Engagement in educational games: An exploration of the interaction between game features, players’ perceptions and learning.

Dissertation zur Erlangung des akademischen Grades Dr. phil. der Fakultät für Bildungswissenschaften an der Universität Duisburg-Essen

vorgelegt von

Michael Karlheinz Filsecker Wagner

geboren am 25.05.1976 in Los Angeles, Chile

Tag der Disputation: 21. Noviembre 2013

Erster Gutachter: Prof. Dr. Michael Kerres
Zweiter Gutachter: Prof. Dr. Dominik Petko

ABSTRACT

Educational games' effectiveness relies on games' engagement power. However, this concept has seldom been examined. To fill this gap, this dissertation explored engagement by using multiple sources of data such as interviews, questionnaires and eye movements. In particular, this dissertation examined the effect of manipulating individuals' perception of the demand characteristics (PDC) of playing an educational computer game (i.e., to learn versus for fun) on individuals' content knowledge, cognitive engagement (CE) (i.e., mental effort and information processing strategies) and behavioral engagement (i.e., eye tracking data), while exploring the influence of self-efficacy and the perceived general mental effort (General AIME) on individuals' actual CE. Data analysis consisted of student's T (one tailed), bootstrap confidence interval, winsorized correlations and Pearson correlation coefficient comparisons. Results showed that participants increased their recall of content knowledge, but contrary to the expectations, no effect of the PDC manipulation on individuals recall test and CE was found. As expected, a positive effect of PDC on behavioral engagement was established. Likewise, a positive correlation of recall with behavioral engagement and one measure of CE was found. A positive correlation showed CE and emotional engagement. Finally, the CE employed was influenced by individuals' initial General AIME.

For the general lack of effect of PDC on CE, it is suggested that some CE measures might have not been sensitive to the PDC manipulation. Competition for participants' cognitive resources coming from both the game (i.e., cognitive overload) and the participants (i.e., volitional judgments), and a relinquishment to cognitively engage with physics content may have hindered the PDC manipulation. The lack of relationship between CE measures and recall may be due to the *inappropriate* cognitive processing employed. In this sense, it was suggested that 1) CE measures tapped different processes and learning outcomes, 2) participants may have not allocated the adequate strategies, and 3) CE may have affected "inferential activity" instead of the information processing strategies. Finally, the positive relationship between CE and emotional engagement suggested either an integrated (i.e., not disrupted) subjective experience - as expected - or the presence of competing goals as suggested by the games for entertainment literature. Implications for theory building, game design and research were provided.

ACKNOWLEDGEMENTS

I would like to thank Prof. Dr. Michael Kerres for having trusted on me and having provided me with ideas, support and critical appraisals of the previous drafts that led to the present dissertation. His critical perspective on educational games together with his knowledge of cognitive psychology and instructional design turn out to be a fertile ground for my ideas as they are expressed through these pages.

I would like also to thank Prof. Dr. Dominique Petko from The Schwyz University of Teacher Education for having accepted to be the Korreferent and for having provided thoughtful suggestions for improving both the overall quality and the discussion section of this dissertation.

I want also to thank my colleagues at the Duisburg Lab for helping and encouraging me through the different stages of this process. Special thanks to Dr. Judith Budgen-Kosten and Andreas Schmidt for helping me with the translation of some instruments used in this dissertation and for their comments on early versions of the introduction of this dissertation. I would also want to thank Tatjana Steinhaus for her help in the transcription and coding of the interviews, and the coding of the pre and posttest for reliability purposes. I also want to thank my friend Paula Grez for her willingness to provide me with material central for this dissertation.

Last but *not least* I want to thank my wife Daniela and more recently my little one Dominique. Daniela is just one of a kind woman and I feel blessed to have her with me. If Love had a face, that's hers. My beautiful and smart Dominique surprises me every day with her accomplishments and her love towards me and her mother. Thanks and I love you. For these two women I have finished this dissertation on the time scheduled.

TABLE OF CONTENTS

List of Tables

List of Figures

List of Equations

1. Introduction

> *The educator who associates difficulties and effort with increased depth and scope of thinking will never go far wrong* (Dewey, 1913, p. 59).

This dissertation addresses from a psychological perspective the link between effort and thinking and its influence on learning as suggested by Dewey more than a century ago. This link is explored using the multidimensional construct of engagement which entails behavioral, emotional and cognitive components. Under the concept of cognitive engagement, the present study examines the role of mental effort and information processing on learning from an educational game[1]. It is argued that this construct has been seldom explored in the field of educational games although educational research has already established its central role in learning.

In education important issues are sometimes discussed in terms of extreme opposites. For example, the issue of the effectiveness of different modes of instructions has taken the form of a contrast between direct instruction and discovery learning as succeeding or failing (Tobias & Duffy, 2009). Research on educational games has also had its own opposite terms. A typical contrast is usually made between "edutainment" as simply "drill and practice" associationist inspired games versus modern games inspired by "experiential learning" approaches. Similarly, the discussion on effective game design has taken the form of extrinsic versus intrinsic game design, which attempts to reflect the extent at which the content and the fantasy elements of the game are closely "coupled" or "interwoven". Likewise, it has been suggested that educational games involve by definition two opposite "modes" or scripts (Schank & Abelson, 1977) (i.e., learning mode versus playing mode).

However, drill and practice, discovery learning, experiential learning, and game playing represent all an *experience*. What it is, is that these experiences might be *qualitatively* different. The same should be true concerning extrinsic or intrinsic game design, edutainment or experiential games. They all provide an experience to the

---

[1] The term "educational games", as used in this dissertation, refers to a very specific software designed and developed with explicit learning goals to teach specific content knowledge. Although a more precise name could be "instructional game", the popularity and visibility of "educational game" is far higher in the field.

individual. What is qualitatively unique to the experience of playing games for entertainment, and the reason why they seem useful for learning purposes, is the *engagement* they produce on individuals. This engagement is usually described in terms of long hours of play in which players struggle to learn how to achieve a particularly difficult goal given the current skills they have and the constrains imposed by the game. The ability of succeeding during the game is also accompanied by feelings of satisfaction, enjoyment and "fun". This sustained engagement can be considered a goal striving-type of behavior and seems difficult to find in school settings or learning oriented activities more broadly. Some researchers have explicitly suggested that learning presupposes "…active, passionate, and engaging people…participating in the act of learning" (Hattie, 2009, p. 22). How this "engaging people" actually looks like while learning from an educational game is the focus of this dissertation.

However, educational games research has seldom examined the quality of engagement from the perspective of educational psychology (e.g., Fredricks, Blumenfeld, & Paris, 2004). For instance, research on cognitive psychology and memory has shown that what individuals do when encoding a unit of information can have important consequences on their later recall of such information. In particular, how conscious or controlled, as opposed to automatic, is the processing of information has a direct impact on the probability of later recall. In a similar vein, visual attention has been portrayed as an effortful process of resource allocation to information areas that need to be processed. This includes a conscious choice to what to pay attention and a certain level of intensity of such attentional process. From research on learning, it has been shown that *wha*t individuals understand from a text depends on *how* they approach the task of reading the text. Likewise, self-regulation, being cognitive engagement its highest form, has already shown how different cognitive strategies during learning may lead to different learning outcomes.

Corno and Mandinach (1983) in their attempt to unify motivation research and learning processes coined the term *cognitive engagement*, which entails different acquisition and transformation processes and a determined overall cognitive activity or effort. Echoing Bandura's reciprocal determinism (1982), the authors highlighted how motivational theories have proposed certain cognitive interpretations individuals engage in concerning the self and the environment, which in turn influences the amount and kind of effort they expend in a classroom task. They also pointed out the need to connect these interpretive processes with what individuals actually do *cognitively* while “engaged”. More recently, Berthold and Renkl (2010) also recognized how cognitive and motivational factors interact and affect learning. A similar argument can be found in the literature on technology, cognition and learning. As a case in point, some authors have claimed that individuals’ perceptions play a central role on the quality of their engagement with a computer tool (e.g., Salomon, Perkins & Globerson, 1991). The authors suggested that technologies with which we work (e.g., word processors, simulations, games), might offer an intellectual partnership to the individual. However, such a partnership requires *mental effort* or *mindful* engagement – i.e., cognitive engagement. Other instructional technologists have also considered cognitive engagement and mental effort as key for learning with technology (e.g., Hannafin, 1989). In particular, how individuals perceive a medium in terms of its cognitive demands and how they perceived the demands of the tasks have shown to be key factors affecting the amount of mental effort invested during a learning task (Cennamo, 1993; Glaser, Garsoffky, & Schwan, 2012; Salomon, 1984). The purpose of this study is to examine these psychological processes in the context of an educational game for learning physics: *Genius Unternehmen Physik* (Cornelsen, 2004).

1.1. Statement of the Problem

A major issue in learning by playing an educational computer game has been to help individuals to move beyond the initial “fun” toward a more appropriate engagement with the instructional content. However, as suggested by Kerres, Bormann, and Vervenne (2009) the game + instructional content combination might involve different scripts or modes (i.e., learning versus playing modes) experienced as opposed and as an interruption turning the “learning mode” superficial. In other words, a poor engagement with the instructional content. Therefore, there is a need to assess whether or not and to what extent individuals are “appropriately” engaged with the instructional content while playing an educational game.

Educational games are inspired by “experiential” learning approaches that produce a number of issues concerning learning. First, in experiential approaches (Hmelo-Silver, 2004) such as project-based learning (Blumenfeld, Soloway, Marx, Krajcik, Guzdial, & Palincsar, 1991), problem-based learning (Savery & Duffy, 1996) and knowledge building (Scardamalia & Bereiter, 1993) the main problem has been the tendency of individuals to engage in activities unrelated to the instructional goals. As Blumenfeld, Kempler, and

Krajcik (2006) suggested, these environments require considerable amounts of persistent mental effort and it is still an open question whether or not individuals are willing to invest the required effort. This implies that the motivational properties found in commercial games might not lend themselves for fulfilling educational goals, or for cognitively engaging individuals with predefined learning objectives. Second, even though computer games can produce a state of flow, this state by its own does not inform whether individuals are engaged with the game mechanics, with the underlying disciplinary content, or the fictional world of the game. Third, in the best of the cases, the virtuosity of the game cycle *user action – system feedback – user judgment* (Garris, Ahlers, & Driskell, 2002) can be considered as an "opportunity to be taken" (Perkins, 1985) that has been suggested to depend on the quality of the intellectual partnership (Salomon, et al., 1991; Salomon & Perkins, 1989) between the individual and the computer game.

These issues point to the *quality* of individuals' psychological involvement while learning from an educational game. By the systematic examination of this involvement, it could be possible to explain the reasons under the modest evidence of educational games effectiveness (e.g., Connolly, Boyle, MacArthur, Hainey, & Boyle, 2012; Hays, 2005; Randel, Morris, Wetzel, & Whitehill, 1992) and to determine under what circumstances and for whom a particular educational game could be more effective. The present dissertation examines this involvement under the general concept of engagement with an emphasis on *cognitive engagement.* This constructs provide a lens to explore how individuals differ both quantitatively and qualitatively when engaged with technological tools such as an educational game, regardless of the currently hypothesized source of their effectiveness or lack of it (i.e., design and pedagogical approaches, game quality, or technological sophistication). In particular, this study integrates Salomon's (1984) model of amount of invested mental effort (AIME) with Corno and Mandinach's (1983) model of cognitive engagement in classroom contexts. The former informs about the quantitative aspects of cognitive engagement and the factors that influence it, and the latter informs about the qualitative aspects of effortful information processing during a learning task.

1.2. Purpose of the Study

The purpose of this study is to examine the effect of manipulating individuals' perception of the demand characteristics of playing an educational computer game either for fun or to learn on individuals' cognitive engagement and learning when their perceptions of games and their self-efficacy to learn from them are controlled. The educational computer game corresponds to *Genius Unternehmen Physik* (Cornelsen, 2004). This game corresponds to a business simulation with several tasks related to basic concepts of physics. By proposing a multidimensional conceptual framework for understanding and studying engagement in games, the study operationalizes cognitive engagement and the other dimensions – behavioral and emotional engagement– by using multiple sources of data such as interviews, questionnaires and eye movements. Within this framework the study explores how these dimensions relate to each other and to learning in two different groups of individuals: One group instructed to play for fun and the other group instructed to play to learn. The learning outcome is assessed by a recall test of specific content embedded in *Genius Unternehmen Physik*.

1.3. Significance of the Study

The concept of engagement has been postulated to be the central mechanism underlying the effectiveness of games, but it has seldom been defined and measured. Although research has proposed theoretical models, such as Kolb's (1984) experiential model (Ketbrichi, 2008) and Garris et al. (2002) "input–process–outcome" gaming model, they have remained theoretical and apparently no attempt at measuring the postulated mechanism has been conducted so far. Therefore, a conceptual framework that argues which concepts are important to include and which may not could help to organize the data collection of different independent studies. From here it will be possible to organize the building blocks for a future theory of educational games and other technologies. On the other hand, research on educational games has not yet explored engagement from a multidimensional perspective that emphasizes cognitive engagement, a variable that has shown consistent relationships with learning and important motivational variables (e.g., Greene, Miller, Crowson, Duke, & Kristine et al., 2004; Gregoire, Ashton, & Algina, 2001).

Research has mostly been oriented toward measuring effects rather than processes or possible mechanisms – the "black box" approach. Likewise, when motivation is explored, it is mainly measured by questionnaires with no connection to the learning processes (e.g., Tüzün, Yılmaz-Soylu, Karakus, Inal, & Kızılkaya, 2009). Apparently, research has not yet explored the learning process during game play and which variables may have an impact on the final outcomes and almost no possible mediator variables have been proposed (Ennemoser, 2009) other than flow (Mattheiss, Kickmeier-Rust, Steiner, & Albert, 2009; Kiili, 2005; Pavlas, 2010). On the other hand, game research (i.e., games for entertainment) has focused on studying the user experience while playing games using more diverse sources of data collection such as eye tracking and physiological measures in addition to questionnaires (e.g., Nacke & Lindley, 2008; Nacke, Drachen, & Goebel, 2010). Many studies in this field are concerned with measuring the experience of immersion (e.g., Jennet et al., 2008). Research on educational games has employed similar measures to explore the behavior of individuals while playing a game and its relation to learning (Filsecker, Bormann, & Kerres, 2011; Kerres et al., 2009; Kickmeier-Rust, Hilleman, & Abert, 2011). However, these studies although promising have not used different sources of data integrated within a conceptual framework that goes beyond usability issues in order to understand individuals' experiences while playing an educational game.

Cognitive engagement reflects investment in learning, which includes mindfulness or the amount of mental effort invested (AIME), and information processing or the quality of that mental effort. Under the label of mental effort, cognitive engagement has been studied in the context of television viewing (e.g., Glaser et al., 2012; Salomon, 1984; Salomon & Leigh, 1984), interactive video (Cennamo, 1993), virtual worlds (Heers, 2005) and classroom environments (Brookhart & Durkin, 2003). With the exception of Cennamo's studies, this research has neither examined Salomon's complete model in general nor in the context of educational games in particular. Likewise, research has not yet integrated this model into the broader concept of cognitive engagement. When Cognitive engagement has been studied as such, it usually has entailed questionnaires on deep and surface strategy use in classroom contexts (e.g., Greene et al., 2004) and online environments (e.g., Richardson & Newby, 2006). In particular, Corno and Mandinach's (1983) model of cognitive engagement has been explored in the context of classroom tasks and internet search using either traces (i.e., students written notes concerning cognitive processes) (e.g., Howard, 1989) or observation of direct behavior (e.g., Rogers & Swan, 2004). This research has not yet explored both components of cognitive engagement, that is, mental effort invested and cognitive processes together. Likewise, research has not yet explored Corno and Mandinach's[2] model through interviews in the context of an educational game.

Finally, the present research might also represent a starting point to help understand individuals' behavior in games and other environments such as, intelligent tutors, multimedia learning, or e-learning environments. In particular, this dissertation can provide with valuable information to understand how individuals may "game" the system in this tutoring systems or how individuals seem "disengaged" or present "careless" behavior in online environment and adaptive systems (e.g., Cocea & Weibelzahl, 2009; Muldner, Burleson, Van De Sande, & Vanlehn, 2011). The approach of this research can also inform recent attempts at understanding the effects of *conative* and emotional feedback on individuals' performance and learning (e.g., Economides, 2009).

1.4. Overview of the Method

This study employed a randomized experimental pretest-posttest control group design. Individuals were randomly assigned to either the playing to learn or the playing for fun condition. The experimental design involved the administration of pretest questionnaires, an introduction to the experimental session, the treatment (introduced by a specific instruction to the participants), the recording of participants' eye movements, the administration of questionnaires, the interview, the posttest, and the participants' debriefing. The analysis of data consisted of descriptive statistics of the control and dependent variables and the quantification of verbal data obtained from the interviews. The test of hypothesis was conducted using traditional and robust statistic methods of group differences and association with the purpose of avoiding the undesirable effects of non-normality and outliers in the data common in social research, in particular when sample sizes are small.

1.5. Organization of the Dissertation

This dissertation is organized into six sections. Section one provides an introduction, the problem statement with purposes, and the relevance of this research. It also summarizes the main method and data analysis employed. Section two provides the

[2] Following the American Psychological Association (2009), when the author is part of the narrative and the name appears within the same paragraph, there is no need to include the year again. Only in cases that can lead to confusion. This criteria is followed throughout this dissertation

theoretical background and literature review that informs the current research: the concept of computer games and educational computer games, the educational research on educational games, engagement and cognitive engagement, and the description of the conceptual framework for studying engagement in educational computer games. Section three presents the process of selection of the game used and the pilot study. Section four presents the method of the study and section five presents the results of the study. Finally, section six discusses the results in terms of their meaning, limitations and implications for future research.

2. Literature Review

The aims of this literature review are the followings: 1) to examine how two research communities understand their object of study, that is, games and educational games; 2) to identify central psychological processes that play a role in learning from educational games, 3) to discuss the main claims about the effectiveness of educational games and the main research approaches to study them; and, 4) to identify the nature of the concept of engagement as central for games and beyond.

To achieve these aims, it is of central importance to describe the *medium* used in this dissertation, that is, an *educational game*. For that an overview is provided with the main definitions of games and their central features that distinguish them from other technologies. In particular, a discussion of the defining properties of games (i.e., rules and fiction) and how, together with players' actions give rise to "meaningful play" or "gameplay" is provided. Then a concrete description of "gameplay" is provided from the perspective of Artificial Intelligence and Game Design Patterns. The section closes with a discussion of the sense in which games can be considered simulations.

With this section as a background the review continues with the analysis of the concept of educational games. In particular, this section discusses a set of concrete definitions of games and educational games and then compares them with simulations. Then, a new definition of educational game is proposed. Next, the main arguments for using games for learning purposes and the main assumptions that the field holds concerning educational games are discussed. The section continues with a review of the pioneer work of Malone (1981) in games that seems to have greatly influenced the field of educational games both in terms of design and research. Therefore, the main game design approaches and research design are reviewed, followed by a synthesis of the empirical literature reviews on the effectiveness of games for learning.

The review continues with an overview of the different meanings of the concept of engagement. After identifying several types of engagement, the review presents a deeper discussion of a specific type of engagement: *cognitive engagement*. This variable and its role on learning is the central concern of this dissertation. In order to operationalize this variable, two models are identified. One corresponds to Salomon's (1984) model of Amount of Invested Mental Effort (AIME) and the other to Corno and Mandinach's (1983) model of cognitive engagement as the highest form of self-regulation. The former provides insights into the quantitative aspect of cognitive engagement. The latter informs about the qualitative aspects of information processing involving higher levels of invested mental effort. Finally, a framework for studying engagement in educational games is proposed. What follows is a summary of key findings from the literature review that provides the background of the present empirical study.

The results of the literature review showed that the game community has done a rigorous work at conceptualizing and defining games as a study object. On the other hand, the educational game community does not provide a distinct definition of educational games, and have presented a rather peculiar manner for understanding what a game should be. This reflects clearly how these two communities go on parallel tracks. For example, while the game community has understood games as producing a particular type of experience (i.e., meaningful play or gameplay), the educational community has seemed to understand games as an object with a set of discrete features, losing sight of the centrality of gameplay, and putting immersion as the central quality of games. Likewise, while the game community has distinguished different levels of analysis of a game (i.e., the game, the player, and the culture) and between the concepts of "play" and "game", the educational community has shown a tendency to mix these descriptive levels and/or concepts. Consequently, trying to define and limit the object of inquiry of the field of educational games becomes a major challenge if the goal is to have a coherent body of knowledge to advance the field forward.

Secondly, while the game community has considered rules and fictions as intimately intertwined and complementary, the educational community has seemed to understand fiction as a mere context or background to be changed at will. Together with scholars' emphasis in merging content either with the mechanics or with the narrative, the rule/mechanic part of games has been relegated to a secondary place. In summary, what is a meaningful relationship (rules and fiction) and a central property of games (gameplay) for the game community, it turns out to be independent, somewhat unrelated and overlooked aspects of games for the educational community. This tendency could have negative consequences for the design and research on educational games.

Thirdly, the idea of immersion highly touted in educational technology concerning games has been considered by some prestigious game theorists as a fallacy. This fallacy, though, is shared by researchers from both communities. Similarly, the concept of engagement though not optimally defined, it has been studied within both communities.

However, neither of them has studied the construct from the perspective of educational psychology and has mainly done so under other theories and constructs (e.g., flow and immersion) that have not been consistently related to learning processes and outcomes. This is a worrisome state of affairs given that in the discourse engagement appears as the driving force to consider games for educational purposes.

Fourthly, both communities but the educational one in a more stronger sense, has worked under a “factoring assumption”[3], that is, the belief that it is possible to remove or add discrete “features” of a game and study their impact without affecting what makes a game unique and without knowing how the removed or added features might have interacted to bring about a specific outcome. For several reasons that are discussed in this review, this dissertation contends that this approach to research on games limits our understanding of how games produce such motivation and engaging experiences in players, and whether those lead to the desired learning outcomes. Likewise, this dissertation assumes that the effectiveness of games depends to a high degree on how individuals use it. However, research has barely considered learner characteristics or other mediator variables that might affect the learning outcomes expected. Therefore, research should focus more in exploring the psychological processes on games under more controlled settings such as randomized clinical trials, clearly underrepresented in game research. In summary, the examination of learners’ characteristics, their psychological process during gameplay in a controlled setting represent an exception in the vast research on educational games.

