Designed by Other Children." *American Educational Research Association, Seattle, WA*.
<http://www.nowhereroad.com/kiddesigner/research/aera2001-rieber.pdf>.
- Rieber, L. P. 1992. "Computer-Based Microworlds: A Bridge between Constructivism and Direct Instruction." *Educational Technology Research and Development* 40 (1): 93–106.
- Rittel, H. W. J., and M. M. Webber. 1973. "Dilemmas in a General Theory of Planning." *Policy Sciences* 4 (2): 155–169.
- Robertson, Margaret. 2010. "Can't Play, Won't Play." *Hide&Seek - Inventing New Kinds of Play*.
<http://www.hideandseek.net/2010/10/06/cant-play-wont-play/#comment-772>.
- Rocketship Education. 2012. "Rocketship Education: FAQ."
http://www.rsed.org/campusuite/modules/faq.cfm?grp_id=9614&main=0#q_2229.
- Rodriguez, Hector. 2006. "The Playful and the Serious: An Approximation to Huizinga's Homo Ludens." *Game Studies* 6 (1) (December). <http://gamestudies.org/0601/articles/rodrires>.
- Roth, Susan. 1999. "The State of Design Research." *Design Issues* 15 (2): 18–26.
- Ryan, R. M., and E. L. Deci. 2000. "Self-Determination Theory and the Facilitation of Intrinsic Motivation, Social Development, and Well-Being." *American Psychologist* 55 (1): 68.
- Salen, Katie. 2007. "Gaming Literacies: A Game Design Study in Action." *Journal of Educational Multimedia and Hypermedia* 16 (3): 301.
- Salen, Katie, and Eric Zimmerman. 2003. *Rules of Play: Game Design Fundamentals*. MIT Press.
- Sawyer, Ben, and Peter Smith. 2008. "Serious Games Taxonomy."
<http://www.dmill.com/presentations/serious-games-taxonomy-2008.pdf>.
- Sawyer, R. K. 2005. *The Cambridge Handbook of the Learning Sciences*. Cambridge University Press.
- Schaller, D. 2005. "What Makes a Learning Game." In *Adapted from a Presentation at the Web Designs for Interactive Learning Conference in Ithaca, NY*.
<http://www.educationalwebadventures.com/schaller-games.pdf>.
- Schell, Jesse. 2008. *The Art of Game Design: A Book of Lenses*. Morgan Kaufmann.
- . 2012. "Bad Simulations Educate Better than Good Simulations. They Raise More Questions." Tweet. *Games 4 Change Festival Twitter Feed*.
<http://new.livestream.com/g4c/festival/statuses/1561385>.
- Schön, Donald A. 1983. *The Reflective Practitioner: How Professionals Think in Action*. BasicBooks.
- Sharma, M., M. Holmes, J. Santamaria, A. Irani, C. Isbell, and A. Ram. 2007. "Transfer Learning in Real-Time Strategy Games Using Hybrid CBR/RL." In *Proceedings of the Twentieth International Joint Conference on Artificial Intelligence*.
<https://www.aaai.org/Papers/IJCAI/2007/IJCAI07-168.pdf>.
- Shearer, James D. 2011. "Development of a Digital Game-Based Learning Best Practices Checklist". Thesis. http://etd.ohiolink.edu/view.cgi/ris.cgi?acc_num=bgsu1303865257.
- Shelton, Brett E. 2007. "Designing Educational Games for Activity-Goal Alignment."
http://digitalcommons.usu.edu/itls_facpub/115/.

- Shelton, Brett E., and Jon Scoresby. 2010. "Aligning Game Activity with Educational Goals: Following a Constrained Design Approach to Instructional Computer Games." *Educational Technology Research and Development* 59 (1) (November 3): 113–138. doi:10.1007/s11423-010-9175-0.
- Singer, Dorothy G., Roberta M. Golinkoff, and Kathy Hirsh-Pasek. 2006. *Play Equals Learning*. Oxford University Press.
- Sitzmann, Traci. 2011. "A Meta-Analytic Examination Of The Instructional Effectiveness Of Computer-Based Simulation Games." *Personnel Psychology* 64 (2): 489–528. doi:10.1111/j.1744-6570.2011.01190.x.
- Slater, S. F., and J. C. Narver. 1998. "Research Notes and Communications Customer-Led and Market-Oriented: Let's Not Confuse The Two." *Strategic Management Journal* 19 (10): 1001–1006.
- spek. 2011. "Trying to Recreate Old-School Platformer Feel". Community site. *GameDev.Net - Physics*. <http://www.gamedev.net/topic/613988-trying-to-recreate-old-school-platformer-feel/>.
