FAKULTÄT FÜR !NFORMATIK
Faculty of Informatics

Groundwork for a new perspective on the evaluation of serious games

DIPLOMARBEIT

zur Erlangung des akademischen Grades

Diplom-Ingenieur

im Rahmen des Studiums

Medieninformatik

eingereicht von

Christian Knapp
Matrikelnummer 0926133

an der Fakultät für Informatik
der Technischen Universität Wien
Betreuung: Ao.Univ.Prof. Dipl.-Ing. Dr.techn. Peter Purgathofer

Wien, 19. April 2017

Christian Knapp

Peter Purgathofer

FAKULTÄT FÜR !NFORMATIK
Faculty of Informatics

Groundwork for a new perspective on the evaluation of Serious Games

DIPLOMA THESIS

submitted in partial fulfillment of the requirements for the degree of

Diplom-Ingenieur

in

Media Informatics

by

Christian Knapp
Registration Number 0926133

to the Faculty of Informatics
at the TU Wien
Advisor: Ao.Univ.Prof. Dipl.-Ing. Dr.techn. Peter Purgathofer

Vienna, 19th April, 2017

Christian Knapp	Peter Purgathofer

Erklärung zur Verfassung der Arbeit

Christian Knapp
Schützengasse 7/10
1030 Wien

Hiermit erkläre ich, dass ich diese Arbeit selbständig verfasst habe, dass ich die verwendeten Quellen und Hilfsmittel vollständig angegeben habe und dass ich die Stellen der Arbeit – einschließlich Tabellen, Karten und Abbildungen –, die anderen Werken oder dem Internet im Wortlaut oder dem Sinn nach entnommen sind, auf jeden Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht habe.

Wien, 19. April 2017

Christian Knapp

Danksagung

Als erstes möchte ich mich bei meinem Betreuer Peter Purgathofer bedanken, nicht nur für das Betreuen meiner Diplomarbeit, sondern ganz besonders für die wertvollen Einsichten die ich aus all seinen Lehrveranstaltungen mitnehmen konnte. Mein Studium wurde stark durch seine Lehre geprägt.

Ein besonderer Dank geht an meine Eltern, dafür, dass ihr mich während meiner Studienzeit immer unterstüzt habt, und mir nie einen Weg vorgeschrieben habt. Ein weiterer besonderer Dank geht an Geraldine. Danke für deine Geduld und deine Unterstützung in den Momenten, in denen das Schreiben an dieser Arbeit am schwersten war.

Und zum Schluss noch ein Dank an alle, die mich während meines Studiums in irgendeiner Form begleitet haben.

Kurzfassung

Serious games sind Spiele, die nicht nur einem reinen Unterhaltungszweck dienen. Die Industrie und das Forschungsgebiet der serious games befinden sich in starkem Wachstum. Allerdings basierte die Forschung über ihre Effektivität lange auf theoretischen Annahmen über ihr Potenzial, ohne dafür empirische Nachweise zu erbringen. Studien, die serious games evaluieren, bringen ergebnislose, inkonsistente oder sogar widersprüchliche Ergebnisse hervor. Ein Grund dafür ist das Fehlen von standardisierten Methoden und Richtlinien zur Vermeidung verbreiteter Fehler, und das unhinterfragte Übernehmen von Methoden aus anderen Forschungsgebieten.

Diese Arbeit versucht Probleme in den verbreiteten Methoden zur Evaluierung von serious games zu identifizieren. Dazu gehört das Finden von intervenierenden Variablen, sowie von Eigenheiten in der Evaluierung von Spielen. Randomized controlled trials, die meist als Methode der Wahl genannt werden, stellen sich als oftmals ungeeignet heraus. Außerdem wird von der Anwendung der weit verbreiteten Medienvergleichsstudien abgeraten, da deren Annahme über die Vergleichbarkeit von Lehrmethoden in unterschiedlichen Medien nicht standhält.

Evaluierungen von serious games treffen oftmals Annahmen, die auf einem falschen Verständnis von Spielen beruhen. Um diesen Missverständnissen entgegen zu treten, bedarf es einer einheitlichen Perspektive zur Unterstützung von Gestaltung, Anwendung, Evaluierung und Weiterentwicklung von serious games. Die Voraussetzung dafür, nämlich das Verstehen von Spieldynamik, Lerntheorien und derer Anwendungsmöglichkeiten auf Spiele, werden in dieser Arbeit betrachtet.

Die vorgeschlagene neue Perspektive auf das Design und die Evaluierung von serious games wird beispielhaft an den Spielen EnerCities, Spent und Portal 2 angewendet. Die Spiele werden aus lerntheoretischer Sicht analysiert, um Ergebnisse aus vorangegangenen Studien zur Effektivität der Spiele zu erklären.

Abstract

Serious games are games that contribute to a purpose other than pure entertainment. The serious games industry and research community are growing rapidly. However, research on their effectiveness has long been based on theoretical assumptions about their potential, but has been missing empirical evidence. Studies attempting to evaluate serious games show inconclusive, inconsistent and even contradictory results. This is caused by a lack of standardized methods and guidelines to avoid common mistakes in the evaluation of serious games, and the unquestioned adoption of methodologies from other domains.

This thesis attempts to identify issues in common methodologies used in the evaluation of serious games. This includes the identification of mediating variables as well as some peculiarities regarding the evaluation of games. The use of randomized controlled trials - promoted as method of choice in the field - is found to be unnecessary or even inappropriate in many cases. The common use of media comparison designs is discouraged, primarily due to their assumption that instructions in different media can be designed to teach identical content, and therefore the impact of the instructional medium can be isolated and measured.

Many evaluations of serious games make assumptions based on misconceptions about games. To develop a common understanding of the domain, a unified perspective is proposed to aid design, application, evaluation and iteration of serious games. The prerequisites for it, namely the understanding of game dynamics, of learning theories and of their possible applications in games, are covered in this work.

The proposed new perspective on the design and evaluation of serious games is exemplified on the games EnerCities, Spent, and Portal 2. All of them have been evaluated in instructional contexts in previous studies. In this work, the games are analyzed from a learning theoretical standpoint to explain evaluation results from previous studies.

Contents

CHAPTER 1

Introduction

48% of Europeans were playing any type of digital games in 2012. At the age of 16–24, numbers are even higher at 80% of males and 61% of females playing. (Ipsos MediaCT & ISFE, 2012) With the rise of games on ubiquitous mobile platforms, i.e. smartphones, digital games reach an even broader audience. Today, video games are the fourth biggest entertainment market in the world, bigger than movies and music. In 2015, video games had a global value of $91.5 billion. (BusinessTech, 2015)

"Entertaining", "good at providing escapism" and "fun" are the most commonly associated terms with gaming. But only 7% associate gaming with "informative/educational", while 20% assign it to film and 47% to television. (Ipsos MediaCT & ISFE, 2012)

Nevertheless, the significance of serious games is rising, and the research community is growing with its industry. (Laamarti et al., 2014) With increasing academic interest in serious games, and many studies evaluating games on their potential for learning, came an increase in critique on methodologies and lack of sound empirical evidence in the young research area. (Mayer, 2012; Connolly et al., 2008; Soekarjo and van Oostendorp, 2015) A cause for problems in the serious games industry and research is a lack of understanding, that different stakeholders in the process of evaluating serious games have of each other's domains. Experienced instructors, researchers and game designers need to interwove their knowledge closely to advance the field of serious games, and to avoid misinterpretations about how games can be used for learning.

This thesis aims on exposing issues in serious games research by analyzing literature in the field, and addressing these issues with the proposition of a new perspective on the development and evaluation of serious games, which considers the role of instructors, researchers, game designers, and learners in the evaluation process.

The following chapter introduces the notion of games, poses the difficulties in defining what games are, and explores what it is about games, particularly digital games, that is appealing to players.

In the third chapter, the foundations of games for learning are explored by analyzing scientific literature on arguments speaking for and against the use of games in learning environments. Then, common terminologies are introduced, and the term *serious games* is discussed in detail. The chapter concludes with a section presenting classifications of learning outcomes. Following leading literature, four dimensions of learning are proposed.

The fifth chapter offers an overview of the wide range of literature on the evaluation of serious games. Based on four meta reviews, the main themes in literature are identified. A further analysis on methodologies used in the evaluation of serious games offers a thorough overview about the state of the art in the research area, including case studies to illustrate the application of some common methodologies. The chapter concludes with a critique on the prevalent methods, offering detailed argumentation about misconceptions in the research of serious games that lead to confounded research results.

Chapter six provides an overview on learning theories in different epistemological frames, to build up the basis for the introduction of a new systematic perspective on serious games research.

The seventh chapter introduces important principles of game design and analysis. Without a basic understanding of how games work on a systemic level, a comprehensive evaluation of serious games is impossible.

Chapter eight synthesizes findings on the methodological and conceptual issues in the evaluation of serious games, principles from learning theories, and insights in game design theory into a systematic perspective on the development and evaluation of serious games. It describes how theoretical principles can be applied in instructional methods in games, and how they interplay with the structure of games. This leads to the proposition of a model, that should be viewed as groundwork for a new, more holistic, and interdisciplinary perspective.

In chapter nine three serious games are picked to be analyzed, using the proposed model.

The last chapter poses final conclusions, and suggestions for future work and for development of the proposed perspective.

CHAPTER 2

Games

This chapter introduces games, and discusses what they are and what makes them appealing to their players.

2.1 What is a game?

To be able to discuss games for learning, it needs to be clarified what games are.

In general, it is intuitively clear to humans what a game is. When people play a game, they just know they are playing. Games and play occur in nearly every known culture. (Juul, 2005) Therefore, one might be tempted to think that it would not be difficult to give a precise definition of what a game is.

So it might come as a surprise, that the first philosopher that is noted to try to precisely define games was Ludwig Wittgenstein in 1953, and he failed to do so. (Wittgenstein, 1953)

First it is of importance to note the differentiation between the terms play and game. An attempt to define play has already been made by John Huizinga in his work *Homo Ludens*, published in 1938:

> Summing up the formal characteristic of play, we might call it a free activity standing quite consciously outside 'ordinary' life as being 'not serious' but at the same time absorbing the player intensely and utterly. It is an activity connected with no material interest, and no profit can be gained by it. It proceeds within its own proper boundaries of time and space according to fixed rules and in an orderly manner. It promotes the formation of social groupings that tend to surround themselves with secrecy and to stress the difference from the common world by disguise or other means. (Huizinga, 1938)

Just as for finding a definition for the term game, there were several subsequent attempts to define play, and just as with games there is not a single, consensual definition to date.

Back to Wittgenstein and his failed attempt to define the notion of games:

> What is common to them all? —Don't say: 'There *must* be something common, or else they would not be all called "games"'—but *look* and *see* whether there is anything common to all. (Wittgenstein, 1953)

Wittgenstein was not the last to struggle to define games. In 1963, Roger Caillois published his work *Man, Play, and Games*. He extensively discusses several qualities that might be defining characteristics of games and explicitly notices that Huizinga in his work only defined the term play and omitted a definition of games. Interestingly enough, he himself concludes with only a definition for play:

> An activity which is essentially: free (in which playing is not obligatory), separate (circumscribed within limits of space and time), uncertain (the course of which cannot be determined), unproductive (creating neither goods, nor wealth, nor new elements of any kind), governed by rules (under convention that suspends ordinary laws), make-believe (accompanied by a special awareness of a second reality or of a free unreality, as against life). (Caillois and Barash, 1961)

The next prominent attempt to define games was made by Bernard Suits in his work *The Grasshopper*, which is a widely used definition to date:

> To play a game is to engage in activity directed towards bringing about a specific state of affairs, using only means permitted by rules, where the rules prohibit more efficient in favor of less efficient means, and where such rules are accepted just because they make possible such activity. (Suits, 1967)

Or more briefly:

> Playing a game is the voluntary attempt to overcome unnecessary obstacles. (Suits, 1967)

In 1982, game designer Chris Crawford attempted a definition of the term *game*, being one of the first to consider video games.

> I perceive four common factors: representation, interaction, conflict, and safety. (Crawford, 1984)

A more recent definition comes from game researchers Katie Salen and Eric Zimmerman:

> A game is a system in which players engage in an artificial conflict, defined by rules, that results in a quantifiable outcome. (Zimmerman, 2004)

Games advocate Jane McGonigal related her definition to the one of Bernard Suits:

> All games share four defining traits: a goal, rules, a feedback system, and voluntary participation. (McGonigal, 2011)

A very comprehensive and thorough examination of the definition of games was made by games researcher Jesper Juul. (Juul, 2005) He importantly notes that a good definition should describe the formal system of the game, the relation between the player and the game, and the relation between the playing of the game and the rest of the world. He suggests a new definition of games, based on six features:

> A game is a rule-based system with a variable and quantifiable outcome, where different outcomes are assigned different values, the player exerts effort in order to influence the outcome, the player feels emotionally attached to the outcome, and the consequences of the activity are negotiable. (Juul, 2005)

Juul uses a number of examples to test several definitions as well as his own, following the argument of Bernard Suits that "the simplest way to test a game definition is to test it for being either too broad or too narrow." (Suits, 1967)

As Juul describes, a problem in the definitions by Huizinga and Caillois is their description of games being "outside ordinary life" or "separate withing limits of space and time". Examples for countering this presumptions made by Juul are playing chess by mail (which overlaps with daily life in time and space), or live action role-playing games (which take place in the same space as ordinary life). Other examples are augmented reality games, which actually use real life as part of the game by definition, or the current trend of gamification, which explicitly transforms part of ordinary life into a game.

McGonigal shares with Caillois' definition that participation in games should be voluntary. Juul criticizes this, because human motivation is too complex to use voluntariness as a feature in the definition of a term. (Juul, 2005)

The problem in Suits definition is the depiction of "less efficient means". If someone contests a friend to get to a certain other point in a city as quick as possible, with all means of transportation allowed, the rules do not call for less efficient means, and still they are playing a game. The game then *is* to find the most efficient means. The notion completely falls apart when taking video games into account, since less efficient means lose their meaning in a virtual, constructed world.

McGonigal's definition is the only one of the aforementioned which was not analyzed by Juul, but has its deficiencies nevertheless. Imagine a digital interactive application, which consists only of the screen showing the text "Click the button below to win", and a big button underneath it. If you click the button, the button disappears and the text changes to "Congratulations, you won". Intuitively, this would not be considered to be a game. But, following McGonigal's definition of games, it fulfills every aspect of it. It has a clear goal which the player will work to achieve. The rules are clearly defined and limiting the players possibility space. There is a clear feedback system, telling the player exactly what is missing to achieve the goal. And finally, the participation is voluntary, accepting the goal, rules and feedback.

What is missing here, is what Juul describes as "variable outcomes". Imagine the same situation as before, but with the text "Click the small button", with two buttons below, one slightly smaller than the other. If the player clicks the small button, the text changes again to "Congratulations, you won", if clicking the big button, the text says "You lost!". With this modification, the activity becomes something that would more likely be described as a game, even if a very rudimentary one. The modification breaks linearity, introduced variable quantifiable outcomes with different values, and therefore fits Juul's definition.

The example of a duel, in the sense of an arranged combat between two people using guns, emphasizes the importance of Juuls feature of negotiable consequences. (Juul, 2005) According to McGonigal's definition (and many other definitions as well), a duel would be a game. It has clear goals, well defined rules limiting the possible actions, a feedback system that immediately (and in fact dramatically) shows the result, and the duelists voluntarily choose to participate, even if there might be underlying reasons that pressure them for participation, e.g. societal pressure. In Juuls defnition, the duel would not be a game, because the consequences of it are not negotiable but necessarily result in serious injury or even death of one of the duelists. Would the potentially lethal weapons be switched to harmless or imaginary weapons, the very serious situation would turn into a game, as consequences are now negotiable. This example also sheds some light on why the notion of voluntary participation in a definition is problematic, as it is almost impossible to determine if an act is voluntary to what degree.

Another aspect that some scholars like Prensky (2003) ascribed to games is a story element. But there is a wide range of games to prove that story or representation is not a necessary feature for a game (e.g. Tetris).

Overall, the remaining question is how broad or narrow a definition should be, which depends on what one wants to be included in the definition. As mentioned in the beginning, people have an intuitive understanding of what a game is, and linguistics also play a role in what is described as a game and what is not.

For the purpose of this thesis, the definition by Juul is sufficiently accurate to include what is commonly understood as a game, but exclude cases which fit other definitions but are generally not considered to be games.

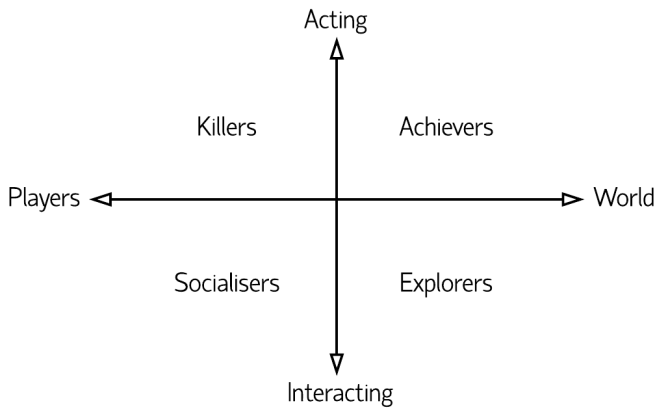

Figure 2.1: Dimensions of player types, according to Bartle (1996).

> A game is a rule-based system with a variable and quantifiable outcome, where different outcomes are assigned different values, the player exerts effort in order to influence the outcome, the player feels emotionally attached to the outcome, and the consequences of the activity are negotiable. (Juul, 2005)

2.2 Why do people play games?

What is it about games that is so appealing to their players? Why do people play games?

The first important contribution to answering this question goes back to Bartle (1996). He identified four player types: killers, achievers, socializers, and explorers. These player types are defined by their placement on two dimensions (see fig. 2.1).

Yee (2006) identified three overarching player motivations with several subtypes, derived from factor analysis in a study with over 3000 players of MMORPGs, using an extensive questionnaire. The motivation types were:

- achievement (desire for progression, power, and for challenging and competing with others, interest in underlying system to optimize performance)
- social (helping and chatting with others, forming relationships, satisfaction from being part of a group effort)
- immersion (finding and knowing things others do not, role-playing a persona, customizing appearance of the character)

Yee disagrees with Bartle, in that he views player motivations as not mutually exclusive, while Bartle's player types suppress each other.

Hoffman and Nadelson (2010) conducted a mixed-method study, encompassing an extensive questionnaire and interviews, to identify motivational factors to play games. They found three main factors: engagement in video games satisfies socialization needs, the perception of control and challenge influenced engagement, and cognitive and affective outcomes lead to feelings of satisfaction, accomplishment, and contentment.

Self-determination theory by Ryan and Deci (2000) has also been used to explain motivations for playing games. According to the theory, competence, autonomy and relatedness are defining factors for intrinsic motivation in individuals. Przybylski et al. (2010) supported this theoretical basis for playing games, as they found competence and autonomy to be important factors for player motivation across all game genres and player demographics. They explicitly note, that violent content has not been found to be a motivator of play.

CHAPTER 3

Games and learning

This chapter addresses the relation between games and learning. How are games used for learning? Why can games be successful tools for learning? How is the academic community looking at games as learning tools? And how can games for learning be categorized?

3.1 Arguments for using games for learning

Play is considered to be an integral part in children's cognitive development, especially to develop multiple representations of objects which supports the development of symbolic thinking. (Plass et al., 2015) But play is still often ascribed as something children do, not adults. And often the intuitive understanding of games for learning may be directed towards analog games, as digital games still suffer from an image of being a waste of time, or even promoting asocial and violent behaviours.

Under the term of edutainment, video games have already been used for educational purposes two decades ago. Games are fun - this was the basic assumption for the benefit of using games in education. But fun is not only a term difficult to define, but especially difficult to measure in terms of its effect on successful learning.

From a scientific point of view, scholars hypothesize that game features are highly consistent with modern learning theories. (Gee, 2003; Shute, 2011; Boyle et al., 2011; Squire, 2011; Prensky, 2007; Linehan et al., 2011)

As described in section 2.2, people are attracted to games for numerous reasons. This motivational appeal is often considered as the most important feature for games as learning tools. (Abt, 1970; O'Neil et al., 2005; Wouters et al., 2013; Hays, 2005; Kebritchi and Hirumi, 2008; Arnab et al., 2016; Breuer and Bente, 2010; Boyle et al., 2011; Ryan et al., 2006; Linehan et al., 2011) The motivating characteristics of video games make players come back to a game over a long period of time, something that conventional

educational methods often struggle to achieve. Also, students' motivation and engagement is considered a key factor in efficient learning. (Linehan et al., 2011)

Another important factor is action instead of explanation. Video games are an ideal environment for active learning due to their highly interactive nature. (O'Neil et al., 2005; Romero et al., 2015; Wu et al., 2012; Kebritchi and Hirumi, 2008; Lieberman, 2006)

Feedback is enabled by assessment, and is a core part of games. Every game measures players' performance in some form, at least by assessing if a task was finished successfully, or keeping any kind of score. This assessment is integrated in games, meaning that direct and immediate feedback can be provided without breaking immersion. Direct feedback is considered to be an important element in game design and at the same time an important aspect for successful learning. (O'Neil et al., 2005; Shute and Ke, 2012; Hays, 2005; Lieberman, 2006)

Additionally to feedback, assessment also enables games to dynamically adapt challenge in the game to fit players' skill levels. The right degree of challenge is crucial for engagement and learning, to avoid frustration or boredom, and to increase players' self-efficacy and perceived competence. (Ryan et al., 2006; Shute, 2011; Hays, 2005)

Modern, constructivist learning theories stress the importance of situated learning, which means that learning should be situated in an authentic context. This is reflected in modern video games, which can offer highly complex, immersive and fantastic virtual worlds. It makes them ideal "playgrounds" for the exploration of identities and situated understanding of matters. (Shaffer et al., 2005; Shute and Ke, 2012; Lieberman, 2006)

Many video games offer highly complex, immersive virtual worlds and demand players to solve complex tasks situated in this world. The skills required in these virtual environments are high level cognitive abilities, therefore supporting skills like situated understanding, critical thinking, problem solving, decision making and argumentation. In games, the application of these skills can be explored by players without facing real-world consequences. (Abt, 1970; Shaffer et al., 2005; Shute, 2011; Boyle et al., 2011; Kebritchi and Hirumi, 2008; Romero et al., 2015)

Video games are system, and they represent real-life systems directly (simulation games) or metaphorically. This offers players the opportunity to interact with a system and its parameters and build a better understanding of corresponding real-life systems. (Shute and Ke, 2012; Hays, 2005) Additionally, the representation of real-life systems offers the possibility for expression through video games. This creates opportunity for not only games as art, but also to confront players with their attitudes and behaviours and foster changes in affective states of players. (Bogost, 2007; Frasca, 2007; Shaffer et al., 2005)

With multiplayer video games, social components are getting increasingly important in games. This is in line with modern learning theory, which emphasizes the importance of cooperation and socially embedded learning. (O'Neil et al., 2005; Shute, 2011) Players of multiplayer games can develop social practices, explore group identities, and develop shared values. (Shaffer et al., 2005; Steinkuehler, 2004)

A concise argument for games as learning tools is that every game is about learning in the first place, as every game teaches at least one thing - how to play the game. Especially since manuals have gradually disappeared as close companions of game, game designers gained a lot of experience in developing methods to interactively teach players how to play the game.

Games contend the classical classroom approach of teachers trying to transfer their knowledge to the students. This conception is based on a information delivery view, meaning that learning happens by adding information to memory, mostly through words. This view is not only long disproved in research, but is also increasingly found to be unattractive by students. (Boyle et al., 2011)

Games for learning advocate Marc Prensky cites media theorist Marshall McLuhanwith with the following words:

> Anyone who makes a distinction between games and education clearly does not know the first thing about them. (Prensky, 2007)

3.2 Arguments against using games for learning

Given all the excitement about the theoretical possibilities of using games to educate and change behaviour, there is also some skepticism about the applicability of games for learning.

One of the key factors for the motivating and engaging nature of games is that playing games is a voluntary act. As discussed in section 2.1, even some definitions of play and game include voluntariness. (Huizinga, 1955; Caillois and Barash, 1961) Serious games, however, are often not played voluntarily, especially if they are incorporated in a classroom curriculum. This could undermine players' motivation to play the game, harm engagement and therefore reduce learning effects. (Garris et al., 2002)

The potential of games for learning has often been misunderstood or looked at superficially. A symptom of this was the edutainment movement in the 1990s. Educators and producers of edutainment games assumed that motivational pull of games could be used as an incentive to persuade students to practice. This resulted in games for learning with strictly separated playing and learning parts, using playing parts as rewards for successfully finishing practice.

Educators came to understand that it is necessary not only to include gameplay parts, but to intertwine these parts with learning and meld the two into a unified experience. But still, serious games developed in the last decade often suffer from neglecting quality of gameplay. (Egenfeldt-Nielsen, 2006; Breuer and Bente, 2010) Most of serious games are acclaimed to be inferior to entertainment games in terms of gameplay, with some notable exceptions. Bruckman (1999) described this as sugar coating education with fun, and compares it to "chocolate-dipped broccoli". It is too simplistic to think that just

because something is a game, it actually is motivating and fun to play. (Linehan et al., 2011)

A problem many games for learning have is the transfer of knowledge, skills or attitudes acquired in the game's context to a real-life context. (Arnab et al., 2016) One example for prevented transfer is when simulation games do not communicate their underlying models, and therefore players understand the game model on the surface but are not able to transfer this knowledge because they gain no understanding of underlying systems. (Hays, 2005).

Another problem is that teachers often do not have a deep understanding of games. They sometimes use games in classrooms for external reasons, or because they think games are worth implementing in a classroom curriculum, but do not play games themselves. This limited understanding of games can result in choosing games which teach something else than what instructors intended, or fail to promote transfer of the learning content. (Squire, 2003)

Following the argument that games provide opportunities to influence players' attitudes and behaviours, it seems natural to worry about the effect that the many violent games on the market have on millions of people who play them. This question has been asked numerous times in research, and there is still no clear, concluding answer to this. (Ryan et al., 2006; Connolly et al., 2012) However, it is a concern for many parents and educators when considering the use of games in education.

An argument against using games for learning comes from a school of thought, which argues that media by itself has no influence whatsoever on the efficiency of learning. (Clark, 1994) According to Clark, the difference in instructional efficiency are moderated entirely by the instructional methods employed, and any instructional method can be used in any medium. So, the medium itself is no determining factor, and since games are a very expensive medium to produce, their cost-efficiency factor does not make them a good option for instructional use. This controversial argument will be examined in more depth in section 4.4.6.

3.3 Terminology

To enable a scientific discourse about the use of games for learning, there is a need for a common vocabulary. The field suffers from a fragmentation in terminology, so stakeholders involved in the creation, application and evaluation of games for learning have no shared language.

> Every scientific discourse needs an appropriate language. At least, one should be able to ask questions and to formulate hypotheses. The digital games science or what some day might be called the digital games science is nowadays still lacking its language of discourse. It is not yet a science. (Jantke and Gaudl, 2010)

There is a number of terms used in the scientific community and in the industry to describe games with a purpose other than entertainment. The most widespread one is *serious games*, which was introduced into the research community in 2003 by Sawyer (2003). Other terms that describe the same or a similar category of games are: *game-based learning*, *edutainment*, *social impact games*, *educational games*, *games for change*, *persuasive games*, and *positive impact games*. This section discusses the use of the term serious games and how it relates to other terms used in literature and practice.

3.3.1 What is a serious game?

The term *serious games* is assumed to have originated from the book with the same name by Abt (1970), and first applied to digital games by Sawyer (2003). There have been several, quite similar, definitions of the term:

> [Serious games] have an explicit and carefully thought-out educational purpose and are not intended to be played primarily for amusement. (Abt, 1987)
>
> [Serious games are] mental contests played with a computer in accordance with specific rules that uses entertainment to further government or corporate training, education, health, public policy, and strategic communication objectives. (Zyda, 2005)
>
> A serious game is a game in which education (in its various forms) is the primary goal, rather than entertainment. (Michael and Chen, 2005)
>
> Serious games are games that use the artistic medium of games to deliver a message, teach a lesson, or provide an experience. (Michael and Chen, 2005)
>
> [Serious games are] Entertainment Games with Non-Entertainment Goals. (Prensky, 2003)

All of these definitions have in common that they describe serious games as games which do not have entertainment as their primary purpose.

Two questions are raised by the term serious games:

- Can a game be serious?
- Are not all games serious?

Using definitions of games that include the concept of fun, enjoyment or something related, the term serious game can be considered an oxymoron. (Breuer and Bente, 2010) But at the same time the idea of marking down a set of games as serious games implies that all other games are non-serious. (Frasca, 2007) As Michael and Chen (2005) describe, most game developers and players see all their games as serious. This would leave the term serious games as a tautology. (Breuer and Bente, 2010) To overcome this issue, the

serious in *serious games* should be seen related to the context of use of a game instead of using it to define a fixed set of games.

Jantke and Gaudl (2010) criticize that most definitions are based on the intention of the developers, because these intentions are not always accessible. In contrast, they propose a more pragmatic approach to define serious games as "any game that may be used for more than entertainment" (Jantke and Gaudl, 2010). Backlund and Hendrix (2013) go in the same direction, defining serious games for their work as any game that contributes to a defined purpose other than pure entertainment, no matter if the formulation of the purpose comes from game designers or users themselves. This is also supported by Djaouti et al. (2008), who suggest to introduce the term *serious gaming*, to point to the context-dependent interpretation of when a game is serious.

An important implication of this approach is that it allows for the inclusion of games which were not developed for a specific purpose other than entertainment, but get applied in a context to fulfill a non-entertainment purpose. This is not uncommon in practice (see Egenfeldt-Nielsen, 2006) and should therefore be considered in research.