Finally, research on engagement has struggled to find a clear conceptualization. Several researchers have proposed different dimensions or levels. It was also noticed that the place of engagement within the educational and psychological theories available has not been explicitly determined. Based on these results, the present dissertation summarized and synthesized the dimensions proposed into three (i.e., cognitive, behavioral and emotional) and has explicitly proposed engagement as a volitional, not motivational construct –issue frequently disregarded in the literature – which is of central importance when proposing models and conceptual frameworks for future research. The present conceptual framework proposed for the study of engagement it is supposed to integrate previous models of engagement in games with models of engagement from the

[3] This “factoring assumption” as used here is an adaptation to the use given by Greeno (1997) and Anderson, Reder, and Simon (1996) in their debate between cognitive and situative perspectives on learning.

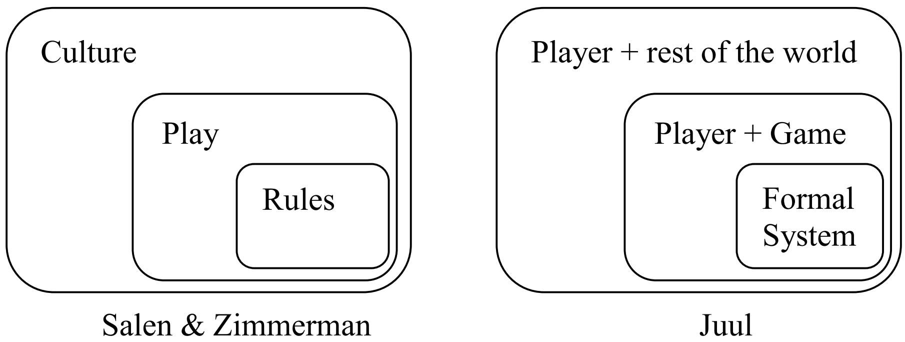
educational literature and proposes an operationalization of the models so that they can be empirically examined.

2.1. Computer Games for Entertainment: What makes them Special as a Medium?

This section does not seek to provide a new extended revision of the different theorizations about games or play, but to identify from already conducted theoretical considerations key properties that make a game a *game*. This section begins by revising the two most influential definitions of games. Then it continues with a detailed description of the defining properties of games, that is, rules and fiction. Finally, the relationship between simulations and games is discussed. In order to illustrate the discussion on what is a game, the game of *Pac-Man* was selected as an example of the different elements of the discussion on games.

The strategy used for developing a definition of games has been to put side by side different things normally call *games* and see if it is possible to established some commonalities that most games seem to share. Salen and Zimmerman (2003) and Juul (2005) attempted a new definition of game using such a strategy. They gathered a set of previous definitions and attempted to find similarities and differences among them. However, in doing that both took slightly different approaches and contributed in a very specific way to the issue of defining games. On the one hand, Salen and Zimmerman (2003) attempted to articulate what makes a game unique and different from other forms of play and, specially, from the *activity* of playing a game. They differentiated between rules (i.e., the game itself), play (i.e., the experience of playing the game), and culture (i.e., games and their broader contexts) as dimensions of a conceptual framework to understand games. On the other hand, Juul (2005) – deliberately or not – applied Salen & Zimmerman's (2005) framework (Figure 1) when he distinguished between the game itself as a formal system, the game and the player as an experiential system and the game and the rest of the world as a cultural system. In addition, Juul (2005) reinterpreted some key concepts apparently different and gave them a new meaning.

Figure 1: Games' Theoretical Frameworks by Salen & Zimmerman (2003) and Juul (2005)



Salen and Zimmerman's (2003) definition started by revising eight previous definitions from a variety of fields. Then the authors articulated the uniqueness of games that makes them different from other forms of play and distinguished between games themselves and the act of playing a game. After listing the fifteen elements described in each of the eight definitions reviewed, the authors noticed that with a few exceptions (e.g., Crawford, 1984), none of the authors belonged to the field of game design. Among the agreements that seemed to exist among the authors was the idea of *rules* (see Appendix A). The second agreement was about the existence of *goals*. Finally, Salen and Zimmerman realized that not all the elements listed were pertinent for a definition of games. For example, the idea of voluntariness did not seem to be applicable to all games, the idea that games created social groups was considered to be an effect of games that did not say anything about the games themselves, and, the make-believe quality could be also found in other media (e.g., films), and therefore was not exclusive of games. That having said, the authors defined games as: "…a system in which players engage in an artificial conflict, defined by rules, that results in a quantifiable outcome" (Salen & Zimmerman, 2003, p. 80). A brief description of the elements of this definition is provided below.

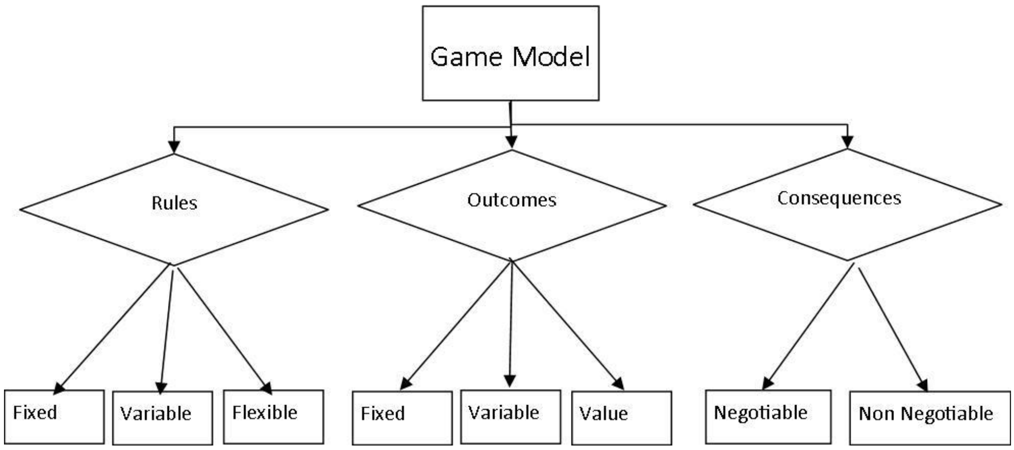
By *system* Salen and Zimmerman (2003) meant a set of things affecting one another and forming a pattern different from any individual part. Players interact with the game system to experience the play of the game. A system is made of objects, attributes, internal relationships and an environment. For example, in the case of *Pac-Man* the objects are *Pac-Man* itself, the four ghosts, the fruits, the pellets and the power pills; the attributes are the original position of the objects within the maze and the ways each object moves; the internal relationships correspond to the actual strategic relationship of *Pac-Man*, the ghost and the power pills; finally, the environment would be the actual playing of the game.

Artificial refers to the fact that games maintain a boundary from "real-life". *Conflict* refers to games as embodying a contest of powers that can take different forms (e.g., cooperation or competition). In *Pac-Man* the player's overall goal is to get a high score, which depends on how many pellets *Pac-Man* can eat, but the ghosts will chase *Pac-Man* making the goal harder to achieve. *Rules* limit what the players can and cannot do providing the structure to support the play of the game. *Pac-Man* can move up-down-left-right, but certainly cannot jump over the ghosts. Finally, *quantifiable outcomes* refer to the fact that at the end of a game a player either wins or loses or receives some numeric score.

In *Pac-Man*, players never win the game, but receive a numeric score. This quantifiable property distinguishes games from other forms of play (e.g., Hide-and-Seek).

Juul's (2005) classic game model started also from the analysis of previous definitions (e.g., Caillois, 1961; Crawford, 1984; Huizinga, 1949; Salen & Zimmerman, 2003). The author categorized them as belonging to three aspects of games: games as rule-based systems, the player relationship with the game's outcomes, and the playing of the game relationship with the rest of the world. In other words, games as rules, outcomes and consequences. Through their variable properties, Juul used these three elements to distinguish among games, borderline cases and no games. As shown in Figure 2, rules can be fixed, flexible or variable. Outcomes can also be variable or fixed, but with or without value as well. And consequences can be negotiable or non-negotiable. A game should have fixed rules, valuable outcomes and negotiable consequences.

Figure 2: A Summary of Juul's (2005) Classic Game Model



This framework is also useful to differentiate between play and game. Regardless of the theorization of what is play and the language issues involved, play is generally a free-form activity with *variable* rules, while games are *rule-based* activities. The latter being the focus of Juul's (2005) understanding of games.

Juul (2005) developed his "classic game model" as an attempt to synthesize theoretical ideas apparently diverse, but that shared common notions:

- Rules and outcomes: The existence of rules in a game – Crawford's (1984) "formal systems"– that are above discussion (i.e., "fixed") seems to be a common agreement among game theorists. From Crawford's (1984) idea of conflict it follows that games have variable outcomes.

- Goals and conflicts: The idea of games representing conflicts (Crawford, 1984; Salen & Zimmerman, 2003) implies also the idea of games as having goals in so far as conflicts presuppose contradictory goals. However, there are famous cases of "video games" with no explicit goals (e.g., The Sims and SimCity). Given that goals describe the relationship between the player and the game, goals can be better understood if divided into three different categories: a) Valorization of the outcome, b) Player effort, and c) Attachment of the player to an aspect of the outcome. This is one of the most important contributions of Juul's (2005) definition of games, that is, of having included the role of the player in the definition of a game.
- Voluntary: This issue coming from Caillois (2001) is difficult to define and settle in a clear cut manner in the context of games. For Juul (2005) playing a game included by necessity the acceptance and following of the rules of the game. Similarly, Salen and Zimmerman (2003) characterized the rules of any game as being *binding* (see below Section 2.1.1).
- Separate and unproductive: Huizinga (1949) and Caillois (2001) described games as separated from the normal human world in space and time and occurring in a "magic circle" (Huizinga, 1949) and therefore as being unproductive (Caillois, 2001). The idea of time/space separation is clearly difficult to sustain given nowadays games played on the internet for months, through email or the so called augmented reality games that involve real activities in individuals' daily life. The idea of unproductivity is even more difficult to set given that some individuals can earn money by accumulating valued "in-game" elements that they can sell through, for example, ebay. In other words, whether or not someone makes money from a game should not necessarily change the status of a game as such. For instance, if an individual play World of Warcraft (WoW) and get some precious and transferable items, WoW should not stop to be a game if this individual sells later the items and earns a few dollars. This is what led Juul (2005) to understand these issues under the idea that games are activities with "negotiable consequences".

- Less efficient means: This idea, coming from Suits (2006), falls apart in the case of video games. For example, the video game FIFA 2002 is easier to play than real soccer and in any case Juul (2005) argued that it is difficult to determine what it means to use less efficient means when playing FIFA2002, less efficient compared to what is the question unanswered. The notion then represented for the author a mix of other features of games such as the player's effort and the player's acceptance of the rules.
- Fiction: Some games do have a fictional element but not all of them. For example, if we compare WoW with Tetris, the fictional aspect is quite different in the two. WoW has a strong fiction, while Tetris is rather abstract.

From the above considerations, the author proposed the following definition of a game:

> 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, p. 36).

Rules should be specified in such a way as to be able to be programmed on a computer or to be above discussion. Even though the rules are unambiguous the game activity assumes that the player will *respect* and follow the rules. Rules should also provide for *variable outcomes*. There are different ways in which players and designers provide for different outcomes. In *Pac-Man* players can flee away from the ghosts or wait for them in the near of a power pill to bring some uncertainty to the game activity. Likewise, designers have different design patterns (see Section 2.1.1.1) to use in order to even or uneven players' skills. In *Pac-Man* the four ghosts could have different abilities (see below the Section 2.1.1.2 on Artificial Intelligence, where the ghosts are discussed as "state machines") that may produce more challenging gameplay. On the other hand, the outcome should be *quantifiable*. For example, the goal of *Pac-Man* is to get a high score, not to move in a pretty way. To *valorize the outcome* refers to the idea that some outcomes are better or more desirable than others. Positive outcomes are normally more difficult to reach than negative ones and there is also the issue of having two or more conflicting positive outcomes. These opposing outcomes instantiate the conflict nature of games, a conflict that is perceived by players as a challenge. In doing so, players experience the tension necessary to be engaged in the game activity. *Player effort* means in this context that games are challenging (or contain a conflict). Rules are designed so as to allow the player's action to change the game state and the game outcomes, which should lead players to develop an attachment to the outcome they have contributed to produce. In *Pac-Man* eating the power pill turns the ghosts into blue, which *Pac-Man* can eat and produce more points. In that manner, the player is emotionally attached to the outcome. Finally, games can be assigned real-life consequences. In order for a game to have *negotiable consequences*, the operations and actions in the game should be harmless. If *Pac-Man* is touched and a live is lost, nothing really happens to the player. These consequences concern what the players can consciously control. As special case, professional sports have generally been considered to be work and not a game precisely because the consequences of the activity have been already negotiated and defined in terms of how much money one can get.

2.1.1. Rules

As already suggested, rules are the most agree upon constitutive characteristic of games. Rules give games their formal identity as such. Understanding how rules construct a game and how they relate to other non-rules aspect of games (e.g., narrative or fiction), makes possible to tackle persistent design problems (Salen & Zimmerman, 2003). However, rules are not unique to games. It is possible to find rules for getting a tenure position in a university, rules to get a bachelor degree, or rules to buy groceries in the supermarket. According to Salen and Zimmerman (2003), rules are "the deep structure of a game from which all real-world instances of the game's play are derived" (p.120). For Suits (2006) rules are defined as proscriptions of useful means to achieve *prelusory* goals (i.e., a specific achievable state of affairs such as crossing first a finish line, but not necessarily fairly). Suits also suggested that rules provide both *ends* (e.g., to be the first one in crossing the finish line) and *means* (e.g., the accepted ways of crossing the finish line). The result is the tension of having to reach a goal through *inefficient* means. Now, what makes the rules of the game special? Salen and Zimmerman (2003) proposed the following six general characteristics of all game rules:

- Rules limit players' action: In the dice game *Yatzee* the rules force players to do a very specific activity with the dice and not whatever they feel like it. They need to follow the instructions that the rules embody. However, this represents only half of the picture concerning rules. Rules can also be creative in the sense that they "set up potential actions, actions that are meaningful inside the game but meaningless outside" (Juul, 2005, p. 58). In a sense, rules represent limitations and affordances, so that rules prohibit certain action but give meaning to the allowed ones. This idea is closely related to the Salen and Zimmerman's (2003) notion of "meaningful play" (see Section 2.1.3).
- Rules are explicit and unambiguous: The instructions need to state specifically what the player should do in a particular moment in a game. For example, when playing a board game, it should be clear what to do when arriving at a particular space in the board.
- Rules are shared by all players: When players are more than one, the understanding and application of the rules should be the same for all the players involved and interpreted in the same manner by everybody.
- Rules are fixed: While playing the game no change to the rules are permitted. Even in games in which changing the rules is part of the game, this changing is highly regulated by more fundamental rules.
- Rules are binding: The rules are fixed and shared because they are binding. They are supposed to be followed by the players. For this reason Juul (2005) argued that it was very difficult to apply the freewill and voluntary ideas of Huizinga (1949) and Caillois (2001) to games.
- Rules are repeatable: This basically means that the rules can be repeated from game to game. If players play *Pac-Man* today, when they do it again the rules will be the same.

Juul (2005) highlighted some characteristics of rules. Rules seem to be the main source of enjoyment, although fiction also plays a role on it (see Section 2.1.2). Rules are easy to learn but challenging to master, requiring ingenuity to overcome the challenges. Finally, rules are more than the sum of all of them: the strategies for playing the game are more complex than the rules themselves.

In what follows, two complementary perspectives of how to understand rules and how to apply the knowledge about rules to analyze and create games are presented. The value of the coming analysis resides in that it introduces a language and a framework to understand and talk about games as a dynamic activity. In other words, a framework to talk later about *gameplay* or *meaningful play.*

2.1.1.1. Design Patterns

A useful manner to look at rules from a more pragmatic perspective is the use of patterns. When describing patterns, one is basically describing a set of rules that affect gameplay in particular ways: “rules dictate the flow of the game…there are rules that govern what the game elements are, how they behave, what actions players can perform, and so on” (Björk & Holopainen, 2005). Specific parts of these rules have been called “game mechanics” and Björk and Holopainen (2005) converted these mechanics into game design patterns. More generally though, the patterns approach to game design has broader goals for the game community. As Kreimeier (2002) pointed out, game design shares the same need of any other design activity, mainly that of documenting, planning and discussing everyday design decision and activity in a formalized manner. Kreimeier also acknowledged that devices and strategies borrowed from other fields, such as narrative media, are limited to help designers capture the essence of *interactive* sequences central to any game design project. Therefore, how to best capture the essence of interactive media such as games might be accomplished by the use of patterns.

From a problem-oriented approachm Kreimeier (2002) postulated that game design patterns are conventions that help document the design decisions in the design of a given game. These patterns collections could provide a specific design vocabulary to assist designers in: 1) communicating efficiently with each other and with other professionals (e.g., software engineers and instructional designers), 2) documenting their insights and experiences, and 3) analyzing their own design and the designs of others for comparison, criticism, re-engineering or maintenance.

On the other hand, Björk and Holopainen (2005) developed their activity-based framework to describe gameplay. The framework is conceived as a tool to analyze and create games and also to describe the “first order” of game design, that is, the physical and logical components of games that affect gameplay. Four components are identified: Holistic, Boundary, Temporal and Structural. These components describe, respectively, 1) the activity of playing a game as separated from other activities, 2) the limits of the possible player actions within the game, 3) the flow of the game, and 4) the objects (physical and logical) required to change the game state. According to Björk and Holopainen, these components support the existence of the "second order" of game design, that is, the *game design patterns* (Table 1).

Table 1: Component Framework and Examples of Game Design Patterns: *Pac-Man*.

Component	Sub-components	Pac-Man	Design pattern
Holistic	Game session	Setup session = inserting the coins and choosing number of players. Set down= "game over" message	Lives
Boundary	Mode of play	Avoiding the ghost and chasing the ghost	Role reversal
Temporal	End conditions	End condition of eating a pill: *Pac-Man* moves over it and the evaluation function adds 10 points.	Movement
Structural	Game facilitator	The software and hardware behind the game.	

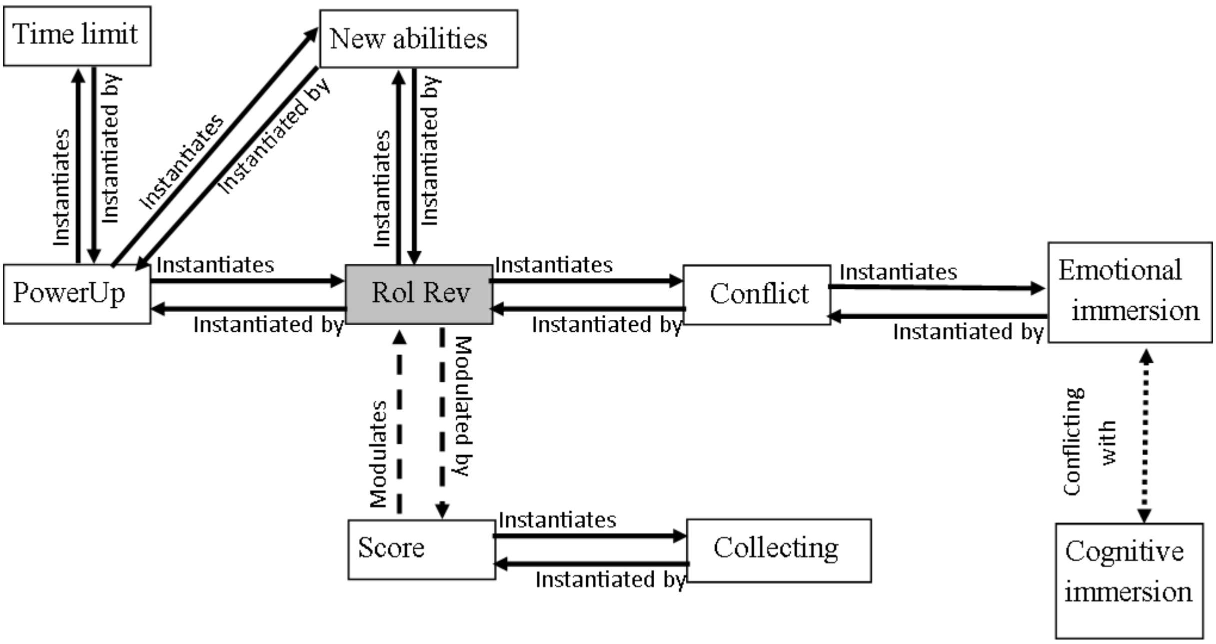
However, Björk and Holopainen (2005) claimed that a focus on the use of patterns as tools to solve problems, Kreimeier's (2002) approach, risks to be too limiting as design tools (see Appendix B for a comparison of the patterns templates). First, defining patterns from problems might focus its use only in removing unwanted features of current designs instead of as tools for creative design. Second, the problem identified in a pattern can be solved by applying a related and more specific pattern. Third, design patterns are not precise tools to solve a problem, given that the introduction of a single pattern might affect different aspects of gameplay. For these reasons, Björk and Holopainen considered the game design patterns as models to construct knowledge about gameplay so that it can be used in the analysis of existing games and the design of new ones.