- Squire, Kurt. 2004. "Replaying History: Learning World History through Playing Civilization III." *Indiana University, Indianapolis, IN*.
- . 2007. "Open-Ended Video Games: A Model for Developing Learning for the Interactive Age." *The John D. and Catherine T. MacArthur Foundation Series on Digital Media and Learning*: 167–198.
- . 2010. "Civilization III and Whole-Class Play in High School Social Studies1." *Civilization* 34 (1). http://www.fi.ncsu.edu/assets/research_papers/new-literacies-collaborative/civilization-iii-and-whole-class-play-in-high-school-social-studies.pdf.
- Squire, Kurt, M. Barnett, J. M. Grant, and T. Higginbotham. 2004. "Electromagnetism Supercharged!: Learning Physics with Digital Simulation Games." In *Proceedings of the 6th International Conference on Learning Sciences*, 513–520. International Society of the Learning Sciences.
- Squire, Kurt, and Henry Jenkins. 2003. "Harnessing the Power of Games in Education." *Insight* 3 (1): 5–33.
- Squire, Kurt, Henry Jenkins, W. Holland, H. Miller, K. P. Tan, and K. Todd. 2003. "Design Principles of Next-Generation Digital Gaming for Education." *Educational Technology* 43 (5): 17–23.
- Squire, Kurt, and Nathan Patterson. 2009. "Games and Simulations in Informal Science Education". The National Academies Board on Science Education.
- Starr, P. 1994. "Seductions of Sim." *The American Prospect* 17: 19–29.
- Stegman, Melanie. 2010a. "Building on Success - Development of Immune Attack 2.0." *FAS Public Interest Report* 63 (1): 4.
- . 2010b. "Immune Attack: Teaching Kids About Science and Researchers About Designing Learning" presented at the AAAS 2010 Annual Meeting, February 21, San Diego Convention Center. <http://aaas.confex.com/aaas/2010/webprogram/Paper2494.html>.
- Steier, F. E. 1991. *Research and Reflexivity*. Sage Publications, Inc.
- Stout, Mike. 2010. "Evaluating Game Mechanics For Depth." *Gamasutra*. http://www.gamasutra.com/view/feature/5901/evaluating_game_mechanics_for_depth.php?print=1.
- Strommen, E. F., and B. Lincoln. 1992. "Constructivism, Technology, and the Future of Classroom Learning." *Education and Urban Society* 24 (4): 466–476.
- Tate, Richard, Jana Haritatos, and Steve Cole. 2009. "HopeLab's Approach to Re-Mission." *International Journal of Learning and Media* 1 (1) (February): 29–35. doi:10.1162/ijlm.2009.0003.
- Tedlock, Barbara. 2005. "The Observation of Participation and the Emergence of Public Ethnography." In *The Sage Handbook of Qualitative Research*. Sage Publications, Inc.
- TEGD. 2012. "The Educational Games Database (TEGD)." https://tegd.arizona.edu/index.php?title=Main_Page.

- TheSnivy10. 2012. *Lets Play Gamestar Mechanic Quest Part 3*. TheSnivy10.
http://www.youtube.com/watch?v=h-6QV8HNFyA&feature=youtube_gdata_player.
- Toups, Z. O, A. Kerne, and W. A Hamilton. 2011. "The Team Coordination Game: Zero-Fidelity Simulation Abstracted from Fire Emergency Response Practice." *ACM Transactions on Computer-Human Interaction (TOCHI)* 18 (4): 23.
- Tuck, B. 1995. "Issues of Objectivity in Assessment: A Plea for Moderation." *Setting the Standards: The Assessment of Competance in National Qualifications*: 59–78.
- Tufte, Edward. 2012. "Grand Truths about Human Behavior."
http://www.edwardtufte.com/bboard/q-and-a-fetch-msg?msg_id=0002XS.
- Turner, S., P. Turner, L. Dawson, and A. Munro. 2000. "DISCOVERing the Impact of Reality (poster Session)." In *Proceedings of the Third International Conference on Collaborative Virtual Environments*, 209–210. ACM.
- Tuzun, H. 2004. *Motivating Learners in Educational Computer Games*. Indiana University.
- Tyler, R. W. 1950. *Basic Principles of Curriculum and Instruction*. University of Chicago Press.
- Vallerand, R. J., L. G. Pelletier, M. R. Blais, N. M. Briere, C. Senecal, and E. F. Vallieres. 1993. "On the Assessment of Intrinsic, Extrinsic, and Amotivation in Education: Evidence on the Concurrent and Construct Validity of the Academic Motivation Scale." *Educational and Psychological Measurement* 53 (1): 159–172.
- Valve. 2007. *Team Fortress 2*. <http://www.teamfortress.com/>.