The hereby suggested approach to not restrict the term serious games to developer intentions is in line with Frascas approach to define games from a perspective of their use by players instead of mere artifacts. (Frasca, 2007)

Ultimately, the term serious games suffers from the semantic interpretation of the word *serious*, which opposes the intuitive understanding of games. But, as Bogost (2007) formulated it, it is a term that we are now probably stuck with to describe our discipline.

Considering the above arguments, I want to follow the definition by Backlund and Hendrix (2013) and define serious games for the purpose of this study as a game that contributes to a defined purpose other than pure entertainment.

However, it is important to keep in mind that the term serious games is usually used to describe a set of games explicitly *designed* towards a serious purpose.

3.3.2 Alternative terms

To set the term *serious game* in context with other terms, this section provides an overview about alternative terms and how they relate to each other and to serious games.

Edutainment was the prevalent label for games with a learning purpose in the 1990s. It describes games that use traditional learning techniques, aimed mostly at school children, and used game elements as additive element to the learning component. As such, they can be seen as a subset of serious games, with specific instructional methods, target audiences and in-game structures. (Breuer and Bente, 2010)

Game-based learning is a term coined by Prensky (2003), and has been widely used since. While Prensky offers no clear definition, the descriptions and examples given suggest a focus on educational use. If intended by Prensky or not, the term seems to be used to relate to the use of games to acquire and develop knowledge and skills as opposed to

affective changes. As such, game-based learning is a subcategory of serious games with a focus on using games in education (not restricted to school education), with strictly defined learning outcomes that are expected to be reached.

Persuasive games is a term coined by Bogost (2007) in his influential book with the same title. He defines persuasive games as games which incorporate procedural rhetoric well. Procedural rhetoric in games is the use of game mechanics to persuade players, to ultimately change their attitudes or behaviours. Persuasive games are a subset of serious games, because they fulfill the definition of games being used for a purpose other than entertainment, but they do not encompass the entirety of purposes that games can have besides entertainment.

3.3.3 Classifying serious games

Serious games have been classified in various dimensions.

- Market in which they are applied (e.g. health care, military, government, education, corporate, political, religious or art) (Michael and Chen, 2005; Laamarti et al., 2014)
- Type of activity (e.g. physiological, mental) (Laamarti et al., 2014)
- Interaction style (e.g. keyboard/mouse, movement tracking, tangible interfaces) (Laamarti et al., 2014)
- Environment (e.g. social presence, virtual environment, 2D/3D) (Laamarti et al., 2014)
- Game genre (e.g. action, RPG, puzzle) (Connolly et al., 2012)
- Digital/non-digital (Connolly et al., 2012)
- Platform (e.g. PC, consoles, mobile) (Connolly et al., 2012)

The categorization by market domain may be useful for instructors, because it makes it easier to find the right game for a specific purpose. In research however, the important question is for criteria that make games effective in a certain learning dimension.

3.4 Dimensions of learning

Learning dimension is the term that will be used throughout the thesis to describe categories to which outcomes of instructional interventions can be assigned. The use of these dimensions is the baseline for finding appropriate instructional methods to fulfill an instructional objective.

The term *instructional objective* describes the targeted learning outcome that is defined when developing a serious game. The term *learning outcome* describes the actual, specific outcome after playing a game. In this context, *learning dimension* describes the category that the learning outcome can be assigned to. To better illustrate the terms, let us assume a serious game that should improve visual perceptual skills. In this case, the instructional objective is improving visual perceptional skills. The learning outcome depends on the actual effectiveness of the serious game. In case of success, it is the improvement of visual perceptual skills, but the outcomes could also include other, not intended effects, such as the acquisition of knowledge. The learning dimension then is the category the learning outcomes are assigned to, which in this case is motor skills.

The definition of learning dimensions is useful, because different learning theories are concerned with different learning dimensions. Knowing the learning dimension can help identify which instructional methods to look for in a game.

Bloom (1956) classified the what he called educational objectives into levels of complexity and specificity, and described the cognitive domain as consisting of the six basic objectives (phases) of remembering, comprehending, applying, analyzing, synthesizing and evaluating. He also mentioned an affective domain, but did not integrate it into the original taxonomy because he described it as too difficult to assess. (Dettmer, 2005)

In recent scientific literature, there is a broad consensus on the basic classification of learning outcomes. It is based on the work by Kraiger et al. (1993), who synthesized previous works into the proposition of three major categories of skill-based, cognitive, and affective outcomes, with each one of them having several subcategories. In later years, additional factors such as collaboration, communication and self-regulation were introduced as additional possible classifications for learning outcomes. (Baker and Mayer, 1999) Some of these skills are now subsumed as *21st century skills*. (Romero et al., 2015) They are described by Romero et al. (2015) as being combinations of knowledge, skills, and attitude, therefore consist of categories that were already defined by Kraiger et al. (1993). But a set of skills that stand for themselves are what Romero et al. (2015) calls *interpersonal abilities*, including social, cultural and communication skills. Other researchers support the suggestion to add communicative and social skills as a category of learning outcomes. (Dettmer, 2005; Baker and Mayer, 1999; Wouters et al., 2009)

3.4.1 Cognitive learning outcomes

Cognitive learning outcomes are subdivided into knowledge and cognitive skills. Knowledge is further divided into subtypes which form a hierarchy in the process of knowledge acquisition.

Declarative knowledge. Declarative knowledge describes explicit knowledge of facts, and is obtained if one is able to recall data or information. This knowledge is a prerequisite for procedural knowledge.

Procedural knowledge. Procedural knowledge is the knowledge of how to perform a task. It is the application of declarative knowledge or of skills to a specific case.

Strategic knowledge. Tactic or strategic knowledge (knowledge about *which*, *when* and *why*) is a more complex set of knowledge, requiring the application of principles and derivation of new principles in novel situations.

Knowledge organization. Knowledge organization means organizing information into mental models. This is achieved by integrating representations of information in working memory with one another and with prior knowledge. (Mayer, 2014)

Cognitive skills. Cognitive skills are developed and can be applied, when enough knowledge is obtained and organized to create sufficiently complex mental models. The application of cognitive skills are complex cognitive processes, such as problem solving and decision making. These skills require a set of underlying skills and knowledge of picking relevant information from a situation and comprehend and organize the information.

3.4.2 Sensimotor learning outcomes

What is called skill-based learning outcomes by Kraiger et al. (1993), includes the development of technical or motor skills. Dettmer (2005) describes this as sensimotor skills, implicating the inclusion of senses and movement. This term is preferable because it allows a clearer distinction to cognitive skills, which are part of cognitive learning outcomes. High sensimotor skills lead to faster, fluid performance with high accuracy and fewer mistakes, and processes are carried out automatically without conscious awareness. If skills get further developed, they allow adaption and origination, so they can be modified to unique situations as needed.

3.4.3 Affective learning outcomes

Affective learning outcomes are differentiated into change of attitude and motivation.

Attitudinal change. Attitudinal outcomes describe the change of attitudes. They influence behaviour, depending on how strong an attitude is. Outcomes include creative individualism, inner growth, self-awareness and change in values. Attitudes are, according to Gagné (1984), composed of both cognitive and emotional components.

Motivation. Motivational outcomes describe motivational state changes. Motivation influences the willingness to perform activities, and the persistence with which they are performed. This includes the motivation for learning. Motivational state changes can for instance result from motivational dispositions, self-efficacy (the perception of ones own abilities), and goal setting.

3.4.4 Social learning outcomes

Social learning outcomes describe effects on the abilities in sociocultural interactions, such as participation as a member of a group or team, communication in a simple or more advanced form, such as negotiation, collaboration to achieve a common goal, or initiation of group efforts. (Dettmer, 2005; Baker and Mayer, 1999; Wouters et al., 2009) Besides individual social development, communicative skills are particularly important in contexts, in which individual performance alone is insufficient to overcome problems. (Wouters et al., 2009)

CHAPTER 4

State-of-the-art in the evaluation of serious games

Are serious games worth the effort?

Research on the effectiveness of serious games has reportedly long been based on opinions about their potential, the quality of research is described as poor, and there has been a lack of empirical evidence on their effectiveness. (Hays, 2005; O'Neil et al., 2005; Bellotti et al., 2013; Mayer, 2012)

The first question that serious games research needs to answer is:

Do serious games "work"?

This inquiry includes the task to define what it means for serious games to "work". What are the essential criteria to decide it? Or in other words, to find appropriate methods for the evaluation of serious game, it needs to be unambiguous what the aim of the evaluation is.

> The overall goal of evaluation is to prove the game's effectiveness and suitability with respect to its designated purpose and application context. The purpose is thereby always in the center of investigation. (Emmerich and Bockholt, 2016)

The second task is to find sound methodologies to find out if these criteria are met. How should be decided which evaluation methods are appropriate, and how should they be applied?

> The primary purpose of an assessment is to collect information that will enable the assessor to make inferences about students' competency states – what they know and can do, and to what degree. (Shute, 2010)

The crucial step is to be able to differentiate which elements of serious games are essential for successful learning outcomes. (Linehan et al., 2011)

4.1 Evaluation results

This section presents an overview about the results reported in literature on the evaluation of serious games.

There is a number of meta-studies on the evaluation of the effect of serious games, and four of them are fairly recent and thorough in that they try to cover a wide range of games, including entertainment games, and are not focusing on a specific learning outcome but include studies with a variety of outcomes. Plus the four meta-reviews transparently describe the methods they use to decide which studies to include in the review, and their inclusion criteria avoid the inclusion of papers with low quality or obviously flawed methodology.

A deep look onto these four meta-reviews serves as a starting point for analyzing the state of the art in the field of evaluation of effects of serious games. Their collection of data can serve as an indicator to problems in the research field.

In the following, the meta-reviews will be introduced to explain their purpose and scope. Thereafter, the results of all four and some additional meta-studies will be presented categorized into several themes.

4.1.1 Meta-reviews

In their review, Connolly et al. (2012) found 129 papers which include empirical evidence related to the impacts and outcomes of playing games, were published from January 2004 to February 2009 and include participants over the age of 14. The studies were categorized in three dimensions with several sub-dimensions:

- *Categorization of games*: digital/non-digital; primary purpose of the game; game genre; subject discipline; platform of delivery
- *Categorization of effects of games*: behavioural and learning outcomes and impacts (knowledge acquisition/content understanding, perceptual and cognitive skills, motor skills, physiological outcomes, affective and motivational outcomes, behaviour change, soft skills and social outcomes); intended or unintended; generic or specific impact
- *Methodological dimensions*: study design; sampling; sample; between-group comparisons; data collection; data analysis; results and conclusions

A quality score based on several criteria was assigned to each paper, and used as weight in the influence on final significance tests.

In 2016, the same research group conducted another meta-review, this time considering papers that were published from March 2009 to February 2014. They identified 512 papers that met the inclusion criteria, compared to 129 in the previous review, showing an increased interest in the evaluation of impacts of games. 143 of the 512 had a high quality rating, and only those were considered in the review.

The coding of the reviewed papers in this meta-study was generally the same as in the previous one by Connolly et al. (2012). The categories of serious games and games for learning were collapsed into one category, as the terms are somewhat ambiguous and often used synonymously in literature.

In his 2009 review, Ke (2009) mentions the wide variety in games research and a missing common link between them, suggesting a rigorous qualitative meta-analysis to show a clearer profile of computer games and meta conjectures or recurring themes in learning effects of computer games. In his meta-study, Ke primarily addresses two questions:

1. What is the cumulative qualitative and quantitative evidence for using computer games for learning, and
2. What are the factors, if any, that weigh in an effective application of instructional gaming? (Ke, 2009)

The meta-analysis uses a grounded theory approach to synthesize theories, methods and findings of qualitative and quantitative inquiries of computer-based instruction games.

In data analysis, low quality paper were excluded, which encompassed quantitative studies not explaining their methodological design features, and qualitative studies if they were biased or failing to provide descriptions of learning context and outcomes. 89 papers were ultimately included in the analysis.

Wouters et al. (2013) present a meta-analysis specifically focused on affective outcomes of serious games.

According to Wouters, previous reviews demonstrated ambiguous results, and in scientific tests serious games do not show to be more effective than other instructional methods.

Following Mayer and Johnson (2010), Wouters describes three categories of games research:

- value-added approach: which features in games foster learning?
- cognitive consequences approach: what do people learn from serious games?
- media comparison approach: do people learn better from serious games than from conventional media?

The meta-analysis is focused on a media comparison approach. In terms of learning dimension, the focus lies on the cognitive dimension, including knowledge and cognitive skills.

Wouters observes that games should theoretically work well for supporting cognitive processes which are crucial for learning. But in most studies, learning is measured immediately after the intervention, which does not reflect sustainable learning. In terms of motivation serious games show high potential in supporting intrinsic motivation.

Deriving from these observations, the study questions if compared to conventional instruction, serious games yield higher learning gains, better retention, and more motivation.

The hypothesized situational and contextual moderator variables were learning arrangement of comparison group, combination with other instructional methods, number of training sessions, group size, instructional domain, age, level of realism, narrative, publication source, randomization and experimental design.

38 studies were identified to meet the inclusion criteria. They either tested on knowledge or cognitive skills and measured motivation in any way (motivation, interest, engagement, attitude towards the topic).

4.1.2 Evaluation outcomes

According to all four meta-reviews, research on positive impact of games for learning, instructional games, or serious games, shows conflicting empirical evidence. Interestingly, in a meta-review about studies which include learning theory principles, Wu et al. (2012) found overwhelmingly positive results.

The learning dimensions of researched serious games were not evenly distributed. Connolly and Boyle provide data in their meta-reviews on the learning dimension of each study (Figure 4.1).

In more detail, the positive impact can be evaluated according to a number of variables, to make a better sense of the overall conflicting results. First, outcomes are presented according to the impact dimension, i.e. according to what people ought to learn from the game. Then, findings from the reviews are presented according to a range of other variables that show interesting results according to the reviews.

Findings by outcome dimension

Learning dimensions are not only unevenly distributed in researched serious games, but they also show diverging results.

Knowledge acquisition/Content understanding. While Connolly et al. (2012) show mixed results, Boyle et al. (2016) demonstrate a tendency to positive results on studies on knowledge acquisition through playing games, for both explicitly designed games for learning and for entertainment games. Wouters et al. (2013) underline this, as

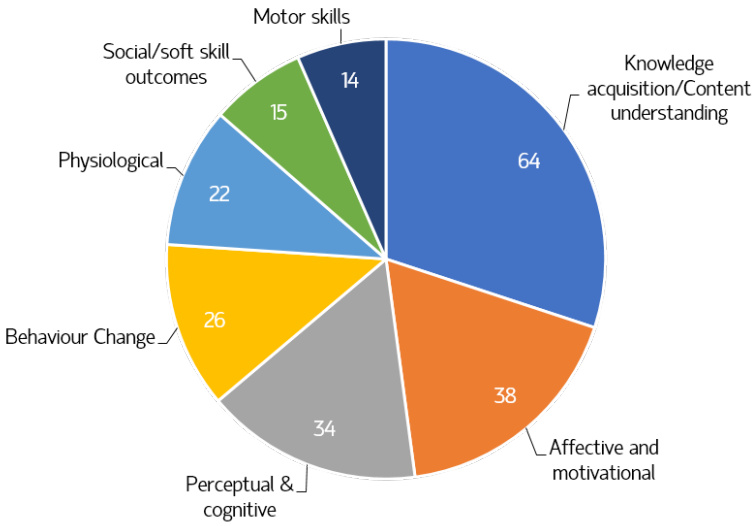

Figure 4.1: Learning dimensions in research studies on serious games. Data from Connolly et al. (2012) and Boyle et al. (2016), only including high quality studies.

the effect size over all analyzed studies shows a significant positive impact of games on knowledge acquisition compared to other instructional methods. Also, while Boyle et al. (2016) report mixed results on long term effects, Wouters et al. (2013) show a superior retention level of serious games over conventional instruction.

Perceptual and cognitive skills. The broad field of perceptual and cognitive skills, including for instance visual perception abilities, spatial thinking and reasoning, but also higher order scientific reasoning skills, shows varying results depending on the specific skill. Connolly et al. (2012) and Boyle et al. (2016) show a strong positive effect of entertainment games, specifically action games, on perceptual skills, and results for entertainment games are better in improving attentional and visual perception skills than in games that were specifically designed to improve them. The results for effect on memory and problem solving skills show mixed results. Real world decision-making and higher order scientific reasoning skills show evidence for positive results, as well as generally higher level cognitive skills, even if showing evidence on it is a difficult task.

Overall, however, there is a stronger evidence base for the effectiveness of serious games on low-level perceptual skills than on higher level cognitive skills. (Boyle et al., 2011)

Motor skills. Studies measuring the effect on motor skills tend to use experimental study designs, and results are mixed in the meta-review by Connolly et al. (2012), and largely positive in Boyle et al. (2016).

Behaviour change. Effects on behaviour change, where transference between virtual and real world took place, has been found, but results are mixed. (Connolly et al., 2012; Boyle et al., 2016) Games that are specifically designed for behaviour change are found to be less successful in their effect than specific entertainment games. (Connolly et al., 2012; Boyle et al., 2016)

Social behaviour. Studies show a positive effect of playing games on social behaviour, and find that gamers are highly social individuals. (Connolly et al., 2012; Boyle et al., 2016)

Affective and motivational outcomes. Studies show mixed results for affective and motivational outcomes. Some studies support flow theory and more intense feeling of presence, better situation awareness and faster performance for game players (Boyle et al., 2016), or find highly positive effects regarding affective learning outcomes, specifically on self-efficacy, attitude and persistence, and facilitate motivation across learner groups and situations. (Ke, 2009) Autonomy and competence were found to significantly account for player motivation and enjoyment. (Ryan et al., 2006) But others do not find psychological needs from self-determination theory to predict motivation for playing games (Boyle et al., 2016), or even do not show a significant advantage of serious games over conventional instructional methods in learner motivation whatsoever (Wouters et al., 2013). But Wouters et al. (2013) also find that serious games are more motivating when they are not combined with other instruction methods, which could suggest that every setting in which there is an interference with basic motivational principles of games, such as the principles of self-determination theory, would cause motivational effects of games to cease, meaning that lab settings to evaluate their impact could decrease motivation.

Physiological outcomes. The results are again mixed, but with a trend to positive effects of exergames in some populations such as elders, inactive adults, stroke patients or people with depression. (Connolly et al., 2012; Boyle et al., 2016)

4.1.3 Mediating variables

There are studies showing positive effects in each learning dimension, but results were also mixed in each one, and it is not yet clear what determines success or failure. However, Ke (2009) and Wouters et al. (2013) found a number of mitigating variables with influence on the effect of games for learning:

- **Researchers' theoretical perspective**. Results are dependent on the perspective taken by researchers, and therefore on their choice of inclusion criteria and outcome variables considered. (Plass et al., 2015; Wouters et al., 2013)
- **Research method**. Qualitative studies tend to report positive effects. (Ke, 2009)

- **Gender**. According to Plass et al. (2015), research shows no significant gender differences in learning or motivational outcomes. But, qualitative studies found gender differences in the effect of serious games, as opposed to quantitative studies. This suggests that differences occur in learning processes instead of results. (Ke, 2009)
- **Socio-economic status**. Learners with lower socio-economic status enjoy games most, but have more difficulties in extracting the target knowledge. (Ke, 2009)
- **Instructional Support**. Games for learning show better effects when they are supported by additional instructional means, e.g. teacher facilitation or embedded feedback and multimodal information presentation. (Ke, 2009; Wouters et al., 2013)
- **Session number**. Only in multiple sessions games show to be more effective than conventional instructional methods. (Wouters et al., 2013)
- **Groups**. Learning gains are higher if games are played in groups. (Wouters et al., 2013)
- **Narrative**. Games without narrative show better results for learning than those with narrative. (Wouters et al., 2013)
- **Study design**. Studies that use randomization show no significant effect in increasing learning gains in serious games. (Wouters et al., 2013)
- **Reflection**. Reflection processes during game play are important to yield better effects of serious games. (Wouters et al., 2013)
- **Immersion**. Jennett et al. (2008) find that immersion is an important element in games, and high immersion leads to longer time required to get back into other tasks after playing a game.

4.2 Methodologies

The methods used to evaluate the impact of games are very fragmented, and there is a variety of different variables regarding context, players and game under evaluation. (Dempsey et al., 1994; Ke, 2009)

The meta-reviews by Connolly et al. (2012) and Boyle et al. (2016) captured the study designs used. Figure 4.2 shows the distribution of study designs of both meta-reviews combined.

Ke (2009) registered 69% quantitative methods (experimental, quasi-experimental, correlational-causal, or descriptive) in studies that evaluated the effectiveness of serious games, while 15% used qualitative ethnography and another 15% used mixed-method designs.

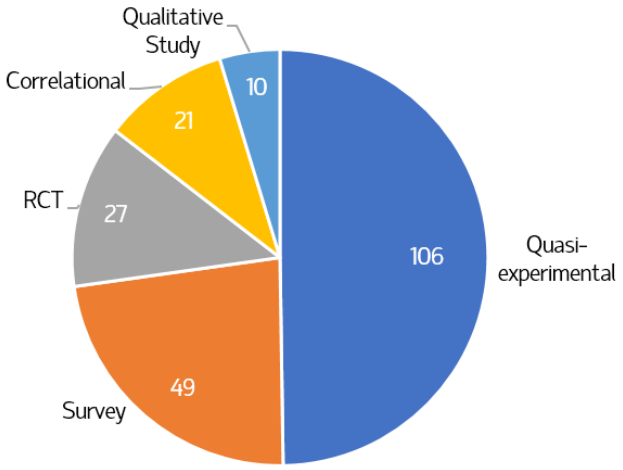

Figure 4.2: Study design used in research studies on serious games. Data from Connolly et al. (2012) and Boyle et al. (2016), only including high quality studies.

According to Connolly et al. (2009), randomized controlled trial (RCT) is the method of choice in game impact evaluation and should therefore be used more, but he also recognizes the difficulty of finding a suitable non game-based control condition. Also, there is a lack of qualitative studies. Especially when it comes to affective and motivational outcomes, qualitative designs would be well suited because of the experiential nature of motives and emotions. (Connolly et al., 2012; Boyle et al., 2016) Moreover, according to Michelsen et al. (2002), "[...] not every intervention strategy lends itself to an experimental evaluation."

Cited by Ke (2009), Dempsey et al. (1994) see a movement to look at incidental learning using more process-oriented inquiry. This statement suggests that it might make more sense not to simply measure outcomes, but to investigate the process of learning, and to analyze a game for its theoretical learning dynamics instead of the impact outcome alone.

Wouters et al. (2013) expressed concerns about usual methods for measurement of motivation. In 30 of 31 pairwise comparisons in his meta-analysis the method used to assess motivation was a survey or questionnaire, conducted after gameplay. The problem is that motivation may cease quickly after gameplay, and self-assessment of affective states may be difficult. Observation during play, and methods to measure physical processes like eye tracking and skin conduction during gameplay could be better suited to measure motivation. (Wouters et al., 2013)

4.2.1 Study designs

Randomized controlled trial (RCT). The randomized controlled trial is the "gold standard" and considered the most powerful experimental design in clinical research.

Participants are randomly assigned to an experimental (game) group, or to a control (non-game) group. Their performance (knowledge, skill, behaviour) is measured before and after the game intervention. The big advantage of this research design is that differences in outcome can be directly attributed to the intervention, because it reduces allocation bias by distributing baseline characteristics equally between groups. (Sullivan, 2011; Hauge et al., 2013)

Quasi-experimental design. A quasi-experimental design refers to a research design which is lacking one of the main features of an RCT, e.g. if participants are not randomly assigned to a group, or the control group is missing entirely. Although RCTs are considered more powerful, they cannot always be applied due to ethical or practical reasons, and therefore quasi-experimental designs are often used instead. (Hauge et al., 2013)

Correlational design. A correlational design measures two or more variables to report a correlation coefficient for them. Reported correlation between variables does not allow to conclude causality, but can help identify variables that could be interesting to examine further.

Quantitative-descriptive design. Descriptive study design uses the collection of a (usually) large amount of data and appropriate processing of the data to describe a current state. Descriptive research can serve as basis for a correlational research. It can also be qualitative as opposed to quantitative.

Qualitative design. Qualitative methods do not examine data for statistically significant effects, but operate in a more interpretative domain with single data sets. They can incorporate interviews, focus groups, or psychological observations. (Bellotti et al., 2013; Hauge et al., 2013) While they do not allow for direct conclusions inferred from statistical tests on gathered data, they can help for a deeper understanding of observed or tested phenomenons. Therefore, they are often used complementary to quantitative methods.

Pre-post-test. Measurements in research studies can occur before, during, and after interventions. Measuring before and after tests allows for comparison of results and therefore conclusions about the direct effect of the intervention. Figure 4.3 shows a typical pre-test, post-test evaluation process, in this case using an RCT.

A long term follow-up post-test can be useful to eliminate short term effects in the evaluation and get more reliable results.

4.2.2 Data gathering

Besides deciding which research design to use, an appropriate method to gather data needs to be chosen.

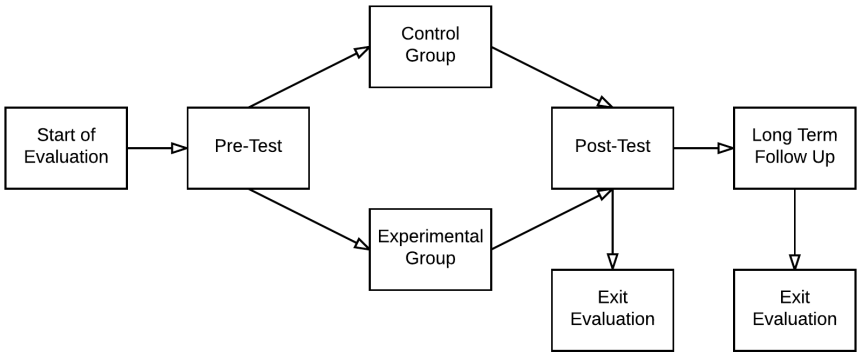

Figure 4.3: Typical pre-test, post-test design, adapted from (Connolly et al., 2008)

Of course, depending on the learning outcome, the data gathered will be different in each study. Hauge et al. (2013) presented a comprehensive overview (table 4.1) about what can be measured when and with which methods.

4.3 Application examples

After the overview about research design and methods for gathering data in the previous section, this section presents several case studies, some using standard methodologies taken from other research fields, and some using methods specifically tailored to the research of serious games.

Quasi-experimental study with knowledge test

Soekarjo and van Oostendorp (2015) conducted a study about the persuasive potential of a game changing peoples' attitude related to energy use, compared to a slideshow presenting the same information.

The game in question was EnerCities, which was designed to foster sustainable energy use in its players, by letting them build a city which requires them to balance environmental and economic needs. They assumed the game to be well-designed, because it won national and international awards. Referring to a previous study showing that the game was effective, they criticized the methods, as no control condition was used.

The study includes a control condition with highly similar information, presented in a slideshow. The authors admit that making the information 100% the same is difficult, but they assure to have taken considerable effort to make it as similar as possible. Additionally, they criticize about the previous study that it only measured micro-level effects in energy saving attitudes, while the game is mostly directed towards macro-level effects. Therefore, they introduce a measure for macro-level attitudes as well.

How	What?	Pre-Game	In Game	Post-Game
Self-reported				
Qual.	Personality, player, experiences, context, etc.	Interviews, focus group, logbook	Logbook, interview or small assignments as part of the game	Interviews focus group, after-action review
Quant.	Social/demographic, opinions, motivations, attitudes, engagement, game-quality learning, power, influence, reputation, network, centrality, learning satisfaction, etc.	Survey, questionnaire, individual or expert panel	In-game questionnaires	Survey, questionnaire, individual or expert panel
Tested				
Qual.	Behaviour, skills, etc.	E.g. actor role-play, case-analysis, assessment, mental models	Game-based behavioural assessment	Game-based behavioural assessment
Quant.	Values, knowledge, attitudes, skills, personality, power	Psychometric, sociometric tests: e.g. personality, leadership, team roles, IQ	Game-based behavioural performance analysis	Game-based behavioural performance analysis
Observed				
Qual.	Behavioural performance of student, professionals, player and/or facilitator, others; decisions, strategies, policies, emotions, conflicts, etc.	Participatory observation, ethnographic methods	Video, audio, personal observations, ethnography, maps, text, figures, drawings, pictures, etc.	Participatory observation, ethnographic methods
Quant.	Biophysical-psychological responses, like stress (heart rate, perspiration)	Participant observation, network analysis, biophsical-psychological observation	In-game tracking and logging, network analysis, data mining, biometric observations	In-game log file analysis, network analysis

Table 4.1: What can be measured when, and with which method in serious games? Table from Hauge et al. (2013).

46 undergraduate students participated and were randomly assigned to either the experimental group, playing the game for 20 minutes, or the control group, receiving a slideshow which they could navigate freely, and which took an average of 20 minutes to complete.

Participants completed an online pretest and post-test questionnaire. The questionnaires measured participants' attitude towards sustainability-related topics on a micro-level and on a macro-level, and a knowledge measure on the sustainability topic in form of open questions. The pretest was completed at home, a week before post-test measurements. After playing the game or reading the presentation, participants' had to conclude a short distractor task before completing the post-test.

Results. No significant differences were measured for changes in micro-level and macro-level attitudes between groups. Also, no significant difference in increase of knowledge was found between groups. The credibility of information presented in the slideshow was judged significantly higher than in the game group.

Quasi-experimental study with knowledge test and observation

Huizenga et al. (2009) conducted a larger scale study on a location-based mobile game, developed to teach 12 to 16 year old students knowledge about medieval Amsterdam. The study assesses the effect of the game on pupils' knowledge about medieval Amsterdam and their motivation for the study of history in general and the middle ages in particular.