Björk & Holopainen defined game design patterns as "semiformal interdependent descriptions of commonly reoccurring parts of the design of a game that concern gameplay" (Björk & Holopainen, 2005, p. 34). The main shared characteristics that are either internal to the pattern or that depend on the relationship among patterns are the followings:

- Non-quantitative descriptions: Game design patterns refer to particular areas of gameplay without using quantitative measures. This implies that the presence or effect of a pattern cannot be accurately measured, but they can be distinguished from each other and their relationships among them identified.
- Interrelationships: Patterns always relate to each other in some form. The following relationships have been identified:
    - Instantiates & Instantiated by: The presence of one pattern causes a second pattern to be present. That is, the first one "instantiates" the second one. The same is true when two or more patterns limit gameplay in such a way that another one emerges. Consequently, the pattern that emerges out of these relationships has the "instantiated by" relationship back to the first pattern. More generally, a pattern can be instantiated by ensuring the presence of a related pattern, which has specific consequences on gameplay.
    - Modulates & Modulated by: One pattern modulates another one when the former affects aspects of the latter in such a way that influences gameplay. Likewise, when one pattern has this relationship with a second pattern, that pattern has the "modulated by" relationship to the first one. More generally, "modulated by" relationships exemplifies how the application of additional patterns can fine tune a pattern's effect on gameplay.
    - Potentially Conflicting with: The presence of a first pattern make impossible the presence of another one.
- Hierarchy: The relationships described above (e.g., instantiate and instantiated by) create different hierarchies of patterns.

Figure 3 depicts in the form of a diagram the relationships mentioned above (e.g., *Instantiates*, *Modulated* and *Conflicting with*) between game design patterns such as *Role Reversal* and *Power-Up.* Basically, when players of *Pac-Man* eat the power pill, this pill represents a *Power-Up* (i.e., a game element that gives time-limited advantages to the player), which *instantiates* (solid arrows) both *Time Limit* (i.e., a time limit for completing an action, reaching a goal, or staying in a certain mode of play) and *New Abilities* (i.e., the gain of new abilities in the game such as chasing ghosts). In parallel, this change in abilities are instantiated by the *Role Reversal* (i.e., shifting between two opposite roles), in which *Pac-Man* can now chase and kill/eat the ghosts (one mode of play) instead of running away from them (another mode of play). The diagram also shows how *Role Reversal* can *instantiate* another pattern, *Conflict*, which in turn *instantiates* the pattern *Emotional Immersion*. This type of immersion, however, may be *conflicting with* other types of immersion such as *Cognitive Immersion*. Finally, the pattern *Score* (i.e., numerical representation of the player's success in the game) can modulate the consequences of *Role Reversal* on gameplay. When ghosts turn blue and *Pac-Man* can chase them, for every ghost captured the player becomes 200 points. Considering that eating each of the pellets give only 10 points, the 200 points make a huge difference in helping the player reach the goal of achieving a high score. If *Score* would not modulate the impact of *Role Reversal* in this way, say no score added for hunting Ghosts, the resulting gameplay might be less interesting since the activity of hunting does not help to achieve the player's overall goal. In other words, it would not matter if the player chases or not the ghosts.

Figure 3: Key Game Design Patterns Affecting *Pac-Man*'s Gameplay



A discussion of rules from the Artificial Intelligence perspective comes next. Again the case of *Pac-Man* is analyzed in terms of how the behavior of the ghosts in *Pac-Man* framed as “state machines” make possible the gameplay described using the game design patterns approach.

2.1.1.2. Artificial Intelligence

When considering games as formal systems with a mathematical and logical model underlying all gameplay (Salen & Zimmerman, 2003), the field of Artificial Intelligence (AI) is an interesting approach to understand how the rules are actually produced and organized and how they produce interesting gameplay and solve different design challenges.

Generally speaking, AI in computer games refers to how machines (i.e., non-player characters or NPCs such as the ghosts in *Pac-Man*) are programmed to show specialized intelligent qualities. For example, in the case of the ghosts, when they should chase *Pac-Man*, evade *Pac-Man* or simply roaming around are problems that AI normally solves. Other general problems of AI are the classic chasing/evading and pattern movement for assessing the threat of the enemy (Bourg & Seemann, 2004). In general terms AI’s techniques are divided into deterministic and non-deterministic. The former produces predictable and specified behavior, such as a particular chasing behavior. The latter produces uncertainty and unpredictable behavior. It is responsible for the degree of learning and adaptability that a game entails, for example, when a NPC learns to adapt to the fighting strategies of a human player. This type of learning uses specific techniques such as neural networks, Bayesian techniques or genetic algorithms. In this section the basic AI techniques reviewed are Scripting, Finite State Machine and Bayesian techniques. These are the most ubiquitous in games development and its understanding can provide new possibilities for understanding game design.

Scripting. It corresponds to a simple programming language tailored to a specific task within a game. It is usually used for altering attributes, responses and game events. For example, scripting verbal interaction is very useful if one is interested in providing hints or to create compelling stories without altering the core game program. Verbal interaction seems intelligent when is related to the current game situation. These interactions involve checking several parameters and then respond accordingly. Figure 4 depicts the code for scripting a dialog between a Giant NPC and a human player in which the parameter to check is simply the type of weapon used by the player.

Figure 4: An Example of Scripted Dialog between a Non-player Character and a Player from Bourg and Seemann (2004)

```
If (Creature= =Giant) and (player= =Human)
    begin
        if (playerArmed= =Staff)
          Say ("You will need more than a staff, puny human!");
```

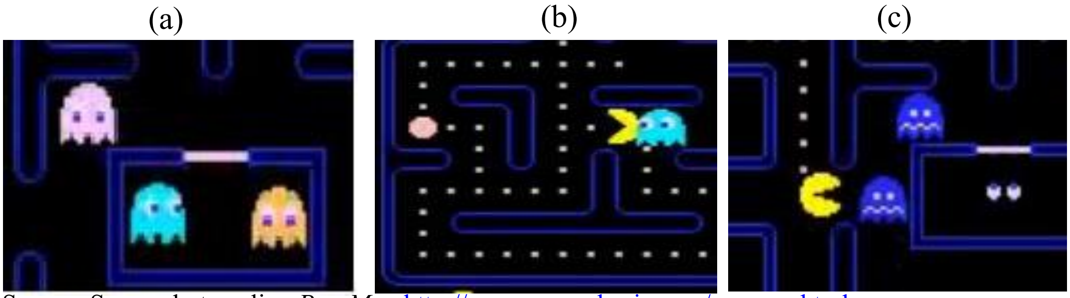
This type of scripting can go from friendly messages, helpful hint and feedback to warnings and taunts. Another example concerns to scripting particular events within the game. The script below (Figure 5) shows how it can trigger in-game events that might not be necessarily related to NPC, but to some actions or position of the players themselves. In this case, when the player's location, which could be walking on an area with grass, takes a particular value, then the player is trapped (e.g., a hole on the grass). The parameter to check here is the actual position of the player compared to a predetermined value. If they are equal, it triggers the trap. These scripts, together with scripts controlling NPCs' behavior can be used and orchestrated in more complex manners. Another possibility is to script written text entered by the players that can affect NPC s' behavior or trigger in-game events.

Figure 5: Example of Scripting of Written Text from Bourg and Seemann (2004)

```
If (PlayerLocation(120,76))
    Trigger (kExpositionTrap);
If (PlayerLocation(56,16))
   Trigger (kPoisonTrap);
```

Finite state machines (FSM). They represent mathematical models of computation of a hardware and software system and can take a finite number of states. A FSM can define the conditions that determine when it should change to another state. In *Pac-Man* the ghosts are FSM with three possible states (i.e., roam, chase, and evade). In each state the ghosts behave differently and the transitions from one state to the other depend on the players' action (i.e., whether they eat or not the power pill). Figure 6 illustrates how the reversal role from being chased to chase occurs in the actual play of *Pac-Man*.

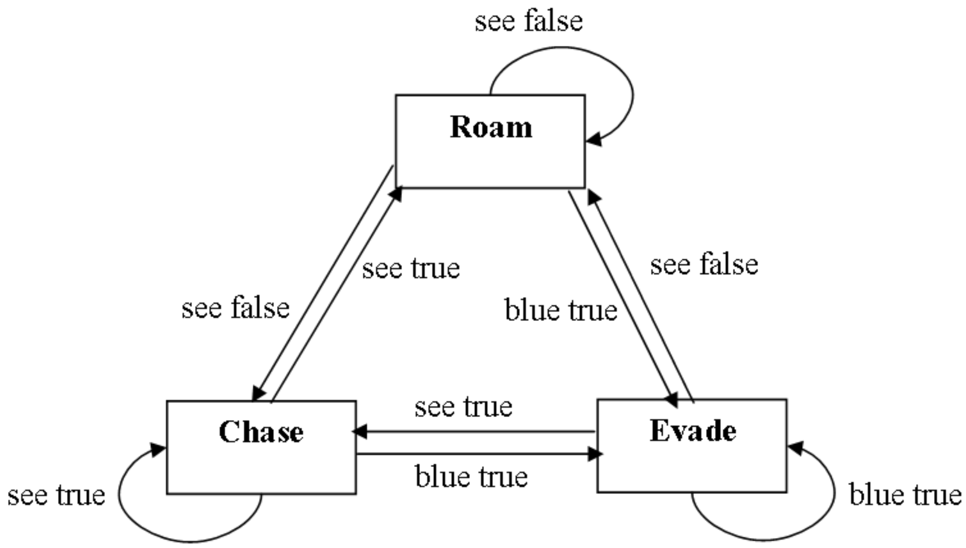
Figure 6: Pac-Man as a Finite State Machine (FSM)



Source: Screenshots online *Pac-Man* http://www.gamesbasis.com/pacman.html

Figure 7 depicts the AI diagram of the ghosts' FSM. The squares represent the three possible states (i.e., Roam, Chase, and Evade) and the arrows the possible transitions among states.

Figure 7: Diagram of Pac-Man's Finite State Machine (FSM) from Bourg and Seemann (2004)

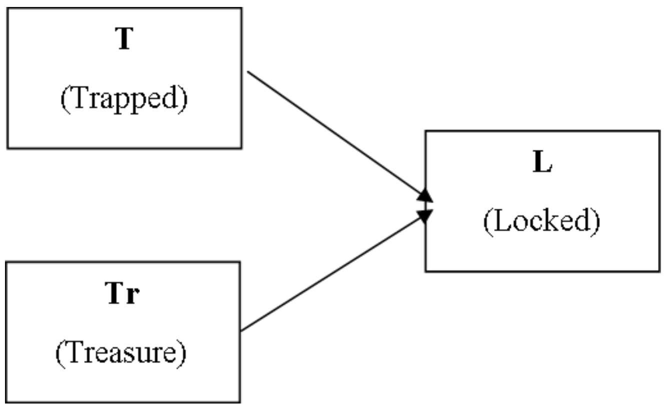


Note: Diagram re-drawn from Bourg and Seemann (2004).

Transitions show the conditions under which the states can change or remain the same. There are three conditions or functions: 1) seefalse, when the ghost see no *Pac-Man*, 2) seetrue, when the ghost see *Pac-Man*, and 3) bluetrue, when ghost has turned blue (because the player ate a power pill). At the beginning, the computer controls the ghosts in the initial state of "Roam" and remain so if the condition see=false (Figure 6a). Two conditions can change this: either see=true, so that the ghost sees the player (Figure 6b), which change from Roam to Chase or blue=true (i.e., the ghost turned blue) (Figure 6c), which means changing from Roam to Evade. When the ghosts are in the Evade state, they remain the same as long as it is blue, otherwise the state changes to Roam if the player cannot be seen and to Chase if the player is seen.

Bayesian techniques. For AI purposes Bayesian techniques allow NPCs to make decisions when the game state is uncertain. For example, an NPC that has to guess whether or not a chest contains a valuable (i.e., a treasure) he can steal, and if so, determine also if the chest is locked and/or trapped. Figure 8 represents a typical Bayesian network containing nodes or random variables (i.e., T or Trapped, Tr or Treasure, and L or Locked) and links representing the causal relationships among the variables. The strength of these relationships is made by assigning probabilities to each of them, based on statistics obtained during gameplay. That is, every time a NPC opens a chest, the frequencies of the chest being trapped can be recorded as well as the frequencies of having or not a treasure inside. Then using the Bayes' rule it is possible to determine the probability of the chest being trapped given that is locked. The Bayes' rule states that the probability of the occurrence of an event B (e.g., the chest being Trapped) given that A has occurred (e.g., the chest being Locked) is equal to the probability of B occurring times the probability of A occurring given that B has occurred divided by the probability of A occurring. Figure 8 depicts this situation.

Figure 8: An Example of Bayesian Rule Applied to a Game Situation from Bourg and Seemann (2004)



Note: Diagram re-drawn from Bourg and Seemann (2004).

But how the Bayes' rule helps a NPC to decide whether or not a chest is trapped and whether or not the NPC should try to open it? Assuming that the NPC has open 100 chests, Table 2 shows the frequencies of different events or states in the game:

Table 2: An Example of Inputs for a Bayesian Rule from Bourg and Seemann (2004)

	Locked	No Locked	Total
Trapped	29 (.78)[3]	8	37 (.37)[1]
No Trapped	18 (.29)[4]	45	63 (.63)[2]
Total	47	53	100

Note. Parenthesis represents the probabilities of different events or states as follows:
[1] Probability of the chest being trapped: P(T) = 37/100 = .37
[2] Probability of the chest not being trapped: P(~T) = 63/100 = .63
[3] Probability of the chest being locked given that is trapped: P(L|T) = 29/37 = .78
[4] Probability of the chest being locked given that is not trapped: P(L|~T) = 18/63 = .29

Without the use of Bayes' rule it is possible to say that there is a 37% chance that a given chest is trapped. But if the NPC also notices that the chest is locked, what is the probability of being trapped? In other words, the question of the conditional probability of the chest being trapped given that is locked can be answered by applying the following formula:

Equation 1: Example of Conditional Probability

$$P(T|L) = [P(L|T)P(T)]/[P(L|T)P(T)+P(L|{\sim}T)P({\sim}T)]$$

Using the data from Table 2, the probability that the chest is trapped is:

$$P(T|L) = [(.78)(.37)]/[(.78)(.37)+(.29)(.63)] = .61$$

In this case the P(T) goes from 37% to 61%, increasing the NPC's belief that the chest is trapped. This could lead the NPC to not opening the chest during gameplay. This is how the Bayesian approach helps computer-controlled objects to make decisions and seem smarter or more intelligent. More broadly, these techniques introduce randomness, define character's abilities, combine probabilities with state transitions in the FSM and also update certain probabilities during gameplay to facilitate NPC to learn or adapt.

2.1.2. Fiction

Although rules are part of all games, not all games project a fictional world: the game can take place in Germany during 1945, or in a city, or simply the player controls a character on the moon. According to Juul (2005) rules and fiction compete for the attention of the players and are complementary. It is possible to discuss rules in games without mentioning fiction. However the opposite is not true: when discussing games' fictional world it is necessary to refer to the rules of the game. Therefore, fiction depends on rules. For example, the manner in which objects behave in games (governed by rules) influences the fictional world imagined by the players (Juul, 2005). For example, in an action game the presence of an ugly monster cues the player of a possible danger and makes him either stay away of the monster or shoot at him.

Fictions are extremely subjective, ambiguous, evocative and subject to discussion. Games project fictional worlds imagined by the players, who fill in the gaps in the fictional world. Some of these worlds are optional, others are contradictory and incoherent. In some games the fictional world contradicts itself or prevents the player to imagine a complete fictional world. In these cases, players usually consider the rules of the game for filling the gaps. That is, an incoherent world is one where some events in the game cannot be explained without referring to the rules of the game. For example, who is *Pac-Man*? It is possible to imagine that *Pac-Man* is a hungry creature hunting for food (the fiction) or that simply is something that moves forward and eats "stuff" all the time (the rule). Another feature of fictional worlds is that they are incomplete. This leaves players with different choices to fill in the gaps and imagine the world. Players fill in the gaps by using their knowledge of the real world and knowledge of the genre convention.

After having played a game several times, players become less interested in the fictional level of the game focusing more on the rules of the game. What is important however is how rules and fiction actually play together as source of enjoyment and interesting gameplay. Given that the rules are hidden from the player, the only way the player can have a sense of the rule system of the game is by experiencing the fictional world or the game representation. As a rule of thumb, only parts of the fictional world are actually implemented in the rules of a game. For example, when playing Black Stone (1993) it seems fair to assume that monsters have no good intentions and might be dangerous (Figure 9). Indeed, monsters can kill the player by spiting a sort of fire out of their mouths. But they are fairly slow and easy to avoid if the player can move aside quickly enough. If the fire reaches the player the damage is not serious. However, there is nothing implemented in the rules of *Blake Stone* game concerning other behavior of these green monsters. One might be cued to believe that if too near, those arms could grab the player or that the big mouth actually eats the player. But in the game when the player

Figure 9: Screenshot from the Game Blake Stone (1993)

Source: http://1morecastle.com/2012/06/blake-stone-aliens-gold/

touches the monster none of these expectations are built into the rules of the game. This is how the fictional world can cue the players into making particular assumptions about the rules of the game. Finally, fiction and rules can also be incongruent. For example, in the fighting game *Tekken 6* from all the possible characters or fighters to choose from, there are also nice looking, delicate girls, which seem weaker as other fighters represented as enormous men with muscle all over (Figure 10). However, the rules show that the girl is as strong as any of these powerful guys.

Figure 10: Screenshot from the Game Tekken 6

Source: http://www.pcgameshardware.de

2.1.3. Gameplay: Rules and Players

Having reviewed the importance of rules and the role of fiction in games, a closer look at how these elements operate to produce an interesting gameplay or meaningful play follows. This section describes how rules create an interesting gameplay by applying Salen and Zimmerman's (2003) notions of *constitutive* and *operational* rules together with Juul's (2005) process of *rules at work*. Then how the game design patterns approach uses this knowledge to describe how particular patterns affect specific aspects of gameplay is presented. Finally, how these patterns might be supported by specific AI techniques is exemplified.

Meaningful play emerges when the relationship between the players' actions and the system outcome are *discernible* (e.g., if I catch the blue ghost in *Pac-Man* I will get 200 points) and *integrated* into the game as a system (e.g., if I keep capturing ghosts my score will rise a lot) (Salen & Zimmerman, 2003). The actual actions of the players correspond to the *operational* rules, while the system's response corresponds to the *constitutive* rules of the game. The former refers to the concrete instructions that players should follow and are usually written-out. The latter refers to the formal structures (logical and mathematical) under the "surface" presented to the players. Meaningful play comes then from the tight *coupling* of the players' actions and the outcome of the system. In other words, from the interaction of the constitutive and operational rules of the game (see Table 3). The meaning of playing *Pac-Man* resides then in the tight coupling of action and outcomes, that is, the action > outcome "choice molecule" (Salen & Zimmerman, 2003): the players' action and the game object at a particular moment are stored in the "state machine" of the game, which then produce the particular outcome.

For Juul (2005) on the other hand, rules operate in the following manner in order to bring about interesting gameplay:

- Rules describe what a player can and cannot do and what should happen in response to the player's action (action-events)
- Rules construct a "state machine" (see Section 2.1.1.2) responsive to players' actions.
- The "state machine" provides a branching game-tree of possibilities, so that playing the game is exploring this game tree.

- If players decide to reach for the positive outcome (usually harder to achieve), they face a challenge.
- When players try to overcome the challenges, this activity is usually call "gameplay" and emerges from the interaction of the rules of the game with the players' attempt to play as well as possible (i.e., reaching the positive outcomes of the game).
- The players have a repertoire of skills and methods to overcome the challenges. A good game continually challenges the players' repertoire.
- The players improve their skills while playing the game, in this sense games are learning experiences.

Table 3: Operational and Constitutive Rules of *Pac-Man*

Operational rules[4]	Constitutive rules[5]
Play occurs in a square maze full with pellets, four power pills and four ghosts. The player has three lives. The player guides *Pac-Man* through the maze, eating the pellets/dots while avoiding the ghosts. If the player eats a Power Pill, s/he can eat the ghosts. If the player eats a fruit, s/he gets a bonus score. If all the lives are lost, the game is over.	If an enemy touches *Pac-Man*, a life is lost. *Pac-Man* is awarded a single bonus life at 10,000 points. Four power pills temporarily provide *Pac-Man* with new abilities (to eat enemies). The enemies turn deep blue, and change into the "evade" state and usually move more slowly. When an enemy is eaten, its eyes remain and return to a box in the center of the screen where they are regenerated in its normal color. When a specific amount of time has passed the blue ghosts turn into their original colors, changing their states.

[4] Extracted from http://www.experienceproject.com/play-free-games/Pac-Man/21
[5] A mix from the section 2.1.1.2 on AI and http://en.wikipedia.org/wiki/Pac-Man and http://www.cnbc.com/id/41888021

	Ghosts roam or chase depending on whether they “see” or not *Pac-Man.*
	Blue enemies flash white before they become dangerous again and the length of time for which the enemies remain vulnerable varies from one stage to the next.
	The red ghost pursues *Pac-Man.*
	The pink and blue ghosts position themselves always at a point that is 32 pixels in front of *Pac-Man*’s mouth.
	The orange ghost moves randomly.

Any specific game can be more or less challenging or provide different types of challenges. This means that different games provide different enjoyable experiences. Juul (2005) considered gameplay as how the game is actually played. Gameplay represents the dynamic aspect of games, their interactivity. Where does gameplay come from? As a consequence of the game rules and the disposition of the players. According to Juul (2005), gameplay results from the interaction of three things:

- The rules of the game
- The players’ pursuit of the goal of the game, seeking actively strategies that work
- The players’ chosen repertoire of strategies and skills

  And as a consequence:
- The strategies used are enjoyable to execute.