- Venn, J. 1880. "On the Diagrammatic and Mechanical Representation of Propositions and Reasonings." *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 10 (59): 1–18.
- Vogel, J. J., D. S. Vogel, J. Cannon-Bowers, C. A. Bowers, K. Muse, and M. Wright. 2006. "Computer Gaming and Interactive Simulations for Learning: A Meta-Analysis." *Journal of Educational Computing Research* 34 (3): 229–243.
- Vrasidas, C. 2000. "Constructivism versus Objectivism: Implications for Interaction, Course Design, and Evaluation in Distance Education." *International Journal of Educational Telecommunications* 6 (4): 339–362.
- Vredenburg, K., J. Y Mao, P. W Smith, and T. Carey. 2002. "A Survey of User-Centered Design Practice." In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Changing Our World, Changing Ourselves*, 471–478.
- Vygotsky, L. 1978. "The Role of Play in Development." *Mind in Society the Development of Higher Psychological Processes*: 92–104.
- Vygotsky, L. S. 1978. *Mind and Society: The Development of Higher Mental Processes*. Cambridge, MA: Harvard University Press.
- Walker, A., and B. E. Shelton. 2008. "Problem-Based Educational Games: Connections, Prescriptions, and Assessment." *Journal of Interactive Learning Research* 19 (4): 663.
- Walsh, Anita. 2007. "An Exploration of Biggs' Constructive Alignment in the Context of Work-based Learning." *Assessment & Evaluation in Higher Education* 32 (1): 79–87.
 doi:10.1080/02602930600848309.
- WeWantToKnow. 2012. *DragonBox - The Multi-Platform Math Game*. <http://dragonboxapp.com/>.
- White House. 2012. "Educate to Innovate." <http://www.whitehouse.gov/issues/education/educate-innovate>.
- Williams, B. 2006. "Microsoft Flight Simulator as a Training Aid." *Aviation Supplies & Academics, Newcastle*.
- Wilson, G. M. 1917. "The Motivation of School Work." *The Elementary School Journal* 17 (5) (January 1): 362–367.
- Wilson, K. A., W. L. Bedwell, E. H. Lazzara, E. Salas, C. S. Burke, J. L. Estock, K. L. Orvis, and C. Conkey. 2008. "Relationships Between Game Attributes and Learning Outcomes: Review and Research Proposals." *Simulation & Gaming* 40 (2) (May): 217–266.
 doi:10.1177/1046878108321866.

- Winn, Brian, and Carrie Heeter. 2008. "The Design, Play, and Experience Framework." *Handbook of Research on Effective Electronic Gaming in Education* 3: 1010–1024.
- Winn, W., M. Windschitl, R. Fruland, and Y. Lee. 2002. "When Does Immersion in a Virtual Environment Help Students Construct Understanding." In *Proceedings of the International Conference of the Learning Sciences, ICLS*, 497–503.
- Wood, D., J. S. Bruner, and G. Ross. 1976. "The Role of Tutoring in Problem Solving." *Journal of Child Psychology and Psychiatry* 17 (2): 89–100.
- Woods, Stewart. 2004. "Loading the Dice: The Challenge of Serious Videogames." *Game Studies* 4 (1) (November). <http://www.gamestudies.org/0401/woods/>.
- World Peace Game Foundation. 2012. "About the World Peace Game." <http://www.worldpeacegame.org/world-peacegame-foundation/about-the-game>.
- "YCombinator Discussion of DragonBox." 2012. <http://news.ycombinator.com/item?id=4105397>.
- Young, W., L. W. Peng, S. Carroll, T. Franklin, C. Lui, and D. Chelberg. 2009. "An Investigation of Self-Efficacy Using Educational Video Games Developed by the GK-12 STEAM Project." In *Society for Information Technology & Teacher Education International Conference 2009*. http://vital.cs.ohiou.edu/steamwebsite/downloads/site09_self_efficacy_final.pdf.
- Z2Live. 2011. *Battle Nations*. <http://battlenations.com/>.
- Zecher, Andreas. 2008. "Understanding Games Website." <http://www.madebypixelate.com/games/understanding-games>.
- Zecher, Andreas, and Martin Straka. 2007. "Understanding Games, Episode 4." *Understanding Games*. <http://www.madebypixelate.com/games/understanding-games/episode/4>.
- Zepke, N., and L. Leach. 2010. "Improving Student Engagement: Ten Proposals for Action." *Active Learning in Higher Education* 11 (3) (October 26): 167–177. doi:10.1177/1469787410379680.

Every reasonable effort has been made to acknowledge the owners of copyright material. I would be pleased to hear from any copyright owner who has been omitted or incorrectly acknowledged.