The game is called Frequency 1550, and puts groups of players in the role of either a beggar or a merchant in medieval Amsterdam, with the goal to attain citizenship. Each group of pupils is split into two subgroups. One is the city team, moving around in the city to conduct small, location-based media tasks to explore, map and gain knowledge of an area of medieval Amsterdam. The second team is the headquarter team, operating on a computer in the main building. The headquarter team follows the city teams' route and guides them through learning tasks using various sources of information. The two teams need to cooperate to fulfill assignments of an area, which concludes with a final multiple-choice or open-ended question. At the end of the day, the groups are invited to present their results to the other groups.

The aim of the study was to find out if knowledge gains and motivation of pupils playing the game differed from a group of pupils getting traditional instructions on the topic, and if previous history abilities and education level will influence these factors.

A quasi-experimental study was conducted with pupils from 10 classes playing Frequency 1550 for a day as experimental group, and pupils from 10 other classes getting a regular project-based lesson series forming the control group. The regular lessons were specifically designed for the purpose, making the content of both conditions similar.

Each team was accompanied by a guide, who recorded observations about the teams on things like specific game activities and how actively pupils were involved in the activities,

and made notes about pupils' engagement. A questionnaire similar to an available questionnaire on motivation for math was used to measure motivation, once directly before playing the game, and once one week after the play-session. Three multiple-choice questions and two open-ended questions were used to measure historical knowledge of medieval Amsterdam.

Transcriptions of oral reports, and observation notes from researchers and guides were analyzed with a grounded-theory approach.

Results Pupils in the headquarter teams were found to be slightly more engaged, according to the quantitative observation data from guides. Surprisingly, regarding motivation for the subject of history and specifically the middle ages, no significant differences were found between the experimental and the control group. But pupils who played the game generally scored higher on knowledge tests. Pupils from a higher educational level benefited more from playing the game than pupils from a lower educational level, but surprisingly, students from a lower educational level getting the lesson-based instruction scored higher than students with higher educational level. Additionally, students with low initial history ability benefited most from playing Frequency 1550.

Quasi-experimental study with mixed data gathering methods

Yang (2012) conducted a study in an authentic classroom setting to compare the effectiveness of a serious game to traditional instruction on learning motivation, problem solving skills, and academic achievement. The study is very thorough in the description of its methods and in contrast to most other studies in this field, studies an authentic setting in classroom as opposed to a lab study.

The study design was quasi-experimental, assigning one class to the experimental condition of playing two games, and another class to the control condition of receiving traditional instruction on the topic. The students were 15–16 years old and enrolled in a *Civics and Society* course. Over a period of 23 weeks, the topics of *daily economics* and *global issues* were covered. Both groups received traditional instruction in the first two lecture units per week. In the third lecture unit per week, the experimental group played a game, while the control group again received traditional instruction. The game used for the topic of daily economics was Tycoon City: New York, and for the topic of global issues SimCity Societies. Both games are complex, commercial strategy games.

A *Civics and Society Test*, based on the textbook used in the course, was used to test academic achievement in a pretest (scores from a previous semester) and post-test. The *New Test of Problem Solving* was used as pretest, mid-test and post-test for problem solving abilities, consisting of subscales for finding causes, finding solutions and avoiding problems. The *Motivated Strategies Learning Questionnaire: Motivation Scales* was used as pretest and post-test to assess students' motivation, which was designed to evaluate learners' motivational orientations.

Results The study measured no improvement of performance in problem solving between pretest and mid-test for the experimental group, but it found improved performance in problem solving in the post-test. No performance increase was found for the group getting traditional instruction. Improved motivation was found over the course of the full semester for the game group, and was significantly higher than for the traditional instruction group. Also, qualitative observation indicated great curiosity in the game group. The emphasis of instructors on the necessity to learn basic concepts to succeeding in the games' challenges lead to students believing in the value of learning *Civics and Society*. Additionally, by succeeding in gaming tasks, students' self-efficacy improved. Lastly, for academic achievement, no significant difference between game group and traditional instruction group was found. This was accounted to the nature of the *Civics and Society Test*, which had most questions about recognizing and restating information, and almost none about applying knowledge or organizing principles.

RCT with A/B test

Shute et al. (2015) examine the effect of the popular commercial video game Portal 2 on problem solving and spatial skills and on persistence.

The study employs an RCT, but unusually does not use a group with no treatment or with another instructional medium as control condition. Instead, another game is used as control condition, namely Lumosity, which is a serious game specifically designed to improve the cognitive skills measured in the study.

77 participants (university students) were randomly assigned to either the Portal 2 or the Lumosity group. Each one played the respective game for a total of 8 hours in three sessions over a week in a laboratory.

Before the first playing sessions, the subjects completed an online set of pre-tests. In a separate fourth session, they completed an online set of post-tests.

To measure problem solving skills, *Raven's progressive matrices* and a suite of insight problems and creative problem solving tasks were used. Spatial cognition, specifically figural, vista, and environmental spatial skills, were measured with mental rotation tasks, *Spatial Orientation Test*, and a self-developed *Virtual Spatial Navigation Assessment*. Persistence was measured using a performance-based measure (picture comparison task) and a self-report survey.

Additionally to the pre- and post-test measures, for both conditions in-game measures were collected to directly assess gameplay performance. In case of Portal 2, this was the total number of levels completed, average portals shot and average time to complete levels. For Lumosity the performance measure was already implemented in the game as the so called "brain power index" (BPI), consisting of the total score of all activities and constructs in the game. After completion of gameplay, participants were asked if they enjoyed playing the game.

Results For all three constructs, PORTAL 2 performed better than LUMOSITY. But, neither of the two games significantly improved problem solving skills. There is strong evidence that PORTAL 2 causes improvements in small- and large-scale spatial skills, suggesting the possibility that video games can improve environmental spatial skills and that improvements can transfer to actual real world navigation. PORTAL 2 lead to a higher persistence measure after gameplay, consistent with findings that exerting hight effort in one task will lead to high effort in a subsequent task. This finding is relevant in that it suggests, that confronting players with difficult challenges in video games can result in increasing their effort in real world challenges. Another important finding was that in-game performance in PROTAL 2 predicts scores in several measurements after controlling for pre-test knowledge, which suggests that increasing performance in games can yield improvements in skills applicable in real world settings.

Experimental design with in-game measurements only

In a study conducted by Orvis et al. (2008), the main research question was if and how task difficulty affects learning outcomes in serious games. The claim about the importance of the optimal difficulty is based on the concepts of zone of proximal development and flow, and will be discussed in more detail in a later chapter.

Games have a number of concepts to deal with the problem to present players with appropriate challenges. One of the concepts is to let player choose difficulty themselves, another one is to dynamically adapt difficulty according to players' performance.

The study has no experimental and external comparison group, but several experimental groups with different models of difficulty adaption, to test which one is most beneficial for learning. The game used is a military training game, in which players shoot on targets in different size and distance. There are eleven difficulty levels in total. In the first variant, the difficulty stays the same over the entire period of the game. In the second variant, the difficulty increases linearly during the course of the game. The third variant is a dynamic adaption according to the players' performance, starting on a medium level. The fourth variant is identical to the third, but starting on a lower level. The goal for the players was to hit as many targets as possible.

The first level played had the same difficulty level for every participant, to act as a pre-test. The last level again had the same difficulty level for all, and acted as post-test. The difficulty level in the levels in between was determined by the variant in action.

Before the start of the intervention, previous experience of participants' playing 3D action games was assessed using a survey.

Noteworthy about the methodology in this study is, that it exclusively relates on in-game measures to evaluate the effect on a learning outcome, varying one variable in the game.

Results All difficulty level conditions showed improvement in performance and motivation, but no single approach was superior to the others.

For inexperienced players, the post-test performance was least beneficial in the forced difficulty adjustment, suggesting that prior experience is important for finding the optimal difficulty adjustment strategy.

Experienced gamers also reported higher initial task self-efficacy, and they set higher personal performance goals. Therefore, it is important to account for lower self-efficacy for non experienced players by giving them additional instructional aid at the beginning. At the end of the treatment, the measure of self-efficacy was not significantly different between experienced and non-experienced players anymore.

Ethnographic research

A very interesting work in the regards of methodology in serious game evaluation was conducted by Ducheneaut and Moore (2005). They conducted a study on the *MMORPG* (Massive Multiplayer Online Role Playing Game) Everquest Online Adventures (or short EQOA), and on its contribution to social learning.

The research method they used was ethnographic field research. This method is about *participant observation*, whereby researchers immerse themselves in others' worlds, becoming unobtrusive participants in a social construct and therefore being able to make valuable observations about the lives in these social communities. They transfer this approach to a virtual world, treating this world as the ethnographic reality.

Data was gathered by videotaping the 100 hours of gameplay spread over a three-month period. The video data was analyzed using a combination of conversation analysis and open-ended coding. Data analysis led to a deep understanding of the natural player practices in the virtual world, identifying primary types of social interactions and observing how players acquire the necessary skills to participate in the community.

Results Playing MMORPGs is about learning and participating in the shared practices of a game community, and as such "provide opportunities for learning social skills such as how to meet people, how to manage a small group, how to coordinate and cooperate with people, and how to participate in sociable interaction with them." (Ducheneaut and Moore, 2005)

The researchers argue that practice in these virtual worlds can actually lead to the important step of transferring learned skills from the game to real life by growing players' self confidence in interacting with and organizing others.

Structural Assessment compared to traditional knowledge tests

Wouters et al. (2011) published a study in which they evaluate how a game can foster building of structural knowledge about primary triage, a process about prioritizing patients based on the severity of their condition in situations when not all can be treated immediately.

The importance of knowledge organization is emphasized by current thinking in cognitive science, and the concept of structural assessment is the identification of knowledge structures which are considered an important part of learning complex cognitive skills.

Several different techniques have been suggested for the assessment of structural knowledge, and Wouters et al. (2011) implement the *Pathfinder* technique in their study. The aim of the study is to explore if structural assessment with *Pathfinder* can be used to assess learning of complex skills in the serious game Code Red: Triage. The game puts players in a 3D environment, in which a bomb explosion occurred and players impersonating a medical officer have to perform a primary triage. Player invoke information about the victims when they approach them, and need to classify the victims accordingly.

The approach for structural assessment in the study comprises of three steps. The first step is knowledge elicitation and is realized by methods such as card sorting, ordered recall or numerically rating the degree of relatedness between concepts (based on Goldsmith et al., 1991). From the elicited knowledge, in the second step a scaling procedure is applied for a better representation of the underlying knowledge structures. This is achieved in the *Pathfinder* technique using a graph-theoretic distance technique to represent the elicited knowledge in a graphical network structure. In the third step, the knowledge evaluation, a referent knowledge structure is built by a domain expert or from analysis of instructional material. The knowledge structure of learners is then compared to the referent knowledge structure, and the similarity between them is assumed to represent learner's understanding of a domain. The comparison is not solely based on similarity scores, but can also look for specific core structures in the knowledge structure.

19 people participated in the quasi-experimental study, which included a pre- and post-test, but no control condition. The concepts for the knowledge structures were derived by the authors from instructional material on the domain. Three domain experts created the referent knowledge structure, and rated 78 concept pairs on their level of connection. Each participant in the study rated the same concept pairs, once before and once after playing Code Red: Triage. Each participant also completed a "traditional" knowledge test before and after the game intervention. The knowledge test consisted of 10 multiple choice items on declarative and procedural knowledge. In the evaluation, a group of advanced learners and a group of novices on the concept were distinguished to compare their results.

Results Both novices and advanced learners improved performance in the knowledge tests from pre- to post-test. For novices, the improvement can only be seen in procedural items, whereas advanced learners only improved in declarative items. Before playing the game, advanced learners performed significantly better than novices, but after playing there was no significant difference in performance between advanced learners and novices.

In structural assessment, the similarity of knowledge structures of novices to the referent knowledge structure was significantly higher after playing the game than before playing the game. For advanced learners, there was no significant effect. Before the game the

similarity of knowledge structures to the referent knowledge structure was higher for advanced learners than for novices, after playing the game the difference disappeared.

Stealth Assessment and Evidence Centered Design

Shute (2010) follows a distinct approach in her study on the game Quest Atlantis: Taiga Park, using an Evidence Centered Design model and stealth assessment.

The game Quest Atlantis: Taiga Park is a 3D role-playing game, set in a beautiful park with a river running through it. The fish population in the river is declining, and players need to figure out how to solve this problem, considering the interests of all involved stakeholders. Players should therefore develop system thinking skills.

For the evaluation of the game, the *Evidence-Centered Design* (ECD) approach is used. The underlying claim of it is, that often knowledge, skills, behaviour or other learner attributes cannot be measured directly, and therefore another approach is required. ECD consists of three steps to create a model which can be used to measure desired outcomes. The first step is to identify what knowledge, skills and behaviour should be assessed, which constitutes a *competency model*. The second step is to find the right observations in a situation that can constitute evidence for the knowledge, skills and behaviour in the competency model. This constitutes the *evidence model*. The third and final step is to identify or construct situations and tasks, which elicit evidence from the evidence model, and therefore make them measurable.

The second important concept in the study is the one of stealth assessment. While ECD constitutes a general model for identifying what outcome should be evaluated and which variables need to be measured therefore, stealth assessment is concerned with how these variables are actually measured. Stealth assessment describes assessment which is embedded in the game in an entirely unobtrusive, virtually invisible way. This is important for several aspects, including maintenance of flow and reducing test anxiety.

The study applies the concepts of ECD and stealth assessment in Quest Atlantis: Taiga Park. The intended measurements are an understanding of water quality indicators and their relatedness to activities along the river, and a systemic understanding of the variables involved in the causal loop that lead to reduction of the fish population. This builds the competency model. The evidence model consists of observable actions which are evidence for the competency variables. In the case of Quest Atlantis: Taiga Park this contains outcomes from assigned quests in the game, rules for scoring them and weights on these outcomes depending on their degree of influence on the competency variables. Shute (2010) renames the task model to action model in the case of games. It defines a sequence of actions, things that students can do. In this case, these are essays and causal loop diagrams that students can submit to complete missions in the game.

Results As a result of the study, a complex ECD model was crafted for the game Quest Atlantis: Taiga Park. This model could then be used to assess students performance

in the game, using the developed methods for the automatic assessment of causal loop diagrams and the integration of direct feedback into the game.

4.4 Methodological issues

The assessment of the effectiveness and efficiency of instructional interventions is a very complex task. In the comparably young field of serious games research, the choice of appropriate methods is still a matter of discussion. The field itself is broad with various fields of application and learning dimensions. The previous chapter demonstrated the variety of methods with the aid of case studies.

This section elaborates on the issues in serious games evaluation. It uncovers flaws in the application of methodologies found in literature, and confounding variables in serious games evaluation that are often forgotten. It also renders conceptual problems which are more difficult to address and to fix. And finally, it discusses the core of the evaluation, namely if what is actually being measured is in fact what researchers think is being measured.

4.4.1 Issues in standard methodologies

Randomized controlled trial

Randomized controlled trials (RCTs) are considered the gold standard, but that choice seems to be often unquestioned. The argument is that it is the standard in clinical trials and is also widely used in evaluation of instructions, but if the method actually affords application in the evaluation of serious games is never discussed.

When applying RCTs for the evaluation of educational interventions, some important characteristics of RCTs in clinical trials are lost.

In medical RCTs, participants usually do not know if they are assigned to the control group or to the experimental group. This can be achieved in clinical trials using placebos, but there is no equivalent for this in instruction. Participants know which treatment they get, and in some scenarios, like authentic experiments in a school environment over a longer period of time, participants are very likely to also know which treatment the other group gets. Also, double blindness cannot be ensured, as it is virtually impossible that researchers do not know which students are in which group. (Sullivan, 2011) This could lead to the effect that students act in a way that is desired by researchers, or it may even make students in the control group feel disadvantaged. (Hauge et al., 2013) Consequently, this is not only a big problem for the validity of the results, but also raises ethical problems.

There are also more general concerns with applying methods from clinical research to education studies. Education is a highly complex system, as are serious games. They do not afford the requirements that RCTs and other models in clinical research have. (Norman, 2010) Not only, as mentioned above, is blinding of subjects impossible,

unethical, or both, but variables in these complex systems cannot be controlled as tightly as in clinical studies. (Sullivan, 2011)

The primary advantage of randomization - to control for differences in baseline parameters of participants - may not even be the biggest confounder in educational interventions, but others which RCTs cannot control for. (Sullivan, 2011)

Norman (2010) describes six characteristics of RCTs in clinical research, and argues that these should be present to make an RCT feasible. These characteristics are:

1. The underlying mechanism of what is tested is well understood, because if it is not, interpretation of measured effects are difficult to interpret.
2. The endpoint is easily objectified, meaning that the possible outcomes are well defined and can be easily measured.
3. It should be easy to identify subgroups who benefit most from the treatment.
4. The treatment effect is very small, because then it is most important to rule out baseline differences of participants that could bias the results if no randomization is applied.
5. The effect of the treatment is independent of who is administering it, because if personal factors are important (e.g. in surgery), interpretation is difficult as it is unclear what determined the outcome.
6. The results can potentially have a big impact, because RCTs are expensive to conduct and therefore the cost needs to be justified by impact.

It is striking to read these criteria and think about what this means for the application of RCTs in educational interventions. It leaves a tiny subset of possible scenarios in which RCTs are appropriate. The points will be briefly discussed in their relation to the evaluation of serious games (point 6 is left out as it only concerns the economic value of a method, which should not be of importance here).

1. The underlying mechanisms of serious games and their potential for learning are hypothesized, but not well understood yet. Research on serious games is just in its beginning in making sense and gain understanding of the mechanisms of games for learning.
2. The endpoint in the evaluation of serious games is rarely easily objectified, as objectifying learning outcomes is complex and outcomes are very diverse.
3. Identifying subgroups that have the highest response to an educational intervention is an important undergoing in the research of serious games, as the adaptability of them affords the accommodation of different needs of learners. However, this is, again, a very complex and difficult undertaking.

4. Effect sizes in educational interventions are large, according to literature. (Norman, 2010) This suggests that RCTs are not crucial as baseline differences in participants may be neglectable.

5. Serious games as artifacts are the same, every time they are played. But they are often applied in a context with an instructor involved. Therefore, the instructor may have a bigger influence on the outcome than the instruction itself.

According to Norman (2010), "RCTs have seen a very limited adoption in education, for very good reason".

It is remarkable that these considerations are not taken into account in the evaluation of serious games. There is continuous preaching for the use of RCTs, or experimental studies at least, without even questioning if they are appropriate. (see for example Emmerich and Bockholt, 2016; Hays, 2005) This is not to say that experimental studies have no value at all in the evaluation of serious games, but these arguments need to be considered when deciding for an evaluation methodology.

Self-reported measures and surveys

Self-reported measures, in the form of questionnaires or talk-aloud protocols, are a common method in the field of human-computer interaction, and have been adopted and applied in serious games evaluation, especially to measure motivation and engagement. (Shute et al., 2015) While often useful, it is important to keep its drawbacks in mind. People tend to say what they think is favorable, they have different understandings of questions, and they often do not have a sufficient understanding of their own skills, knowledge, or behaviour. (Jennett et al., 2008; Shute et al., 2015; Loh et al., 2015) Questionnaires additionally are very sensitive to weak methodological design. (Connolly et al., 2009) Therefore, if questionnaires are used, it is crucial to use validated questionnaires and to have experienced researchers, who can adopt and apply them accordingly. This is often not fulfilled in serious games research. (Mayer et al., 2014)

4.4.2 Assessment changes play

The application of standard methodologies to the evaluation of serious games not only suffers from issues regarding the methods themselves as described before, but also raises additional problems closely related to the act of playing a game. Specifically, some standard methodologies may be questioned for being suitable, because they interfere with playing and thus cannot produce reliable results.

Teaching to the test

"Teaching to the test" describes the close connection between learning and assessment, making students feel like the purpose of learning is doing well on a test. This can potentially decrease students' interest in learning. (Bellotti et al., 2013) In the case of

serious games, it may also undermine the nature of playing. As discussed in section section 2.1, the lack of real life consequences in games is even part of some definitions of games. If students get the feeling that they play a game for the purpose of passing a test, according to these definitions, it ceases to be a game. But even if it is not considered a part of the definition of games, direct real life consequences at least change the act of playing.

Participation pressure

When games are used in formal learning environments, either as pilot study or as part of a curriculum, students do not participate voluntarily, because playing them is mandatory or students feel pressured into participating. (Hauge et al., 2013) It would be well worth to investigate if missing voluntariness influences students' motivation to play serious games or influence learning outcomes in experimental studies.

Intervention length

Games are played extensively, for long hours and over a long period of time. Many studies in literature use very short playing session lengths and often only a single or just a few sessions. (Loh et al., 2015) This is problematic as it is very different from actual habits when playing games, and therefore could leave the potential of games unused in studies and the conclusions invalid.

Obtrusive assessment

Playing games is a highly engaging activity, and engagement may be one of the determining factors of the effectiveness of serious games for learning. If assessment disrupts engagement, it can distort evaluation outcomes. This demands for unobtrusive ways to assess learning in serious games. (Shute and Ke, 2012)

4.4.3 Confounding variables in serious games evaluation

In the evaluation of instructional interventions, it is crucial to account for influencing factors on the effectiveness of the intervention, other than the intervention itself. This is generally true for any evaluation of instructional methods, and most studies take confounding variables into account. But there are some additional variables for serious games that are often overlooked.

Prior gaming experience

The prior experience of playing games can be influential to the effectiveness of learning with serious games in both positive and negative ways. (De Freitas et al., 2010; Van der Spek, 2012) Experienced players of a specific game genre know typical genre conventions and are used to the controls of the game. This means that they have to expend less cognitive resources for playing than people who are not experienced in the type of a

game. New players may even be entirely invested in learning how to control the game and understand its functionality, so that they do not have any free cognitive resources for the actual instructional objective. On the other hand, experienced players may be prejudiced by game conventions, and hence may not be open to an unusual exposure of mechanics. Also their motivation might be negatively influenced if they play a serious game that is technologically or gameplay-wise inferior to games they are used to. In other words, experienced players do not forgive mistakes in game design as easily.

Learner preferences

Additionally to the problem that experienced game players have higher expectations of games, they may also have strong preferences for a specific type or genre of game. Some might be used to AAA games (pronounced "triple A games", an informal classification for video games with high development budgets, analogous to the "blockbuster" classification in film), and for that reason be reluctant to simple or casual games. On the other hand, non experienced game players may have no experience in playing for exactly the reason that they do not like to play video games. Both of these factors can influence the motivation of playing, and accordingly influence learning outcomes. (Shute and Ke, 2012)

4.4.4 Conceptual challenges for assessment in games

Besides the issue of picking the right methodology and applying it correctly, there are also more conceptual issues in evaluating serious games.

Complexity of measuring high level cognition

Complex cognitive skills, like problem solving, communication, or system thinking are difficult to measure directly, because they consist of a combination of low-level skills and knowledge organization. (Bellotti et al., 2013; Chen and Michael, 2005; Shute and Ke, 2012) Accordingly, novel approaches are required to find ways to measure them.

Methods for traditional learning environments

Common assessment methods are usually focused on traditional learning environments, such as classroom teaching. Therefore, these methods may not be appropriate for other settings. (Arnab et al., 2012)

Long term measurements

Learning outcomes are often measured directly after playing a serious game. (Wouters et al., 2013) This is less than optimal, because it does not focus on sustainable learning. Therefore, long term assessment would be beneficial. (Emmerich and Bockholt, 2016) The challenge in long-term measurements of learning is to consider influencing factors. It is already difficult in a short-term experiment to control for external factors which might influence learning besides the learning intervention itself, and it gets increasingly hard

the longer the period between the learning intervention and point in time of measurement is. (Loh et al., 2015)

Unintentional learning

In the evaluation process, the first step is usually to define what it is that should be measured. Games are very complex learning environments, and it is likely that the instructional outcomes exceed the defined instructional objectives. (Shute and Ke, 2012) Consequently, the potential of games for learning can easily be underestimated by neglecting "incidental" learning.

4.4.5 Measuring the right thing

Even if the choice about methodology is made, and the evaluation process is technically done right, the question remains if one is measuring the "right thing" (i.e. what was intended). (Bellotti et al., 2013; Wouters et al., 2009; Emmerich and Bockholt, 2016) This problem can be viewed from different perspectives, which will be elaborated in the following.

Learning transfer

Learning outcomes of playing a game can be measured in two ways: either in the context of the game, or in another, possibly real-world, context. In the first variant, even if a positive outcome is measured, it is not clear if the acquired knowledge, skill, or behaviour is contextualized. Most games, if they are not completely abstract, are situated in a context, and thus every learning outcome resulting from playing them may be contextualized as well. (Wouters et al., 2009) Only if the learning outcome is also measured outside the context of the game, transfer of learning can be assumed to have happened. Therefore, it is important to consider if measured learning outcomes are contextualized.

Causes of success and failure

Most experimental studies evaluating serious games examine if the game "works", i.e. if instructional objectives get achieved. But they often do not examine why a game does or does not work. Is there a specific determining element in the game? If a single aspect in a game changes, do learning outcomes change? Some studies use A/B testing to find out how a specific element in a game influences the learning outcome (see Plass et al., 2014; Orvis et al., 2008)), but they are the exception rather than the rule. In other words, most studies follow a media comparison or cognitive consequences approach, instead of a value-added approach.

What are feasible methodologies to find out why a serious game is successful or not? The following section describes and discusses a debate in educational literature, that is

essential to understand how determining factors for success and failure of serious games can be identified.

4.4.6 Media comparison

The pitfalls of using RCTs to evaluate serious games were described in section 4.4.1. Another issue about using RCTs to evaluate serious games is the choice of a proper control group. Designing a control group is far more complex in educational experiments than in clinical trials. Often, the control group of choice is an educational intervention that comprises the same content as the game played in the experimental group. The approach is based on the assumption that games can be broken down into content and the efficiency of presenting the content can be compared between games and other media. However, section 3.1 described many possible advantages that games can have over traditional instructions, some of them include the systemic and situated nature of games. These are characteristics which other media can hardly imitate with comparable means.

By using instruction in other media as control group, researchers try to measure if a specific game is more successful in achieving instructional objectives than instruction in another medium. This however offers no insight on what it is that makes the game more or less successful. Only by getting this insight will game developers, instructors and researchers be able to create better serious games, decide in which circumstances it is beneficial for learning to use a serious game, and which game to use.

Clark (1983) postulated that "media do not influence learning under any conditions". The metaphor Clark uses is the one of vehicles delivering groceries, which do not influence the nutrition but only the cost or extent of distribution. Media are like the vehicles, only determining cost of distribution but not influencing achievement, or in this case, learning, which is only influenced by content.

Instead of differences in media, Clark argues that it is the method of instruction that determines learning outcomes. He strongly argues for separating method from medium to be able to explain learning variance, and discourages future media comparison research.

Clark supports his claim by referring to meta-analyses, which found that studies reporting connections between media and learning outcomes were confounded. The most common reason for this were uncontrolled effects of different instructional methods in the media under comparison, and the novelty effect of new media, which disappears over time.

Accordingly, what is measured in the evaluation of serious games is not the efficiency of the medium of games compared to other media, but the efficiency of the instructional methods implemented in the games compared to the instructional methods implemented in the medium used as control group.

The takeaway message is, that it is the instructional method which determines learning outcomes, regardless of the medium it is delivered in. Media at most influence learning by attitudes learners have towards a medium. (Clark, 1983)

The implication however is not that the medium of instruction is arbitrarily interchangeable and thus it is not worth studying serious games as a medium. Different media afford different instructional methods in various degrees. This does not contradict arguments for using games as educational tools. In the contrary, these arguments are usually about characteristics of games that afford the implementation of instructional methods. Direct, immediate feedback can be provided by a teacher in a classroom, but one teacher can only assess performance of a limited number of students' in real time, and resources in schools are limited. Situated learning environments can be created without games, for instance through real-life role playing, but games can create such environments in a complexity that is practically impossible in real world.

> Whether we refer to video games as media or simply 'delivery methods' does not alter the fact that games and game technology offer a mechanism for 'learning by doing' that in many cases would be too expensive or dangerous to do in real life (Becker, 2010)

Clark's arguments against media comparison studies have been picked up by scholars in the field of serious games, who discourage the use of media comparison studies in serious games research. (Mayer, 2003; Loh et al., 2015; Erhel and Jamet, 2013)

Consequences

What is the consequence of this media comparison critique? If instructional methods are the determining factor of learning outcomes, and if measuring the outcome of serious games is really measuring the effect of the instructional methods delivered in the game, then a way to identify instructional methods embedded in serious games is required.

The development of instructional methods is based on learning theories. Media may be vehicles for delivery of instructional methods, but every medium has its characteristics that shape the implementation of the methods. Therefore, to enable developers, instructors and researchers to analyze games for instructional methods, they need to be aware of how games implement them. The prerequisite is an understanding of learning theories and game design, both of which are introduced in the following two chapters.

CHAPTER 5

Learning theories

This chapter presents an overview about how people learn. It explains overarching epistemologies in learning and gives a brief overview about influential learning theories.

5.1 Behaviourist theories

Behaviourism is an epistemology of learning based on works of Watson (1913), Thorndike (1913) and Pavlov and Anrep (1927), and expanded rapidly during the 1950s. Although widely criticized in modern educational theory, behaviourist theory is still influential today, especially in learning basic skills such as reading, writing, or spelling. (Egenfeldt-Nielsen, 2006)

In the behaviourist worldview, learning is manifested by a change in behaviour. Behaviour is shaped by the environment, or more precise, by external stimuli. Repetition is a central principle of behaviourist learning.

Behaviourism views the human brain as a black box, so it is not concerned with the internal operation of the brain to understand learning. Rather, learning is viewed as a simple input-output process.