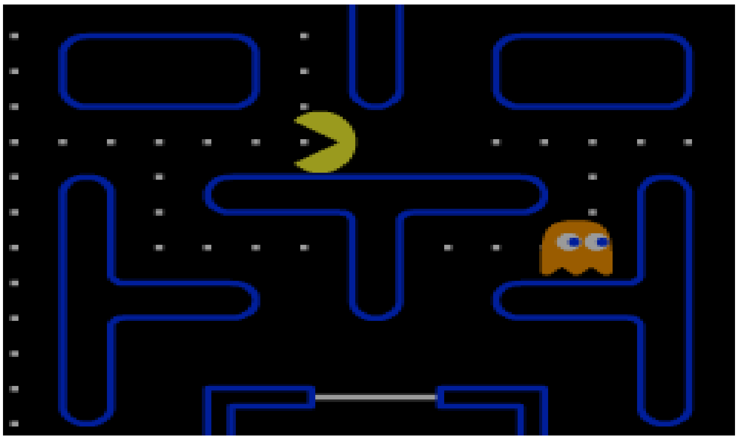
In order to put these two concepts of meaningful play and gameplay together, it is central to see the parallels of both. The “state machine” mentioned by Juul (2005) belongs to Salen and Zimmerman’s (2003) constitutive rules of the game, while the players’ actions are performed based on the operational rules of the game. It is true that the same constitutive rules may take different operational rules, but its specific coupling makes possible the emergence of meaningful play or gameplay. The coupling of these type of rules supports the action <outcome “choice molecule”. Table 4 shows a dissection of these chains of choices as they occur in *Pac-Man*.

Clearly, *Pac-Man* fulfill successfully with all the stages of the choice process linking action and outcome. In Figure 11 the player sees that the orange ghosts comes upwards

Table 4: Design Issues in each Stage of Choice in *Pac-Man*

Anatomy of choice	*Pac-Man*	Design problem associated
1. What happened before the player was given the choice? (internal event)	It is represented by the current trajectories of the ghosts.	
2. How is the possibility of choice conveyed to the player? (external event)	The possible actions are conveyed through the arrow keys and the screen indicating the relationships among the ghosts and *Pac-Man.*	Not knowing what to do next…
3. How did the player make the choice? (internal event)	The player presses one of the four arrow keys.	Not knowing if an action has an outcome.
4. What is the result of the choice? How it will affect future choices? (internal event)	Each arrow key affects the system differently by positioning *Pac-Man* in different locations within the maze.	
5. How is the result of the choice conveyed to the player? (external event)	Via the screen graphic and audio.	Losing a game without knowing why…

Figure 11: Screenshot Depicting the Action> Outcome "Choice Molecule" in *Pac-Man*



Source: Screenshots online Pac-Man http://www.gamesbasis.com/pacman.html and will turn to the right (Stage 1 and 2 of the anatomy of choice), then the player presses the down and left arrow to go to the left (moving to the right would have likely meant to crash with the ghost) (Stage 3). *Pac-Man* moves toward the power pill (Stage 4). Finally, *Pac-Man* is away from the orange ghost as intended. In this manner, *Pac-Man* provides action and outcomes that are discernible and integrated, that is, it provides for meaningful play. In Juul's (2005) language the gameplay just described emerges from the choice of the player to pursue the positive outcome of *Pac-Man* (i.e., get points and achieve a high score). To achieve this goal, necessary sub-goals should be pursued such as avoiding colored ghosts. As ghosts are state machines, they have a programmed behavior that changes depending on their spatial relationship to *Pac-Man* and whether or not they are blue. Among the strategies to avoid the ghosts are the use of the two tunnels in the middle of the maze, the skillful use of the corners and also waiting for the ghosts near a power pill.

2.1.4. Games as Simulations

The field of Game Studies has defended itself from being studied under the same paradigm of other media such as film, books and television. The classic paradigm for these media has been called "narrativism", the belief that everything is a story (Aarseth, 2004). Against this tendency, some theorists have reacted and searched for other modes of discourse about games. For Aarseth (2004), the hidden structure behind the pleasure of games is not narrative or interactivity, but *simulation*. According to Aarseth, the computer is the art of simulation and all computer games contain simulations. In contrast to stories that are top-down and pre-planned, simulations are bottom up and emergent during gameplay. Interestingly, for Frasca (2001), it is narrative that works bottom up and simulations are top-down: "…narrative works in a bottom-up sequence: it describes a particular event from which we can generalize and infer rules (this is why narrative is used so much in education). On the other hand, simulation is usually top-down: it focuses on general rules, which then we can apply to particular cases". In any case, the important point is that both authors agree on the centrality of simulation in computer games. For Frasca (2001), the introduction of the computer has allowed simulations new ways of portraying both reality and fiction, beyond the traditional "representation as depiction" typical of other cultural forms such as narrative, literature, theater or cinema. Simulations, contrary to narrative, have the ability to not only represent systems, but also *model* their *actual behaviors*. In this manner, game theorists have found an alternative mode or rhetoric to describe games beyond the more common idea of games "telling stories". Even though these issues are still open, they provide rough background to understand the interest of an influential part of the game community on the idea of simulation.

Uriccio (2005) reviewed the term "simulation" and suggested that its meaning evolved from "false pretense/deception" to a tendency to "resemblance" something else, and more lately – after World War II – to a tendency of modeling the dynamic behavior of a situation or process, analogous in relationships and with a pedagogical goal. As Uriccio writes: "Unlike a representation, which is fixed in nature, a simulation is a process guided by certain principles" (p. 333). He argued that simulations are able to generate countless encounters that later on can be "fixed" by means of an image or narrative. In other words, simulations can contain narratives, but not the other way around. Uricchio then pointed out that simulations have their own history independent of video games. This conceptualization is closer to the one usually used within the educational community (see Section 2.2.1.1).

In concrete terms, within the field of games simulations have been understood as the "act of modeling a system A by a less complex system B, which retains some of A's original behavior" (Frasca, 2001, p. 3). Frasca's definition is attempting to differentiate between a simulation and a representation. For the author, what is key of a simulation is its ability to replicate some behaviors of the system modeled. He put as an example Sim City. Frasca suggested that Sim City can be considered as a dynamic system which "behaves" like a city. On the other hand, a painting of a city can only show its "fixed" characteristics, but cannot reflect how the city functions, that is, its behavior. Similarly, Salen and Zimmerman (2003) defined a simulation as "a procedural representation of aspects of 'reality'" (p. 423). For Salen and Zimmerman, simulations are located between representations and dynamic systems. Both simulations and games create representations, but according to the definition a particular kind of representation, that is, a *procedural* representation. With procedural the authors meant an "on-going" process that emerges from the interaction of the players with the game. Therefore, in games, simulations are built from a set of procedures, behaviors or forms of interaction. As a summary, an exercise can be a simulation if:

- The exercise represents an actual situation of some sort, from real life or imaginary (e.g., an extraterrestrial being visiting the earth).
- The exercise must be operational, that is, it must constitute an on-going process.

From the first criteria it follows that a simulation refers to something in the real world. However, a simulation cannot depict all the features of what it attempts to simulate, but a subset of characteristics on which the procedural representation is built. Salen and Zimmerman (2003) also distinguished between game and non-game simulations. According to the authors there are of course simulations that are not games, but they assert that any game, by fulfilling the two criteria above (i.e., procedural representation of aspects of reality) is a simulation. In this sense, any game is a *simulation game*. This term is normally used to depict a particular genre that includes video games that "simulate sports, flying and driving, and games that simulate the dynamics of towns, cities, and small communities" (Apperley, 2006, p.11). The simulation genre "remediates" common activities or the depictions of these activities on other media such as cinema or television. However, Salen and Zimmerman (2003) go beyond games "remediating" common activities and see, for example, Tic-Tac-Toe, as a simulation game: Tic-Tac-Toe can be "framed as representations of territorial conflict, in which simulated units war for control of a stylized battlefield."(p .424). It should be acknowledged though that seeing Tic-Tac-Toe as a territorial conflict and as a "stylized" battlefield requires a considerable amount of imagination. However, the key here is the word "stylized". How stylization works in games was described later by Juul (2005):

> low-fidelity simulation is when the player enters a car in Grand Theft Auto III. Simply being near the car and pressing △ on the Playstation 2 controller makes the protagonist run to the nearest car door, open the door, remove any person in the car, get in, and close the door…we are unlikely to feel any significant loss here, since entering a car is generally not considered a very interesting activity. (Juul, 2005, p. 170).

This process represents a simplification, a stylization that can be found in games with fictional worlds (see Section 2.1.2). Juul (2005) also acknowledged that games' fiction and rules are not a perfect simulation of the real world, but provisional in nature. By using the idea of simplification as implemented in the world of comics, Juul suggested that games focus on a specific idea of what a game is about (e.g., racing, tennis, etc.): "A game does not as much attempt to implement the real world activity as it attempts to implement a specific stylized concept of a real-world activity…" (Juul, 2005, p. 172). In the case of *Grand Theft Auto*, the simulation of the activity of getting into a car is not interesting in the context of the whole game. In this sense, simulations in games are oriented toward the *perceived interesting parts* of being a criminal who steals cars. The author also suggested that video games are metaphors of activities in the world, such as playing tennis. The author asked himself for the connection between serving in real tennis (with all the complexities associated to concentration and precise body movements) and serving in a game (by pressing a button). He concluded that both activities are *difficult* and games usually substitute one difficult task for another, meaning that "in games that emphasize a fictional world, there has to be a metaphorical substitution between the player's real-world activity and the in-game activity performed." (Juul, 2005, p.173).

In summary, simulations have been evoked as alternatives ways to study games beyond narrative and representation. Against this background, simulations entail an *on-going process, an action and activity experienced as procedures that attempt to represent some (interesting) aspects of reality in a stylized manner*. Therefore, all games can be portrayed as simulation if one allows the term "reality" to mean almost everything imaginable and if one keeps in mind that games are "stylized" simulations that simulate only the interesting parts of what is being simulated. The issue of simulation is later discussed in the context of educational games. The definitions are similar, but the issue of what "reality" entails, is what, it can be argued, distinguishes both perspectives (see Section 2.2.1.1.).

2.2. Educational Games: What are they?

Educational games emerged in the context of PLATO[6] (Tobias & Fletcher, 2011). The project PLATO attempted to solve the problem of transmitting huge amounts of

[6] For more information about PLATO, see http://faculty.coe.uh.edu/smcneil/cuin6373/idhistory/plato.html

information to increasingly huge number of students through effective, expandable and inexpensive means. In other words, through technology (Bitzer & Johnson, 1971). Within PLATO, Dugdale and Kibbey (1975) designed *Darts*, an educational game for teaching fractions, which provided the setting for the first systematic research on the intrinsically motivating features of games (Malone, 1981; Malone & Lepper, 1987). Malone's research triggered the interest of other scholars on the potential of games for skills' cultivation (Loftus & Loftus, 1983; Greenfield, 1984) and later for learning and instruction (Gee, 2003; Habgood, 2007; Prensky, 2001a).

The bulk of research on games and learning can be organized mainly in two groups of scholars. One group sees games as fostering 21st century skills which are difficult to acquire in schools, and are mostly interested in studying COTS (i.e., Commercial off the shelf software), such as *World of Warcraft* (e.g., Steinkuehler, 2006), in more informal settings and whose sources of inspiration come from scholars such as James Paul Gee (2003) embracing a socio-cultural (e.g., Vygotsky, 1978) or situative perspective (Greeno, 1998) to teaching and learning. Another group is interested in harnessing the engagement of games to support curricular and subject-matter knowledge acquisition whether in the context of schools or at home. They are interested in the instructional design of games that can secure the quality of the engagement already happening around COTS while explicitly addressing learning goals related to specific subject-matters such as Biology, Ecology, Algebra, Physics, History, etc. However, it has been recognized that the quality of the educational games produced by this group are modest and it can be argued that they are seldom games in the first place.

In fact, the key lesson from the previous section is that in order to better understand the field of educational games and what these games are or should be they need to be primarily *games* (see Fortugno & Zimmerman, 2005). Even though this might sound obvious, in actuality, the myriad of "questionable" educational games have led some scholars to ironically call these products either *shavian reversals* (Papert, 1998), *chocolate-dipped broccoli* (Bruckman, 1999), or simply *crap* (Brenda Laural as cited in Fortugno & Zimmerman, 2005). Others have looked down upon these games and grouped them under the general label of *edutainment*, that is, "bad" games characterized by drill and practice, simple gameplay, curriculum oriented and behaviorist-based principles as opposed to more "experiential" oriented games (Egenfeldt-Nielsen, 2005). This critique is only valuable to the extent that considers the quality of gameplay. Its weak side concerns the reference to the curriculum and behaviorist approaches. In themselves these two aspects do not seem to be necessarily "bad" than other alternatives. In fact, from a pedagogical perspective, none of the theories, that is, behaviorism, cognitivism or social constructivism are better by themselves. Kerres, Ojsterstek and Stratmann (2009) pointed out that the quality of a specific learning delivered depends more on the adequacy of the conceptual solution to the specific demands of the learning situation than on the particular learning theory on which the learning delivered is based. As Kerres et al. writes: "Insofern hängt die Qualität des Lernangeboten auch nicht davon, ob ein bestimmtes lerntheoretisches Modell verfolgt wird(…) Es kommt vielmehr darauf an, die richtige konzeptuelle Lösung für genauer zu spezifizierende Anforderungen einer Lernsituation zu finden" (Kerres et al., 2009, p. 265). Therefore, the quality and effectiveness of an educational game is not determined by its explicit theory of learning.

What follows is an overview of the main issues in the field of educational games. First a brief description of the different manners in which different stakeholders use the term "game", "educational game" or "game for learning" is provided and their proximity to simulations is presented. Then a discussion of the main arguments and assumptions behind the interest on the use and design of games with learning purposes is presented. Next, the main approaches to the design of educational games are discussed, followed by a critical revision of Malone's (1981) seminal piece on games for learning. The section closes with a summary of the main features and limitations of current research on educational games, highlighting the gaps that this dissertation conceptually and empirically attempts to fill.

2.2.1. Definitions of Educational Games

It has been pointed out that the field of educational games seems fragmented and with lack of consensus concerning a definition of what an educational game is, which arguably could preclude the field to advance in the systematic accumulation of knowledge (Honey & Hilton, 2011). Definitions are important to the extent that they allow the communication among specialists, the research with a distinct objet of study, and, ultimately, the design of these technologies for enhancing learning (Clark, 2007; Honey & Hilton, 2011). As shown in the previous section, trying to reach a definition of what is a "game" is a highly challenging endeavor given the many uses of the term "game" and "play". In the educational arena things can be even more complicated because of multiple reasons: the mix of concepts with different meanings, such as “educational”, “instructional” and “learning”, misleading assumptions concerning the term “game” (e.g., games as spontaneous or free activity, games as immersive experiences, games as having “explicit” rules, etc.), and play and game used as synonyms. An example of this last point can be found in Shaffer’s (2007) book *How Computer Games Help Children Learn*. Shaffer attempts to define games, but he never gets to a definition, and continuously mixes play, game and learning. In other words, he mixes the three different areas identified by Salen and Zimmerman (2003) and Juul (2005) (Section 2.1), and that are fundamental to start defining the concept. More interestingly, in a following chapter Shaffer continues with a description of how children learn the notion of *center of gravity* through a “game” that really looks more like a simulation (concerning issue of games and simulations see sections 2.1.4 and 2.2.1.1).

For other researchers the problem lies on contradictory definitions of games. For example, de Freitas (2006) exemplified the diverse definitions of games by naming Salen and Zimmerman (2003), Caillois (1961) and Huizinga (1949). However, as shown in previous sections, Juul (2005) and Salen and Zimmerman (2003) have already distilled all the possible definitions and have come out with a synthesis, so that the contradiction of multiple definition is not entirely accurate. In other words, the purpose of the work of these game theorists and designers was precisely to shed light on the contradictory nature of the concept of games and provide an extensive discussion and a meaningful definition of what is a game. Secondly, de Freitas (2006) considered Huzinga’s (1949) definition a problem for the field of serious games:

> Huizinga defined games as a free activity standing quite consciously outside ‘ordinary life’, as being ‘not serious’ (1980), following this definition games cannot be serious. Caillois (1961) similarly defined games as voluntary and therefore also conflicts with the notion of serious games. (de Freitas, 2006, p. 10).

As discussed in the Section 2.1 neither Caillois (1961) nor Huizinga (1949) defined a “game” itself, but the activity of “play”, which is not the same. It has also already mentioned how the idea of “free play” does not apply to games simply because games are rule-based system with rules being *fixed* and *binding*. Finally, de Freitas (2006) chose to see games as “immersive worlds”, which does not help in defining the object of study of the field of educational games, given that the notion "immersive" is problematic and ambiguous (cf. Callejas, 2007; Salen & Zimmerman, 2003).

With these examples as background Table 5 provides a sample of definitions of games, educational games and serious games. It can be seen that the definitions vary in terms of how close they match the definitions of games discussed in section 2.1. For example, Randel et al. (1992) included most of the elements of a definition of game, such as competition (in the sense of conflict), rules, goals to be achieved and the role of skills to achieve those goals. However, they included chance as part of the definition, which in Juul's (2005) position, games are about acceptance, commitment and effort investment with the goals and rules of the game. The definitions also vary in terms of the substantive that define a game. For example, most of the authors have considered games as either an "environment" or an "activity" which is then qualified by diverse adjectives such as interactive, competitive or immersive. These qualifiers, except for "competitive", have no clear meaning and their role in games is still being discussed. On the other hand, only five definitions predicate of games to have or to be based on rules or constraints. Finally, the most important point to notice here is the lack of explicit reference to the notion of system and conflict, which are central to understand what makes games a unique medium and, therefore, how they should be appropriately designed. This state of affairs makes difficult for the field to achieve the necessary coherence to start building a knowledge base that could describe how one learns from an educational game and under what circumstances and for whom a particular educational game could be more effective. It also makes difficult to operationalize the concept of "educational game" in empirical studies (Clark, 2007) so that it does not overlap with simulations, virtual or multimedia environments.

Table 5: Definitions of Games and Educational Games from the Educational Research Field

Author	Definition
Randel et al. (1992)	"Games are competitive interactions bound by rules to achieve specified goals that depend on skill and often involve chance and an imaginary setting" (p. 262).
Dempsey et al. (1994)	"…a set of activities involving one or more players. It has goals, constraints, payoffs and consequences. A game is rule-guided and artificial in some respects. Finally, a game involves some aspect of competition, even if that competition is with oneself" (p. 6).

Gredler (2003)	“Games are competitive exercises in which the objective is to win and players must apply subject matter or other relevant knowledge in an effort to advance in the exercise and win. An example is the computer game Mineshaft, in which students apply their knowledge of fractions in competing with other players to retrieve a miner’s ax” (p. 571).
Kirriemuir & McFarlane (2004)	“For the purposes of this report, we will define a digital game as one that: • provides some visual digital information or substance to one or more players • takes some input from the players • processes the input according to a set of programmed game rules • alters the digital information provided to the players” (p. 6).
Egenfeldt-Nielsen (2005)	“Educational computer games: Computer games developed for educational use or titles often finding their way to educational settings both the fake, bad, ambitious and superb. This includes edutainment but is not limited to it. In this dissertation educational computer games often implicitly exclude edutainment when used” (p.23).
Zyda (2005)	“Serious game: a mental contest, 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” (p.7).
Hays (2005)	“A game is an artificially constructed, competitive activity with a specific goal, a set of rules and constraints that is located in a specific context” (p.15).
de Freitas (2006)	“Applications using the characteristics of video and computer games to create engaging and immersive learning experiences for delivering specified learning goals, outcomes and experiences” (p.9). “Educational game: Games for learning are often imaginary (e.g. fantasy) interactive and immersive environments in which role play, skills rehearsal and other learning (e.g. collaborative or problem-based) may take place individually or in teams” (p.69).
Federation of American	“Educational games are fundamentally different than the prevalent instructional paradigm. They are based on challenge, reward, learning

Scientists (2006)	through doing and guided discovery, in contrast to the "tell and test" methods of traditional instruction" (p. 6).
Habgood (2007)	"…an interactive challenge on a digital platform, which is undertaken for entertainment" (p.18).
Klopfer et al. (2009)	"Learning Games are differentiated from Games for Training in that they target the acquisition of knowledge as its own end and foster habits of mind and understanding that are generally useful or useful within an academic context" (p. 21).
Pavlas (2010)	"A serious game is a representative task that harnesses play to convey knowledge, skills, or attitudes to one or more learners" (p. 8).
Sigmund & Fletcher (2011)	"Our emphasis is on a subset of simulations – interactive, computer-based games and the learning environment they create" (p. 6).
Leemkuil & de Jong (2011)	"Games are competitive, situated, interactive (learning) environments based upon a set of rules and/or an underlying model in which, under certain constrains and uncertain circumstances, a challenging goal has to be reached" (p. 355).

2.2.1.1. Simulation and Games: Simulating a Model versus Playing a Game

The simulation literature have had in the past the same concern as this dissertation: in order to understand a technology and assess its effectiveness, it is necessary first to know what this technology actually is, and second *what exactly happens* when this technology is being used (Crookall, Oxford, & Saunders, 1987). Similarly as the game community, this section attempts to search for what differentiates a simulation from a game to better understand what a *simulation game* is. But first, some commonalities between simulations and educational games in terms of research and the rationale behind their use are provided.

Simulations have faced in the past the same limitations that face educational games today. For example, Gredler (1996) mentioned a few of the problems with simulations that easily resonate with that of educational games. She pointed out that the lack of learning principles supporting the design of simulations has led to "truncated" exercises often called erroneously "simulations". Similarly, the lack of well-designed research that could overcome persistent weaknesses on simulation research related to measurement and design. Concerning measurement, Gredler mentioned the failure of the studies to describe the nature of the tests used to measure student learning. Concerning design, the comparison of simulations with regular classroom was considered problematic because both methods might be effective for different instructional goals, and because studies are usually insensitive to individuals characteristics and their interaction with instruction (e.g., prior knowledge and ability), and more importantly, studies normally have failed to document *how students interact* with the subject matter through a simulation. The same drawbacks seem to apply to educational games (e.g., Honey & Hilton, 2011).

The reasons for using simulations for education and training are also similar to the ones stated in relation to games for learning. For example, de Jong (1991) summarized a set of reasons that can be classified as referring to the learning and motivational process, learning outcomes and economic and practical reasons. The first category involved the learning processes expected to be evoked by simulations such as hypothesis generation and testing, planning and monitoring together with the general motivational appeal of simulations. The second category, learning outcomes, entailed enhancement of cognitive learning of factual information, better understanding of processes, improvement of critical thinking, transfer, and the development of positive attitudes, among others. Last but not least, economic and practical reasons related to how some real situations might be too expensive to implement, too dangerous or simply too time consuming. More succinctly, Crookall et al. (1987) gave three main reasons for using simulations in education: 1) simulations are motivating and fun; 2) they are more congruent with the learning processes; and, 3) they are more like the "real world".