Important behaviourist learning theories are direct instruction, operant conditioning, programmed instruction and social learning theory.

5.1.1 Operant conditioning

Operant conditioning is a theory developed by Skinner (1963). Learning is hereby defined as a function of change in behaviour. The basic principle of the theory is that reinforced behaviours tend to reappear, while punished behaviours tend to disappear. The individual associates a behaviour with a consequence, and operant conditioning can shape this association.

The four principle elements of operant conditioning are positive reinforcers, negative reinforcers, positive punishment and negative punishment. Reinforcers are consequences used to increase behaviours, while punishments are consequences used to let behaviours disappear. Skinner himself strongly argued for using reinforcement instead of punishment in education, as it shapes students behaviour as opposed to show them what they cannot do.

A positive reinforcer is what is generally understood as a reward. It is a consequence given to the individual, when she or he conducted a behaviour that is desired. A negative reinforcer on the other hand is the removal of an unpleasant consequence after a desired behaviour. A positive punishment is an unpleasant consequence after an undesired behaviour, while a negative punishment is the removal of a favorable consequence after undesired behaviour.

Another important aspect in Skinner's operant conditioning theory is scheduling of reinforcements. If reinforcments or punishments are used every time after a desired or undesired behaviour occurs, as soon as the reinforcement stops, the behaviour will also gradually stop (it gets extinct). A way to avoid this problem is to use intermittent schedules of reinforcements. In these reinforcement schedules, behaviour will not be reinforced every time it occurs, but on a defined schedule. Therefore, individuals will not expect to be rewarded every time they elicit a behaviour, and thus not stop the behaviour when they do not get reinforcement for it anymore, as they keep waiting for the reward. Consequently, intermittent schedules of reinforcements result in a higher maintenance of the behaviour compared to continuous reinforcement. Reinforcement schedules can either be in time-based intervals, or in ratios. Intervals and ratios can be either fixed or variable.

Reinforcements are given for a specific stimulus, but they are generalized by individuals to also apply to similar stimuli. This is called secondary conditioning, and describes conditioning that occurs on a stimulus that was not intentionally reinforced but is similar enough to a stimulus that was reinforced.

Principles

1. Positive reinforcement of behaviour will result in conditioning the behaviour.
2. Intermittent reinforcement is more effective to maintain a behaviour than continuous reinforcement.
3. Using varying environments can lead to secondary conditioning.

5.1.2 Programmed instruction

Programmed instruction was developed by Skinner, and is based on his operant conditioning theory. It is in a way just the application of the theory, but at the same time contains some important principles that extend operant conditioning, and is thus worth treating as its own theory.

Skinner introduced the concept with his teaching machine in the 1950s, which is a computer program presenting fill-in-the-blank tasks. The use of a computer allows direct, immediate and individual feedback, which in this case simply consists of the information if the answer was correct or not. Skinner considered this approach often superior to a teacher, because the feedback is immediate and individual.

The core steps of Skinner's programmed instruction start with the breakdown of contents in very small pieces, to get simple statements or questions which are true or false. Then the learner has to submit an answer - usually in the form of filling in blanks. The feedback is immediate, in the form that the learner gets notified if the answer was correct or not, and if it was, the system progresses to the next task. Another important element is that instruction is self-paced and learners are active.

This process in strictly linear. A further development of programmed instruction introduced a branching system. In this, responses are in multiple-option format. If the correct response is given, the system progresses like the linear version. In the case of an incorrect response, the system progresses to another branch, that should correct the wrong assumptions that lead to the incorrect answer.

Special forms of programmed instruction in which repetition of the same concept is a key element, are called drill and practice.

Principles

1. Content should be broken down into small pieces, called frames, so that single questions or statements with answers that are either true or false can be deducted (and therefore behaviour can be shaped).

2. Immediate and individual feedback is crucial for learners.

3. Active learning is preferable to passive learning.

4. Drill and practice style repetition is used to strengthen specific content.

5.1.3 Direct instruction

Direction instruction originated in the 1960s from Engelmann (1968), and has been and is still used widely in education. A good overview about the theory and its development was given by Kinder and Carnine (1991).

Direct instruction, just like other behaviourist theories, is based on the principles of reinforcement, regular assessment, and breakdown of content into small components. The distinct feature of direct instruction is the focus on the stimulus, which should be as precise and unambiguous as possible, to leave the learner as only variable. This can solely be achieved by perfectly controlled environmental variables through "faultless communication".

In the initial stages of instruction, rules and strategies, and their application, are explicitly taught in detail. Thereafter, examples are presented in which the rules are applied. Varying examples are selected to allow students to generalize the rule or strategy. The sequence of examples is important to clarify in which examples different strategies are needed. In successive instruction, the explicit, overt mentioning of every single step of a rule or strategy is gradually reduced, to allow students to apply the rules independently and automatically. Feedback is crucial, and should be adapted to the type of error students make. If lack of information is the cause of an error, then the information should be provided by the teacher. If false application of strategy is the cause of error, then the steps of the strategy should be restated by the teacher.

Principles

1. Clear learning objectives should be established by teachers, and communicated to learners.
2. Rules and strategies of knowledge and skills taught should be presented step by step at the beginning of instruction.
3. Examples should be used to strengthen understanding of the application of rules.
4. Variation in examples allows generalization of learned content.
5. Feedback should be adapted to the error students make, and not simply be a true or false response.

5.1.4 Social learning theory

Social learning theory, also called social cognitive theory, was introduced by Bandura (1977). It is not strictly a behaviourist theory, as it also encompasses attention, memory and motivation and therefore lies in between behaviourism and cognitivism. The difference to classical behaviourist theory is, in one sentence, that Bandura believed that the world and an individual's behaviour cause each other, while classical behaviourism assumes that this causal relationship is bidirectional from environment to individual.

According to social learning theory, individuals learn by observing and modeling behaviours, attitudes, and emotional reactions of others.

There are four core elements in observational learning:

- Attention: the amount of attention paid towards significant features of the modeled behaviour.
- Retention: remembering and retaining the observed behaviour from long term memory.
- Reproduction: physically reproducing the observed.

- Motivation: the reasons to imitate behaviour, motivated by reinforcements, punishments, self-regulation and self-esteem.

Principles

1. Individuals can learn by observing behaviour of others and modeling and imitating the behaviour.
2. In the ideal process of observational learning, the observed is organized and rehearsed symbolically (by means of words, labels or images).
3. Behaviour is more likely modeled by individuals if they identify with the observed person and value the outcome of the behaviour.

5.2 Cognitivist theories

As a reaction on critique on behaviourism, especially on its denial of the importance of cognitive processes on learning, a new theory of learning emerged and replaced behaviorism as dominant paradigm in the 1960s. The concern with the act or process of knowing, the cognition, lead to cognitivist learning theory, to overcome the black-box thinking of behaviourism.

In cognitivist learning theory, the individual mind is important and is seen as an active, organized processor of information. The learner is the center of attention in cognitive learning theories, which focus on individual's construction of representations of the world. (Egenfeldt-Nielsen, 2006)

The cognitivist paradigm includes a number of influential theories: elaboration theory, stage theory of cognitive development, theory of conditioned learning, cognitive load theory, cognitive theory of multimedia learning, dual coding theory and gestalt theory.

5.2.1 Elaboration theory

Elaboration theory goes back to Charles Reigeluth in the 1970s. (Reigeluth, 1992) It is more an instructional design model than a theory, and helps in sequencing content in the optimal way. The theory does not aim at affective content, but on medium to complex cognitive and psychomotor learning.

The basic reasoning of the elaboration model is to sequence learning components in an increasingly difficult way, so that learners can develop meaningful context into which subsequent ideas and skills can be integrated.

Wilson and Cole (1992) describe in their review eight basic strategies of elaboration theory:

1. A single guiding organizing structure should be determined out of conceptual, procedural or theoretical structures.

2. The simple and fundamental parts should occur in the first lessons, successive lessons add layers of complexity depending on the organizing structure, and integrating into the context built buy the underlying layers.
3. Depending on the organizing structure, the sequence of content within a lesson is constructed. For conceptually and theoretically organized instruction, the easiest concept should come first. For procedurally organized instruction, the concepts should be ordered chornologically. Additionally, supporting content should be placed immediately after the related content, coordinate concepts should be presented simultaneously, and principles should be taught before procedures.
4. Summarizers present content reviews after each lesson and unit.
5. Synthesizers help learners integrate content into a meaningful context and prior knowledge. This is often done in diagram form, e.g. flowcharts.
6. Analogies should relate content to learners' prior knowledge, and multiple analogies should be used to accommodate a divergent group of learners.
7. Cognitive strategy activators help triggering cognitive strategies to process material. Tools for this can be embedded in form of pictures, diagrams or mnemonics.
8. Learners should be encouraged to exercise control over content and the instructional strategy.

Principles

1. The type of content that should be taught needs to be identified to develop appropriate instructional strategies. The three types are conceptual, procedural and theoretical.
2. Learning components should be sequenced in increasing difficulty to allow the integration of subsequent content into prior knowledge and context.
3. Within lessons, for conceptual and theoretical content, the simple concepts should come first. For procedural content, the steps of the procedures should be taught chronologically.
4. Principles related to a procedure should be taught before the procedure itself.
5. Content reviews to summarize content should be used after lessons and units.
6. Synthesizers should be used to foster organization of knowledge, e.g. in form of flowcharts.
7. Analogies can help relate new content to learners' prior knowledge.
8. Using multiple varying analogies can help accommodate different types of learners.
9. Learners should have control over the content and the instructional strategy.

5.2.2 Theory of conditional learning

The theory of conditional learning, and as part of it the nine events of instruction, were developed by Gagné (1972). He describes 9 subsequent steps in instruction, which should be satisfied for learning.

The nine events of instruction according to Gagné (1972) are:

1. Gain attention
2. Inform learners of the objective.
3. Stimulate recall of prior learning.
4. Present stimulus material.
5. Provide learning guidance.
6. Elicit performance.
7. Provide feedback.
8. Assess performance.
9. Enhance retention and transfer.

Principles

1. Learners should know the learning objectives.
2. Learners should be guided through the learning process.
3. Learners should be provided with feedback on their performance.

5.2.3 Cognitive load theory

Cognitive load theory (CLT) dates back to the 1980s and became an important theory which expanded and was developed in the 1990s.

Cognitive load can be defined as

> a multidimensional construct representing the load that performing a particular task imposes on the cognitive system of a learner. (Paas and van Merrienboer, 1994)

The assumption of cognitive load theory is, that the human cognition consists of a very limited working memory, but has a comparably unlimited long term memory. This demands for considerations in the instructional design.

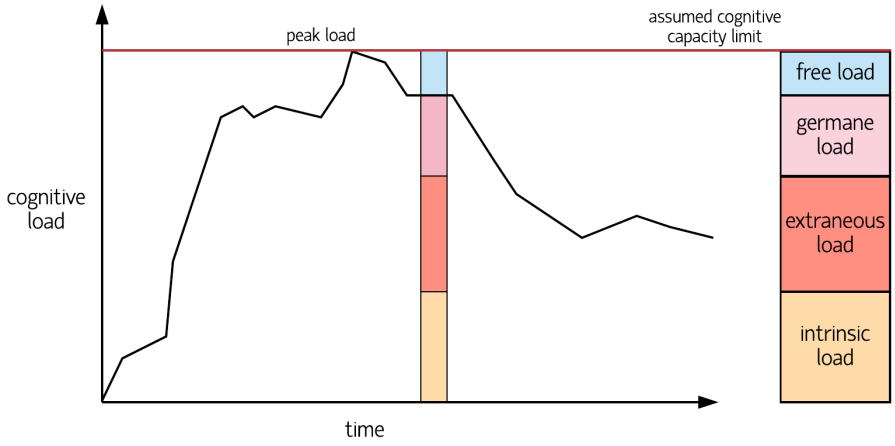

Figure 5.1: Conceptual framework for cognitive load and its attributes by Xie and Salvendy (2000). Figure adapted from Paas et al. (2010).

A key aspect in the theory is schema construction and automation. They work as a mean to free up working memory by chunking multiple elements of information as one single element in a cognitive schema, which can be automated and therefore do not occupy working memory anymore. To store information in schematic form in long-term memory, it is necessary to manipulate it in the working memory first.

Instructional method that were designed for efficiently using the working memory have been proved to be successful in a wide variety of domains for children, young adults and older adults alike. (see Paas et al., 2004; Sweller, 1998)

The efficiency and complexity of what is learned is affected strongly by the amount of working memory resources that are devoted to a task. If working memory is either excessively overloaded (exceeding the available cognitive capacity) or underloaded (inadequate allocation of cognitive resources), learning performance degrades. (Paas et al., 2004)

CLT knows three types of cognitive load: intrinsic load, extraneous or ineffective load, and germane or effective load.

Intrinsic cognitive load is the load directly associated with the learning content itself, and is partly in control of instructors by how they structure content. Extraneous cognitive load is additional load that is added by the instructional material and therefore in full control of instructors. Germane cognitive load is related to the processes that allow for the construction and automation of schemas. Terms and concept are illustrated in Figure 5.1.

Resulting from the theory that overloading and underloading working memory decreases learning and the concept of the three types of cognitive load is the assumption that

instructional material can be designed so that it moves more cognitive load from intrinsic and extraneous to germane load, i.e. it keeps a low cognitive overhead in the learning materials and rather fosters the construction and automation of schemas.

Principles

1. Content should be subdivided into small chunks to minimize intrinsic cognitive load.

2. Instruction should be designed to minimize cognitive overhead, i.e. cognitive load that occurs from the instructional material itself.

5.2.4 Dual coding theory

Dual coding theory was developed by Paivio (1991), and suggests that cognition is specialized for dealing simultaneously with language and nonverbal objects.

Paivio suggests that there are two cognitive subsystems, one for processing of language and one for processing of nonverbal objects, e.g. images. Both subsystems have referential connections to each other. The use of both parallel systems can improve the recall of information.

Principles

1. Information should be presented in both visual and verbal form to optimize chance of recall.

5.3 Constructivist and socio-cultural theories

Constructivism considers learning to be an active, constructive process. (Wu et al., 2012) Knowledge is not transferred from an instructor to the learner, but a learner actively constructs the knowledge in the learning process.

Learners are not viewed as "blank slates" on which any knowledge, skill or attitude can be layered on, but as individuals who are shaped by their past experiences and by cultural factors. Newly constructed knowledge, skills and attitudes are always integrated into these previous experiences.

In the socio-cultural perspective of learning, knowledge is the tool that mediates activity, as opposed to the more traditional view of knowledge as information that is memorized and can be recalled. (Egenfeldt-Nielsen, 2006) It emphasizes the importance of cultural and social context in learning.

In both constructivist and socio-cultural theories, instructors act as facilitators of learning as opposed to sources of knowledge which they try to transfer to learners. Hence, teaching becomes student-centred.

Learning theories in constructivist and socio-cultural epistemology are often not unambiguously classifiable to one or the other, and the epistemologies overlap in many points. Consequently, they will be treated as one category for this work.

5.3.1 Social development theory

Social development theory was proposed by Vygotsky (1978) and was one of the foundations of constructivism. The core assertion of the theory is that social interaction is of high importance in the development of cognition.

The theory encompasses the principal of the zone of proximal development. It is the distance between the actual ability of a learner related to a specific task and the ability which the learner has with additional support. The optimal way of learning would then be to present learners with activities that lie in the zone of proximal development. (Hamari et al., 2016) In other words, a facilitator (teacher, parent, peer, or tool) should guide learners "from an actual point of development to a potential point of development". (Egenfeldt-Nielsen, 2006)

Vygotsky generally emphasized the role that social interaction plays for learning and development.

> Every function in the child's cultural development appears twice: first, on the social level, and later, on the individual level; first, between people (interpsychological) and then inside the child (intrapsychological). This applies equally to voluntary attention, to logical memory, and to the formation of concepts. All the higher functions originate as actual relationships between individuals. (Vygotsky, 1978)

Vygotsky also characterized play as taking a vital role in creating a zone of proximal development for a child. He describes play as an always social activity, that allows them to succeed in things that are slightly beyond their capabilities. (Plass et al., 2015)

Principles

1. Activities lying in the zone of proximal development of the learner optimize learning.
2. Social interaction is a crucial factor in learning and development, especially with people who have a higher ability in the instructional objective.

5.3.2 Situated learning theory

Situated learning theory originated from Lave (1991), and is rooted in social development theory. It hypothesizes that learning is embedded in activity, context and culture. In other words, learning is situated.

Following situated learning theory, learning content should be presented in authentic contexts. These should be the contexts that typically involve the learning content in reality. They are, so to speak, the real-world contexts of the content that is to be taught.

Social interaction is an important part of situated learning. Learners get involved in a community of practice, which is a group of people who are engaged in common activities. (Wenger, 1998) This community of practice constitutes the context of learning, in which learners interact with values, norms and culture of the community. Through participation in a community of practice, learners form meaning about the content. (Kebritchi and Hirumi, 2008) Learners will gradually move to the center of this community which makes them take the role of experts in the content field.

Principles

1. Content should be taught in an authentic context.
2. Engagement in a community of practice lets learners move closer to the role of experts and thereby strengthens their feeling of competence.

5.3.3 Cognitive apprenticeship

Cognitive apprenticeship was proposed by Brown et al. (1989), and is a further development of situated learning. It is closely related to situated learning in conception, but emphasizes the role of instructors.

In cognitive apprenticeship, learning takes the form of three steps, based on Brown et al. (1989):

1. Instructors overtly model their strategies to complete a task.
2. Learners complete the tasks with support by instructors. This is called scaffolding.
3. Learners are encouraged to complete tasks independently.

Principles

1. Learners should not be left alone to discover how to solve a task, instead instructors should make their methods to complete a task explicit to learners.
2. When learners start to solve tasks themselves, they should be supported by experts.
3. Learners should articulate their knowledge, reasoning and problem-solving strategies to learn more effectively and expose weaknesses.
4. Learners should compare their problem-solving processes to those of other learners or experts to reflect on their approach and to motivate them to approach experts' performance.

5. The support of instructors to solve a task should gradually disappear to encourage learners to independent exploration.

5.3.4 Discovery learning

Discovery learning is based on the work of Bruner (1961) on constructivist theory, and emphasizes self-directed learning.

Learners should work in an environment in which they can discover concepts on their own, which is assumed to improve deep learning, meta-cognitive skills (problem solving, creativity, ...), and engagement.

Originally, the concept of discovery learning promoted an environment in which students are confronted with challenges they have to solve without any guidance, just drawing from their past experiences and existing knowledge. By exploring and manipulating the environment, learners are supposed to identify the problem, search for relevant information that can help them solving the problem, and independently develop a solution strategy. The environment should provide feedback to learners' actions and allow them to make mistakes.

However, research has shown that discovery learning is most effective if some form of guidance is involved. (Kirschner et al., 2006; Mayer, 2004) Specifically, learners should have the prerequisite knowledge to successfully develop strategies to solve a problem. This could be implemented by using a form of direct instruction previous to the discovery learning stage, or it could suggest a combination of discovery learning and cognitive apprenticeship.

Problem-based learning is a pedagogical approach that is directly based on discovery learning. The core aspect in it is the use of ill-structured problems, i.e. open-ended problems which do not have one right answer. In problem-based learning, learners usually work in small groups. Teachers are facilitators for learning only, but do not transmit their own knowledge. The biggest challenge in problem-based learning is that students may develop misconceptions, if they are completely unguided and do not know what it is that they should learn.

Principles

1. Deep learning and complex meta-cognitive skills are promoted if students discover problems and strategies to solve these problems on their own.
2. Prerequisite knowledge should be taught before learners approach a problem space independently.
3. Independent development of problem solving strategies should be supported by feedback.
4. Instructors act as facilitators, and can take a passive or active role in guiding the learning process.

5.3.5 Experiential learning

Experiential learning, as developed by Kolb (1984), is based on the assumption that "learning is the process whereby knowledge is created through the transformation of experience". (Kolb, 1984) Experiential learning theory is sometimes categorized as a humanist learning theory. (Wu et al., 2012)

In experiential learning, learners are confronted with real experiences, and encouraged into a reflective process about the experience made to develop new knowledge, skills and values. (Kebritchi and Hirumi, 2008)

Kolb modeled a cyclic experiential learning process, consisting of four steps:

- concrete experience
- reflective observation
- abstract conceptualization
- active experimentation

Learners can start the process in any of the stages, and move through them in a cycle. In the stage of concrete experience, learners are confronted with a real experience. Through consciously reflecting on their experience, learners are expected to conceptualize what they observed. This new conceptualization should then be actively tested through experimentation, leading again to a concrete experience.

Experiential learning can start out with direct instruction, to help learners in making sense of their experiences and effectively conceptualize their observations. This is supported by the claim that additional guidance is required for effective learning. (Kirschner et al., 2006; Mayer et al., 2004) Experiential learning can then be used to immediately integrate what was learned in an authentic context.

Principles

1. People learn best by doing, therefore they should be put into practice to learn.
2. Experiential learning is a cyclic process of experiencing, reflective observation, conceptualization, and experimentation.
3. Prerequisite knowledge should be provided before experiential learning.
4. Acquired knowledge can be integrated in practice through experimental learning.

5.3.6 Transformative learning

Transformative learning is a theory first introduced by Jack Mezirov in 1978, focusing on adult education. It describes "the process of effecting change in a frame of reference" (Mezirow, 1997) and states autonomous thinking as its main objective.

Frames of reference are described as "structures of assumptions through which we understand our experiences". (Mezirow, 1997). They are constructed through experience and shape ones expectations, perceptions, cognition and feelings, and thereby delimit mental and behavioural activities.

Frames of reference can be transformed "through critical reflection on the assumptions upon which our interpretations, beliefs, and habits of mind or points of view are based." (Mezirow, 1997)

Points of view may be intensified if we find evidence for our assumptions. They may be newly established, based on our habits of mind which dictate certain assumptions. They may also be transformed if a critical reflection is triggered by an unexpected result of our actions, a disentanglement of our assumptions, and if this happens continuously it may lead to a transformation in a habit of mind. And finally, our habit of mind may be transformed by self-aware critical reflection which leads to realization of biased assumptions and therefore to a transformation in our habit of mind.

The aim of transformative learning is not a more efficient acquisition of knowledge, but rather to think as an autonomous and responsible agent who is critically reflective of one's own assumptions. This is crucial as a citizen in a democratic system, for the ability of socially responsible thinking and for successful adaption at work. (Mezirow, 1997)

> Becoming critically reflective of the assumptions of others is fundamental to effective collaborative problem posing and solving. Becoming critically reflective of one's own assumptions is the key to transforming one's taken-for- granted frame of reference, an indispensable dimension of learning for adapting to change. (Mezirow, 1997)

Principles

1. Learning is a change in frames of references, which are structures of assumptions that set up learners' understanding of their experiences and limit mental and behavioural activities.

2. Frames of reference can be transformed through critical reflection on assumptions.

3. A continuous disruption of expectations and assumptions can lead to the transformation of frames of references.

5.4 Motivational theories

Motivational theories are concerned with the underlying psychological processes of why people do things, and why they prefer some activities over others. Early theories of human motivation, like the hierarchy of needs by Maslow (1943), tried to identify basic human needs which shape people's behaviour, as they try to satisfy these needs and act accordingly.

In behaviourist theories, motivation has been explained as simple reactions to certain stimuli which can be conditioned. Socio-cultural theories contended this view, and focus on the role of social interaction to motivate behaviour.

In principle, two types of motivation are distinguished: extrinsic and intrinsic motivation. The former is a form of motivation that results from external factors, like rewards or social pressure. Intrinsic motivation on the other hand is inherent in humans. Malone (1981) researched why people are intrinsically motivated to play games. Self-determination theory by Deci and Ryan (1985) is an influential theory about the requirements for intrinsic motivation, and how they relate to extrinsic motivation.

5.4.1 Intrinsically motivating instruction

The theory of intrinsically motivating instruction was proposed by Malone (1981). The theory is based on Malone's curiosity to find out what makes computer games so captivating and how these features can be used to make learning interesting and enjoyable. Not surprisingly, the study has been very influential in the field of serious games.

Opposing the behaviourist position of external reinforcement, cognitivist theorists emphasized the importance of intrinsically motivated play-like activities in learning. Intrinsic motivation in learning leads to more time spent and higher effort in learning and more positive feelings towards it. It may even lead to the ability of "learning how to learn" through the modification of more fundamental cognitive structures. (Malone, 1981)

Intrinsic motivation means that people are engaged in an activity without receiving any external reward. Malone suggests three categories that make up intrinsic motivation of an activity: challenge, fantasy, and curiousity.

To make an environment challenging, it requires goals, which are personally meaningful to the learner. Also learners need to be uncertain if they will reach the goal, but need to be certain that there is a chance for reaching it. This can be ensured by variable difficulty levels, multiple level goals, hidden information, and randomness. Challenge should engage learners' self-esteem. Success in challenging activities makes people feel good about themselves, on the other hand failure can lower their self-esteem and even decrease their interest in the activity. Therefore, variable difficulty levels can be beneficial to allow learners to work on an appropriate one, and performance feedback should be provided in a way that does not damage self-esteem, while still being clear enough to enhance learning.

Fantasy was defined by Malone following the American Heritage Dictionary: "mental images of things not present to the senses or within the actual experience of the person involved." He distinguishes between intrinsic and extrinsic fantasies. Extrinsic fantasies are ones "where the fantasy depends on the use of the skill but not vice versa." (Malone, 1981) In intrinsic fantasies the skill also depends on the fantasy, meaning that problems are somehow embedded in the fantasy, and feedback is more natural. Generally, intrinsic fantasies are more interesting and more instructional, as they have the advantage that the fantasy often indicates usage of a skill. On the other hand, extrinsic fantasies can easily be adapted for different games. Intrinsic fantasy also has cognitive advantages, because they offer metaphors that can help learners to apply old knowledge in understanding new things, and they can improve memorizing material through vivid images.

Learners' curiosity is evoked if environments are not too complicated and not too simple compared to learners' existing knowledge. There should be novel and surprising events, but it is important that learners can have some expectation in what will happen and do not find events too unpredictable. Two kinds of curiousity can be distinguished. The first type is sensory curiousity, in which changes in sensory stimuli of an environment trigger one's attention. This can take the form of audio and visual effects, either used as decoration, to enhance fantasy, as a reward or as a representation system that supports learners' cognition. Cognitive curiousity taps into the wish of forming one's knowledge structures in a certain way. Malone identifies these characteristics that learners' wish to have in their cognitive structures to be the same as the characteristics of a well-formed scientific theory, which is completeness, consistency, and parsimony. Learners' curiousity is then raised if their existing knowledge does not meet one of these characteristics.

Informative feedback is a final important element of motivation, mentioned by Malone (1981). Environments should be responsive to feel interestingly complex. This can be achieved by certain feedback qualities. Feedback should first be surprising to engage learners' curiosity. This can be due to randomness, or due to how a system operates which has to be uncovered by learner's to understand the feedback. Second, feedback should be constructive to be educational. It should explicitly tell learners why something was not right, and what they need to change to make their knowledge complete, consistent, or parsimonious.

Principles

1. Three important factors of motivation are challenge, fantasy and curiousity.

2. Providing clear goals is important for challenge.

3. If goals are emergent (they arise during the interaction between person and environment), the environment should be structured in a way that supports the creation of goals of appropriate difficulty.

4. Feedback should clarify how far away a learner is of reaching her or his goal.

5. Challenge ceases if the outcome of an activity is certain.
6. Tools to prevent that the outcome gets certain too early are varying difficulty levels, multiple level goals, hidden information, or randomness.
7. The difficulty level of instructional activities should be variable, so that learners can work on the right level compared to their abilities to strengthen self-esteem.
8. Feedback should be presented in a way that avoids damaging self-esteem.
9. Instrinsic fantasies (fantasies which are intertwined with skills) are more favourable for motivation than extrinsic fantasies (fantasies in which the fantasy depends on skills, but not the other way around).
10. Intrinsic fantasies can use metaphors that help learners to integrate newly learned things into previous knowledge structures.
11. Intrinsic fantasies can be used to improve memory of instructional material.
12. Fantasies satisfy emotional needs, but need to be carefully selected as different people find different fantasies appealing.
13. To raise learners' curiousity, environments and events should be novel and surprising, but still not completely incomprehensible.
14. Audio and visual effects draw attention by sensory curiousity, and can be used as a reward.
15. Audio and visual effects can be used to support learners' cognition.
16. Pointing out that learner's knowledge is incomplete, inconsistent or unparsimonious can stimulate their curiousity.
17. Feedback should be surprising to raise curiousity.
18. Feedback should be constructive to be educational, in the sense that it should help learners improve.

5.4.2 Flow theory

Mihály Csíkszentmihályi, who formulated the concept of flow after studying the creative process of artists in the 1960s, developed the desire to understand intrinsically motivated activities, which are rewarding in and of themselves without needing any form of extrinsic incentive. (Nakamura and Csíkszentmihályi, 2002)

Flow is a state of mind, in which an individual devotes its full attention to a single task, and coming to a peak of enjoyment, focus and concentration. In this state, individuals operate at full capacity.

In a flow state, experiences unfold from moment to moment (actions during flow are responsive to what happened immediately before) and proximal goals arise during the interaction.

The unfolding experience is shaped by the individual and the environment, but in an asymmetrical way. The person's experience is subjectively perceived, it is important to understand that information and the opportunities of action are selectively attended. Likewise, the own capabilities and thus capacities of action are defining the experience. One cannot speak about the person's skills and the task's challenges in merely objective terms.

> It is the subjective challenges and subjective skills, not objective ones, that influence the quality of a person's experience. (Nakamura and Csíkszentmihályi, 2002)

For entering and staying in flow, it is important that attention is entirely absorbed in the activity. To reach this, activities should have challenges that are on an appropriate level to one's capacities, have clear proximal goals and give immediate feedback about progress. The complete absorbed attention leads to an absence of spare attention and accordingly to a loss of the sense of passage of time, and of self-consciousness. (Nakamura and Csíkszentmihályi, 2002)

If challenges begin to exceed skills, one first becomes vigilant and then anxious. If skills begin to exceed challenges, one first relaxes and then becomes bored.

In flow state, people raise their skill levels, implicating that the perceived difficulty of challenges also needs to increase in order to avoid boredom which leads to ceasing flow. This leads to a continuous growth process of skills and challenges. From a learning perspective, this process is very desirable. It is also desirable from an individual's perspective, since the state of flow is intrinsically rewarding and therefore a self-justifying experience. (Nakamura and Csíkszentmihályi, 2002)

In the original model of flow, Csíkszentmihályi saw three regions of experience (see Figure 5.2) : a flow channel along which challenges and skills matched; a region of boredom, as opportunities for action relative to skills dropped off; and a region of anxiety, as challenges increasingly exceeded capacities for action.

In subsequent development, Csíkszentmihályi saw the additional prerequisite for flow that skills and challenges not only must be in balance, but that challenges must be above the levels that the individual encounters on average in their daily lives, and they must have the skills to engage these challenges.

When challenges and skills are in balance but the challenges perceived are lower than the average challenge an individual faces in their daily life, they get into a state of apathy. The phenomenological map can be refined further, and eight experiential channels can be defined (see Figure 5.3).

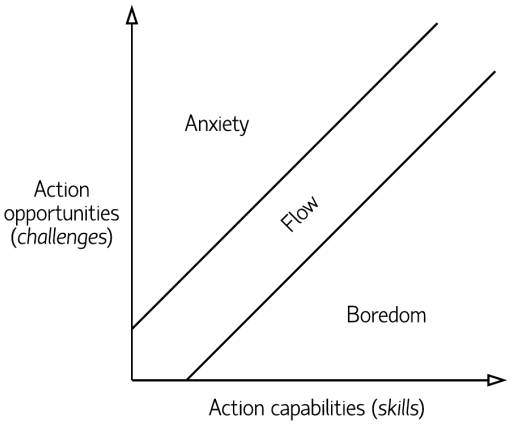

Figure 5.2: The original model of flow, adapted from Csíkszentmihályi (1975).

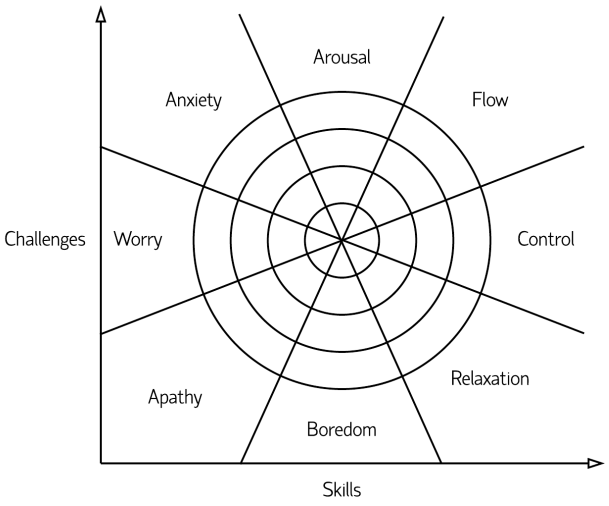

Figure 5.3: Advanced flow model with eight experiential channels. Adapted from Csíkszentmihályi (1997).

Principles

1. Reaching flow state is desirable for learning, because it leads to a growth process in which skills and challenges continuously increase.
2. Reaching flow state leads to a motivational benefit, as activities are intrinsically motivating when learners are in flow.

3. Activities should provide challenges, with a perceived difficulty matching the ability of learners.

4. Activities should have concrete proximal goals or allow the emergence of those.

5. Activities should provide immediate feedback on how learners are doing towards reaching a goal.

6. The perceived challenges of an activity should be above the challenges that learners face in everyday life to avoid apathy.

7. The learning environment should screen out distraction as good as possible to support entering flow state.

5.4.3 Self-determination theory

Self-determination theory (SDT) is a psychological theory, that evolved around the upcoming research on differences between intrinsic and extrinsic motivation. It was formally introduced in the 1980s by Deci and Ryan (1985) and has gained a lot of attention.

Based on a research article by Ryan and Deci (2000), the theory states three psychological needs as essential basis for self-motivation: competence, autonomy, and relatedness.

SDT states that these needs are also essential for personal development and well-being. It assumes that humans naturally strive towards activity and integration, but with vulnerability to passivity. The psychological needs of competence, autonomy, and relatedness then need to be fulfilled to support people's natural activity and keep them from sliding into passivity.

Motivation is often seen as a singular construct, although it is obvious that people get their motivation through different factors. A distinction is made between intrinsic motivation (people value an activity itself, regardless of measurable outcomes) and extrinsic motivation (there is an external influence which drives people to act, focusing on the outcome of the activity). SDT tries to give a more nuanced view on motivation types.

SDT is not about the cause of intrinsic motivation, which is seen as inherent. Cognitive evaluation theory (CET), a sub-theory of SDT presented by Deci and Ryan (1985), is about the conditions influencing the expression of intrinsic motivation, so which factors sustain or diminish the inherent intrinsic motivation of people.

Competence. Events, which give conductors of an activity a feeling of competence, enhance intrinsic motivation. E.g. positive feedback on a task can have positive effects, while negative feedback diminishes motivation.

Autonomy. Autonomy is perceived if learners experience an activity as self-determined and not relying on an external causality. For instance, material rewards like money for well-doing an activity can give a feeling of competence, but can take away the feeling of self-determination and autonomy and thus undermine intrinsic motivation. Also threats and deadlines can have a diminishing effect on intrinsic motivation, because they give the feeling of an external "locus of causality".

Relatedness. Intrinsic motivation is enhanced in contexts with a feeling of relatedness, e.g. students are less intrinsically motivated if they characterize their teachers as cold and uncaring.

These principles of CET so far only apply to activities, which people are already intrinsically motivated in. This is where SDT comes in, by looking closer at extrinsic motivation. When a person is to be pushed into a certain behaviour, the motivation of the person can vary, depending on to which degree the person internalizes and integrates the requested behaviour to her/his own values. So extrinsic motivation can still become self-determined, depending on the social environment.

For example, students doing homework they are not intrinsically motivated to do, can do it with different extrinsic motivations. They could do it because they would otherwise be punished, which then does not feel self-determined. But they could do it because they see the tasks value for their future, and consequently perceive it as a more self-determined task because of the higher level of autonomy.

Relatedness, competence and autonomy are viewed as the key factors for the integration of regulations, because it gives the possibility to grasp the regulations' meaning and to see them coherent with ones own goals and values. They give a sense of choice and freedom from external pressure.

Principles

1. Learners should perceive engagement in activities as self-determined to enhance motivation and avoid sliding into passivity.
2. The feeling of competence positively influences motivation, and can be enhanced by the right level of difficulty and by appropriate feedback.
3. The feeling of autonomy is a prerequisite for positive effects of competence, and decreases if learners feel that the activity is determined by external causalities.
4. Relatedness enhances motivation, and can be improved by connecting an activity to actors that learners feel related to.
5. Regulations which act as extrinsic motivations should be designed to align to learners' own goals and values, so they are perceived as self-determined and therefore share qualities with intrinsic motivation.

CHAPTER 6

Basics of game design theory

To understand how games can be designed to have a value for learners, it is inevitable to have a basic understanding of theoretical concepts of games and game design. After all, not only developing serious games but also analyzing serious games and how learning theories can be represented in them relies on the structures that games are built on.

6.1 MDA framework

The *MDA framework* is an influential approach is understanding games, and one that has been noticed in the game development industry more than any other theoretical model. It was first presented at the Game Developers Conference in 2004 and published as a research paper later that year (Hunicke et al., 2004).

Hunicke et al. (2004) argue that a tool for analysis of games is necessary, especially because the field of games research and development is a highly interdisciplinary field and thus requires a common understanding of the subject matter.

The MDA framework is based on the principle that the game as an artifact stands in different relationships to its designers and players. The designers create a game, but how the game will be played by players is relatively unpredictable. Therefore, they suggest to break games down into three distinct components, or levels of abstraction (fig. 6.1).

Mechanics are the components of a game on a data representation and algorithms level.

Dynamics emerge from mechanics, depending on players' input and on how mechanics act on each other over time.

Aesthetics are the emotional responses by the player when they are playing the game. It is what players eventually experience. Hunicke et al. (2004) present a non exhaustive taxonomy for these experiences, including sensation, fantasy, narrative, challenge, fellowship, discovery, expression, and submission.

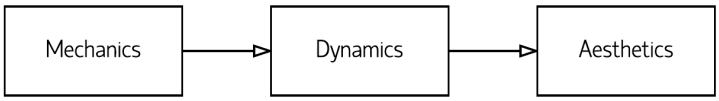

Figure 6.1: Basic model of the MDA framework. Adapted from Hunicke et al. (2004).

Players and designer are looking at games from different perspectives. The designer crafts the mechanics of the game, but players first experience aesthetics. Changing mechanics cascades into the other layers.

The important point of this breakdown is that a game cannot be analyzed by solely looking at it as an artifact, as which it only consists of mechanics. The emergent nature of video games, making them complex systems, requires analysis of the game in action to allow drawing conclusions about its effects.

6.2 From MDA to DDE

The MDA framework has mostly been recited in its original state. It is widely recognized as being useful, especially for pointing out the "second order nature" of game design. (Lantz, 2015) This second order nature is the indirectness between what developer put into the game and how players experience this, which leads to sometimes complex, unpredicted behaviours of games.

However, there has also been critique on the MDA framework, specifically on its usage of ambiguous terms and its, most likely unintentional, implied exclusion of elements like narrative. (Winn, 2009; Lantz, 2015; Walk, 2015)

So in the following, the three levels of abstraction of the MDA framework will be revisited and discussed in more detail.

6.2.1 Mechanics

In the MDA framework, mechanics describe the level of data representation and algorithms, so essentially everything that designers have control over. In game development however, the term mechanics usually refers to something more specific, namely to the units of operation and interaction and how these are connected to each other. There are many more things that developers have control over in the game, including graphics, sound, and level design, which is what MDA presumably refers to as "data representation". Following this line of argument, Winn (2009) and Walk (2015) in unison propose the term *design* for what MDA tries to describe with the word mechanics. This not only solves the ambiguity of the term mechanics and its collision with the actual use of the term in game development, but it also emphasizes that it is the part that is in direct control of the game developers and designers.

6.2.2 Dynamics

Dynamics are what emerges through interaction of players with the mechanics. Walk (2015) decomposes this layer in more detail, suggesting that dynamics consist of three types of interactions:

- player to game
- player to player
- game to game

Also Walk introduces the concept of the player-subject, based on the works of Miguel Sicart. According to this theory, it is not us who are playing the game, but subsets of ourselves. A character with a different set of abilities and ethics, explaining why we are able to take actions in games which we would consider highly unethical in reality. As player-subject, we are in a safe space while ethically and mentally challenging situations are explored. (Walk, 2015)

According to Walk, the player therefore interacts with the dynamics of the game mediated by the player-subject, while the designer interacts indirectly through the mechanics.

Winn (2009) proposed to use the term *play* as a term to encompass all elements of a game in this layer, as dynamics are closely related to gameplay and not to, e.g., storytelling.

The term dynamics however is more precise in highlighting its emergent nature, and the term is shared with dynamic systems, which are a good analogy to the complex, emergent behaviour of games.

6.2.3 Aesthetics

Aesthetics are described in the MDA framework as players' emotional responses to playing the game. The term suffers from a similar problem as mechanics. It is intuitively connected to something different than it tries to communicate, and is used differently in game development. Aesthetics is usually thought of as visual aesthetics. (Lantz, 2015) In game development and analysis, aesthetics are usually discussed as a counterpart to gameplay, to hold the two concept apart or talk about connections of both. Moreover, aesthetics is also a term with different definitions in philosophy and psychology, making it even more confusing. (Walk, 2015)

Winn (2009) and Walk (2015) agree to use the term experience instead of aesthetics. It is what is essentially described by MDA under the term of aesthetics, namely the experience that the dynamics of the game provide when they are *experienced* by the player through interaction with the game.

The experience of the game is out of direct control of game developers. That is not to say that they have no influence on it, because they can shape it through changing the

mechanics, which will cascade up to the experience of playing. But, there is no immediate control for the developers, as experience is influenced by the person that is playing.

6.2.4 DDE framework

Critique on terminology of the MDA framework suggests changing the term mechanics to design, and aesthetics to experience. This is in line with the suggestion of Walk (2015) and results in the *Design-Dynamics-Experience (DDE)* framework.

The terminology of design, dynamics and experience is much more consistent with everyday language of game developers than mechanics, dynamics and aesthetics, and communicates the core concepts of the levels of abstraction by their name.

The key concept of the framework is that games cannot be analyzed by solely looking at them as artifacts, or as a set of mechanics. They need to be analyzed by playing, because only then the emergent nature of games unfolds, and often unpredicted dynamics develop during gameplay. What players experience results from dynamics of the game and the previous experiences and characteristics of individual players. Thus, when analyzing a game, it is not sufficient to look at the game's mechanics, visuals, narration, audio, or level design. In the end, what matters is the experience that players have when they are playing the game, and that needs to be taken into account in the design and evaluation of serious games.

CHAPTER 7

Opening the black box: learning theories applied to games

In this chapter, the principles derived from learning theories presented in chapter 5 will be discussed from a game design perspective. This leads to basic guidelines for a new perspective on serious games which closely integrates learning theories with serious game design and evaluation.

7.1 What's in the box?

Many studies on the evaluation of serious games assess learners' performance after playing a game (see section 4.2). The game is treated like a black box, used as an intermittent tool that is placed between learners and instructional objective, and leads learners towards a learning outcome. The game itself however receives no attention exceeding a basic description of its mechanics. The application context of the game is often artificial. Consequently, the gained insights in evaluation studies are often limited to the finding if a specific game is or is not useful to reach a defined instructional objective in an unnatural environment. As long as the game is treated as a black box, as it is impossible to know which aspects of the game cause measured learning outcomes.

Developers and researchers of serious games alike are often not considering any coherent learning theories in the design or evaluation process. (Shaffer et al., 2005; Wu et al., 2012) In a literature search by Wu et al. (2012), they found that only 91 of 567 studies on serious games based their investigations in learning theory. Kebritchi and Hirumi (2008) reported that less than half of serious game designers reported pedagogical foundations of their games. The assumption often seems to be, that having a game is sufficient to improve learning. (Hays, 2005) But literature reviews and theoretical studies emphasize the significance of using established learning theories in the design of serious games. (O'Neil et al., 2005; Kebritchi and Hirumi, 2008)

Which learning theories are most appropriate to apply in games has been discussed in literature. (Plass et al., 2015; Egenfeldt-Nielsen, 2006; Kebritchi and Hirumi, 2008) Some serious games advocates argue that games are so distinct from other forms of learning that they should form a unique learning theory. (Gee, 2003; Prensky, 2003; Bogost, 2007) However, Plass et al. (2015) argues that the uniqueness of games for instructional purposes is not defined at an epistemological level, and games use elements of behaviourist, cognitivist and constructivist theories.

Games are complex systems, and using them as learning environments can be related to a variety of learning and motivational theories. Understanding them requires the consideration of multiple principles of learning and of motivation. (Plass et al., 2015) Studies that focus on one learning perspective neglect other qualities of games. For example, evaluations of serious games that focus only on the cognitive aspects of games as learning environments neglect the motivational and sociocultural aspects. (Plass et al., 2015) Many studies take a motivational perspective, which can lead to games that are engaging but in which no transfer of the learned material occurs, meaning that players only learn how to play the game, and not how to apply the learned content outside of the game.

Therefore, it is important to consider games as tools which can implement any learning theory. Studying the design of games and how learning theories can be applied in it can inform the design process to create better serious games, and can also serve as indicator that games can promote learning in ways that other media cannot.

Kelly et al. (2007) writes:

> Many educational games do not properly translate knowledge, facts, and lessons into the language of games. This results in games that are often neither engaging nor educational.

That does not mean that principles of learning theories can simply be translated into specific game mechanics which then make the game to an effective tool for learning. This misunderstanding does not take into account what has been discussed in chapter 6: as designer, one can change the mechanics of the game, but the mechanics do not deterministically infer the dynamics and ultimately the experience of players, or learners in this case. Because games are complex systems, mechanics can unfold unexpected dynamics and experiences.

Nonetheless, studying the application of learning theories in games can be valuable to understand which dynamics developers of serious games can strive for, and from there it requires the skills of game designers to put mechanics into place in a way that the desired dynamics unfold from them.

Not only the design, but also the evaluation of serious games can benefit from the insights of studying the application of learning theories to games. Based on the arguments about

media comparison studies (see section 4.4.6), what is really evaluated when performance of learners is assessed after playing a game, is how effective the instructional methods implemented in the game were. To make meaningful conclusions from these assessments, it is a prerequisite to know which instructional methods *are* present in the game.

To improve both design and evaluation of serious games using established learning theories, a common perspective of developers and researchers is required. This chapter will provide the basis for the development of such a common language and perspective.

7.2 Models of serious games

Several attempts to create comprehensive models to guide the design and evaluation of serious games have been made. None of these models has been adopted in a relevant number of research papers, but are usually only applied by the creators of the models themselves. Nevertheless, it is worth reviewing the models to identify what their valuable contributions are, and where they fall short.

7.2.1 A theory of game-based learning

Plass et al. (2015) proposed a theory on game-based learning, that suggests the addition of the concept of playfulness as an orthogonal dimension to the learning theory a game is based on. They argue that games in their complexity can apply any learning theory, not only a specific one. But instead of using a meta-theoretical model about all learning theories, their approach is to construct a simple structural model that applies to all games, consisting of the three elements of challenge, response and feedback. The model can then be used to induce which learning theory is used in the design of the game (see Figure 7.1).

The model of Plass attempts to take multiple levels of engagement into consideration, namely affective, behavioural, cognitive and social/cultural engagement. The types of engagement are facilitated through design features, which are themselves influenced by the different learning theories (Figure 7.2).

The model offers a good representation of how learning theories can inform the design of games. It also shows the link between design features and categories of engagement, leading to a good learning experience.

However, the model falls short in explaining how the relations between learning theories to game features, and game features to engagement work in more detail. Therefore, it is useful as a model and theoretical perspective, but its practical use is limited. Serious game researchers miss the clarity about how game design elements influence engagement. Serious game developers on the other hand miss clarity about how learning theory can influence game design.

These shortcomings mean that the links between the elements of the model have yet to be explained, which has been remarked by the authors themselves:

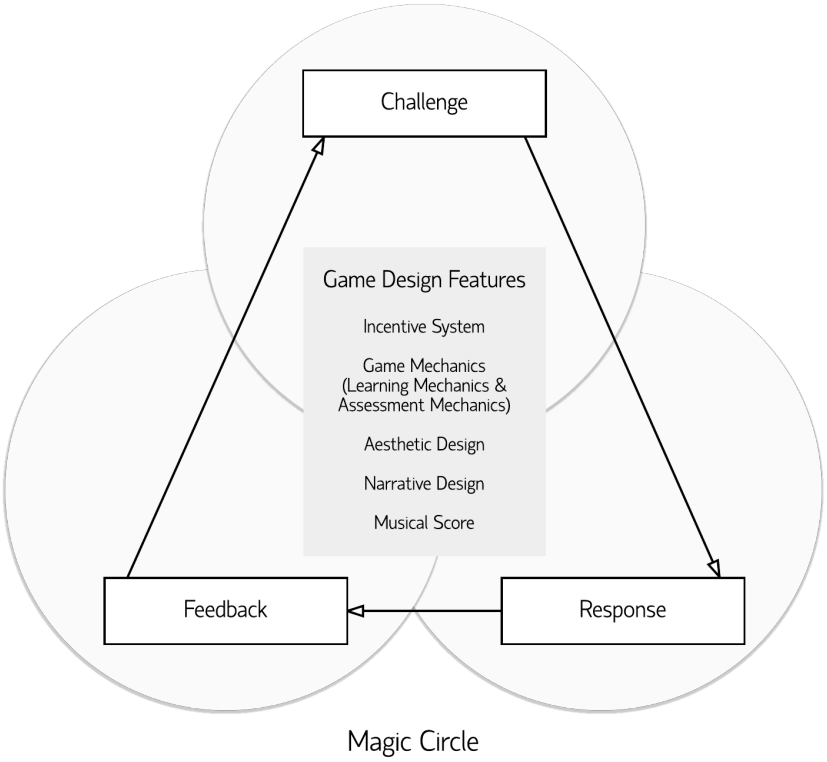

Figure 7.1: Model of game-based learning in the magic circle. Adapted from Plass et al. (2015).

> The design framework [...] leads to several lines of research that should be investigating how specific theories can be implemented through specific design elements, how these designs can lead to specific types of engagement in learners, and how these types of engagement facilitate learning. (Plass et al., 2015)

7.2.2 Matching game characteristics to development of 21st century skills

Romero et al. (2015) seek to identify characteristics of games that facilitate the development of so called 21st century skills, which relate to complex cognitive skills like problem-solving, communication, decision making, planning and adaptivity.

Based on other works, Romero et al. (2015) identify five main characteristics of games: competition and goals, rules, choice, fantasy, and challenges.

The authors conduct a study on the relevance of each of the game characteristics to the

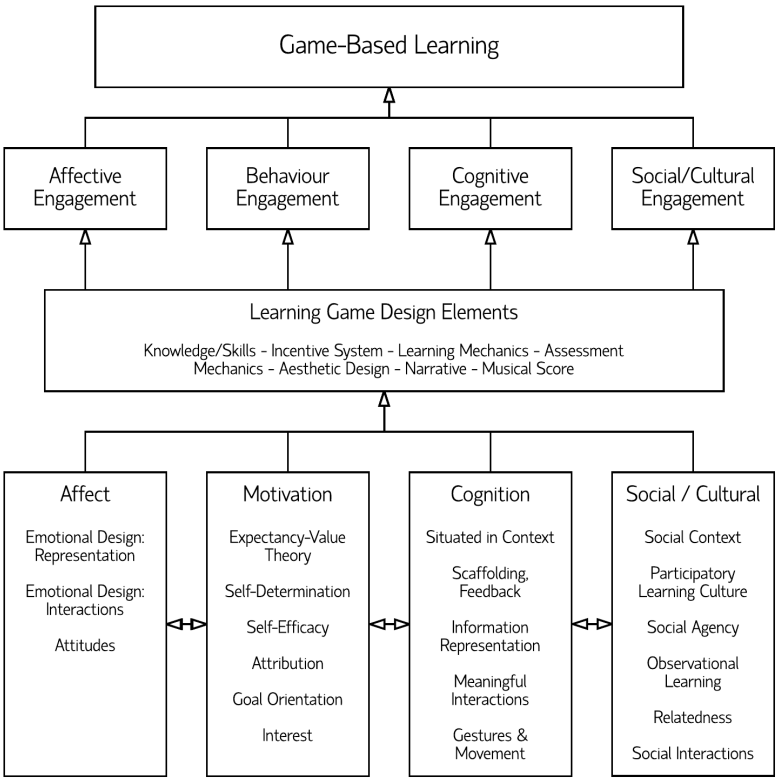

Figure 7.2: Comprehensive model about how learning theories influence serious game design in different dimensions of engagement. Adapted from Plass et al. (2015).

development of 21st century skills, and discuss the advantage of some characteristics to support specific skills.

The identified characteristics of games are one of many possible perspectives on how games can be analyzed. By choosing this limited set of characteristics, the complexity of games is reduced to these elements. The impact of the elements on the development of specific skills can be mediated by many other factors than the ones chosen to examine.

7.2.3 Learning Mechanics-Game Mechanics model

The *Learning Mechanics-Game Mechanics* model (LM-GM model) was proposed by Arnab et al. (2016), and tries to integrate educational and game design principles. It should aid the analysis, design and specification of serious games.

The authors suggest that high-level pedagogical intents can be implemented through low-level game mechanics. The concept of *Serious Game Mechanic (SGM)* is used, which

are defined as the game components that translate a learning principle into a game mechanic.

The model consists of elements in two categories. The first category is a set of learning mechanics, while the second category is a set of game mechanics. Users can then freely relate learning and gaming mechanics to each other to create a representation of the implementation of learning principles in a game.

The learning mechanics have been extracted from literature and are described as non exhaustive. The game mechanics were derived from articles on game mechanics and dynamics.

In the application of the model, the user should identify the mechanics in each game situation and create a map of a game flow of actions. The learning and game mechanics can be assigned to each step in the game flow, to represent how learning principles are applied in the design of the game.

The first issue in this approach is the limited set of learning and game mechanics used, and that the process of choosing them is obscured. The second issue is the seemingly limited understanding of the authors of what game mechanics are. The elements in the game mechanics category are a mix of mechanics and dynamics of a game (see chapter section 6.1), and some are very specific while others are highly abstract. The same is partly true for the set of learning mechanics, which are of various specificity. Additionally, all the mechanics are only described by a name, making it difficult to infer their exact meaning. Lastly, the model seems to suggest that learning principles can be directly translated to game mechanics without the consideration of emergent dynamics.

7.2.4 RETAIN model

The *RETAIN* model was proposed by Gunter et al. (2008), and is based on game design and instructional design principles. The author claims that the concepts in game and instructional design share similar aspects, and a common understanding for both aspects needs to be fostered in game designers and instructional designers.

The model is based on three instructional theories, which were identified to be closely aligned with game design principles:

- Keller's ARCS model
- Gagne's events of instruction
- Piaget's ideas on schema

Drawing from these theories, the RETAIN model identifies six crucial elements in serious game design: relevance, embedding, transfer, adaption, immersion, and naturalization.

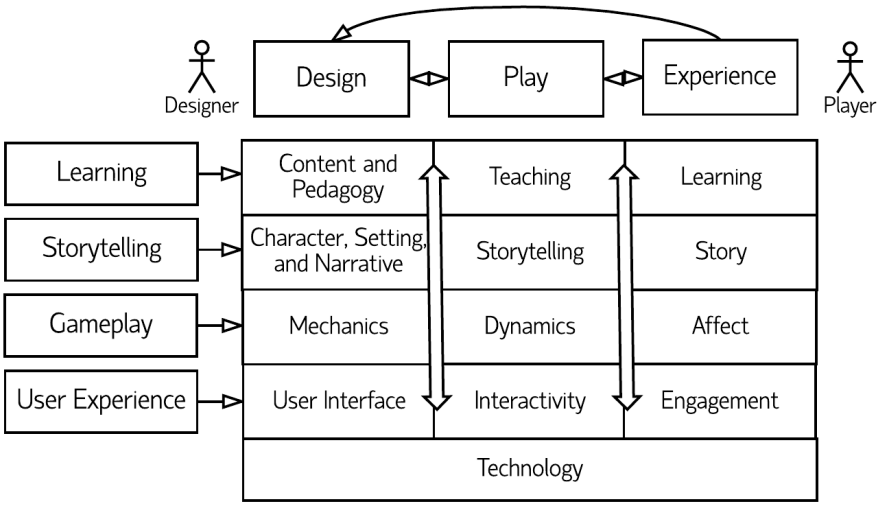

Figure 7.3: Several layers in game design as part of three main components: design, play and experience. Adapted from Winn (2009).

The model can offer guidelines for serious game designers to incorporate learning theories into the design of games. However, the choice of the three instructional theories seems somewhat arbitrary and focuses on cognitive theories, leaving out constructivist approaches. Not only is the choice of these theories limiting to the definition of guidelines, but part of the guidelines themselves stay rather abstract and create a unification of the three theories, without discussing their application in game design sufficiently.

7.2.5 DPE framework

The *DPE framework* by Winn (2009) is an expansion of the MDA framework and applies it specifically to serious games. It emphasizes the crucial importance of promoting the understanding of game dynamics in research and development of serious games.

The framework tackles the semantic issues of the MDA framework (see section 6.1) by changing the terms mechanics, dynamics, aesthetics to design, play and experience. Additionally, it defines subcategories for each of the layers of abstraction which specifically relate the components to learning, storytelling, gameplay and user experience (see Figure 7.3).

The framework is supposed to aid the decomposition of the design of serious games to analyze them, and to define a common language for the stakeholders of different domains that are involved in the development, application and evaluation of serious games.

The framework is an important contribution to understand that the model of design, dynamics, and experience also apply to the specific components of serious games. But

it is limited because it only describes some selected principles of learning theories, like difficulty and reward structure, but does not include comprehensive learning theories in the framework.

7.2.6 Serious Game Design Assessment Framework (SGDA)

The *Serious Game Design Assessment Framework* (SGDA Framework), proposed by Mitgutsch and Alvarado (2012), tries to structure the study of the design of games with an explicit or implicit purpose of fostering learning or social change.

The authors argue that what serious games have in common is that they are designed to have a purposeful impact on players' lives beyond the game, and consequently the analysis and evaluation of the purpose-based design should be the focus of research on the impact of serious games.

The basic concept of the SGDA framework is the study of the configuration of its design elements in relation to the game's purpose. The purpose should be reflected in all aspects of the game: in content, fiction & narrative, mechanics, aesthetics & graphics, the framing, and the coherence and cohesiveness of the relation among these components (Figure 7.4).

The aim of the framework is to offer a structure for game criticism, assessment and evaluation, and to structure design processes. It is valuable for a purpose-centric perspective in the design and evaluation of serious games, specifically to study if the games' purpose is reflected in all of its components. How the purpose can be best reflected in the components however, is not in the scope of the framework. It does not draw principles from learning perspective, but leaves researchers or designers alone in identifying how to achieve reflection of the purpose in game components.