When reviewing some attempts of defining what a simulation is, it is clear that the concept of *model* and *system* are an important part of the definitions. For example, Randel et al. (1992) also pointed out that a simulation models a certain process or mechanism by relating changes in the input to changes in the outcomes within a "simplified" reality that usually do not have an end point. Similarly, for Tobias and Fletcher "simulations model a process or mechanism determined by a specific algorithm; they usually incorporate a system to model complex processes that range from routine to extreme situations" (Tobias & Fletcher, 2011, p. 7). Then the authors considered games as simply a "subset" of simulations and provided the following list of differences:

Table 6: Comparison between Simulations and Games by Tobias and Fletcher (2011)

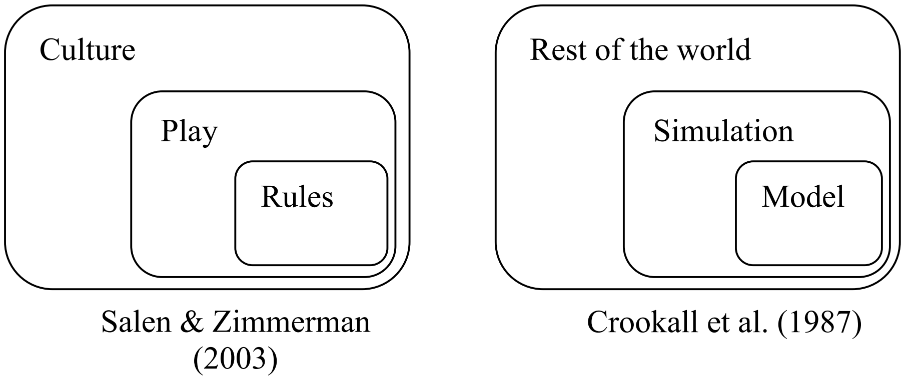
Simulations	Games
Sacrifice entertainment in favor of reality	Sacrifice reality in favor of entertainment
Scenario/tasks	Storyline/quest
Emphasis on task completion	Emphasis on competition
Not necessarily interactive	Necessarily interactive
Focus on (rule) accuracy/detailed	Focus on (rule) clarity/stylized
Not all simulations are games	All games are simulations

Although some statements seem obvious, such as the first one related to “reality” versus “entertainment”, other seems to be more problematic, such as the statement concerning the degree of interactivity or, the last statement “All games are simulations”. It is this last point that needs further elaboration. However, some other definitions do not help that much. For example, Leemkuil and de Jong (2011) stated that “Simulations are also based on a model of a (natural or artificial) system or process.” (p.355). This definition is problematic because it implies that games are also a model of something. Whether this is true or not, will depend on the meaning of the word “model”. Other definitions are simply too broad, and use terms whose generality and multiple meanings make impossible to pin down the nature of a simulation, or introduce vague terms that do not allow for a clear definition of what a simulation is. For example, Lee (1999) defined a simulation as “a computer program in which it temporarily creates a set of things through the means of a program and then relates them together through cause and effect relationships” (p. 72). Similarly, for Rieber (1996) a simulation is an attempt to mimic a real or imaginary environment or system and distinguished between scientific and educational simulations. Scientific simulations are used by scientist to study particular systems of interest, while educational simulations are designed to teach individuals about the system of interest by experiencing the consequences that different actions have within the simulation. In this definition, the idea of “imaginary” seems to be at odds with the basic idea that simulations entail a sort of fidelity to something in the real world.

A more clarifying attempt at defining simulations and differentiating them from games was done by Crookall et al. (1987). The authors distinguished two views of simulations:

the representation view and the “reality” view. The former is the more traditional way of understanding a simulation and basically correspond to the idea that simulations are representations of “real world” systems. The latter, correspond to the view that simulations, from the user perspective, are operating realities in their own right, that is, that users experience them as real. Crookall et al. defined a simulation as a special kind of model *representing* a “real” system. The logical question here concerns the difference between a model and a simulation. It can be argued that the distinction made by Crookal et al., resembles that of Salen and Zimmerman (2003) and Juul (2005) when they differentiated between the game as a rule-based system and the playing of the game, or the relation between the game and the player (Figure 12).

Figure 12: Comparison of Theoretical Frameworks Describing Games and Simulations



As with the case of games, the Model represents the set of fixed rules (i.e., the “thing” or the “noun”), while the Simulation represent the Play (i.e., the “activity” or the “verb”). To some extent “simulating a model” would be the equivalent of “playing a game”. These considerations raise two questions: What does it mean that simulations are a “special” kind of model? And what does it mean that the model “represents” a real system?

According to Crookall et al. (1987) a model is a “map” used to both *represent* features of a real world system and to reduce the *cost of error* within the system. To represent can mean either to “depict” or to “bring to life”. A model only “represents” in the sense of “depiction” and not in the sense of “bringing to life”, which is what a simulation can do. A model embodies a theory of the system it maps and therefore abstracts its features by selecting specific features to be mapped from the system into the model, simplifying the features and their relations. Then the abstract features in the system become symbols in the model. Therefore, a model is a *symbolic representation of a system*.

The cost of error of a model or simulation refers to the fact that risks and their consequences remain within the system and have little impact on the "real-worldly" system represented. Neither models nor simulations are risk free. They invite risk taking, exploration, and "what if…" type of questions, so that individuals can gain confidence through practice. In brief, simulations protect individuals from the consequences of their mistakes.

If models and simulations both represent a system with low error consequences, what makes a model "special" so that it turns it into a "simulation"? Systems, in general, consist of processes and behaviors governed by fixed rules that can yet provide with flexible and variable strategies to be performed by an individual. A model (as a simulation), contains a set of *rules*, but does not consider or contain the variable *strategies* of simulation. This means that a model cannot be "operated" by an individual or "brought to life", that is, a model cannot be made to function in a similar way as the system it represents. In the case of a simulation, individuals operate it and apply all kinds of strategies to it. In this sense, a simulation can be made to function like the real system it represents. This is why, roughly speaking, an individual engaged in a simulation can be said to be "simulating" a model. A simulation is *potentially* the "live" part of a model. For example, a flight simulator is just a model of how things fly. But when used by a pilot and responds to the pilot actions/strategies, then it becomes a simulation. The same is true for the case of games. A game is a rule-based system and when individuals use it to achieve the goals of the game, they are *playing* the game.

Crookall et al. (1987) also proposed that the criteria of *representivity* and *error consequences* are useful to differentiate between simulations and games, given that the common features of games are usually present also in simulations. They stated that in terms of these two criteria, games have the contrary effect of simulation, that is, a game is not intended to represent any real-world system, and the consequences of game errors can be high in the real world. As Crookall et al. put it "a game is a formalized system in its own right, while a simulation is a formalized representation of another system; a game is a "real" system, a simulation a meta-system" (Crookall et al., 1987, p. 161). This might be confusing with the definitions of game already provided (see Section 2.1). For Salen and Zimmerman (2003), games are an "artificial" conflict in the sense of being something different to the real world. But they also pointed out that games are a system. It seems then that the difference is that the system in games is "artificial" in the sense that it has little to do with some real referent in the world (e.g., *Pac-Man*). *Pac-Man* is a system in its own right, but does it have a referent in the real world? Probably not. On the other hand, a simulation is a model of a system with a real referent in the world (e.g., Flight simulator). A Flight simulator has a model of the "system of flying objects". In this sense, it is possible to imagine the possibility of making a simulation out of a game, that is, abstracting key features of this system – the game – into a model and then "bring it to life", that is, simulate it. Crookall et al. (1987) give an example: soccer. This highly popular sport can be simulated in a computer. However, as Juul (2005) already pointed out in this case one is in front of a "stylized simulation". In brief, for Crookall et al. (1987), even in the case of simulation/games with "heavy" game components and low representivity, the activity should be regarded as a simulation if "individuals' perceptions involve inferences to 'real world' referents" (p. 162). Interestingly, the authors seem to consider a game as being part of everyday world, such as the game of soccer, while a simulation for the authors is a "bracket, a hiatus, within the ongoing 'real-world'" (Crookall et al., 1987, p. 161). They further suggested that games, as opposed to simulations are a "full-fledged part of life, a sub-system in its own right embedded in the everyday life systems of the 'real world'." This is why they suggest that cost of error is higher in games, that is, it can potentially have costly consequences in "real life". They put as an example the game of *Poker*. Another example is the Olympic Games. However, the same might be true when one is using a simulation, say a truck simulation, as a training tool in order to get a license to drive heavy vehicles. It is easy to imagine the consequences of making a mistake within this simulation, for example, not to get a good evaluation of the performance on the truck simulation and being advised to try again. Therefore, the idea of cost of error for differentiating between simulations and games is not completely appropriate. On the other hand, distinguishing simulations and games based on whether or not the individuals engage in considerations related to "real world" referents might be confusing. It should be considered the level at which both technologies are being compared. It can be argued that simulations here refer to the activity by which individuals deploy a set of strategies afforded by the model behind the simulation. And games here refer to the level of "rules" as such, not involving the activity of play (see Figure 12 above).

In summary, it can be suggested that when comparing simulations and games, the issue of representivity is the most appropriate criteria for distinguishing the two: games are artificial system in their own right, while simulations represent a real world system. In the case of games being simulated in a computer, as in the case of soccer, it is more appropriate to use the term "stylized simulation". In terms of what a "simulation game" could be, according to Crookall et al. (1987), it is simply a case in which a game that is part of everyday life, such as soccer, is implemented or "simulated" in a computer program.

Later, Gredler (2003) provided a broader notion of simulation than that described above. For her simulations are "evolving case studies of a particular social or physical reality. The goal, instead of winning, is to take a bona fide role, address the issues, threats, or problems arising in the simulation, and experience the effects of one's decisions" (p. 573). This definition, as opposed to that of Crookall et al. (1987), emphasized what individuals are supposed to do and the general attitude they are expected to have in these exercises. Gredler (1996) also proposed some criteria for differentiating academic games and instructional simulations. For her both are "experiential exercises" that allow individuals to interact with a knowledge domain. In order to better understand the nature of both, simulations and games, she coined the concepts of *surface* and *deep* structure. The former refers to paraphernalia and observable mechanics such as moving pieces in a board (in games) or a set of data to be addressed (in simulations). The latter refers to the psychological mechanisms that operate during the experiential exercise. In terms of this deep surface, simulations and games have similarities and differences. Concerning similarities, the author pointed out that both, games and simulations, transport individuals to another world and provide environments in which the individuals are in control. However, they differ in at least three aspects: 1) in games the goal is to win, while in simulations the goal is to execute serious responsibilities; 2) games have a linear sequence, while simulations' sequences are not linear (e.g., branching); and, 3) games consist of rules that provide constrains and privileges, while simulations consist of dynamic set of relationships among several variables that change overtime and reflect authentic causal processes.

According to Gredler (1996), in a simulation participants take on either (1) demanding roles (e.g., concerned citizens, business managers or physicians) or (2) professional tasks (e.g., exploring the causes of water pollution or operating a complex equipment system). Simulations present the property of branching, that is, it allows individuals to face diverse decision points with particular problems, issues or events mainly resulting from individuals' prior actions within the simulation. In addition, she considered a simulation to be based on a dynamic set of relationships among several variables that (1) change over time and (2) reflect authentic causal processes. As a summary, simulations include the following features:

- An adequate model of the complex real-world situation with which individuals interact
- A defined role for each individual with responsibilities and constraints
- A data-rich environment allowing the execution of a range of strategies
- Feedback for individuals' actions in the form of changes in the problem or situation

Gredler (2003) also distinguished between two types of simulations: Experiential Simulations and Symbolic Simulations. The main difference between them, as elaborated below, is that in the former individuals are functionally connected to the simulation and therefore their actions have powerful contingent consequences, while in the latter individuals are external to the simulated situation and, therefore, a mindless set of action from them does not bring about the same consequences. Similarly, de Jong and van Joolingen (1998), distinguished between Conceptual Simulations and Operational Simulations. The former is usually used in business and physics and contains principles and facts related to the system, the latter contains sequences of cognitive and non-cognitive operations (procedures) applied to the simulation (Table 7).

In particular Experiential Simulations are "social microcosms" where learners interact with real-world scenarios in which they enact specific roles within an evolving situation. The key components of an experiential simulation are:

- A scenario of a complex task or problem that unfolds mainly in response to learners actions
- A serious role taken by the learners in which they enact the responsibilities related to their roles within constraints
- Multiple paths through the experience
- Learners are in control of the decision making process

Individuals engaged with an experiential simulation have to apply their knowledge base to solve a complex situation in which they are a component of the simulation. The situation within the simulation evolves and changes partly due to individuals' actions.

Table 7: Types of Simulations by Gredler (2003) and de Jong and van Joolingen (1998)

	Classification of Simulations	
Gredler (2003)	*Experiential simulations* Developed to provide interactions in situations otherwise too costly or hazardous in a real-world setting. Individuals are immersed in a complex, evolving situation and are a functional component of the simulation. Key is the fit between the experience and the reality it represents (fidelity or validity). Consequences for individuals' actions include changing other participants' actions or the whole task (random strategies bring about strong consequences) Examples of high-fidelity simulations are pilot and astronaut trainers	*Symbolic simulations* Dynamic representations of the functioning or behavior of some system, process or phenomenon by the computer. Represent a set of events or processes external to the individual, who tests from outside her conceptual model of the relationships among the variables in the system Individuals are expected to interact with the symbolic simulation as a researcher or investigator, but the exercise cannot divert the learner from the use of random strategies (no strong consequences)
de Jong and van Joolingen (1998)	*Conceptual models* They hold principles, concepts, and facts related to the system being simulated. Examples usually in economics and physics.	*Operational models* They include sequences of cognitive and non-cognitive operations (procedures) that can be applied to the simulated system (e.g, radar control tasks)

On the other hand, a Symbolic Simulation is the dynamic representation of the functioning or behavior of a particular universe, system, set of processes, or phenomena into another system (e.g., the computer). The key features of symbolic simulations are:

- They involve the interaction of two or more variables over time
- Individuals function as investigators by testing their conceptual model of how the multivariate nature of the system simulated works
- The purpose is to discover scientific relationships, explain/predict events and confront misconceptions
- No specific role that commits the individual to the outcome of the simulation is provided

This conceptualization is quite consonant with Crookall's et al. (1987) definition of a simulation as a representation, in terms of "bringing to life", of a real system into a model. In science education, symbolic simulations are used in the context of discovery learning and are usually an alternative to expository instruction or hands-on laboratory exploration. de Jong & van Joolingen (1998) considered conceptual simulations as the most common in discovery learning contexts and left the operational ones as more appropriate to more "experiential learning". Leemkuil and de Jong (2011) also stated that their operational systems and Gredler's (2003) experiential simulations are closer to the notion of games than the other categories (i.e., conceptual and symbolic simulation). This goes in line with Salen and Zimmerman's (2003) idea that a game is a simulation if it entails a procedural representation of aspects of reality (Section 2.1.4).

An important point addressed by Gredler (1996) is the issue of mixing games and simulation and the difference between simulations and "other technology-based exercises". Concerning the issue of "simulation games", these may send contradictory messages to individuals. That is, given that games are competitive exercises with a clear goal of winning, and experiential simulations are interactive exercises where individuals take serious roles, they represent different psychological realities and mixing them can be confusing. In these types of mixed exercises "students will tend to enact those behaviors that are reinforced by winning. In simulation games "these actions may be counterproductive to the expected learning" (Gredler, 1996, p. 532). As a case in point, she described a simulation game called *Business Policy Game*, where the winning firm is the one with the highest return of investment. Some behaviors observed in the participant teams were 1) trying to "crash the system" and preventing others from winning (a typical behavior of game players), 2) charging huge amount of money for some product to catch up with the more advanced firm, and 3) others just wanted to show their prowess. These behaviors are a sample of conducts in which "the focus on being 'the winner' distorts the simulation experience" (Gredler, 1996, p. 532).

2.2.1.2. Defining an Educational Game

After having reviewed the definitions of games, educational games and simulations, the present dissertation conceptually define an educational game as:

> *Digital applications consisting of a fictional world and a character enclosed by a rule-based system that provides a hierarchy of goals to instigate on individuals the voluntary investment of mental effort and deep processing of information in order to acquire knowledge and skills, and preparing individuals to apply those knowledge and skills in future real life situations.*

As it will become clearer in section 2.4, this definition considers a game to be educational if supports individuals' cognitive engagement (i.e., mental effort and deep information processing). The meaning of "rule-based system" and "fictional world" was already described in sections 2.1. The idea that a hierarchy of goals (cf. challenges or obstacles) instigate cognitive and thinking processes was taken from Dewey's (1913) characterization of the experience of effort in section 2.4.1 and the effort and commitment players invest to achieve an outcome (Juul, 2005). This definition represents an alternative to current definitions (see Section 2.2.1) that define this medium as an "activity" without describing the properties of such activity or refer to the medium in broad terms such as "environment" or "immersive tools". It also makes clearer the difference with virtual/immersive worlds, multimedia environments and symbolic simulations (see Section 2.2.1.1). The present definition implies also a particular way to understand the process of learning, beyond any specific technology, epistemology or situation more broadly. The definition of learning as a process proposed in this dissertation is the following:

> *Learning is a goal oriented process of experiencing obstacles towards the acquisition of knowledge in the form of an emotional mix of frustration and desire that instigate the conscious implementation of strategies to reach the always elusive understanding. This process prepares individuals to participate in future experiences in deeper and more expansive ways.*

2.2.2. Educational Games: Claims and Issues

It has been claimed in the last 30 years that the engaging nature of games may facilitate involvement, motivation and interest, and the retention of learned skills (Greenblat, 1981; Greenfield, 1984; Loftus & Loftus, 1983; Malone, 1981). Greenblat (1981) classified the claims supporting the use of games for learning purposes. She distinguished 6 categories: 1) motivation and interest, 2) cognitive learning, 3) changes in later course work, 4) affective learning re subject matter, 5) general affective learning, and 6) changes in the classroom structure and relations. More recently, it has been claimed that games might foster motivation to *learn* (O'Neil, Wainess, & Baker, 2005). It has also been suggested that players of commercial games are developing problem solving and literacy skills (e.g., Gee, 2003; Squire, 2008) and that good commercial games represent good learning principles that provide opportunities for gamers to engage actively and reflectively during gameplay (Gee, 2005, 2003). Others have suggested that games involve individuals in a virtual cycle of action–feedback–reflection (Charles, Charles, McNeill, Bustard, & Black, 2010; Hickey, Ingram-Goble, & Jameson, 2009), which in turn results in an optimal learning experience usually characterized by a state of Flow (Csíkszentmihályi, 1990) and said to explain the "deep engagement" of individuals with computer games. Furthermore, the "immersive" nature of modern educational games has the potential to transform the knowledge outlined in school content standards into "just-in-time" knowledge to solve meaningful problems within interactive narratives (e.g., Barab & Dede, 2007) (see Section 2.2.4.3). These environments are supposed to be more appealing and appropriate to the new generation of individuals. In summary, games are said to support learning through (a) promoting active learning via its problem solving approach, (b) engagement of learners, and (c) fostering collaboration among learners (Gee, 2003).

Some of these arguments are more relevant for the present dissertation than others and are discussed in more detail below. The first one refers to Gee's contention that good games entail good learning principles. The second refers to Prensky's notion of "Digital Natives". Third the issue of "edutainment" versus "experiential" games is addressed. Fourth, the role of motivation, flow and engagement in learning from educational games is discussed.

Gee: Good games = good learning principles. Gee's most relevant claim concerning games is that good games entail good learning principles, learning principles that, in turn, are routed in the learning sciences (Gee, 2003). Gee further proposed 36 learning principles embedded in "good" games. But taken together they barely add something new to the knowledge based that education and psychology have already constructed. Appendix C presents Gee's 36 principles with its main idea as related to learning and the key authors used by Gee to support the principles. It can be seen that the key points concerning learning are already contained in three widely known pieces written by Bereiter and Scardamalia (1989), diSessa (2000), and Bransford and Schwartz (1999). In addition, Gee's key argument for his notion of "good" games is that if the game can be learned, it will sell, and if not, it cannot be sold well and companies can be in trouble. So, by learning from one another, in a sort of "Darwinian process", games increasingly incorporate good learning principles. In summary, a selling game is a good one because entails good learning principles. However, according to Blumenfeld et al.,

> For a learning sciences approach to work, students must invest considerable mental effort and must persist in the search for solution to the problems… (but) it remains unclear whether students are willing to invest the time and energy necessary for gaining the desire level of understanding (Blumenfeld et al., 2006, p. 475).

If we assume this to be accurate for technology-enhanced environments in genetics (Hickey & Kindfield, 1999), problem-based learning (Savery & Duffy, 1996), Multiuser Virtual Environments (MUVEs) (Hickey et al., 2009; Nelson, 2007), and Knowledge Building (Bereiter & Scardamalia, 1989; Scardamalia & Bereiter, 1993), it is reasonable to think that educational games face the same challenges. Basically, these challenges are:

- The difficulties of sustaining the doing in these rich environments. The more meaningful the problems are (typically hands-on problems), the more superficial is learners' understanding of the underlying academic concepts (Blumenfeld et al., 1991; Hannafin, 1989), and the more important is to take into account the learners' goal to understand the impact of an instructional intervention (Tobias & Duffy, 2009);
- Individuals tendency to engage in activities unrelated to the instructional goals and, consequently, their attention is captured by superficial features of the task or the technology used; and
- In the best of the cases, the virtuosity of the cycle action–feedback–reflection is an "opportunity to be taken" (Perkins, 1985) that may depend on the quality of the intellectual partnership (Salomon, et al., 1991; Salomon & Perkins, 1989) between the individual and the computer game.