7.3 Games from learning theory perspectives

In this section, the principles derived from learning theories are described in terms of game design. The resulting design principles are guidelines to aid the process of bringing the desired instructional methods into game dynamics through the use of game design, or to help finding instructional methods in the dynamics of existing games.

7.3.1 Games from a behaviourist perspective

Even if widely criticized, behaviourist learning theories still play a role in certain educational environments today. In general, education moved on to rely on more student-centered methods, but behaviourist theories can at least give some valuable inputs to it. This is especially in the design of serious games, as games commonly use (knowingly or unknowingly) core aspects of behaviourist theory, such as rewards, reinforcement schedules and direct feedback.

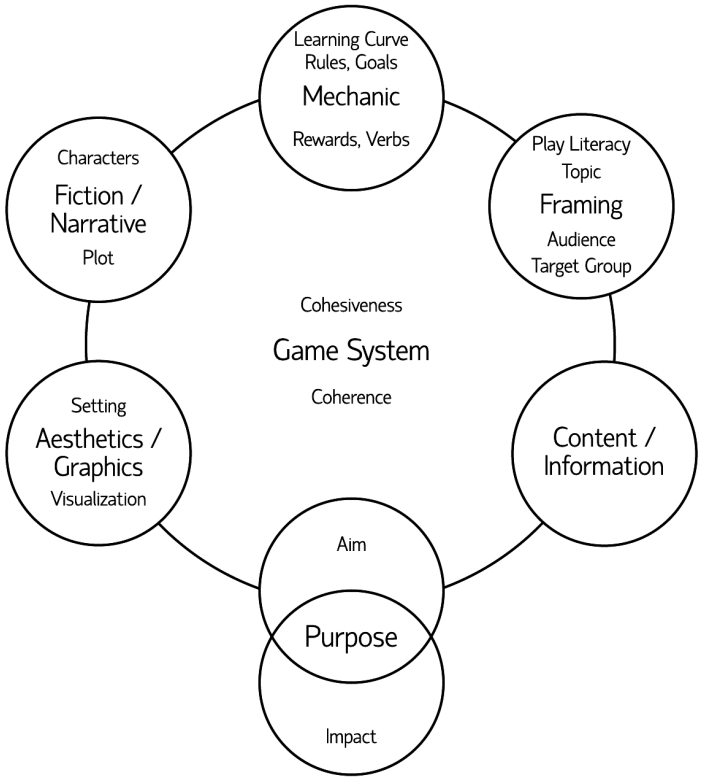

Figure 7.4: The Serious Game Design Assessment framework emphasizes that the purpose of the serious game should be reflected in every other component. Adapted from Mitgutsch and Alvarado (2012).

Feedback. Feedback is one of the central elements in behaviourist learning theory. While feedback in operant conditioning is binary, in that it shows if a behaviour is desired or not, later behavioural theories like direct instruction suggest that feedback should inform learners about their mistakes in more detail.

The concept of feedback is a central element in game design. In good games, feedback should give players immediate or timely information about their performance, in an implicit or explicit form. (Shute and Ke, 2012) It cannot only be provided as text, but comes also in visual and auditive form. Feedback is, at least in a basic form, part of every game, as it enables players to understand what is going on and what the consequences of their actions in the game are. This immanent form of feedback to teach players how to play the game, can be expanded onto other instructional objectives in serious games.

Rewards. A specific form of feedback are rewards. Rewards are important in classical operant conditioning to reinforce certain behaviours. In behaviourist theory, rewards are

usually external.

In games, rewards are usually embedded in the game world, but can take an external or internal form in that world. For instance a very common kind of external reward inside the game's context is a point system, that rewards good performance by giving players points which rank them in high score systems. Instead of points, rewards can also be ingame currency, which can be used to buy items or unlock elements of the game, and are required for the progression in the game. In certain game genres like role-playing games, experience points are rewarded. They enable the development of a player's character in the game. Rewards like in-game currency or experience points are somewhere in between internal and external rewards, as they are an integral part of the game itself and are required to progress through the game, but are still not an immanent part of the rewarded task and are exchangeable. In many games, the main reward is the progression to higher levels or through the storyline. This rewards are internal, as they are integral to the task at hand. Sometimes, rewards can be much more subtle than these quantifiable systems. A specific sound that is played or visual element that appears every time a player elicits good performance in relation to a goal in the game can also act as a reward.

Rewards are therefore a specific kind of feedback for an action that is desired. In the case of serious games, rewards can be used to reinforce a behaviour that is desired in relation to the learning objective.

Reinforcement schedules. Based on the theory of operant conditioning, reinforcements lose their impact when they are rewarded every single time a desirable behaviour occurs.

Balancing the frequency of rewards is a common problem in games. It is often related to regulating perceived challenge, as too frequent rewards can make the achievement of proximal goals seem to easy and consequently negatively influence motivation.

Winn (2009) describes the problem of balancing frequency of rewards related to the learning curve of a game. When the difficulty of challenges rises quickly (steep learning curve), the frequency of rewards should be increased as opposed to sequences with a flat learning curve (see Figure 7.5). This is supported by Linehan et al. (2011), who postulate that the workload required to reach a reward can be increased gradually without negative effects on motivation.

Repetition. In programmed instruction theory, repetition is explicitly mentioned as a core element, and it is implicitly important in all, and not only, behavioural theories. Repetition ought to internalize knowledge or skills and avoid failure. It can bridge the step between descriptive and procedural knowledge (see section 3.4), between changing a point of view and changing a habit of mind (see section 5.3.6), and to mastery of a skill.

Repetition is an inherent element of many games, in which failure leads to the repetition of the last segment, the current level, or even to a complete new start of the game. But it is not only an inherent part of failure in many games, but also in the progression of

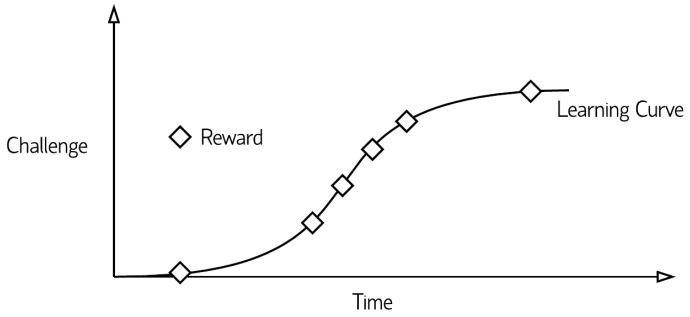

Figure 7.5: Reward schedules should be adapted to the learning curve of a serious game. Adapted from Winn (2009).

the game itself. In what is considered to be good design, mechanics that players learn will be repeated in further progression of the game in the same or in a slightly altered form, to lead players' skill in interacting with the mechanic to mastery and progress to the integration of the next skill.

Breaking it into pieces. Following programmed instruction, the breakdown of content into small pieces is an important concept to apply behavioural strategies. If complex behaviours or knowledge structures are taught at once, it is impossible to give immediate, short and concise feedback on the performance of learners.

In game design, breaking complex processes into pieces is a common method to gradually lead players towards the required skills to overcome complex challenges. The pieces can be learned individually one after the other, each one strengthened by repetition before the next one is introduced.

Defining and communicating learning objectives. The learning objectives of an instructional intervention should, according to direct instruction theory, be explicitly defined and communicated to learners.

This aspect is debatable in relation to serious games. Concerns are that if players do not know what they are supposed to learn, transfer of learning will not occur. On the other hand however, communicating that playing the serious game has the intention to reach certain learning objectives may harm perceived autonomy, and consequently decrease intrinsic motivation and subsequently negatively influence learning outcomes.

It would be worth investigating this further, to get a better understanding when the clear communication of learning objectives to players is beneficial or detrimental to actual learning outcomes.

Varying application environments. Providing application examples of instructional content is a principle in the theory of direct instruction. Applying skills in varying environments or contexts can improve transfer of the learning outcomes to applications outside of the game.

Game systems to achieve this generalization effect are related to repetition and breaking into pieces. After skills are learned, they get enforced by repetition and application in different contexts, until, ultimately, the skill can be generalized to arbitrary contexts.

Social learning. In social learning theory, learning is conceptualized to occur through observation of others. The observed behaviour then gets organized symbolically, trained through repetition and is finally acquired. The chance that one mimics behaviour of another is dependent on the identification one has with the other and the value one assigns to the observed behaviour.

The translation of this process can primarily be related to multiplayer games, because it is (at least today) unlikely that the interaction between humans in this regard can be replaced by the interaction with artificial social actors.

In multiplayer games, players observe the behaviour of others and mimic it, if the outcome of the behaviour fits into their value system. As stated in the theory, the chance to adapt observed behaviour depends on the identification with the modeled individual. In the case of certain multiplayer games, identification can arise from complex social communities around these games.

Active learning. The theory of programmed instruction postulates that learning should best be active, and not passive.

This comes for free in games that exceed a certain threshold of interactivity. Serious games should be designed to offer player enough agency for active learning.

7.3.2 Games from a cognitivist perspective

The cognitive perspective of learning moves the focus on the limitations and potentials of cognitive processes of the human brain. It emphasizes the importance of structuring learning material to support cognitive processes. (Egenfeldt-Nielsen, 2006)

In serious games, the complexity of the game should be aligned with the limitations of human processing capacity. (Wouters et al., 2009; Egenfeldt-Nielsen, 2006)

Minimizing cognitive load. Cognitive load theory differentiates between intrinsic, extraneous and germane cognitive load. Holding the sum of cognitive loads on an optimal level is important, because overloading or underloading working memory can result in mental overload or boredom. The negative effects of it can be viewed from different theoretical perspectives, e.g. decreasing feeling of competence, preventing state of flow,

and decreasing engagement. The distribution of cognitive load to intrinsic, extraneous and germane load should be directed towards maximizing germane cognitive load.

Minimizing intrinsic cognitive load can be achieved by keeping the amount of content that is presented at a time as small as possible. To achieve this, complex content should be subdivided into small chunks.

Extraneous cognitive load results from cognitive overhead generated by the instructional material. Minimizing it is one of the main concerns in serious game design and the most important distinction between designing games for entertainment and designing games for learning. Narrative, visual aesthetics, music and sound design are all elements in the design of entertainment games that contribute to an engaging playing experience. But in serious games, these elements must be handled with care, as they can have a high impact on extraneous cognitive load and thus further limit resources for constructing and automating schemas. On the other hand, they should not be abandoned entirely, as they contribute to engagement and motivation of players.

This tension can be exemplified with the case of narratives in games. On one side, narratives (just as visuals and audio) can enhance authenticity in games, and can present information in a way that is embedded in the game world and help organizing educational material. (Adams et al., 2012; Romero et al., 2015; Wouters et al., 2013; Plass et al., 2015) On the other hand, from a cognitive perspective, narratives may be distracting learners and increase extraneous cognitive load. (Mayer, 2003; Wouters et al., 2013; Adams et al., 2012)

To summarize, serious game designers must handle engaging elements like visuals, audio, and narrative with more care than entertainment game designers. The elements should optimally be directly related to the instructional content, so that they are not distracting from the learning objective, but support it additionally.

Scaffolding. According to the theory of conditional learning, learners should be guided through the learning process. This avoids that learners are overwhelmed if they have to tackle complex problems on their own.

The advice and resources in a learning environment that support learners to reach the learning outcomes are called scaffolding. (Connolly et al., 2008) Especially in games that offer a high degree of autonomy and broad set of choices to learners, scaffolding should ensure to keep them on track and be present to help if it is required. This can be achieved by in-game manuals, or through unobtrusive support that is integrated into the game world and available for players to consult if they get stuck. Scaffolding can ideally also be more subtle, by supporting players without them even noticing through skillful use of game mechanics.

Sequencing learning components. The elaboration theory emphasizes importance of properly sequencing learning components. Thoughtful sequencing of content should allow integrating subsequent content into prior knowledge and context.

From a game perspective, it is a usual process in game design to consider the capabilities of players and align content with increasing difficulty. Subsequent content should be presented in a way that makes connection with previously learned content easy and fosters the integration of the previous content with new one.

Analogies. Analogies are related to the sequencing of learning components and support integration of subsequent content into previous knowledge.

Analogies in games can take many forms. They can be verbal, visual, auditory, or systemic. The combination of those, or context-dependent choice of one, can help to vary analogies to accommodate different learner styles.

Reviewing content. Suggested by elaboration theory, lessons and units should conclude with a review of the newly introduced instructional content.

Building on the sequencing of content in games, a segment of the game can conclude by requiring players to apply cumulated knowledge and skills they acquired in a concluding sequence. Only by integrating all chunks of content that have been acquired previously can a concluding challenge be overcome. A classic example of this form of reviewing in games are boss fights.

Present visual and verbal information. The contiguity principle, presented by Mayer and Anderson (1992) and based on dual coding theory, suggests that words and pictures should be presented contiguously in time and space to increase learning effectiveness and retention of learned content.

In games, it is common practice that textual, visual and even auditive information are closely connected to each other. In serious games, special care should be taken that relevant information for learning fulfills the contiguity principle of close connection between words and pictures in space and time.

Control. The final postulate of elaboration theory is that learners should be in control over content and instructional strategy. Games can provide a high level of control through interactivity and embedded assessment. Learners can maintain control over the pace of learning.

Games can facilitate learner control by giving players freedom of choice on multiple levels. This can be the autonomous choice of proximal goals, or the choice between multiple ways to solve one problem. It is important to balance this freedom with some form of guidance, to keep players on track and not let them get overwhelmed or feel helpless.

7.3.3 Games from a constructivist perspective

Principles from constructivist learning theories have been used as key arguments for the advancement of games as learning tools (see section 3.1). Accordingly, many principles from constructivist learning theories can be applied in serious game design.

Zone of proximal development. The zone of proximal development is the zone between actual and potential capabilities of learners. Effective instruction relies on activities based in this zone. Instructors or other peer individuals can support learners in achieving their potential capabilities.

Serious games can make use of the insights from social development theory by integrating the zone of proximal development in their design. In practice, activities in serious games should present challenges that are slightly above learners capabilities, and provide support to close the gap between actual and potential capabilities.

Situated learning. Situated learning emphasizes the importance of authentic learning environments, in which situated knowledge and skills can be acquired.

Games in general provide, with the exception of abstract games, environments and virtual worlds that represent reality in various degrees of realism. The virtual environment can either embed learning content into the world and so create an optimal environment for situated knowledge acquisition, or it can make the understanding of underlying mechanisms of the game to the core experience to foster understanding of systems and application of skills in the authentic environment.

Social interaction. In situated learning theory, the principle of communities of practice focuses on the acquisition of knowledge and skills embedded in a community. The community builds a cultural context, and learning includes understanding and adapting cultural traits and practices of the community in a certain domain.

By presenting authentic contexts, serious games can represent communities of practice to present cultural elements of a domain to players of the game. By actively participating in the community through the game, players can adopt cultural habits and domain specific language of a community. Players can be put into the role of experts in a domain in the virtual world, fostering the feeling of competence and affiliation to the community.

Scaffolding. Scaffolding has been introduced as important element in the cognitive perspective on games (section 7.3.2). It is also a key element in some constructivist theories.

Social learning theory and cognitive apprenticeship emphasize the role of scaffolding in developing and constructing skills and knowledge, by drawing from the support of domain experts. In serious games, scaffolding can take the form of an instructor that is external to the game, or virtual instructors in the game. Virtual instructors can be either represented as characters in a game, or sources of information that are readily accessible at all times during gameplay. Many science fiction and fantasy entertainment games like the The Legend of Zelda: Ocarina of Time and the Halo series depict the role of instructors as either an AI or a transcendental entity, that supports the player in learning the relevant skills to progress in the game.

Discovery. The theory of discovery learning is based on the assumption that learning is promoted best if students discover problems and strategies to solve them on their own, especially for complex meta-cognitive skills. As research suggested, discovery learning should be guided to be successful, to help learners find right directions and not letting them get overwhelmed.

Game environments can be designed to support discovery learning. Games can provide a wide range of choices and afford different actions, and players can use different strategies to solve problems in complex games. However, this can contradict game design principles of limiting player choice to avoid the so called paradox of choice. Situational scaffolding can help countering the problem by offering guidance as suggested by guided discovery theory, to help players in finding relevant choices to solve a problem.

Experimentation. Experimentation is one of the four steps of experiential learning. Experimentation allows learners to try approaches to problem solving in a save environment, in which they can authentically observe the results of their experimentation, but experience no real life consequences. This allows learners to make mistakes without being afraid of consequences, and they can integrate their acquired knowledge into practice without carrying responsibilities for their actions.

Games are an ideal medium for experimentation. Players can follow their custom problem solution strategies in a game, applying them in an authentic environment that can present the results of their strategies in various ways. Players do not have to suffer real-life consequences if they make mistakes, but can simply try again with a new strategy.

Experience. Experience is a key factor in experiential learning and transformative learning. Experiential learning is based on experience, as it argues that learners should make real-life experiences to learn. The assumption is based on the learning by doing principle. It is the step after experimentation, in which learners experience the consequences of their actions. In transformative learning theory, experiences confront individuals with their assumptions of how the world works, and can either strengthen points of view if the results of an experience support their assumptions, or challenge the assumptions if the result of an experience contradicts them.

The real-life experiences can be represented in serious games. It may be difficult in some cases to make an experience in a game as compelling and rich as if it would happen in real-life. On the other hand, no other medium can get as close in representing experiences than games do.

Game mechanics are the construction elements of the real-world representation, which are enriched with visuals, audio and narrative. How game mechanics can be used to confront players with their assumptions and thus lead to a change of attitudes in behaviour has been extensively discussed by Bogost (2007) in his book titled *Persuasive Games*, and by Frasca (2007) in his thesis *Play the Message*.

Reflection. Reflection is important in constructivist learning theories for several reasons. Experiential and transformative learning rely on reflection as a subsequent step to experience, to identify issues in problem solving strategies, and to critically review one's assumptions and habits. In transformative learning theory, a continuous disruption of expectations and resulting reflection can lead to a change in frames of references. But reflection is also important to transfer what has been learned in a situated environment to other environments, in order to generalize learning outcomes.

Reflection can be promoted by games either explicitly or implicitly. Implicitly, games can use unexpected events in the storyline, or in the perspective of rhetoric games, twists in the game's mechanic that come unexpected to players. Explicitly, games can directly address players to start a reflection process, or (as suggested by cognitive apprenticeship) bring them to explicitly articulate their knowledge to start reflection. The issue with explicit method to promote reflection is that they can be detrimental to the flow and immersion of the game, and might undermine players feeling of autonomy.

Feedback. From a constructivist perspective, feedback plays an important role in guiding players through their learning experiences. Feedback can help players to find an initial focus in discovery learning to support the development of problem solving strategies.

To guide players in learning processes in serious games, they can implement immediate feedback throughout the problem solving processes. That means that players should not only get to know if they are following a valid problem solution strategy when they fully applied it, as it is often the case in other learning environments, but they can get feedback on the way of applying a strategy and adapt it accordingly.

7.3.4 Games from a motivational perspective

Motivational benefits are often argued to be the most important factor in using games as learning tools (see section 3.1). The previous sections reported how learning theories can be applied in games. This section presents how principles of the most important motivational theories for games can influence serious game design.

Intrinsic motivation is something that is inherent in learners, but that can be diminished by external conditions. (Deci and Ryan, 1985) Considerations need to be taken to keep intrinsic motivation on the highest level possible, something that self-determination theory is concerned with. If activities are not perceived as self-determined, learners are more likely to slide into passivity.

Apart from the obvious benefit of motivational aspects in serious games to draw players to play serious games, there are also some direct benefits to learning efficiency. According to flow theory, the state of flow leads to a continuous growth process of skills and challenges.

To ensure keeping intrinsic motivation intact and enable a state of flow, serious games need to consider several aspects.

Clear goals. Malone's theory on intrinsic motivation emphasizes the importance of clear goals for the perception of challenge. If goals are not known from the beginning but emerging during gameplay, the creation of proximal goals with an appropriate level of challenge should be supported.

The principle of clear goals is generally important in game design, and it is usually considered bad game design if players ever feel that they do not know what to do in a game. Serious games can adopt this relative obsession with clarity of goals, but need to take special care that goals and proximal goals are related to the learning objectives.

Feedback. Feedback can also be viewed from a motivational perspective. Leaving players without any feedback of their actions can deter them from a game, because they may feel lost or even may get a feeling of lack of competence.

Feedback is important to let players know how far they are from reaching their goal. In serious games, this means that learners know if they are moving closer towards the learning objective. To support that, feedback should be constructive, so it should clearly communicate to learners what it is that they do right or wrong. However, it should be presented in a way that does not damage learners' self-esteem.

Ambiguous outcome. If it is clear to players that they cannot reach their goal anymore, or that they are sure to reach their goal, engagement can be lost as challenge ceases.

Games in general, and serious games in particular, can prevent this loss of engagement by offering several difficulty levels, so that players with higher capabilities can play on higher difficulty levels and do not lose interest because outcomes get certain too early. Difficulty levels can take different forms, they can either be freely chosen by players, or they can adapt automatically depending on the performance of players. Another way to prevent losing suspense is to include random elements, to keep a small chance of victory or loss independent of players performance. This element should be used with care, because random elements that cannot be balanced by players' skills can make games feel unfair. Finally, especially multiplayer games can implement mechanics to give advantages to weaker players, so that they do not get too far behind better players. A well known example for this is the "rubber-band" mechanic in racing games like Mario Kart 64.

Intrinsic fantasy. Fantasy is an important aspect of engagement in games, and is used extensively in entertainment games.

In terms of serious games, fantasy can fulfill even more purposes, as it can help integrate newly learned things into previous knowledge structures by using analogies.

One thing to consider in serious games is that they usually want to target a range of people with varying preferences regarding fantasy settings. That means that it might be dangerous to deeply base a serious game on a fantasy, if the target group of the serious game consists of different type of learners who do not share the preference for the same

fantasy setting. So it is important to weigh in which cases a fantasy might deter a group of learners from using a serious game.

Curiousity. Curiousity is considered to be a core aspect for intrinsic motivation in games, according to Malone (1981).

Serious games should include novel and surprising environments and events to raise learners' curiousity. But, a certain amount of predictability should be maintained, as challenging expectations too frequently could lead to disbelief in the game content. Additionally, serious games can assure to convey the feeling of incompleteness, inconsistency and unparsimoniousness to make learners explore deeper to fulfill their desire to accomplish these traits.

Challenge. A key component not only of learning theory perspectives, but also in the motivational perspective on serious games, is the right level of challenge.

Challenges in games can, as described for difficulty levels before, either be fixed, changed in a fixed learning curve, freely chosen by players, or adapt according to players' performance. The right level of challenge is crucial to motivation and engagement, as well as to reach flow. Challenge should be kept at a high level, to avoid boredom or apathy, but not too high, to avoid making players feel overwhelmed.

Distraction. To promote reaching flow state, distractions should be screened out as good as possible. External distractions can usually not be controlled by serious games, but distractions in the game can just as well prevent entering flow.

If a game constantly requires players' attention on information or tasks that are not interesting, not relevant, or tedious, it can distract them from achieving flow. This can lead to difficulties in serious game design, as instructional methods sometimes require intervention at certain moments, e.g. to initiate reflection on events or actions. These interventions need to be designed with care, and whenever possible should be part of the usual playing experience and not feel external to the game.

Autonomy. Autonomy is one of the three factors of motivation in self-determination theory. It is determined by the feeling of an activity being self-determined and not relying on external causality.

Serious games in certain circumstances have the problem of not being able to fulfill this need. If serious games are applied at school or in other formal education contexts, the "locus of causality" lies external to the learners, meaning that autonomy is not given.

Relatedness. Relatedness is also part of the three factors of motivation in self-determination theory. The principle states that relatedness can be improved by connecting an activity to actors that learners feel related to.

In a narrow sense, this applies to multiplayer games, or games which include an instructor in the learning process. But in a broad sense, relatedness could also be applied between the player and non-player characters in a game. This fact is used in game design already, specifically in narrative design in games, by creating characters that people feel related to. Players who feel connected to characters in a game get more engaged in the game's story and thus in the game itself.

Extrinsic motivation. Opposed to intrinsic motivation, extrinsic motivation has its locus of control outside of learners themselves. There are, according to self-determination theory, several types of extrinsic motivation, some of which share attributes with intrinsic motivation and are therefore worth striving for. The more assimilated to one self and congruent to one's values and needs external regulations are, the more they feel as self-determined and assume qualities of internal motivation.

For serious games, this means that even if motivation to play serious games feels external because they are prescribed at school or in another context, it can feel self-determined to play them if they are close enough to one's values and needs. They could be perceived as self-determined by people who like playing games, as long as they reach a quality that learners feel like they would play the game even in a voluntary context. Accordingly, aiming for high quality in the production of serious games to make them as engaging as possible can also contribute to their positive perception of students who have to play the game in a formal educational context.

7.4 Model for a new perspective on design and evaluation of serious games

To tackle the issues identified in serious games research, I propose a new systematic perspective on the research and development of serious games. The main step forward over existing models for serious games research is the consideration of the role of instructional methods in serious games, and of emergent dynamics of games. The model is presented in Figure 7.6.

Figure 7.7 and 7.8 depict how the proposed perspective differentiates from common approaches in serious games research.

7.4.1 Process for developing and evaluating serious games

Defining instructional objective. The instructional objective is the starting point in the model, and is explicitly defined by instructors or game designers. Based on the instructional objective and which learning dimension it is part of, learning theories can be derived that support achieving the objective.

Designing the serious game. Learning theories suggest instructional methods that should be put to use. Learning theories are not translated directly into instructional

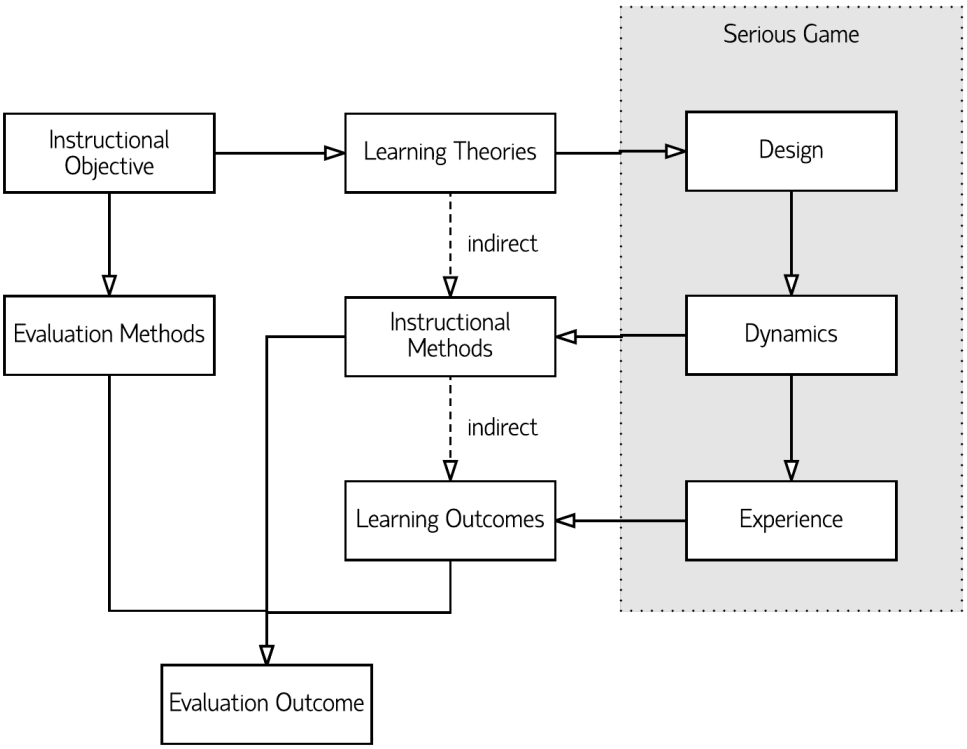

Figure 7.6: Comprehensive perspective on the design and evaluation of serious games.

methods, but are mediated through the instructional medium in use, games in this case. As explained in chapter 6, games have the peculiarity of emergent dynamics, which means that game designers cannot simply use a set of rules to translate learning theory principles into instructional methods in games. They need to design mechanics which lead to the desired instructional methods emerging from them.

Experiencing the serious game. The application of instructional methods leads to learning outcomes, which are again mediated through the learning medium, or serious game in this case. The player interacting with the game has an experience, which is created by game dynamics (and hence instructional methods) and the player's previous knowledge and experience. The interactive experience generates learning outcomes.

Evaluating the outcome. The last step is the evaluation of the learning outcomes. The instructional objective informs the choice of methods, because objectives from different learning dimensions afford different evaluation methods. The application of the evaluation method results in measurements, which need to be interpreted. The interpretation of the measurements in enabled by analyzing the implemented instructional methods in the

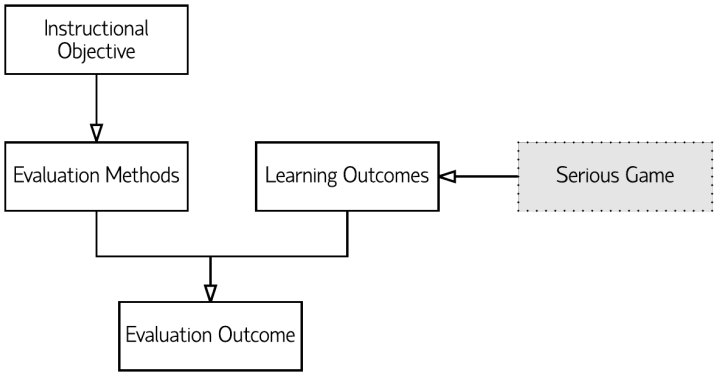

Figure 7.7: Common approach in serious game evaluation, not considering the role of learning theories and instructional methods.

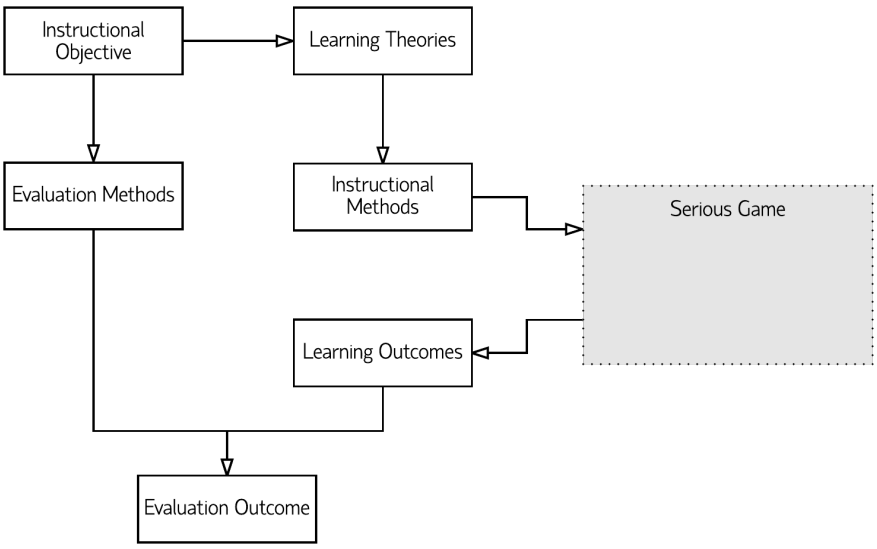

Figure 7.8: Common approach in serious game evaluation, not considering emergent game dynamics.

serious game. The consideration of this additional aspect leads to a concluding evaluation outcome.

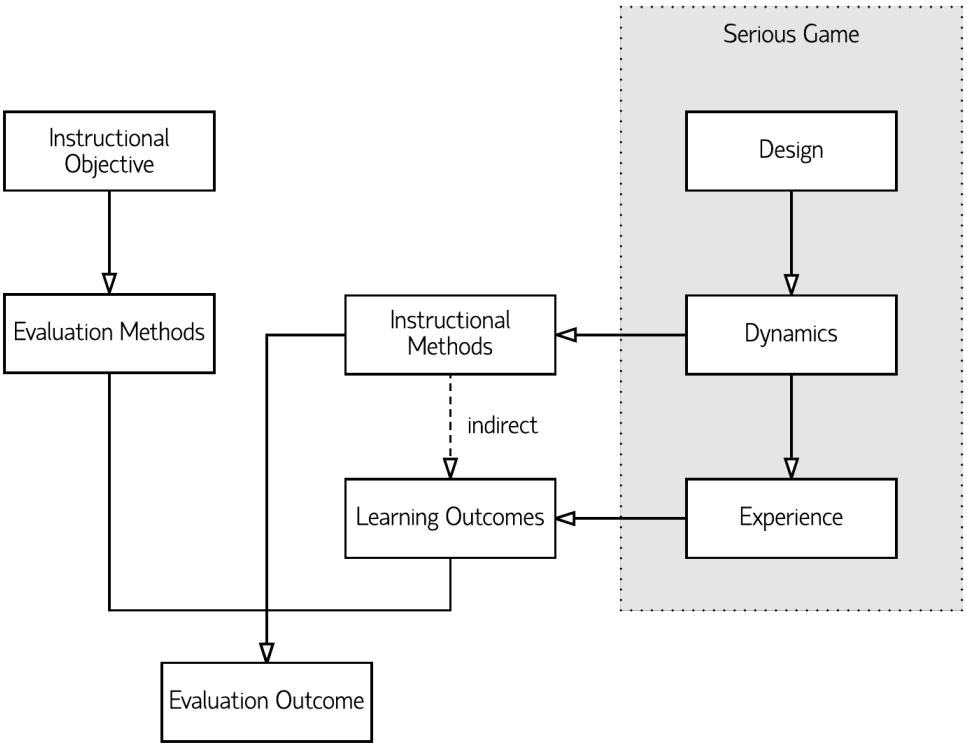

Figure 7.9: Model for evaluating existing serious games, without influence on the game.

7.4.2 Process for evaluating existing serious games

If the proposed perspective is applied in the analysis and evaluation of existing serious games, the process of designing the game is omitted. The design of the serious game and the resulting emergent instructional methods cannot be influenced, but instructional methods can be identified and inform the interpretation of evaluation results. (see Figure 7.9)

CHAPTER 8

Case studies

In this chapter, three games are analyzed using the proposed new perspective on design and evaluation of serious games.

The games EnerCities, Spent, and Portal 2 were selected, because all of them were evaluated in studies that are published in peer-reviewed journals.

The games were analyzed, following these guidelines:

1. Playing the game to the end at least once, ideally multiple times.
2. Creating a comprehensive description about the core mechanical features of the game.
3. Identifying the instructional objective of the game.
4. Analyzing the existing scientific evaluation of the game.
5. Analyzing the applied methodology in the evaluation, following section 4.4.
6. Using the guidelines from section 7.3 to identify instructional methods in the game, and to find weak spots in the game system related to the learning objectives.
7. Compiling the evaluation outcome by considering instructional objectives, measured learning outcomes, evaluation methodologies, identified instructional methods, and issues in game design.

8.1 EnerCities

EnerCities by Paladin Studios is a game about energy and sustainability, in which players build a city while balancing environment, economy and energy production. The

Figure 8.1: EnerCities is a city building game, in which players construct a sustainable, energy-efficient city.

game was developed in a project for the European Commission, and aims to improve young people's awareness of energy issues.

The goal in the game is to grow a city to a certain population. There are five stages of growth, in each one new buildings get unlocked for construction. Players can place buildings of different categories on tiles on the map:

- residential: increase city population
- economic: generate money
- environmental: increase environment level
- well-being: keep citizens happy
- energy: generate power

Players need to take care of several factors in the game:

- growing the population of the city

Figure 8.2: At the start of the game, players have a grid of empty tiles to place facilities on.

Figure 8.3: Facilities can be upgraded to improve their energy efficiency or reduce resource consumption.

- raising cash reserve, to be able to construct new buildings and upgrade existing ones
- produce enough energy to power all buildings
- minimizing use of natural resources (oil and gas), as they are limited and cannot be restocked
- keeping economy score positive, e.g. by building industry and commercial buildings
- keeping environment score positive, e.g. by building parks, forests
- keeping well-being of citizens positive, e.g. by building public and entertainment facilities

Buildings and areas can be upgraded to lower their environmental impact, energy consumption and resource demand. Possible upgrades include rainwater storages, bus stops, subway stations, solar roofs, and recycling facilities.

Each category of buildings includes three facilities that can be placed, depending on the current stage of growth. There is no real choice of what to build to follow a certain strategy, except for the energy category. The player can build plants for renewable energies, like windmills, solar panels, hydro power, and near the end of the game nuclear fusion, or generate environmentally inferior but economically more effective energy from coal or nuclear fusion.

The game finishes when a population of 200 is reached, and a final score is assigned by summing up the different scores from each growth stage.

Figure 8.4: When natural resources (oil and gas) run out, operations of facilities are drastically limited.

Figure 8.5: At the end of the game, points are awarded depending on how well players did on several metrics.

8.1.1 Instructional objective

The defined instructional objective of the game is to improve young people's awareness of energy issues. This relates to the affective learning dimension, as the goal is to change people's behaviour. Behaviour change should be supported by knowledge acquisition.

8.1.2 Learning outcome

Soekarjo and van Oostendorp (2015) conducted an evaluation of EnerCities. The game was chosen for evaluation because it has been evaluated before in a study without control condition reporting positive results, and because the game had won several awards which lead to the assumption that it is a well-designed persuasive game.

The study includes a slideshow as informative control condition, containing highly similar information to the game. 46 undergraduate students participated, and completed an online pre-test and post-test questionnaire on micro- and macro-level topics about energy consumption and production in the form of attitude and knowledge measures. The post-test included measures about participants' attitude towards the instructional medium. Participants were allowed to play the game for 20 minutes.

The study found no significant affects in micro-level or macro-level attitudes in neither of the conditions. However, on micro-level attitudes, measures suggest a trend for higher increase in the control condition, while on macro-level attitudes, the majority of measures improved more in the experimental condition.

Knowledge measures show increased knowledge for both conditions. The difference between the conditions are not significant, but show a trend in favour of the control condition. Additionally, the credibility of information was rated significantly higher in the control group. Participants that were playing the game did not find it significantly more engaging or enjoyable than participants reading the slideshow.

8.1.3 Methodological analysis

The study uses a media comparison method to evaluate the effect of EnerCities. The authors assume that the slideshow contains highly similar information, but they do not discuss differences between the two media in experimental and control condition related to instructional methods. Conclusions about the relative effectiveness of the media compared to each other is meaningless, without considering implemented instructional methods.

The study treats the game as a black box. Game features are not analyzed for their instructional quality. The only reference to the game's quality is the authors mentioning that the game won awards, which leads them to the conclusion that the game is of high quality.

The sampling of the study is questionable, as undergraduate students are presumably not the primary target group of the game. The baseline level of awareness of ecological themes can be expected to already be high on undergraduates. Also, the study did not control for participants' game playing habits.

The choice of a questionnaire to assess attitudes is vulnerable to biased results, because students presumably notice what is tried to assess, and may therefore tend to respond in line with researchers expectations.

The authors limited play time of EnerCitiesto 20 minutes, based on the average time it took participants in the control group to read through the slideshow. The potential consequences of this time limit are not mentioned in the study. It can happen that participants are not able to finish the game and therefore miss important parts of the experience. In contrast, all participants in the control group were allowed to read the slideshow to the end. Moreover, limiting the time of play may negatively influence motivation by reducing perceived autonomy.

8.1.4 Instructional methods

From a constructivist perspective, EnerCities provides a learning environment in a situated context. Players learn about advantages and disadvantages of different types of buildings and energy production, by building a city with these elements and getting feedback about the influence of each building on the city's environment, citizens' happiness, and economic values. Players can experiment by placing buildings, and experience the consequences. However, the feedback on the experimentation is very limited, as it is mostly only noticeable in numerical scores. The game environment itself does not provide any meaningful feedback, other than buildings being limited in their functionality if certain requirements are not fulfilled. There is no real opportunity for reflection about players' actions in the game. The scaffolding in the game is very rudimentary, in the form of an adviser who gives hints on urgent issues, but misses the opportunity to trigger deeper reflection.

From a cognitive perspective, the game does a good job in minimizing cognitive load. The elements in the game are all related to the topic that players are supposed to learn about. The controls of the game are approachable and keep cognitive overhead to a minimum. Information in the game is not very extensive, but the visuals support communication of information well. Sequencing of learning components is achieved by locking parts of the buildings at first, meaning that players can focus on a limited number of buildings, and get more options later in the game when they learned the basic mechanics.

In motivational terms, EnerCities struggles to provide engaging challenges. Progressing through the game is easy, and there is more than enough time to react to issues. No feeling of urgency emerges from the game, and the problems to solve are rather simple. If one of the variables that players need to balance slips in a dangerous level, it is easy to figure out which building or upgrades are required to fix the problem. The game does not provide interesting choices or foster curiousity in a notable way. And the lack of real challenge prevents getting into a state of flow.

A major issue of EnerCities is its shortcomings in procedural rhetoric. At the end of the tutorial, the player is confronted with the following questions:

> What type of EnerCities player are you? Are you a tycoon hungry for money or a true nature lover? Do you want the best for your population, or are you only interested in the latest technologies? How you play EnerCities is up to you!

The problem with these questions is, that they actually do not carry over into the game mechanics. The game tries to make players balance environment and economy, but in the game mechanics, it is easy to take care of both of them. Reaching the end of the game with a fully developed economy and flourishing environment is the default. This communicates the message that it is easy to take care of the environment and at the same time arbitrarily expand industry.

The second problem in the game's rhetoric are the natural resources. They are limited in the game, and players need to take care of decreasing use of natural resources like oil and gas, in order not to run out of them and suffer the consequences. Presumably, the intention is to communicate that humanity will run out of fossil fuels eventually, and should be prepared by having adapted economy and energy production accordingly. But, the consequences of running out of natural resources in the game are just that buildings work less efficient than before. This makes progress in the game difficult, but does not provide players with information about what running out of fossil fuels would mean in reality.

The third issue in the procedural rhetoric of EnerCities is its purely techno-optimist perspective. By advancing in the game and unlocking new buildings, players get the option to build super solar plants and wind mills, and eventually nuclear fusion. This conveys the message that, ultimately, the advancement of technology will safe the planet,

as these technologies (in the game) produce so much energy, making it easy to power the city without any harm to the environment.

Finally, the game surprisingly does not mention climate change in any way. Oil and gas consumption only influence the remaining stock of natural resources, but do not directly influence the environment or climate. In-game time starts in 2010, and players can advance to the year 2110 (or until they reach a population of 200). In all this time, there is no noticeable impact by climate change. This is a missed opportunity or even a counterproductive message, because EnerCities never conveys the urgency of the situation.

8.1.5 Evaluation outcomes

Combining evaluation outcomes by Soekarjo and van Oostendorp (2015) and game analysis provides a clearer pictures of EnerCities and its potential as a learning tool.

The information presented in the game on energy sources and ways of energy consumption is very basic, especially for undergraduate students. The procedural rhetorics of the game are faulty, in that they convey the message of easily manageable problems of energy production and usage by the development of new technologies. Climate change is not implemented in any way into the games mechanics, nor is it brought up as additional information. Dealing with problems in the game is no challenge, and accordingly playing the game is not very engaging.

These factors can explain the missing positive effect of EnerCities on attitudes towards energy-related issues according to the review by Soekarjo and van Oostendorp (2015).

8.2 Spent

Spent is a simple text based game commissioned by Durham Homeless Coalition, and developed by McKinney. Players take the role of a person who loses her home, and has to survive for one month.

Players start with 1,000$ and chose one of three jobs (Figure 8.7). They get weekly payment from the job, and are confronted with an important decision every day.

The first decisions are on health insurance (either higher monthly fee or higher co-pay) and housing (living closer to job is more expensive but saves transportation costs). From then on, every day a new decision on expenses or resolving of a dilemma has to be made. Decisions include:

- pay leasing rate of the car (losing the car means losing the job)
- pay old credit card loans
- pay student loans

Figure 8.6: Beginning of Spent.

Figure 8.7: Choose a job.

Figure 8.8: After making decisions, the game presents related facts and statistics about homelessness in the US.

- decide whether to fix a tooth
- buy groceries
- start smoking to relieve stress
- decide to take or deny a neighbours offer to take his coat
- decide what to do after character starts feeling depressed

Most of the events and decisions in the game are linked to some facts and statistics about poverty in the US, which are presented to the player after making a decision (see Figure

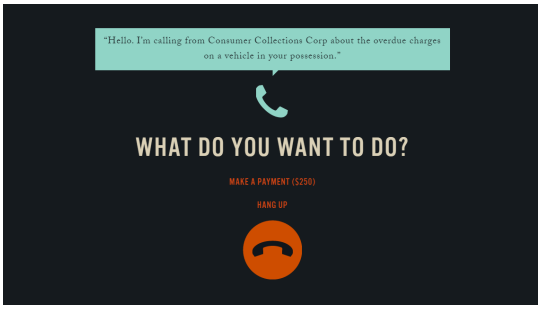

Figure 8.9: Unexpected expenses decrease your already low balance. Not paying can have serious consequences.

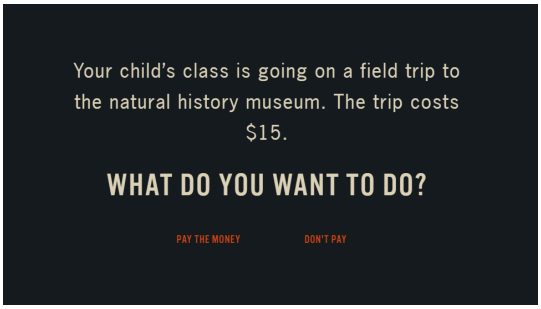

Figure 8.10: Decisions about financial contribution for social activities of your child have to be made.

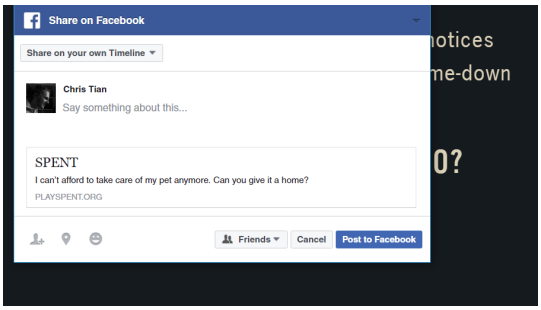

Figure 8.11: Some choices open up a Facebook popup to post a game related text on ones Facebook timeline.

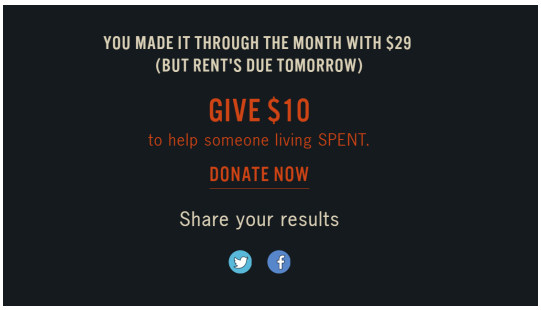

Figure 8.12: At the end of the game, SPENT calls players to action to spend for an organization aiding homeless people.

8.8). This creates a connection between the role of players in the game, and the situation of people who are living in relative poverty in real life.

The game confronts players with the difficulties to keep up their social life in a financially difficult situation. Friends can ask the game's protagonist to come to a free concert with them, but the protagonist would need to pay a babysitter for her child, which is most likely not in her budget.

Another dimension that has an important part in the game is the protagonist's ability to care for her child. For instance, not paying for the child to take place in a school play will not have direct negative consequences in the game, but conveys the bad feeling of letting the child down and not providing it with opportunities.

In certain situations, players can ask friends for help, e.g. to let them wash their cloths at their place, take their pet, or borrow them money. The friends always help out, but in these situations a browser window pops up, directed to the player's real Facebook page, prefilling the status update field with a message that asks their friends for help. Players can click this away and still get the money, but this twist successfully conveys the feeling of unease to admit being in a troubling financial situation to everyone in one's social

environment, and begging them for money.

If the player makes it through the month, they are presented with the info that rent is due the next day, which exceeds the budget left. The game ends prematurely if the balance drops to 0$ before the end of the month. On the final screen, the game shows a link to donate 10$ for "someone living SPENT" (Figure 8.12).

8.2.1 Instructional Objective

The game aims to change attitudes of people towards homelessness, by trying to make them understand what it means to suffer from a severe event that brings one in financial trouble, and how difficult it is to deal with the situation on different levels. It calls players to action at the end of the game, to donate for an organization aiding homeless people. Spent puts players in the situation of being close to homelessness, in order to change their attitude, as well as educate them with facts about homelessness in the US. Accordingly, the instructional objectives lie in the affective and knowledge acquisition domains.

8.2.2 Learning outcomes

Ruggiero (2015) conducted a quasi-experimental design to evaluate the impact of playing Spent, using one control and two treatment groups. The participants were 5,139 school students at the age of 12–18. Classes were randomly assigned to a game, reading or control group with pre-test and two post-tests, one immediately after treatment, another one three weeks after.

The *Affective Learning Scale (ALS)* was used to measure students' agreement with statements. The *Attitude Towards Homeless Inventory (ATHI)* is a questionnaire to measure knowledge and attitude in four dimensions of homelessness: personal causation, societal causation, affiliation, and solutions.

The reading group used an article by the '*Huffington Post*, written in first person. It portrays the experiences leading to and following living on the edge of financial ruin.

Both play group and reading group had approximately 30 minutes.

Playing Spent increased affective scores of students over control and reading group right after the treatment, and retained the effect three weeks after. Attitude towards homeless people did not change immediately after treatment in the game group, but was significantly more positive than reading and control group three weeks after the treatment.

8.2.3 Methodological analysis

The study uses a media comparison design, having a second treatment group reading a narrative on homelessness. Content-wise the conditions are similar as they both offer a first person perspective on homelessness. Nevertheless, the game uses procedural rhetorics

and therefore different instructional methods. Consequently, determining factors can only be inferred by analyzing the instructional methods.

The study controls for playing habits of participants, but not for preferences of genre or types of games.

A questionnaire is used to assess participants' attitude towards homelessness. The danger in using this methodology is that students may be aware of researchers' expectations and respond in line with what is expected from them.

8.2.4 Instructional methods

The game applies methods from a cognitive, motivational, and especially constructivist perspective. It employs procedural rhetoric to persuade players towards a desired attitude and behaviour.

From a constructivist point of view, Spent uses situated learning by placing players in the role of a person who lost her home and is now struggling to afford a flat and health care, keep her social life intact, and care for her child. Facts about related themes are presented to players after some decisions, making them more receptive to the statistics and facts after they just experienced a concrete situation about the same theme in the game. Feedback about players' actions are immediate and feel drastic, because they either push them closer towards becoming broke, or illustrate the impact on the character's social life and child. There is no real iterative experiential learning process, because impact of decisions is definitive, and the game is designed such that there is usually no optimal choice - each choice has its advantages and disadvantages and there is no way to know in advance which one is better in a certain situation. This would in most cases be considered bad game design, in the case of Spent it is part of the procedural rhetoric. It conveys to players that no matter what they do, they cannot be sure to make the right decisions, just as people in this situation in real-life constantly face difficult decisions with no "right" way to respond. Negative social consequences or neglecting one's child in the game does not have any consequences in the game mechanics - the mechanics are just about surviving with the money at hand. But by subtly using narration and facts, players feel bad to neglect the child and therefore do not necessarily make choices that are optimal from a game progression perspective.

In cognitive terms, the game keeps extraneous cognitive load small by using a very limited interaction model. Players are confronted with one decision at a time, can take as long as they want to make the decision, and usually have not more than four options. Visually, the game has a clean appearance that does not distract from the content.

From a motivational standpoint, Spent has the advantage of being a short experience and therefore does not have to tie players to the game for a long time. The game's outcome stays ambiguous until the very end, which maintains curiousity. On the other hand, the game is challenging in the sense that it is hard not to get broke. But it does not foster the feeling of competency through challenge, because the consequences of the decisions are too unpredictable.

Figure 8.13: The gap is too wide to reach the other platform by jumping.

Figure 8.14: Shooting a second portal on the wall connects the two portals.

Overall, the game's procedural rhetorics are its strength, because it highlights the issues of people in deep financial trouble. These issues might be hard to understand from an outside perspective, and Spent ambitiously tries to convey situations by confronting players with decisions that cannot be made right. The presented facts and statistics are informative, and their relation to a previous decision makes them seem more relevant.

8.2.5 Evaluation outcomes

According to the evaluation by Ruggiero (2015), Spent succeeds in its instructional objectives. The well implemented procedural rhetoric and constructivist learning principles can account for the advantage that players of Spent have over the control group and the group reading an article about homelessness.

8.3 Portal 2

Portal 2 is a 3D, first-person perspective, puzzle game, developed and published by Valve and released in 2011. It is the successor of the successful Portal, released in 2007.

The core mechanic of the game is the portal gun, with which players need to find their way through the levels of the game. The gun shoots projectiles, which, when hitting a solid surface, open a portal. There are two types of portals, a blue and an orange one. The player character can walk into one portal to come out from the other one. Physical force that acts on the character when entering a portal will be maintained when exiting through the second one, allowing for complex physics based puzzles.

The puzzles include only a couple of elements to form increasingly complex challenges. Metal cubes can be picked up by the player and moved through portals, and need to be placed on switches to open doors. Lasers have to be redirected with redirector cubes and through portals, to point them to receptors. Automatic turrets shoot on the character when she gets in their viewing range. Light beams - which players can walk on - need to be redirected through portals to overcome abysses or to stop lasers, metal cubes, and

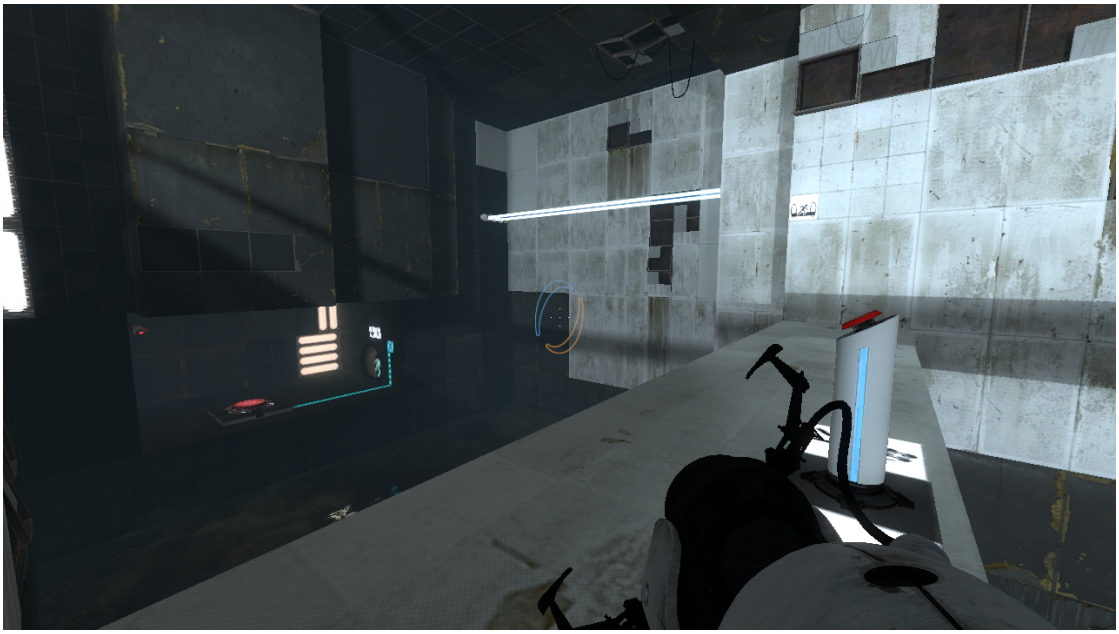

Figure 8.15: This is one of the test chambers, which require solving puzzles to reach the exit. The player tries to get an overview of the elements in the chamber to develop problem solving strategies.

Figure 8.16: Cubes need to be placed on switches to open doors.

Figure 8.17: Mechanics in the game get explained with little hints like this one.

turret shots. Blue gel that covers floors and walls lets objects and the character bump off to reach higher platforms.

Portal 2 is, just like it's predecessor, famous for its dark humor. The narrative is critically acclaimed to be unique in the medium of games, and exceeds competitors in originality. The character finds herself in a facility, in which her capabilities are assessed in so called *test chambers*. Each test chamber usually consists of one puzzle. The character is accompanied by the deeply sarcastic commentaries of an AI.

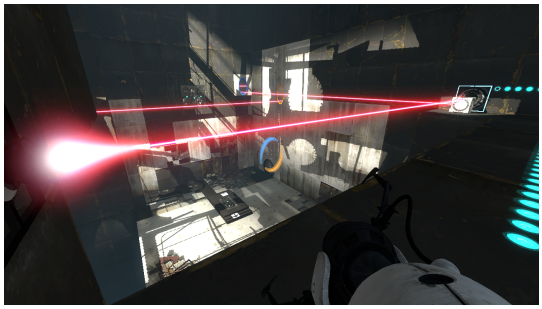

Figure 8.18: Lasers can be redirected, blocked, and directed though portals to reach receptors.

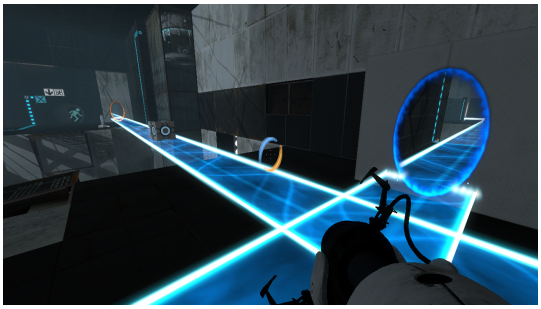

Figure 8.19: Light bridges are flexible bridges that can be directed through portals.

8.3.1 Instructional objective

Portal 2 is a commercial game designed for entertainment, not for instructional use. Therefore, there is no predefined instructional objective. The objective depends on the context in which the game is applied to generate a learning outcome.

Instructional objectives for this analysis will be defined in line with objectives in the evaluation study by Shute et al. (2015). They research the effect of the game on problem solving skills, spatial skills, and persistence.

Problem solving is a complex cognitive skill, and according to Shute et al. (2015) consists of rule application, problem decomposition, flexibility, and resource management. Spatial skills are lower-level cognitive skills, and can be distinguished into figural, vista, and environmental spatial skills. Persistence describes the dispositional need to complete difficult tasks with high performance, that pertains in frustrating moments. Consequently, persistence can be viewed as part of the motivational learning dimension.

8.3.2 Learning outcome

Shute et al. (2015) conducted a study measuring the impact of playing Portal 2 and the dedicated learning game Lumosity on impact on problem solving, spatial skills, and persistence. They hypothesized that Portal 2 would get equal or higher outcome scores than Lumosity, despite not being explicitly designed for learning.

77 participants, all students on a major US research university, were randomly assigned to either playing Portal 2 or Lumosity. Each participant played for 8 hours, plus completing tests for 2 hours, over four sessions in 1–2 weeks. Before the first play session, participants completed a pre-test, and after the last session they completed a post-test. Additionally, in-game metrics were collected from both games. In the case of Portal 2, they consisted of total number of levels completed, average portals shot, and average time to complete levels.