Concerning the "Darwinian process" suggested by Gee (2003) that should have created the good games with good learning principles, Buckingham suggested that things might not be that simple: "research on the games industry suggests that the processes through which games are produced, marketed and distributed are rather more complex – and significantly less benign – than Gee implies…" (Buckingham, 2007, p.111). This suggests that a game that sells might not imply "only" good learning principles. Finally, it has been recognized that even though Gee's work is interesting in itself, his idiosyncratic nature and, more importantly, his omission of the bulk of research from game studies and educational games limit his contribution to the field.

Prensky's "Digital Natives". Prensky's (2001a,b) claims, in a stronger version than Gee's, represent a type of techno-centrism, by which technology is seen as a decontextualized force affecting an entire generation in terms of learning styles and brain plasticity (Buckingham, 2007). The core argument is that the generation that has grown up digital is used to and more willing to deal with technology as opposed to the older generation normally composed of parents and teachers. This generation or "digital natives" has fundamentally changed and education should now accommodate its methods to the new skills and interest of the digital natives. This generation lives immersed in technology "surrounded by and using computers, videogames, digital music players, video cams, cell phones, and all the other toys and tools of the digital age" (Prensky, 2001a, p. 1). According to the author, these individuals are "active experiential learners" with multitasking skills, and highly relying their social interaction and information acquisition on the communications technologies.

Bennett, Maton, and Kervin (2008) revised the evidence concerning Prensky's main claims, that is, 1) digital natives possess sophisticated skills with information technologies, 2) they have particular learning preferences that differ from previous generations, and 3) education must change to meet these digital natives' preferences. Concerning claim 1, the authors reported that research has found that a minority of the students engage in creating their own content in the web and that an important proportion of students present lower level skills than might be expected of digital natives. They also reported that the frequency and nature of children's Internet use differs in terms of age and socio-economic background, suggesting that the skills and experience related to technology are far from universal. Similar findings can be also found in the study *the dumbest generation* by Mark Bauerlein (2008), who portrayed the current generation as being stupefied by technology through a significant contraction of youth's horizons to themselves and peers. Following Bauerlein, the time they invest playing video games and participating in social networks does not seem to help them know aspects of history that are key for responsible citizenship. Concerning the issue of learning styles (claim 2), it is difficult to see whether this might be accurate or not given the ambiguity of the term "learning styles" and the inconclusive research relating learning with a particular style. This is so because, among other things, these "styles" are not supposed to be fixed, but variable among individuals and changing depending on how the task is perceived and how useful a particular approach has been in the past. Therefore, ascribing such a style to a whole generation is thus questionable. In a similar vein, a case for multitasking is hard to build. Research has shown that the "net generation" does not "multitask" as often as expected (Judd & Kennedy, 2011) and even if they do, research on cognitive load has suggested that "multitasking" could easily undermine individuals' performance and likelihood to achieve some learning goal (e.g., Kirschner, Sweller, & Clark, 2006). Similarly, from a neuroscience perspective multitasking can result in the acquisition of less flexible and therefore applicable knowledge (Foerde, Knowlton, Poldrack, 2006). This implies that multitasking or task-switching might not always be desirable. In summary, the one thing that is sure is that youth has grown up with technology as part of their everyday lives and that this is also true in the case of games. Therefore, it seems reasonable to expect that these individuals have developed a set of perceptions and preconceptions about games in general.

Edutainment versus experiential learning. Researchers on educational games usually contrast “edutainment” with “experiential learning”, where the latter is assumed to be superior to the former. In his doctoral thesis Egenfeldt-Nielsen writes: “Edutainment titles are characterized by using quite conventional learning theories, providing a dubious game experience, relying on simple gameplay and are mostly produced with strict reference to a curriculum” (Egenfeldt-Nielsen, 2005, p. 9). Edutainment refers to any combination of educational and entertainment use on different media platforms (e.g., computer games) (Egenfeldt-Nielsen, 2005). The author characterized edutainment games as using behavioristic learning theories which provide poor experiences, simple gameplay and with reference to the curriculum. He writes: “…the computer game will ask a question and the player will answer. When the question and answer are linked enough times resulting in reward, learning will occur” (Egenfeldt-Nielsen, 2005, p. 79). As a summary, edutainment games have:

- Little intrinsic motivation: extrinsic arbitrary rewards, for example, getting points for completing a level, instead of the feeling of mastery for completing the level.
- No integrated learning experience: this lack of integration makes the player concentrate on playing the game rather than learning from the game.
- Drill-and-practice learning principles: drill-and-practice thinking rather than understanding. Players get problems such as 2+2 memorizing the results and not necessarily understanding the underlying principles under 2+2 = 4.
- Simple gameplay: simple gameplay inspired on classic arcade titles or simple adventure games.

However, this characterization represents the extremes points that do not help to understand how to improve the design of educational games. For example, in games points (or Scores) to get to a next level correspond to a set of highly used game design patterns (Section 2.1.1.1). According to Egenfeldt-Nielsen’s (2005) characterization even *Pac-Man* could be considered edutainment. Finally, getting points and having feelings of mastery should not necessarily be two exclusive situations. The dichotomy over integrated versus not integrated learning experience is further explored in sections 2.2.3 and 2.24. Only to mention that what can be found in current titles of educational games (see Appendix D) are not integrated versus nonintegrated, but “more or less” integrated.

Secondly, the “drill and practice” approach is more related to “direct instruction” and is considered to be too narrow and related to rote memorization of meaningless facts (Egenfeldt-Nielsen, 2005). This picture is an oversimplification that simply makes a caricature of students learning the 2x2=4 in a mindless manner. This of course could be the case, but also could be the opposite case. It is also argued that these experiential approaches embedded in games for entertainment and “good” games promote higher order thinking skills similar to the ones performed by experts. However, it is also true that those skills do not operate in a vacuum and that they require a suitable knowledge based, in other words, an organized body of knowledge of facts and principles (Bransford, Brown, & Cocking, 1999) that are not necessarily easy to grasp through experiential approaches and where directed instruction seems to be more appropriate. The problem is that some of the facts and procedures have not direct practical application or have no apparent direct relationship to the learners’ everyday interest or activities. In this sense games have been regarded as a potential solution to the extent that they can “situate” the knowledge in a “context” of use (Barab & Dede, 2007). This argument is similar to the Premarck Principle, which states that high probability behaviors (e.g., those freely performed), can be used to reinforce low probability behaviors. In this context games represent the high probability behavior which becomes the reinforcer for engaging in learning, the low probability behavior.

Experiential learning is not a property of some games and not others. Dewey (1938) referred to “traditional education” as being also an experience, but pointed out that the *quality* of the experience is what counts. Similarly, “edutainment”, being the villain because it is based on associationist theories inspiring drill-and practice also depends on the interpretation of the terms. For instance, Wertheimer (1945) suggested that an „association“ can mean either „Die Verkettung von Einzelsachverhalten in einer Undsumme von Verbindungen…“ or „ Das Bemerken strukturellen Zusammengehörens, bei welchem die Einzelsachverhalte einander gegenseitig fordern…“ (p.239). If the former is meant of course any instructional design based on it will be at least narrow. But if the latter is meant, then the distinction between edutainment and experiential games seem to become a bit fuzzy. In summary, the claim that games represent a more appropriate pedagogy based on experiential learning principles depends on the goals that one wants to achieve. Secondly, the experiential approach is difficult to implement and faces serious challenges as the one already mentioned concerning the role of the learner and her willingness to invest the necessary effort and approach in a mindful way to the task of learning. Finally, it has been shown that for individuals with a minimum of prior knowledge experiential learning can hardly be the most appropriate method of instruction. Therefore, when choosing a particular educational game the criteria of whether it corresponds to an edutainment or a more "experiential" approach might be misleading.

Learning from games: Motivation, Flow and Engagement. Another issue of importance is the role different construct might play in learning from an educational game. First, games are said to foster individuals' intrinsic motivation (i.e., doing an activity for its own sake). There are three problems with this characterization of intrinsic motivation: 1) intrinsic motivation as a concept has multiple meanings and also coexist with more extrinsic sources of motivation; 2) intrinsic motivation has moved from individuals reasons for doing an activity to individuals "liking" or "enjoying" an activity, 3) game design is about creating a rule-based system that provides interesting challenges using a handful of reward systems. For this reasons, the phrase "for its own sake" seems at least confusing. Furthermore, it has been shown that what individuals enjoy is not from which they learn the most (Clark, 1994, 2001). Second, concerning flow some proponents (Kierremur & McFarlane, 2004) assert that by understanding flow and how the "deep" structures of games contribute to flow, it would be possible to use these structures to design learning environments. This proposal leaves open the question of how flow relate to cognitive processes, performance and learning. At the empirical level, the evidence that flow is central to learning is still to be produced (e.g., Admiraal, Huizenga, & Akkerman, 2011; Kickmeier-Rust, et al., 2011; Kiili, 2005; Pavlas, 2010), which is an example of a more general case for intrinsic motivation and learning (Brophy, 2010; Kebritchi, Hirumi, & Bai, 2010; Renkl, 1997). For example, for Kiili (2005), games by having clear goals and feedback should foster in individuals a "flow state" consisting of sense of control, concentration, loss of self-consciousness, time distortion and autotelic experience. This complex flow state, in turn, should lead to a set of "flow consequences" such as learning and exploratory behavior. The author suggested that learning occurs "from the process of working towards the understanding and resolution of a problem" (Kiili, 2005, p. 476) in the game, that is, learning as occurs in problem-based learning. As exploratory behavior the author meant the “experimentation of game features and generated playing strategies” (Kiili, 2005, p. 476). However, the hypothesis that flow state led to learning and exploratory behavior could not be supported. More generally, it has been argued that some design approaches should be better than others precisely because they can better foster and protect the flow state of the game from the intrusion of content knowledge. In fact, there are three main approaches to educational game design (see Section 2.2.4) that borrow a natural rejection of the edutainment-like extrinsically designed games, despite the minimal evidence for the role of flow on learning.

From a more skeptic position, Kerres et al. (2009) have questioned the initial attempt of mixing the activity of playing a game with a learning activity. Kerres et al. (2009) suggested that this mixture implies the disruptive experience of two different modes embedded in educational games, the learning mode and the playing mode. The authors also suggested that these two modes, which appear interchangeably during a game session, activate two different scripts (i.e., learning versus gaming) that are experienced by individuals as two opposite and interrupting experiences. According to the authors the learning mode, in the context of an educational game, means to be “taken out of the play mode” and switch to a learning mode that can be felt as an obligation to attend and process the instructional content in order to continue or be able to switch back to the playing mode. Finally, the “learning mode” means to individuals the end of the immersion in the flow of the game. The authors provided preliminary evidence in the form of a case study based on individuals’ eye movements. In an example of one individual they showed the visual behavior of the individual during the learning mode within an educational game. The example showed a superficial scan of the instructional content, which led the authors to conclude that the search for information relevant to the task is rapidly abandoned in order to get immersed again in the game. Although the example is enticing, no additional information in terms of individuals’ subjective experience is provided (i.e., interviews, questionnaires, etc.).

On the other hand, a characteristic of games is their power to deeply engage individuals in “gameplay”. This engagement has captured the attention of instructional designers for decades, Malone (1981) being one of the pioneers in the area (Section 2.2.3). The basic idea is to *harness* the power of engagement of videogames and use it to support individuals’ attainment of specific learning goals. However, in the literature on educational games it is rare to see an attempt of definition and operationalization of the concept engagement, which hinders the understanding of its role in learning from educational games. As a case in point, Egenfelfdt-Nielsen (2005) mentioned the term "engagement" 56 times in his dissertation and never addressed the issue of what engagement meant to him and how it could be measured. However, he made an important point for the purposes of this dissertation: he noted that the students playing the strategy game *Europa Universalis II* did show progressively more engagement, but "not always educationally relevant" (p. 202). This implies that at least two kinds of engagement do exist, an educationally relevant one and a non-educationally relevant engagement. Therefore, identifying what type of engagement is more effective for learning purposes is a research path worth pursuing.

Dickey (2005) compared engagement strategies of games with features of engaged learning, in order to inform instructional designers about strategies for developing *engaging* learning environments – Malone's (1981) "intrinsically motivating instruction" 25 years earlier. Her review positioned engagement as the cornerstone of gaming yet she did not provide a definition of the concept. She also left unanswered the question whether this structure for design inspired by entertainment industry will be suitable for the design of environments with learning goals and not entertainment goals. However, her review is valuable to the extent that she incorporated the educational literature on "engaged learning" with technology (Jones, Valdez, Norakowski, & Rasmussen, 1994) and the features of engaging "school work" (Schlechty, 1997). Even though these models inform about the value of technology use and tasks and school work for engaged learning, they are described in terms of the design of learning activities and do not provide per se an analytical framework to capture engagement during a particular activity. Similarly, Whitton (2011) developed a model and questionnaire of "engagement with learning", based on Malone's earlier work and the psychology of flow. However, more than a conceptual definition and operationalization of engagement with leaning, the article represents an attempt to measure the factors that might lead to engagement and therefore does not offer an idea of what engagement is and how can be captured.

Garris et al. (2002) represent to some extent an exception in the literature on educational games. The authors proposed an input–process–output model of educational games in which the game cycle (the process), composed of *user judgments*, *behaviors* and *feedback*, is the "hallmark of engagement in game play" (Garris et al., 2002, p. 441). User judgments consist of Interest, Enjoyment, Task involvement (e.g., level of attention, concentration, depth of involvement and cognitive engagement) and Confidence (self-efficacy); user behaviors consist of more time on task, actively pursue of challenge and more commitment to continue "on task". Here the authors described these behaviors as persistence *re-engagement*: more intense effort and concentration, and the willingness to return to game play even when unprompted to do so. Feedback refers to the basic knowledge of results. To the extent that feedback represents a standard, it regulates user's judgments and behavior. The model provides key elements to think about engagement and its connection with learning, even though it does not offer a straight forward definition of it. The main limitation of the model is the overlapped categories first presented as exhaustive. For instance, within the user judgment category is the sub-category Task involvement which is defined by attention and concentration, while the category behavior is also described in terms of effort and concentration. Another drawback of Garris et al.'s model (2002) to use it for the study of engagement is that their model incorporate variables, such as self-efficacy, that seem to be more moderators or at least factors that may influence engagement and not a part of the concept of engagement itself. In Section 2.5 the components Garris et al.'s model (2002) were rearranged to fit the conceptual framework proposed in this dissertation to study engagement in educational games.

Given that the main claim for using games for learning purposes is games' motivational qualities, a review of the seminal work of Malone (1981) is provided. He proposed Challenge, Fantasy and Curiosity as the main motivational features of games and until today it has remained a reference point for the arguments "pro games", and also the inspiration for instructional game designers pursuing the exact coupling of content and fantasy. The review below highlights some problematic issues with Malone's (1981) work in particular in terms of the methodology to study games and his claims concerning the centrality of "intrinsic fantasies".

2.2.3. Thomas Malone's Legacy

Malone (1981) was mainly interested in understanding and avoiding the undesired effects of reinforcement on individuals. This led him to search ways to enhance more "intrinsic" factors to improve individuals' motivation. The main purpose of Malone's studies was to distinguish which of three main competing theories, that is, challenge, curiosity and fantasy, was more likely to be responsible for the fun in games, which of them was more important in making games fun and how they varied for different people and different games. The main point of this section is to highlight the modest theoretical and empirical support of Malone's strongest and most influential claim concerning the role of "intrinsic fantasies" on the fun of games. This claim has been followed by many designers and researchers (see Section 2.2.4).

In order to accomplish his goals, Malone (1981) revised some theoretical approaches to support each of the theories, conducted empirical studies and proposed a theory for the design of "interesting" learning environments. The main theories and theorists reviewed by Malone are presented in Table 8.

The main observations made by Malone (1981) were that 1) Piaget did not mention what features of an environment/activity make it challenging, 2) Csikszentmihalyi did not addressed why his list of features were important and how they related to each other, 3) Freud was concerned with fantasies produced by people, but maybe similar processes underlie the fantasies individuals find appealing in external environments, and 4) Berlyne's "conceptual conflict" could be thought better in terms of "lack of consistency". With these theories and his qualifications, Malone conducted three empirical studies briefly reviewed below.

Study one. Malone (1981) interviewed 65 elementary students (25% of the entire school) that participated in a computer class. The teacher in charge of the class gave students a range of popular games they had to rate (i.e., 0-never played, 1-didn't like it, 2-liked, 3-liked a lot). Once the top games were identified, their characteristics in terms of the degree at which they may affect motivation were examined. Findings showed that among the highest rated games, four features were present: a goal, a scoring, audio effects and randomness. The correlations between these features and the average preference for a game ranged from .48 to .65. It is interesting that "fantasy" was *not* correlated with average preference for a game (r =.06). Furthermore, the author called the destroying of the bricks in the top rated *Breakout Game* (see description below) as a "goal" but later as a "fantasy goal" (see study two below), which seems confusing. If it is the first case, the challenging aspect of the interactive environment is meant, if the second the fantasy aspect is meant.

Study two. The game under study was the *Breakout* game, one of the most popular at the time. The game consists of knocking out all the bricks of a wall by using a bouncing paddle. Each time the paddle knocks one brick out it adds to the score. In trying to find out the secret of its success, the study had 10 undergraduates playing six versions of the game. The Malone (1981) hypothesized that the score, the breaking out of bricks or the bouncing paddle represented the essence of the game. Table 9 summarizes the versions of the game employed in the study.

Table 8: Competing Theories about "Fun" in Games by Malone (1981)

Theory	Main ideas
Challenge	
Piaget	"practice games" or repetitive, pleasureful exercise of recently acquired skills.
Csikszentmihalyi:	Structural features of intrinsically motivating activities: the actor can increase/decrease the level of challenge to her current skills; isolate perceptually the activity to avoid interference; clear criteria for performance; clear feedback; several qualitatively different levels of challenges.
Fantasy	
Piaget	Fantasy as an attempt to "assimilate" experience into existing structures with minimal need to "accommodate" to the demands of reality.
Freud	Symbolic games invented by the people to actively repeat traumatic events experienced passively. It is also the fulfillment of (often unconscious) wishes.
Curiosity	
Berlyne	Novelty, complexity, surprisingness and incongruity. People spend more time looking at the more complex or incongruous stimuli in a pair of similar pictures. For each person there is an optimal level of informational complexity. The principal factor under Curiosity is conceptual conflict, that is, incompatible attitudes/ideas evoked by a situation.

Results showed that V1 & V4 presented the highest average rating (4.8 and 4.1, respectively). The difference between them (V4 had no score) was not statistically significant. On the other hand, V4 obtained a significantly higher average rating than V2. They differed only in that V2 did not have the breaking of the bricks. Malone (1981) concluded that the breaking out of bricks “presents a visually compelling fantasy goal and, at the same time, is a graphic scorekeeping device telling how close the player is to attaining that goal” (p. 348). Versions 3, 5, & 6 with no score or breaking bricks were

Table 9: The Experimental Versions of the Game *Breakout* used by Malone (1981)

Version	Description
V1 & V4*	The original game. There are five balls. The ball bounces back and forth, destroying the bricks. Each broken out brick scores the number of points at the bottom of the screen.
V2 & V5*	The ball bounces back and forth between the wall and the paddle, but never breaks the bricks out. Each bounce is given one point.
V3 & V6*	The ball does not bounce off, it is simple “caught” when the paddle is in front of it.

*Represent the same version, but with no score.

interpreted as not having a clear goal: “Without a clear goal, the game was not really a game at all” (Malone, 1981, p. 348). This clever observation of Malone could not be truer and it also could have been extended to the other, more appealing versions (i.e., V1, V4 and V2). In terms of Salen and Zimmerman’s (2003) idea of meaningful play in the “no breaking bricks” condition players’ actions stop being discernible and integrated into the game system (Section 2.1.3). In other words, the action>outcome “choice molecule” was broken. Similarly, when Juul (2005) defined games involving a player exerting effort to influence a particular outcome in the game, by no breaking the bricks, the player has no outcome to influence. The main point of this discussion is to highlight, as did Malone, that this research strategy used by him and his “followers”, that is, to think in terms of discrete features to be removed at will, risks to stop studying a *game* in the first place.

In any case, it is interesting to observe how Malone (1981) continued to keep alive the idea of fantasy as the explanatory concept of fun in games. In his second study, he suggested that the breaking bricks represented a “fantasy goal”, when in the context of the study 1, it was only a “goal”. On the other hand, it is not clear how Malone concluded that fantasy played a role in the interestingness of the game given that the game was rated the third highest one in the first study and that fantasy did not correlate with it. At the end, the author acknowledged that with 10 individuals it is difficult to “reveal” the secret behind the game, but it might “illuminate” the importance of combining challenge and visual effects in the game design, no mention of fantasy was made though. In summary, the most important result is that it is most important to consider the combination of features, instead of isolated elimination of them, and as Malone himself suggested his method may not have been able to achieve that.

Study three. This study was conducted using an educational game called *Darts* to teach fractions to elementary students. In *Darts* (Dugdale & Kibbey, 1975), the player tries to guess the position of a set of balloons on a number line located on the left of the screen by typing mixed numbers (whole numbers and/or fractions). Each guess is followed by an arrow coming from the right side of the screen to the specified position. If the guess is correct, the arrow pops the balloon, if not the arrow remains on the screen and the player has to try again until all balloons are popped. As with the previous studies, here Malone (1981) also highlighted the role of fantasy. In fact, he saw in *Darts* a good example of an *intrinsic fantasy* “where the fantasy (the positions of arrows and balloons on the number line) is intimately related to the skill being used (estimating fractions)” (Malone, 1981, p.350). As he explains: “Besides the intrinsic fantasy, Darts has a number of other potentially motivational features such as feedback, music, and graphics” (Malone, 1981, p. 350). However, given the modest theoretical background provided for fantasy (namely Piaget and Freud) and the limited evidence for the role of fantasy in the previous two studies, it seems that Malone’s goal was to build a case for his notion of “intrinsic fantasy” regardless of the evidence he himself was producing. As for the empirical part, again the same method was used and 8 versions (i.e., Conditions) of the game were created and 10 students played one of these versions (Table 10). The difference here was that instead of removing features, Malone added one more “presumably motivational feature” (Malone, 1081, p.350) to the previous version of the same game. That is, if the versions are numbered from 1 to 8, the first version is the most basic one, the second has a feature absent in the first one, the third version has a feature absent in the second one, and so on.