For all three instructional objectives measured, the Portal 2 group outperformed the Lumosity group. Within the Portal 2 group, results only show a non-significant trend of improvements in problem solving skills. The hypothesis of improvements on small- and large-scale spatial skills was partially supported in the Portal 2 group. Participants playing Portal 2 reported significantly higher enjoyment ratings.. Additionally, in-game measures predicted post-test scores in several different tests, suggesting that successful learning can be inferred from in-game performance.

8.3.3 Methodological analysis

The study compares the effects of two different games, which implement different instructional methods. Finding determining factors on learning outcomes requires an analysis on the instructional methods used in the games.

The study controls for previous game playing experience by excluding frequent players from the study. However, the study does not analyze cognitive overload weighing on players to learn the controls of Portal 2.

8.3.4 Instructional methods

Portal 2 strongly employs cognitive, constructivist, and motivational learning principles.

The game uses experiential learning principles at its core. After arriving in a chamber, the first thing to do is to get an overview of the situation and all elements that could be relevant for the puzzle. After identifying the problem, previously acquired skills can be used to conceptualize a possible solution for the problem. This strategy can be executed in active experimentation. If the attempt fails, players can observe what happened and construct a new conceptualization of the problem. If the character dies in the experimentation, she gets set right back in the chamber and experimentation can continue without an additional penalty.

In the first levels, the portal gun is restricted to shoot one side of the portal while the other side is fixed in the chamber. The players can slowly learn the ways in which the portals can be used. Puzzles are on a low difficulty level, solely involving metal cubes and simple combinations of elements. Later levels introduce new elements, and integrate them with the previous ones.

Visual clues guide players in finding ways to solve puzzles. They help to identify how elements can be combined. These hints are not obvious to players, but are subtly used in the game to direct them without taking away their feeling of competence by overtly holding their hand.

Portal 2 has an outstanding narrative, and manages to foster players' curiousity for the events to come.

The level of challenge in Portal 2 is fixed and cannot be changed by players. However, players can take as much time as they want in each test chamber. This lack of time

pressure is important for players who are either not used to the control scheme of first person 3d games, or take longer to come up with solutions for the problems. There is no penalization whatsoever for taking longer or having to restart a chamber. Some specific outstanding or creative solutions are rewarded with achievements batches.

The puzzles stay mentally challenging, by continuously increasing difficulty and introducing new elements. Combined with the thrilling atmosphere and short levels which lead to quick achievements and rewards in form of progression in the narrative, all prerequisites to reach flow are fulfilled.

8.3.5 Evaluation outcomes

The high quality properties of Portal 2 match with the measured learning outcomes by Shute et al. (2015). The game successfully introduces learners to new concepts and confronts players with difficult problems without pressuring them. If a player is not able to come up with a solution by conceptualizing the problem, he or she can learn how to solve it by continuous experimentation with the available elements. The experience made in experimentation can then help to understand the problem better and to come up with solution strategies. These factors make it motivating to continue searching for solutions, even if a problem cannot be solved right away. This results in improved persistence.

The improvement of spatial skills through playing Portal 2 is supported by the repeating occurrence of spatial and physics puzzles, which are accompanied by a reinforcement schedule consisting of visual rewards and progression in the story.

Evaluation results show high engagement of study participants when playing Portal 2. The evaluation took place in a laboratory setting and showed successful results. However, Portal 2 is not restricted to classroom use, but is played for entertainment because of its engaging qualities. In fact, millions of people played the game at home. Considering that they on average improved their problem solving skills, spatial skills, and persistence, suggests a massive overall impact by the game.

CHAPTER 9

Conclusion and future work

This chapter summarizes the core contributions of this thesis and suggests future work to further build on these insights.

9.1 Conclusion

Serious game are in a period of massive growth, but they still did not arrive in mainstream education and are not accepted thoroughly as well-invested time. It is crucial to develop research on serious games, to overcome the issue of highly inconsistent evaluation outcomes.

This thesis provided an overview about the state of the art in the evaluation of serious games, and expounded why a new systematic perspective on the development and evaluation of serious games is required. A model has been proposed which represents the suggested new perspective, and it has been analyzed how previous models and perspectives diverge from it.

The first contribution of this thesis is that it highlights the issues of media comparison studies, which are widely used in the evaluation of serious games. The critique is based on Clark (1983), and argues that learning outcomes are determined by instructional methods, and not by media. Instructional methods can be implemented in different media, but certain media may afford specific instructional methods more than others. However, when comparing measured learning outcomes in different media, it must be considered which instructional methods are implemented in them. The effect of a medium itself is mediated by its instructional methods, but this is not controlled for in media comparison studies.

A second contribution is the examination of the applicability of study designs in serious games research, specifically of RCTs. The particular focus on RCTs rests on the common trait to predestine them to the study design of choice. (Connolly et al., 2012) But,

as a matter of fact, scholars promoting the use of RCTs in serious game research do not question its usage in this particular domain, and do not consider critique on the use of RCTs in educational interventions, e.g. by Norman (2010) and Sullivan (2011). Consequently, the use of RCTs in the evaluation of serious games should not be the default choice, but each study should reexamine which study design is appropriate in the given case.

The third contribution makes interrelations between learning theory and game design theory explicit. In previous models about serious games, the peculiarity of emergent dynamics in games has been disregarded. In the proposed perspective, game dynamics are highlighted in the process of creating and analyzing serious games, by stressing the fact that learning principles derived from learning theory cannot be translated to the medium of games as instructional methods by following a set of fixed guidelines. Game developers can only directly influence the design layer of the game, which in turn enable dynamics to emerge that encompass the desired instructional methods.

In general, this thesis is supposed to lay groundwork for a new and comprehensive scientific perspective on serious games, by closely incorporating game design theory and questioning common evaluation methodologies on their adaptability to serious games research. This resulted in the proposition of a new model and its application in case studies. The informal application revealed the potential of the model to explain contradicting results from learning outcome measurements and to elicit suggestions to improve the design of the game to better meet the instructional objectives.

9.2 Future work

The proposed model for a perspective on serious games is only to be seen as a first step.

This work analyzed issues in the evaluation of serious games from a conceptual and methodological perspective. However, it focused on finding a new overarching perspective on how to structurally approach evaluation. The concrete choice of evaluation methodology, informed by the instructional objective, should be elaborated further. An alternative that would be worth studying in the context of serious games is so called *realist evaluation*. (Pawson and Tilley, 1997) In terms of data gathering, Shute and Ke (2012) offer an interesting approach of *Evaluation Centered Design* in games. Guidelines to help choose study design and overall methodology should help researchers to pick the right evaluation method for each case and improve comparability of results.

The application of learning theory principles to games presented in this thesis is not exhaustive, and could be structured more compactly to ease practical application. Additionally, the principle of *game grammar* and other game design theoretical views could be used to improve the guidelines.

The proposed model should be applied in more case studies, to get more experience with it and to find its weak spots.

The process of analyzing games in the case studies was structured by guidelines, but only conducted by one person (the author). To achieve more reliable results, several experts should analyze the game under evaluation, and suitable methods should be defined to synthesize these multiple analyses into a more comprehensive meta-analysis.

In the thesis, the model has only been applied in the analysis of existing serious games. In future work, it would be interesting to apply it in a full cycle of designing, developing, evaluating and iterating a serious game.

List of Figures

List of Tables

List of Games

Code Red: Triage (2009). Developed by *Institute of Information and Computing Sciences, Utrecht University*.

Enercities (2010). Developed by *Paladin Studios*. Commissioned by the *European Commission*.

EverQuest Online Adventures (2003). Developed by *Sony Online Entertainment*. Published by *Sony Online Entertainment*.

Frequency 1550 (2005). Developed by *The Waag Society*.

Halo: Combat Evolved (2001). Developed by *Bungie*. Published by *Microsoft Game Studios*.

Lumosity (2007). Developed by *Lumos Labs*. Published by *Lumos Labs*.

Mario Kart 64 (1996). Developed by *Nintendo EAD*. Published by *Nintendo*.

Portal (2007). Developed by *Valve Corporation*. Published by *Valve Corporation*.

Portal 2 (2011). Developed by *Valve Corporation*. Published by *Valve Corporation*.

Quest Atlantis: Taiga Park (1999). Developed by *Center for Research on Learning and Technology, Indiana University Bloomington*.

SimCity Societies (2007). Developed by *Tilted Mill Entertainment*. Published by *Electronic Arts*.

Spent (2011). Developed by *McKinney*. Commissioned by *Urban Ministries of Durham*.

Tetris (1984). Originally developed by *Alexey Pajitnov*.

The Legend of Zelda: Ocarina of Time (1998). Developed by *Nintendo EAD*. Published by *Nintendo*.

Tycoon City: New York (2006). Developed by *Deep Red Games*. Published by *Atari*.

Bibliography

Abt, C. C. (1970). *Serious Games.* Viking Press.

Abt, C. C. (1987). Serious Games. *Serious games. Mechanisms and effects*, 48–61.

Adams, D. M., R. E. Mayer, A. MacNamara, A. Koenig, and R. Wainess (2012). Narrative Games for Learning: Testing the Discovery and Narrative Hypotheses. *Journal of Educational Psychology 104*(1), 235–249.

Arnab, S., S. D. Freitas, F. Bellotti, T. Lim, S. Louchart, N. Suttie, R. Berta, and A. De Gloria (2012). Pedagogy-driven design of Serious Games: An overall view on learning and game mechanics mapping , and cognition-based models. Technical report, Research Report, Serious Games Institute.

Arnab, S., T. Lim, M. B. Carvalho, F. Bellotti, S. D. Freitas, S. Louchart, N. Suttie, R. Berta, and A. De Gloria (2016). Mapping Learning and Game Mechanics for Serious Games Analysis. *IEEE Transactions on Emerging Topics in Computing PP*(99), 1.

Backlund, P. and M. Hendrix (2013). Educational Games: Are They Worth the Effort? In *Proceedings of the 5th International Conference on Games and Virtual Worlds for Serious Applications (VS-Games 2013)*, Number December.

Baker, E. L. and R. E. Mayer (1999). Computer-based assessment of problem solving. *Computers in Human Behavior 15*(3), 269–282.

Bandura, A. (1977). *Social learning theory.*

Bartle, R. (1996). Hearts, Clubs, Diamonds, Spades: Players Who Suit MUDs. *Journal of MUD Research.*

Becker, K. (2010). Distinctions Between Games and Learning: A Review of Current Literature. *Gaming and Cognition: Theories and Practice from the Learning Sciences: Theories and Practice from the Learning Sciences*, 22.

Bellotti, F., B. Kapralos, K. Lee, P. Moreno-Ger, and R. Berta (2013). Assessment in and of serious games: An overview. *Advances in Human-Computer Interaction 2013.*

Bloom, B. S. (1956). Taxonomy of Educational Objectives. *Educational and Psychological Measurement 16*, 401–405.

Bogost, I. (2007). *Persuasive Games: The Expressive Power of Videogames*. MIT Press.

Boyle, E., T. M. Connolly, and T. Hainey (2011, jan). The role of psychology in understanding the impact of computer games. *Entertainment Computing 2*(2), 69–74.

Boyle, E. A., T. Hainey, T. M. Connolly, G. Gray, J. Earp, M. Ott, T. Lim, M. Ninaus, C. Ribeiro, and J. Pereira (2016). An update to the systematic literature review of empirical evidence of the impacts and outcomes of computer games and serious games. *Computers and Education 94*, 178–192.

Breuer, J. S. and G. Bente (2010, jan). Why so serious? On the relation of serious games and learning. *Eludamos. Journal for Computer Game Culture*.

Brown, J. S., A. Collins, and P. Duguid (1989). Situated Cognition and the Culture of Learning. *Educational Researcher 18*(1), 32–42.

Bruckman, A. (1999). Can Educational Be Fun? In *Game Developer's Conference*, pp. 75–79.

Bruner, J. (1961). The Act of Discovery. *Harvard Educational Review 31*, 21–32.

BusinessTech (2015). The biggest entertainment markets in the world.

Caillois, R. and M. Barash (1961). *Man, Play, and Games*.

Chen, S. and D. Michael (2005). Proof of Learning : Assessment in Serious Games.

Clark, R. (1994). Media will never influence learning. *Educational technology research and development 42*(1), 21–29.

Clark, R. E. (1983). Reconsidering Research on Learning from Media. *Review of Educational Research 53*(4), 445–459.

Connolly, T., M. Stansfield, and T. Hainey (2008). Development of a General Framework for Evaluating Games-Based Learning. In *Proceedings of the 2nd European conference on games-based learning*, pp. 105–114.

Connolly, T., M. Stansfield, and T. Hainey (2009, jan). Towards the development of a games-based learning evaluation framework. In *Games-Based Learning Advancements for Multi-Sensory Human Computer Interfaces: Techniques and Effective Practices*, pp. 251–273.

Connolly, T. M., E. A. Boyle, E. MacArthur, T. Hainey, and J. M. Boyle (2012, jan). A systematic literature review of empirical evidence on computer games and serious games. *Computers & Education 59*(2), 661.

Crawford, C. (1984). *The Art of Computer Game Design.* Osborne/McGraw-Hill.

Csíkszentmihályi, M. (1975). *Beyond boredom and anxiety.*

Csíkszentmihályi, M. (1997). Finding flow. *Psychology Today 30*(4), 46–71.

De Freitas, S., G. Rebolledo-Mendez, F. Liarokapis, G. Magoulas, and A. Poulovassilis (2010). Learning as immersive experiences: Using the four-dimensional framework for designing and evaluating immersive learning experiences in a virtual world. *British Journal of Educational Technology 41*(1), 69–85.

Deci, E. L. and R. M. Ryan (1985). The general causality orientations scale: Self-determination in personality. *Journal of Research in Personality 19*(2), 109–134.

Dempsey, J. V., K. Rasmussen, and B. Lucassen (1994). Instructional gaming: Implications for instructional technology. *Annual Meeting of the Association for Educational Communications and Technology*, 1–21.

Dettmer, P. (2005). New blooms in established fields: Four domains of learning and doing. *Roeper Review 28*(2), 70–78.

Djaouti, D., J. Alvarez, and J. P. Jessel (2008, jan). Play, game, world: anatomy of a videogame. *International Journal of Intelligent Games & Simulation 5*(1).

Ducheneaut, N. and R. J. Moore (2005). More than just 'XP': learning social skills in massively multiplayer online games. (2), 89–100.

Egenfeldt-Nielsen, S. (2006). Overview of research on the educational use of video games. *Digital Kompetanse 1*(3), 184–213.

Emmerich, K. and M. Bockholt (2016). Serious Games Evaluation: Processes, Models, and Concepts. In *Entertainment Computing and Serious Games*, Volume 1, pp. 265–283.

Engelmann, S. (1968). Relating operant techniques to programming and teaching. *Journal of School Psychology 6*(2), 89–96.

Erhel, S. and E. Jamet (2013, jan). Digital game-based learning: Impact of instructions and feedback on motivation and learning effectiveness. *Computers & Education 67*, 156.

Frasca, G. (2007). Play the Message: Play, Game and Videogame Rhetoric. *Unpublished PhD dissertation. IT University of Copenhagen, Denmark.*

Gagné, R. M. (1972). Domains of learning. *Interchange 3*(1), 1–8.

Gagné, R. M. (1984). Learning outcomes and their effects: Useful categories of human performance. *American Psychologist 39*(4), 377–385.

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

Gee, J. P. (2003). What video games have to teach us about learning and literacy. *Computers in Entertainment 1*(1), 20.

Goldsmith, T. E., P. J. Johnson, and W. H. Acton (1991). Assessing structural knowledge. *Journal of Educational Psychology 83*(1), 88–96.

Gunter, G. A., R. F. Kenny, and E. H. Vick (2008). Taking educational games seriously: using the RETAIN model to design endogenous fantasy into standalone educational games. *Educational Technology Research and Development 56*(5-6), 511–537.

Hamari, J., D. J. Shernoff, E. Rowe, B. Coller, J. Asbell-Clarke, and T. Edwards (2016). Challenging games help students learn: An empirical study on engagement, flow and immersion in game-based learning. *Computers in Human Behavior 54*, 170–179.

Hauge, J. B., R. Nadolski, P. Moreno-ger, F. Bellotti, T. Lim, and J. Ritchie (2013). Study Design and Data Gathering Guide for Serious Games ' Evaluation. In *Psychology, Pedagogy, and Assessment in Serious Games*, pp. 394–419.

Hays, R. T. (2005). The effectiveness of instructional games: a literature review and discussion. *Naval Air Warfare Center Training Systems Division*, 1–63.

Hoffman, B. and L. Nadelson (2010). Motivational engagement and video gaming: A mixed methods study. *Educational Technology Research and Development 58*, 245–270.

Huizenga, J., W. Admiraal, S. Akkerman, and G. Ten Dam (2009). Mobile game-based learning in secondary education: engagement, motivation and learning in a mobile city game: Original article. *Journal of Computer Assisted Learning 25*(4), 332–344.

Huizinga, J. (1955). *Homo Ludens: A Study of the Play Element in Culture*. Beacon Press.

Hunicke, R., M. LeBlanc, and R. Zubek (2004, jan). MDA: A formal approach to game design and game research. *Proceedings of the AAAI Workshop on Challenges in Game AI 4*(1).

Ipsos MediaCT & ISFE (2012). Videogames in Europe: Consumer Study. Technical report, Ipsos MediaCT & ISFE.

Jantke, K. P. and S. Gaudl (2010, jan). Taxonomic contributions to digital games science. In *2010 2nd International IEEE Consumer Electronics Society's Games Innovations Conference (ICE-GIC 2010)*, pp. 1–8.

Jennett, C., A. L. Cox, P. Cairns, and S. Dhoparee (2008, jan). Measuring and defining the experience of immersion in games. *International journal of human-computer studies 66*(9), 641–661.

Juul, J. (2005). *Half-real: Video Games Between Real Rules and Fictional Worlds*. MIT Press.

Ke, F. (2009, jan). A qualitative meta-analysis of computer games as learning tools. *Handbook of research on effective electronic gaming in education 1*, 1–32.

Kebritchi, M. and A. Hirumi (2008). Examining the pedagogical foundations of modern educational computer games. *Computers and Education 51*(4), 1729–1743.

Kelly, H., K. Howell, E. Glinert, L. Holding, C. Swain, A. Burrowbridge, and M. Roper (2007). How to bulid serious games. *Communications of the ACM 50*(7), 44–49.

Kinder, D. and D. Carnine (1991). Direct instruction: What it is and what it is becoming. *Journal of Behavioral Education 1*(2), 193–213.

Kirschner, P. A., J. Sweller, and R. E. Clark (2006). Why Minimal Guidance During Instruction Does Not Work. *Educational Psychologist 41*(March 2015), 87–98.

Kolb, D. A. (1984). Experiential Learning: Experience as The Source of Learning and Development. *Prentice Hall, Inc.* (1984), 20–38.

Kraiger, K., K. J. Ford, and E. Salas (1993). Application of cognitive, skill-based, and affective theories of learning outcomes to new methods of training evaluation. *Journal of Applied Psychology 78*(2), 311–328.

Laamarti, F., M. Eid, and A. El Saddik (2014, jan). An Overview of Serious Games. *International Journal of Computer Games Technology 2014*, 1.

Lantz, F. (2015). MDA.

Lave, J. (1991). Situating learning in communities of practice. In *Perspectives on socially shared cognition*, pp. 63–82.

Lieberman, D. A. (2006). What can we learn from playing interactive games? *Playing video games: Motives, responses, and consequences*, 379–397.

Linehan, C., B. Kirman, S. Lawson, and G. G. Chan (2011). Practical, appropriate, empirically-validated guidelines for designing educational games. *CHI '11 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (July 2015), 1979–1988.

Loh, C. S., Y. Sheng, and D. Ifenthaler (2015). Serious Games Analytics: Theoretical Framework. In *Serious games analytics*, Number November, pp. 3–29.

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

Maslow, A. (1943). A theory of human motivation. *Psychological Review 50*(4), 370–396.

Mayer, I. (2012, jan). Towards a Comprehensive Methodology for the Research and Evaluation of Serious Games. *Procedia Computer Science 15*, 233–247.

Mayer, I., G. Bekebrede, C. Harteveld, H. Warmelink, Q. Zhou, T. van Ruijven, J. Lo, R. Kortmann, and I. Wenzler (2014). The research and evaluation of serious games: Toward a comprehensive methodology. *British Journal of Educational Technology 45*(3), 502–527.

Mayer, I. S., L. Carton, M. de Jong, M. Leijten, and E. Dammers (2004). Gaming the future of an urban network. *Futures 36*(3), 311–333.

Mayer, R. E. (2003). The promise of multimedia learning: Using the same instructional design methods across different media. *Learning and Instruction 13*(2), 125–139.

Mayer, R. E. (2004). Should there be a three-strike rule against pure discovery learning? *American Psychologist 59*(1), 14–19.

Mayer, R. E. (2014). Introduction: Taking an Evidence-Based Approach to Games for Learning. In *Computer Games for Learning: An Evidence-Based Approach*, pp. 3–24.

Mayer, R. E. and R. B. Anderson (1992). The instructive animation: Helping students build connections between words and pictures in multimedia learning. *Journal of Educational Psychology 84*(4), 444–452.

Mayer, R. E. and C. I. Johnson (2010). Adding Instructional Features that Promote Learning in a Game-Like Environment. *Journal of Educational Computing Research 42*(3), 241–265.

McGonigal, J. (2011). *Reality is Broken: Why Games Make Us Better and How They Can Change the World*. Penguin.

Mezirow, J. (1997). Transformative Learning: Theory to Practice. *New Directions for Adult and Continuing Education 1997*(74), 5–12.

Michael, D. R. and S. L. Chen (2005). *Serious Games: Games That Educate, Train and Inform*. Thomson Course Technology.

Michelsen, E., J. F. Zaff, and E. Hair (2002). Civic engagement programs and youth development: A synthesis. Technical report, Child Trends.

Mitgutsch, K. and N. Alvarado (2012, jan). Purposeful by design?: a serious game design assessment framework. In *Proceedings of the International Conference on the foundations of digital games*, pp. 121–128.

Nakamura, J. and M. Csíkszentmihályi (2002). The Concept of Flow. In *Handbook of positive psychology*, pp. 89–105.

Norman, G. (2010). Is experimental research passé? *Advances in Health Sciences Education 15*(3), 297–301.

O'Neil, H. F., R. Wainess, and E. L. Baker (2005). Classification of learning outcomes: evidence from the computer games literature. *The Curriculum Journal 16*(4), 455–474.

Orvis, K. A., D. B. Horn, and J. Belanich (2008, jan). The roles of task difficulty and prior videogame experience on performance and motivation in instructional videogames. *Computers in Human Behavior 24*(5), 2415–2433.

Paas, F., A. Renkl, and J. Sweller (2004). Cognitive Load Theory: Instructional Implications of the Interaction between Information Structures and Cognitive Architecture. *Learning and Instruction 32*(1-2), 1–8.

Paas, F., J. Tuovinen, H. Tabbers, and P. W. M. Van Gerven (2010). Cognitive Load Measurement as a Means to Advance Cognitive Load Theory. *Educational Psychologist 38*(1), 43–52.

Paas, F. and J. J. G. van Merrienboer (1994). Instructional control of cognitive load in the training of complex cognitive tasks. *Educational psychology review 6*(4), 351–371.

Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal Of Psychology 45*(3), 255–287.

Pavlov, I. P. and G. Anrep (1927). *Conditioned reflexes; an investigation of the physiological activity of the cerebral cortex*. Oxford University Press.

Pawson, R. and N. Tilley (1997). *Simple Principles for The Evaluation of Complex Programmes*. SAGE.

Plass, J. L., B. D. Homer, and C. K. Kinzer (2014). Playful Learning: An Integrated Design Framework. Technical report, Games For Learning Institute.

Plass, J. L., B. D. Homer, and C. K. Kinzer (2015). Foundations of game-based learning. *Educational Psychologist 50*(4), 258–283.

Prensky, M. (2003). Digital Game-based Learning. *Computers in Entertainment 1*(1), 21.

Prensky, M. (2007). *Digital Game-Based Learning*. Paragon House.

Przybylski, A. K., C. S. Rigby, and R. M. Ryan (2010). A motivational model of video game engagement. *Review of General Psychology 14*(2), 154.

Reigeluth, C. M. (1992). Elaborating the elaboration theory. *Educational Technology Research and Development 40*(3), 80–86.

Romero, M., M. Usart, and M. Ott (2015). Can Serious Games Contribute to Developing and Sustaining 21st Century Skills? *Games and Culture 10*(2), 148–177.

Ruggiero, D. (2015). The effect of a persuasive social impact game on affective learning and attitude. *Computers in Human Behavior 45*, 213–221.

Ryan, R. and E. Deci (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *The American psychologist 55*(1), 68–78.

Ryan, R. M., C. S. Rigby, and A. Przybylski (2006). The motivational pull of video games: A self-determination theory approach. *Motivation and Emotion 30*(4), 347–363.

Sawyer, B. (2003). Serious Games: Improving Public Policy through Game-based Learning and Simulation. Technical report, Woodrow Wilson International Center for Scholars.

Shaffer, D. W., K. R. Squire, R. Halverson, and J. P. Gee (2005). Video Games and the Future of Learning. *Phi Delta Kappan 87*, 105–111.

Shute, V. J. (2010). Conceptual Framework for Modeling, Assessing and Supporting Competencies within Game Environments. *Technology, Instruction, Cognition, and Learning 8*(2), 137–161.

Shute, V. J. (2011). Stealth Assessment in Computer-Based Games To Support Learning. *Computer Games and Instruction*, 503–524.

Shute, V. J., S. D'Mello, R. Baker, K. Cho, N. Bosch, J. Ocumpaugh, M. Ventura, and V. Almeda (2015). Modeling how incoming knowledge, persistence, affective states, and in-game progress influence student learning from an educational game. *Computers and Education 86*, 224–235.

Shute, V. J. and F. Ke (2012). Games, Learning, and Assessment. In *Assessment in game-based learning*, pp. 43–58.

Shute, V. J., M. Ventura, and F. Ke (2015). The power of play: The effects of Portal 2 and Lumosity on cognitive and noncognitive skills. *Computers and Education 80*, 58–67.

Skinner, B. F. (1963). Operant behavior. *American Psychologist 18*(8), 503–15.

Soekarjo, M. and H. van Oostendorp (2015). Measuring effectiveness of persuasive games using an informative control condition. *International Journal of Serious Games 2*(2), 37–56.

Squire, K. (2003). Video games in education. *International Journal of Intelligent Simulations and Gaming 2*, 49–62.

Squire, K. (2011). The Future of Games for Learning. *Video Games and Learning: Teaching and Participatory Culture in the Digital Age. Technology, Education–Connections (the TEC Series)*, 356–376.

Steinkuehler, C. A. (2004). Learning in Massively Multiplayer Online Games. In *ICLS '04 Proceedings of the 6th international conference on Learning sciences*, Number 608, pp. 521–528.

Suits, B. (1967). What Is a Game? *Philosophy of Science 34*(2), 148–156.

Sullivan, G. M. (2011). Getting off the "gold standard": randomized controlled trials and education research. *Journal of graduate medical education 3*(3), 285–289.

Sweller, J. (1998). Cognitive load during problem solving: Effects on learning. *Cognitive Science 12*(2), 257–285.

Thorndike, E. L. (1913). *The Original Nature of Man*. Teachers College, Columbia University.

Van der Spek, E. D. (2012). Effects of Playing a Serious Game: A Comparison of Different Cognitive and Affective Measures. In *Proceedings of Measuring Behavior 2012*, pp. 227–230.

Vygotsky, L. (1978). Interaction between learning and development. *Interaction between learning and development 23*(3), 34–41.

Walk, W. (2015). From MDA to DDE.

Watson, J. B. (1913). Psychology as the behaviourist views it. *Psychological Review 20*, 158–177.

Wenger, E. (1998). Community of Practice: a Brief Introduction. *Learning in doing 15*(4), 1–7.

Wilson, B. and P. Cole (1992). A critical review of elaboration theory. *Educational Technology Research and Development 40*(3), 63–79.

Winn, B. (2009). The design, play, and experience framework. In *Handbook of research on effective electronic gaming in education*, pp. 1010–1024.

Wittgenstein, L. (1953). *Philosophical Investigations*. Macmillan.

Wouters, P., E. D. van der Spek, and H. van Oostendorp (2009). Current practices in serious game research: A review from a learning outcomes perspective. In *Games-based learning advancements for multi-sensory human computer interfaces: techniques and effective practices*, pp. 232–250.

Wouters, P., E. D. van der Spek, and H. van Oostendorp (2011). Measuring learning in serious games: A case study with structural assessment. *Educational Technology Research and Development 59*(6), 741–763.

Wouters, P., C. van Nimwegen, H. van Oostendorp, and E. D. van der Spek (2013). A meta-analysis of the cognitive and motivational effects of serious games. *Journal of Educational Psychology 105*(2), 249–265.

Wu, W. H., W. B. Chiou, H. Y. Kao, C. H. Alex Hu, and S. H. Huang (2012). Re-exploring game-assisted learning research: The perspective of learning theoretical bases. *Computers and Education 59*(4), 1153–1161.

Xie, B. and G. Salvendy (2000). Prediction of Mental Workload in Single and Multiple Tasks Environments. *International Journal of Cognitive Ergonomics 4*(3), 213–242.

Yang, Y. T. C. (2012). Building virtual cities, inspiring intelligent citizens: Digital games for developing students' problem solving and learning motivation. *Computers and Education 59*(2), 365–377.

Yee, N. (2006). Motivations for Play in Online Games. *CyberPsychology & Behavior 9*(6), 772–775.

Zimmerman, E. (2004). Narrative, Interactivity, Play, and Games: Four Naughty Concepts in Need of Disciptine. In *First person: New media as story, performance, and game*, pp. 154–164. MIT Press.

Zyda, M. (2005). From visual simulation to virtual reality to games. *Computer 38*(9), 25–32.