Table 11 shows the main results of the study. Significant effect of Condition on time played and preference and a significant interaction effect between condition and sex on time played. Liking the game did not show differences among conditions. However, Malone pointed out that “the only significant difference when boys and girls were analyzed together was a significant increase ($p < .05$) in "liking Darts" when the extrinsic fantasy of arrows and balloons was introduced (Condition 4 vs. 5)" (Malone, 1981, p. 353).

Table 10: The Experimental Conditions (Versions) of the Game *Darts* used by Malone (1981)

Conditions	Description
C1	Students guess the location of a rectangle on a number line
C2	After each guess, students are told whether the guess was right or wrong
C3	At the bottom of the screen the number of tries and correct answers is displayed.
C4	After a wrong guess, students are told in which direction and how much the answer was wrong.
C5*	After a right guess, an arrow pops a balloon in another part of the screen
C6	Music is played at the beginning and when students guess all three numbers in four or fewer tries.
C7	Correct and incorrect answers are marked by lines on the number line
C8**	The original *Dart* game is used with arrows popping balloons on the number line.

*Condition 5 represented the "extrinsic fantasy"; ** Condition 8 represented the "intrinsic fantasy"

Malone (1981) also highlighted that the most surprising result was that girls found less interesting the intrinsic than the extrinsic fantasy condition. Additionally, boys liked the extrinsic fantasy condition (version 4 versus 5).

Table 11: Results of Malone's (1981) Study using *Darts*

Independent Variable	Dependent Variable	Main effects	Interaction effects
-Condition (8) -School -Sex	-Time playing Dart -Liking Dart -Preferring Dart to Hangman	Condition on Time played and Preference.	Condition by Sex=Time playing Dart.

In the face of these results, it seems surprising the claim that the author put forward: “I would like to claim that: In general intrinsic fantasies are both (a) more interesting and (b) more instructional than extrinsic fantasies.” (Malone, 1981, p. 361). He also portrayed the role of fantasy as providing analogies that help the learners applied their previous knowledge. He suggested that students in the context of Darts “if they make the crucial connection between number size and position on the number line, then they are able to use this old knowledge…” (Malone, 1981, p. 361). However, it does not seem obvious why this *crucial connection* is less likely to occur in his extrinsic fantasy version.

Finally, the author explained girls less interest in the intrinsic fantasy condition than in the extrinsic one in terms of girls disliking the fantasy. In short, girls were not interested in the intrinsic fantasy because they did not like it. Just appealing to “liking” does not seem a satisfactory answer though. However, today in can be warranted by using evidence from neuroscience, evidence that Malone (1981) did not have at the time of his studies. According to neuroscience the response to rewards is important because it is associated to the release of significant amounts of *dopamine*, which helps individuals orientate their attention and also it enhances *synaptoplasticity* (i.e., learning). The mechanism is as follows: dopamine is not released in front of a *predictable* event, but in situations of a 50/50 chance of getting a reward, that is, in front of *uncertain* outcomes. This mechanism of dopamine release when faced with uncertainty is more developed in boys than in girls (Howard-Jones, Demetriou, Bogacz, Yoo, & Leonards, 2011). In the context of Malone’s (1981) study, a careful analysis of the two versions (extrinsic versus intrinsic) reveal a subtle, but important difference:

Version 7 (Extrinsic fantasy): the player's guess is marked immediately on the number line, and an arrow goes across the screen only if the guess is right” (Malone, 1981, p. 354).
Version 8 (Intrinsic fantasy): “an arrow goes across the screen after every guess, and there is a moment of suspense before the player can tell whether the guess is right or wrong” (Malone, 1981, p. 354).

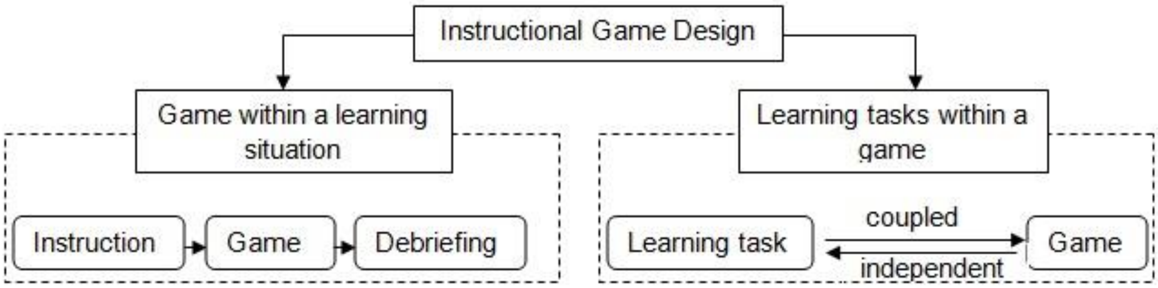
It can be argued that this “moment of suspense” represents an “uncertain outcome” toward which girls do not have the same dopamine mechanism that boys do. This would improve Malone’s (1981) explanation. Instead of girls not liking the fantasy, the difference might be due to the "suspense" and/or "uncertainty" embedded in the fantasy, not the fantasy itself.

In summary, the main conclusion from Malone's (1981) studies is the limited theoretical and empirical evidence of the role of fantasy, and intrinsic fantasy in particular, on the appeal or interest of the participants. Despite this fact, the notion of intrinsic fantasy has been widely applied and used with certain modifications as a design goal for educational games (see Section 2.2.4). In terms of research, it has contributed to the general strategy of adding or eliminating discrete features, regardless of whether such manipulations in actuality hinder the game under study and in doing so, the claims derived from such research (see Section 2.2.5).

2.2.4. Educational Game Design

Kerres et al. (2009) and Filsecker & Kerres (2013) have distinguished two alternatives to harness the potential of games for learning purposes (see Figure 13)

Figure 13: Classification of Instructional Game Design by Kerres et al. (2009)



As Filsecker and Kerres (2013) argued, the alternative on the left represents a game within a learning situation, and is concerned with embedding a game in a particular learning situation as a source of reflection or as a context for practicing the content delivered by instruction. In order to profit from the game a thoughtful and carefully designed de-briefing session is required. In the second alternative depicted on the right, embedding learning tasks within a game, the gameplay seems to be the reward for having solved different learning tasks. The game and the learning tasks can be more or less independent from each other (see below discussion on Intrinsic/Endogenous Fantasy). The general assumption is that the more dependent they are the better the learning effect of the general game experience. No matter what the implemented alternative is, they require a game designed with an explicit learning goal (however, in practice the first alternative usually is implemented with COTS). Similarly, Klopfer, Osterweil, & Salen, (2009) using a “baking” analogy, described two main ways to go about “baking” a game for learning (here the “cake” symbolizes the “game”):

> One recipe takes the yummy calorie laden cake and injects beta-carotene, vitamin D and calcium right into the cake (…) The other recipe simply takes all of the healthy content – wheat germ, oat bran, carrot juice, spinach leaves, etc. – and bakes them into something that looks like a cake. (Klopfer et al., 2009, p. 27).

According to Klopfer et al. (2009) an example of the first type is the well-known game *Math Blaster* and more modern titles such as *DimensionM*. These titles might look interesting but “chances are that you’ll hate the game and the math” (p.27). This fear, although possible, seems to be an exaggeration of the impact of this design on individuals’ final interest on a particular subject matter. The second type is in actuality not a game: “just because it looks like a game, doesn’t make it a game” (Klopfer et al., 2009, p.28). They cited Castranova’s 3D world Arden, presented as a game that “sucked”, using its creator’s own words.

Overall, there seems to be little agreement as to what makes a good entertainment game and in particular what makes a good educational game. However, what seems to be a general perception is that *Math Blaster* is a “bad” one, mainly because it represents either the old behaviorisms’ “drill-and-practice” method as opposed to “experiential approaches” (Egenfeldt-Nielsen, 2007) or it represents something far from a game in the first place (Fortugno & Zimmerman, 2005). According to Gee (2011), the issue of “good” or “bad” games has directly consequences for the evidence and claims one can make about games and learning given that his hypothesis is that *good* games are good for learning. Gee (2011) and Prensky (2011) both suggested that the most important thing in game research is to have good games, which seems to be seldom the case. Both authors recognize that without studying good games, it is not possible to conduct empirically relevant research on games and learning to test the above hypothesis. But with no shared standards to assess high quality games and the inner difficulty of creating good games, empirical research to test the hypothesis of games and learning seems a lost enterprise. For instance, the lack of shared standards is reflected in the perception of *Re-Mission*, a serious game for young people with cancer. On the one hand, Prensky (2011) mentioned this game as a high-quality educational game. On the other hand, a small study showed that *Re-Mission* was barely “average” in a scale of enjoyability due to its mediocre gameplay (Ritterfeld, Cody, & Vorderer, 2009). Given that there is no clarity concerning what is a good game, this issue can rapidly lead to the “chicken or egg” dilemma. Fortunately, instructional design is a pragmatic field and has been designing educational games from several decades regardless of the issue raised by Gee (2011) and Prensky (2011). In this context, the assumption of this dissertation is that more carefully design research oriented toward understanding how learning takes place during an educational game, can inform future design efforts to improve the currently “bad” games studied so far as suggested by Gee and Prensky. What follows is an overview of the main approaches to the design of educational games.

The general approaches encountered in the design of educational games reflect the debate in the broader game community concerning the centrality of either interaction/interactivity or storytelling/narrative (Juul, 1998, 2001). Similarly, the educational community has distinguished between endogenous fantasy/intrinsic integration and narrative as two approaches to effective game design. The former emphasizes the mechanics/interactivity side of games, and the latter emphasizes Bruner’s (1996) idea of Narrative either as a particular structure for organizing knowledge, as a means of representation (Dickey, 2005), or as a powerful metaphor (Barab et al., 2010). Whether mechanics or narrative, both approaches have tried to solve the problem of the effective design of educational games using the same basic idea: *coupling* the academic content with the game. What follows is an overview of the main design approaches to educational games.

2.2.4.1. Endogenous Fantasy

From the endogenous/intrinsic fantasy perspective, the effective instructional design of games is suggested to reside in the way in which the learning content is organized within a game. One possibility is represented by the highly cited work of Malone (1981). He claimed that intrinsic or endogenous fantasy, is a powerful design goal for *intrinsically motivating instruction*. The author defined a “a fantasy-inducing environment as one that evokes "mental images of things not present to the senses or within the actual experience of the person involved" (American Heritage Dictionary)” (Malone, 1981, p. 360), and further distinguished between extrinsic and intrinsic fantasy – later called “exogenous/endogenous” with no change in meaning (Malone & Lepper, 1987). The extrinsic fantasy is exemplified in the game *Hangman* in which the player progresses to avoid the "fantasy catastrophe" of being hung up and it is "extrinsic" because the fantasy "depends on the use of the skills but not vice versa" (Malone, 1981, p. 360), in other words, the skills could be related to algebra, vocabulary and so on and the fantasy will remain the same. On the contrary, in intrinsic fantasies the fantasy depends on the skill, but at the same time the skill depends on the fantasy. Malone (1981) exemplified his idea with the *Dart* game in which players need to estimate distances on a number line on the basis of introducing fractional numbers (the skill) to determine how distant or near to each other are located a set of balloons on a number line, so that if the players apply the skill correctly (i.e., the correct fractional number) they can aim at the balloons with an arrow and destroy them (i.e., the fantasy). The author finally claimed that "In general intrinsic fantasies are both (a) more interesting and (b) more instructional than extrinsic fantasies" (Malone, 1981, p. 361), claim that was not empirically supported by his studies (see Section 2.2.3). As already discussed Malone's (1981) studies only supported two claims: 1) boys played significantly more time the *Dart* game in the "extrinsic fantasy" condition, and 2) girls played significantly less time the *Dart* game in the "intrinsic fantasy" condition.

Despite this, the intrinsic/endogenous fantasy idea has been easily taken up as a central design goal for effective instructional game design (e.g., Gunter, Kenny, & Vick, 2008; Rieber, 1996). For example, Rieber (1996) claimed that games employing endogenous fantasies "weave the content into the game. One cannot tell where the game stops and the content begins" (1996, p. 50) and that they can lead better to intrinsic motivation than exogenous fantasy. However, he warned that these possible effects of the intrinsic fantasy may happen if the players *accept* the fantasy – which did not seem to be the case for the girls in Malone's (1981) studies. Although Rieber's (1996) description is straight forward, his example of an intrinsic game seemed far from the concept of intrinsic fantasy he described. From this work the question concerning what individuals variables can make the "acceptance of the fantasy" more likely is still open.

More recently, Gunter et al. (2008) also proposed intrinsic fantasy as part of the design guidelines for effective instructional game design. According to the authors, an effective designed game should introduce the academic content 1) in a hierarchical manner, and 2) intrinsically coupled with the fantasy context of the game. They think that even though something can be learned from an extrinsic fantasy, the risk of breaking the flow is too high, which would not be the case with the intrinsic fantasy. The authors developed a rubric for the design of "endogenous" games called RETAIN (i.e., Relevance, Embedding, Transfer, Adaptation, Immersion, and Naturalization). Relevance means how close the academic content is coupled with the fantasy of the game; Embedding refers to how closely the academic content is coupled with the fantasy/story content; Transfer refers to whether the game provides opportunities to use the targeted knowledge in new or unique in-game situations; Adaptation relates to the assimilation and accommodation process occurring during the game; Immersion is a hierarchical notion going from interaction to engagement and finally leading to intellectual investment; finally, Naturalization refers to automaticity and spontaneous knowledge that occurs through repeated play. Although the different elements of the rubric can be conceptually confusing (e.g., Relevance and Embedding seem to be address the same idea), the rubric contributes to the assessment of educational games' attributes.

In summary, even though the notion of an endogenous fantasy that integrates the learning content into the fantasy context seems plausible, it may not be certainly the only way to produce better educational games. Even though it is recognized that something can be learned from more extrinsically designed games, researchers assert that this approach compromises the flow of the game, which is regarded as central for the success of educational games (Kiili, 2005). Moreover, edutainment titles are highly criticized and looked down upon without any empirical evidence of its inferiority precisely because they seem to break players' flow. What is not explicitly addressed from this perspective is how designers can address the issue of designing fantasies that are more likely to be accepted and liked by players. Furthermore, it is also open the issue of the interaction of the fantasy and the subject matter: a highly accepted or liked fantasy would lead individuals to engage with an uninteresting subject matter? In any case, intrinsic fantasy has survived the limited theoretical and empirical support that has received.

2.2.4.2. Intrinsic Integration

Among the few that raised a critique to Malone's (1981) initial conceptualization of intrinsic fantasy were Habgood and colleagues (Habgood, Ainsworth, & Benford, 2005; Habgood & Ainsworth, 2011). They concluded that the concept of intrinsic/extrinsic fantasy was conceptually misleading and proposed to focus on the game's "core mechanics" and its assumed relation with flow experience to design effective educational games. Specifically, the authors' basic design guidelines for the integration of learning in digital games were:

> 1. Deliver learning material through the parts of the game that are the most fun to play, riding on the back of the flow experience produced by the game and not interrupting or diminishing its impact.
> 2. Embody the learning material within the structure of the gaming world and the player's interactions with it, providing an external representation of the learning content that is explored through the core mechanics of the gameplay. (Habgood & Ainsworth, 2011, p. 173)

Under these guidelines they developed a math game for children called *Zombie Division* from which they created and compared two versions: an intrinsically and extrinsically integrated version. Without addressing the details of the study, the results showed no differences at the post test, but only in the delayed test. These results do not support the intrinsic integration design as more effective for two reasons. First, the study mixed game sessions with teacher led reflection sessions. Second, the operationalization of the "extrinsic" version of the game (i.e., the comparison condition) was an extreme case of "extrinsic" design that no one would seriously pursue. In a word, it is similar to comparing an average instruction with purposefully designed mediocre instruction. The difference of this proposal to that of Malone (1981) could lie on the relationship established by Habgood & Ainsworth (2005) between "core mechanics" and "type of flow experience". Overall, the authors' proposal of intrinsic integration has turned out to be less different than the original intrinsic fantasy proposed by Malone (1981). Both used the idea of integrating the academic content with the game elements, for both this integration entails also the mechanic of the game and finally, for both protecting flow is a central concern.

2.2.4.3. Interactive Narrative

The third broad design guideline is interactive narrative (Barab, Sadler, Heiselt, Hickey, & Zuiker, 2007). The core argument is that a curriculum would help students better in understanding the meaning and value of the underlying principles of an academic topic (e.g., Erosion), if the academic content is embodied within an interactive narrative, so that the person and the story are *coupled* together. In a game this interaction should be designed as to push back player's understanding of academic concepts, thus becoming games a type of transactive curricula that afford the interplay between player and the story line. Later on Barab et al. (2010) described games as enabling a narrative transactivity by which this person/story coupling is reached through affording agency and a sense of consequentiality to players. Under these premises the authors develop *Quest Atlantis* (QA) a multiuser learning environment with different worlds in it. In this environment players are situated in rich narrative contexts where they adopt particular intentions and where their actions have clear consequences (i.e., during the learning process the narrative "unfolds" based on students' choices). Moreover, the authors claimed that by taking the role (i.e., "identity") of field investigators, mathematicians, etc., students in QA use the academic content embedded in the game so as to make inform decisions that will change the environment.

It is no clear whether the intrinsic fantasy, the extrinsic fantasy or the creation of rich narratives is the best strategy to go about designing an educational game. Certainly, comparing these approaches together and testing their impact on individuals' engagement and learning seems to be a reasonable way to find an answer to the features of effective educational game design. Next section discusses the main features, findings and limitations of current research approaches to achieve a better understanding of effective game design and theory building.

2.2.5. Educational Games Research

What follows is a review of the empirical research on educational games. First, examples of the type of research on the lines of Malone's (1981) seminal work are presented. Second, the main empirical literature reviews on games for learning that encompass more than four decades of research (1963–2011) is summarized. The section ends with a description of the gaps that the present dissertation attempts to fill.

The idea of game attributes, elements, or features. For some scholars, games are a new medium whose features have changed considerably compared to the new generation of games (Becker, 2010), such as their advanced graphic and 3D features (Kebritchi, 2008). Becker used this argument as a rejoinder to Clark (2007), who claimed that serious games have not proved to be better than other instructional methods. Becker found Clark's "vehicle" metaphor (the medium as a vehicle for delivering instruction) inapplicable to games given their interactivity, but mostly because she argued that Clark used examples of games belonging to an “old” generation and not to today’s top technology. For example, *DimensionM* would represent a *modern* educational game due to its technological features. At the beginning of her doctoral thesis Kebritchi writes: “Modern instructional games are significantly different from edutainment game generation as they may use advanced 3-D graphics and interface, multi-player options, high-speed telecommunication technologies (e.g., Quest Atlantis[tm]), immersive 3-D environments and visual storytelling (e.g., Zombie Division[tm])…”(Kebritchi, 2008, p. 19). These expectations on the power of technological features of games have encouraged research on topics such as the effect of high and low immersion (e.g., Heers, 2005) or simply the effects of a highly modern game such as DimensionM on learning (e.g., Ketbritchi, 2008). Similarly, according to Wang, Shen, and Ritterfeld (2009), by identifying the game elements behind the enjoyability of games it is possible to enhance the fun quality of serious games. However, this research fails to acknowledge the role these technological features really play in games for entertainment. Juul (2005) pointed out that the more players play a game, the more they focus on the particular mechanic and less on “secondary features” such as the ones suggested by Kebritchi (2008). The same is true for the fictional elements of games. The more the game is played, the less relevant become the fiction or narrative. In other words, players in time begin to see through the aesthetics of the game and play it for “how it feels rather than how it looks (Klopfer et al., 2009, p. 30). Another drawback of the emphasis on game elements or features is that they fail to account for the engagement power of games. The risks of such approach resides in the likelihood of 1) hiding a misconception of the centrality of a game, that is, the quality of its gameplay, and, consequently, 2) assuming that by removing a feature gameplay can be affected, 3) assuming that researchers know a priori which features are the most important ones and how removing one affects the other features of a game, finally, and 4) ignoring that several central features are nothing but a label for a complex *pattern* of game design (see Section 2.1.1.1). As a case in point, Wang et al. (2009) compared experts’ opinion in two groups of games: fun and not fun games. From the experts responses they categorized which elements were important for fun to be achieved. The most cited category for both groups was “overall game design”. This would mean that when the “overall game design” is positively assessed, the game is likely to be fun, otherwise the game should tend to be boring. But, what does it mean “overall game design”? The authors defined it as “general comments on game design" (Wang et al., 2009, p. 31) – clearly too broad to be useful. Furthermore, some categories found to be core in games, such as Interactivity, Mechanics and Fantasy played almost no role when assessing whether a game was fun or not. It could be the case that the name of the 30 categories selected might have some overlapping or the labels might not be the more appropriate ones. This study might have been more useful if the authors had adopted a more specific categorization and definition of these categories (e.g., "overall design") and had included the discussion of specific game design patterns.

A recent study of Wilson et al. (2009) suggested that particular game attributes are more likely linked to particular learning outcomes. This perspective resembles the type of research strategy followed by Malone in the 80's, but it recognizes that little research has explored how different attributes might work *together* to affect learning and which levels of the attributes (e.g., low, medium, high) are appropriate for a particular learning outcome. The authors assumed that attributes such as challenge, fantasy, and control were quantifiable and, therefore, they could be differentiated into "levels". However, this approach fails to recognize that what might explain how attributes work together is the pattern approach described above (Section 2.1.1.1). For the pattern approach though, patterns are not an issue of quantity, but of qualitative relations among intertwined patterns. For example, the idea of challenge is regarded as a key element behind the engaging power of games and a main attribute from the perspective of Wilson et al. (2009). However, challenge according to Juul (2005) is another expression to state that games represent a conflict, that is, that games have different and opposing outcomes, and when players decide to reach for the right outcome and use their skills for achieving it, it emerges the "epiphenomenon" of challenge. Moreover, from a game design pattern perspective, there are theoretically dozens of patterns to produce the conflict underlying any notion of challenge within a game. For example, the patterns *Combat* and *Enemies* (Björk & Holopainen, 2005) represent an instance of how to produce challenge "by design" (cf. Filsecker & Kerres, 2013). As a case in point, Serrano and Anderson (2004) studied a "game" with a challenge and a "game" with a storyline about the *Food Pyramid*. They described both instances as:

Games with a challenge: "Foods appear and the child clicks on the group of the Pyramid where the food belongs. The game gets faster and more complicated with each additional round" (Serrano & Anderson, 2004, p. 4).

Game with storyline: “A number of meals and menus are presented and the child needs to determine if he/she needs “more” or “less” food from different food groups. The Food Guide Pyramid is reinforced in this game, as well as serving numbers.” (Serrano & Anderson, 2004, p. 4).

In the case of the game with a challenge, the authors offered no explanation of what “faster” and “more complicated” meant in the context of the game. This issue is not addressed in the study, but the hypothesis that challenge leads to motivation and therefore may lead to better learning outcomes was supported. This evidence might give future instructional designers or teachers’ guidelines to include some elements of challenge and therefore this information is useful. However, when it comes to the *design* of educational games, this evidence might not be enough to design better games. Similarly, Pavlas, Bedwell, Wooten, Heyne, and Salas (2009) offered a set of attributes that could be manipulated experimentally. Among these, the attribute Challenge/Conflict in the serious game *Innercell* was described. The first impression from the description of the manipulation of Challenge/Conflict is that it is not about challenge, but about adapting the challenge to the player’s skill. What is it in actuality the challenge or conflict and how it is instantiated within *Innercell*, is not provided by the authors. In summary, the main drawback of these studies is their limitation to provide relevant knowledge about games as challenges: if the treatment means to have different levels of challenges in order to see their effects on a particular dependent variable, it might be possible to know something about challenging situations, but limited understanding of *games as challenges* can be gained. The question is not whether challenge might foster a particular outcome, but what kind of challenge and what kind of instantiation of a challenge might have produced the outcome. However, to address this last issue, research should consider the study of attributes or features but as part of broader game design patterns (see Section 2.1).

General issues in research on educational games. The National Academy of Science (Honey & Hilton, 2011) described persistent limitations in the research on games and simulations that may hinder the advancement of knowledge in the field. According to the report studies of educational games have seldom developed a theory of action about how the designed educational game will reach a particular learning goal. Secondly, most studies have been “proof of concepts” aiming at proving whether or not a developed game “works”– and usually the same developers conduct the research. Thirdly, studies normally have lacked control groups. Fourthly, studies have tended to mix games with other learning activities, which makes difficult to distinguish whether the game or the additional learning activities are responsible for a particular learning outcome. Finally, the field has lacked common definitions and terminology as already described concerning the concept of game (Section 2.2.1) and as discussed later concerning engagement (Section 2.3). According to Honey and Hilton (2011), all these research practices have limited the claims about the effectiveness of educational games and have prevented the development of an empirically based theory of learning from educational games and its concomitant identification of effective design approaches. However, regardless of these research practices, the last 30 years have produced numerous studies focused on the effectiveness of educational games. The following is a brief summary of such endeavor.

From an empirical perspective, several literature reviews have arrived at the following conclusions (see Table 12). First, the evidence of effectiveness of games on cognitive and conative outcomes is modest and games may keep some promise for supporting learning, but until today rhetoric over empirical studies have prevailed. Second, studies show inconsistent patterns of results concerning motivation and learning. These studies can be essentially classified into three groups: 1) studies showing no motivational effect with positive learning outcomes (e.g., Kebritchi et al., 2010), 2) studies showing a motivational effect but no learning outcomes reported (e.g., Egenfeldt-Nielsen, 2005), and 3) studies showing both motivational and learning effects, but without reporting the relationship between the two (e.g., Cordova & Lepper, 1996). These studies based mainly on individuals self-reports, do not address the issue of how motivation and learning could be related (e.g., more motivated learners may use more effective learning strategies than less motivated students, Renkl, 1997). Third, the effectiveness of games depends to a high degree on how individuals use it and the goals they set for themselves. However, research has barely considered learners' characteristics or other mediator variables that might affect the learning outcomes expected (Ennemoser, 2009). Some exceptions can be found though. For example, Elliot and Harackiewicz (1994) investigated the interaction between achievement motivation (i.e., high vs. low) and goals (i.e., mastery vs. performance) on individuals' intrinsic motivation while playing an enjoyable pinball game. The authors found that under the performance goal condition, intrinsic motivation was enhanced for individuals high on achievement motivation. When provided with mastery goals, low achievement-oriented individuals showed higher levels of intrinsic motivation.

Furthermore, task involvement (i.e., time playing and enjoyment) was one variable that mediated the direct effect of achievement motivation and goals on intrinsic motivation: high achievement-oriented individuals showed greater path coefficients with enjoyment under the performance goal condition, while in the mastery goal condition low achievement-oriented individuals showed higher enjoyment than their high achievement-oriented counterparts. Even though the authors setting was not an educational game, but a game for entertainment, it is fair to conclude that the effects they found may be replicated under similar circumstances. In the context of an educational game to teach children arithmetic, Plass, O'Keefe, Homer, Hayward, Stein, & Perlin (2011) carried out an experimental study using a computer-based game adapted to allow solo, competitive or collaborative gameplay. The authors showed that individuals in the competitive condition were found to endorse a stronger mastery goal orientation. Interestingly, the learning outcome favored individuals in the solo condition who did not show a significantly stronger mastery orientation. Even though the design could not establish whether individuals' achievement orientation changed as a result of the experimental conditions, the study represents one of the few addressing different motivational process other than intrinsic motivation.

Finally, research seldom focuses on exploring the psychological processes on games under more controlled settings such as randomized clinical trials. In general terms, this means that the studies on educational games employ a "black box" approach to research. A black box approach refers to the general tendency of postulating theoretical mechanisms (i.e., psychological processes) underlying educational games effectiveness but in actuality only measuring the inputs and outputs (pre-posttest) and leaving the processes unexplored. For instance, Kebritchi's dissertation (2008) focused on the impact of an educational game for mathematics *DimensionM*. The author hypothesized that the game would engage individuals in Kolb's (1984) cycle of experiential learning: concrete experiences, reflection, abstract conceptualization and active experimentation. In particular, it was hypothesized that "...learners (a) completed game missions or concrete experiences, (b) reflected on their learning experiences through game quizzes, (c) developed abstract Algebra concept, and (d) engaged in mathematics class activities and moved to the next experience" (p. 54). However, how individuals went about completing the missions, what sort of reflection the game quizzes prompted in the students and how that ultimately led students to developed abstract concepts, all belong to the "black box" dimension mentioned. What the author provides are pre and posttest of achievement and a questionnaire on motivation. Furthermore, the author hypothesized that at the core of Kolb's (1984) model is motivation, which provided the energy to move through the model's stages. Even though gains in achievement were found in the presence of no motivational effect, it is surprising to see that this result was not further discussed by the author.

As suggested by the review of empirical studies on educational games, it has become increasingly difficult to ignore the investigation of individuals' "engagement" while playing an educational game. Having discussed so far the need for defining and studying engagement, the next section addresses the different ways in which the educational research field has defined it. It will also be addressed what type of engagement seems to be more relevant when the purpose is to impact learning processes and outcomes.

Table 12: A Synthesis of Literature Reviews on Educational Games Effectiveness from 1963 to 2011

Author	Main conclusions	Number of studies included and time frame
Randel et al. (1992)	Areas such as math and physics where specific goals can be stated, simulation/games can be used. However, researchers should pay attention to the role of mediator variables. Measures should match what the game teaches (e.g., measuring problem solving with a Multiple Choice Test makes little sense), possible practice effect when using the same test in the pre and posttest. If the test was constructed by the investigator, reliability should not be too low.	N=67 (1963–1991)
Dempsey et al. (1994)	Division of articles into 5 categories: research, theory, reviews, discussion, development. Learning outcomes mostly ignored. Some studies mention verbal knowledge, concrete concepts, and attitudes. Game function: practicing skills and learning new skills. Gaming environment (e.g., elementary, adult, business, etc.): instructional gaming no more effective in one environment than in another. Measurement: achievement, problem solving and retention the most frequent measures. Learner characteristics: barely studied.	N= 99 (33 research and 43 discussion) (1972–1993)
Hays (2005)	Far fewer articles have documented with empirical data the effectiveness of educational games.	105 (26 rev, 31 theoretical, 48 empirical studies)

	The effectiveness of an instructional game depends on whether it was designed to meet specific learning goals and whether or not was used as intended. That is, as important as the designed game, is how individuals use it. One design goal should be to increase learner involvement. Some evidence that games can interfered with the individuals' learning.	
Vogel et al. (2006)	Meta–analysis: participants using interactive simulations or games report higher cognitive gains and better attitudes toward learning compared to those using traditional teaching methods. It is important to consider the role of mediator variables on these results.	32 studies (1987–2003)
O'Neil et al. (2005)	Games are not in themselves sufficient for learning, if not support and/or instructional strategies are added.	N=19 (1990–2004)
Tobias et al. (2011)	A tentative conclusion: games hold some promise for delivering instruction. Literature seems to be more about the rhetoric (highlighting the affordances of games and their motivational properties) than to conduct research demonstrating those affordances. Attitudes to games are positive, but less widespread as expected. Furthermore, users are not the best judges of what is best for them. Actualization of Randel et al.'s (1992) observations concerning the limited evidence of games effectiveness. Importance of studying the cognitive processes on games. The implication for transfer is direct: for transfer to occur, the cognitive processes engaged in a game should be the same as the ones required by an external task.	N=95 (1985–2010)
Connolly et al. (2012)	More rigorous Randomized Clinical Trials should be conducted.	N=129 (2004–2009)

Young et al. (2012).	Only K-12 Video games and pseudo educational games such as QA were included (in schools).Some evidence for the effects of video games on language learning, history, and physical education. Modest evidence for effects on science and mathematics. Limited evidence that educational games solve the limitations of traditional K–12 schooling and academia. Future research? Methodologies should explore individualized nature of gameplay. One possibility is the use of log file analysis.	39 (out of 363) (2005–2011)
Girard et al. (2012).	Three of the 11 studied games had a positive effect on learning compared with other types of training. The games included *Re-Mission*, *DimensionM*, and *SimCity*. Participants' level of engagement and motivation was assessed by questionnaires.	N=9 (from 30 originally) (2007–2011)

2.3. Engagement

As it was pointed out in the introduction, engagement represents a particular type of experience afforded by games for entertainment. This experience has been described in terms of players' struggles to achieve a goal given their current skills and given the games' constraints over long periods of sustained effort. Normally, this experience is accompanied by feelings of satisfaction and enjoyment. Likewise, engagement has also been proposed to be the central mechanism behind the success of educational games (e.g., Dickey, 2005; Egenfeldt-Nielsen, 2005; Garris et al., 2002; Kiili, 2005). However, this perspective either mentions the concept of engagement without explicitly defining its properties (Dickey, 2005; Egenfeldt-Nielsen, 2005) or equates its meaning with the concept of flow (Garris et al., 2002; Kiili, 2005). As was already suggested in Section 2.2.2, the relation of flow and learning has been barely established and it is an open and interesting line of research. On the other hand, the purpose of this section is to argue for the inclusion of a specific conceptualization of engagement that renders the concept more relevant for understanding learning processes. The concept of engagement discussed here is based on the definitions and research carried out by the educational research community (e.g., Fredricks et al., 2004; Jimerson et al., 2003).

Engagement is understood to be a multidimensional concept, which can provide a richer characterization of individuals' experiences and allow the examination of the antecedents and consequences of behavior, emotion and cognition simultaneously, so that particular configurations can be explored (e.g., Blumenfeld et al., 2005). Second, the general notion of commitment/investment under this concept implies the possibility of qualitative differences on the degree of engagement of each component. For example, cognitive engagement can range from simple memorization to the employment of self-regulated learning strategies. These qualitative differences can be short-term and situation specific or show longer timescales and more stability across situations. Third, from these characterizations engagement is a malleable attribute emerging from individuals' interactions with the environment and it is also sensitive to changes in the environment. Therefore, it represents an efficient target to be changed by instructional activities (Corno & Mandinach, 1983).

In summary, for this dissertation engagement is a three-dimensional concept entailing cognitive, behavioral and emotional dimensions (Fredricks et al., 2004) representing a volitional process (Kuhl & Beckmann, 1985; Corno, 1993) close to the notion of mindfulness (Salomon and Globerson, 1987) and Amount of Invested Mental Effort or AIME (Salomon, 1984). What follows is a synthesis of the main dimensions, the general measurement strategies and the different conceptualizations of the concept of engagement that can be found in the educational literature.

2.3.1. Dimensions and Indicators of Engagement

As with any other construct, a comprehensive agreement on the nature of engagement and the dimensions it entails has been difficult to establish (Appleton, Christenson, Kim, & Reschly, 2006; Jimerson et al., 2003). For example, Jimerson et al. (2003) and Fredricks et al. (2004) identified three dimensions (cognitive, behavioral, and emotional) while Appelton et al. (2006) distinguished four dimensions (cognitive, behavioral, academic, and psychological). Table 13 compares the dimensions and indicators of engagement as proposed by the three groups of researchers. Five points are important to highlight: 1) emotional engagement is similar across researchers and entails the idea of feelings and emotional reactions towards a situation (e.g., teacher, subject matter, etc.); 2) behavioral engagement is also coherence across researchers and entails the core idea of "participation" (e.g., in extracurricular activities, attendance, etc.); 3) Appelton et al. (2006) academic (e.g., time on task) and psychological (e.g., feelings of belonging) engagement can be reorganized in terms of behavioral and emotional engagement, respectively; 4) some dimensions have motivational concepts (e.g., self-efficacy, autonomy, value) , while others seem to refer to different outcomes of engagement (e.g., grades, credits earned, homework completion; and, finally, 5) cognitive engagement shows the most variable characterization across researchers.

The last two points reflect the lack of clarity concerning the psychological status of the concept of engagement as related to motivation (Appelton et al., 2006). According to the authors motivation is related to the why of the behavior, while engagement is related to energy "in action". As described in Fredricks et al. (2004), cognitive engagement has two different roots that might explain the differences among the authors concerning its conceptualization. On the one hand, cognitive engagement entails the idea of investment, similar to constructs such as motivation to learn and intrinsic motivation. On the other hand, this construct also entails the idea of being strategic or self-regulated. Fredericks et al. (2004) considered within self-regulation the construct of volition. As it is shown later, (see Section 2.3.2), the concept of volition (as engagement) is also multidimensional and distinguish, at least, between cognitive and behavioral components. These characteristics have led to conceptualize engagement as a volitional construct rather than as a motivational one.

Table 13: Comparison of Dimensions and Indicators of Engagement

Dimensions/Authors	Jimerson et al. (2003)	Fredricks et al. (2004)	Appelton et al. (2006)
Cognitive	Students’ perceptions and beliefs related to the self, school, teachers (e.g., self-efficacy, motivation, expectations)	Investment, thoughtfulness, willingness to exert the effort necessary to comprehend complex ideas and master difficult skills	Self-regulation, relevance of schoolwork to future endeavors, value of learning, and personal goals and autonomy
Behavioral	Observable actions/ performance: participation in extracurricular activities, completion of homework, grades, grade point averages, and tests’ scores	Participation: involvement in academic and social or extracurricular activities	Attendance, suspensions, voluntary classroom participation, and extra-curricular participation
Emotional	Students’ positive feelings about the school, teachers, and/or peers	Positive and negative reactions to teachers, classmates, academics, and school	
Academic			Time on task, credits earned toward graduation, homework completion,
Psychological			feelings of identification or belonging, and relationships with teachers and peers

2.3.1. Measuring Engagement

The same features that make engagement an interesting construct to examined, namely its multidimensional nature, turn engagement difficult to capture or measure (Fredricks et al., 2004). Table 14 shows common measures and measurement issues of the different dimensions of engagement. Concerning the limitations of current measurement strategies, it is possible to find the followings: 1) the inclusion of conceptually distinct scales for each type of engagement, 2) most measures fail to distinguish a specific source of engagement (i.e., a task or situation), and, finally, 3) measures do not consider qualitative differences in the level of engagement. In addition, researchers face the challenge of measuring in depth each component (i.e., emotional, behavioral, cognitive). The deeper the measurement of each component the longer the questionnaires become. This reflects the tension between conceptual/measurement clarity and practical reality. At the end, all depends on the original goals of the research on engagement. If research's goal is to search for predictions, then mixing the three components seems to be the best decision. However, this decision brings with it the side effect of lack of conceptual clarity. On the other hand, if researcher's goal is to understand a particular construct in depth (e.g., cognitive engagement) the more inclusive measures of engagement might be insufficient.

This dissertation does not intend to solve all these issues, but it has taken some of them as part of the scope in the study conducted. First, the relevance of the cognitive component means an extra effort on conceptual clarity, by identifying specific lines of research such as mental effort and information processing strategies. Second, this dissertation takes a closer look at the qualitative aspects of cognitive engagement, by conducting interviews with the participants. Finally, the measures of cognitive engagement used in the present study are task specific. As described in the Method section (see Section 4), the measures both differentiate tasks features and also are administered several times in relation to the specific tasks embedded in the educational game *Genius Unternehmen Physik* (See Section 3).

Table 14: Behavioral, Emotional and Cognitive Engagement Dimensions and Measurement Strategies

Engagement dimensions	**Measurement**	**Example**	**Issues**
Emotional:	*Positives/negatives reactions to activities, persons, school. It includes interest, enjoyment, enthusiasm, belonging, valuing, and intrinsic motivation.*		
	Self-reports Emotions (happy, bored, interested, etc.) Feelings about the teacher/school Identification with school Orientation toward work and school Intrinsic motivation	"I feel satisfied with school because I am learning a lot"	Items tend to tap emotional and behavioral engagement. Items do not always specify the source of the emotions. Items are more general than items used to measure interest or value.
Behavioral:	*Active participation. It includes effort, concentration, asking questions, persistence, attendance, following rules and avoiding troubles.*		
	Teacher rating & Self-reports Measures of conduct. Completing homework Following norms	The student "is persistent when confronted with difficult problems" "I work hard in my school work"	Scales combined all the aspects together. On-task/off-task does not consider the quality of the effort.

	Work involvement. Effort, attention, concentration, persistence, and participation. *Observation techniques* Attentiveness, doing work, Enthusiasm		
Cognitive	*Investing in learning. It includes self-regulation, thoughtfulness and willingness to go beyond basic requirements to master difficult skills. Deep strategy use and mental effort. Closely related to motivation to learn and learning goals.*		
	Quality of discourse Substantive engagement Procedural engagement *Self-reports* Strategy use: Metacognition, volitional and effort control Cognitive strategy use *Observational techniques* Self-monitoring, exchange of ideas, use of learning strategies, justifying answers, persistence, volitional control	“I went back over things I did not understand” “I skipped the hard parts”	Measures not situation specific. Self-reports do not link strategies to specific tasks. (behavioral)Effort versus mental effort needs to be differentiated.

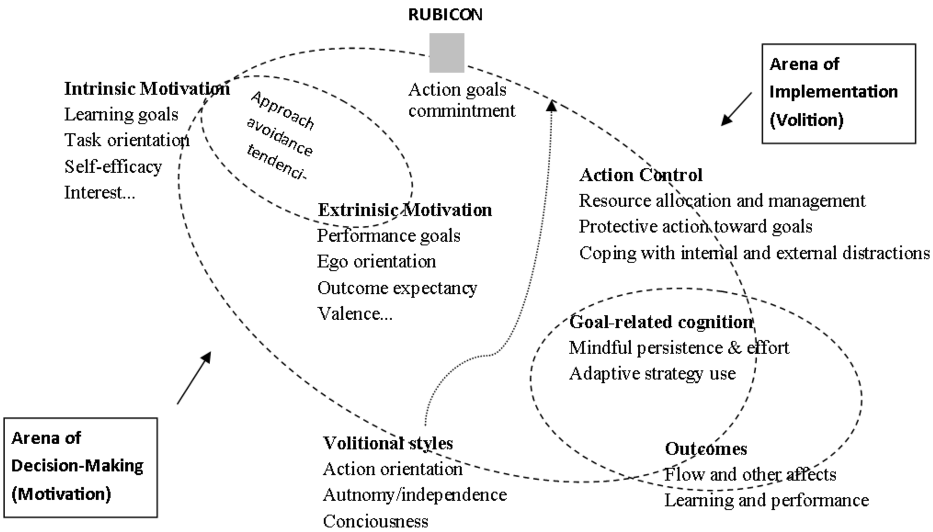
2.3.2. Engagement as a Volitional Construct

Before describing engagement as a volitional construct, a brief definition of motivation helps to set the stage for the differentiation between motivation and volition. *Motivation* is a psychological construct that attempts to explain the initiation and persistence of goal-directed activity and has been normally defined as "the process whereby goal-directed activity is instigated and sustained" (Pintrich & Schunk, 2002, p. 405). By *goals* the individuals are assumed to be conscious of something they are trying to achieve. *Activity* refers to physical (e.g., effort, persistence and other overt actions) and mental activities (e.g., planning, organizing, monitoring, making decisions and assessing progress). *Instigation* refers to the fact that individuals make a commitment and the first step towards achieving the goal. Finally, *sustainability* has to do with the motivational processes behind the "sustained" action (e.g., expectations, attributions, emotions).

On the other hand, roughly considered, volition refers to that sustainability mentioned above. In particular, Kuhl (1987) sustained that the motivational processes only conduct to establish a decision to act, in other words, they explain individuals goal *setting* processes. However, once individuals engage in concrete action to achieve the goal, that is, in goal *striving* processes, volition takes over and determine whether or not the goals are reached. Corno (1993) illustrates the relationship between motivational and volitional processes (see Figure 14). The author defined volition as "…control processes that protect concentration and directed effort in the face of personal and/or environmental distractions, and so aid learning and performance" (Corno, 1993, p. 16). According to the author, the primary role of volition is the *management* and *implementation* of goals, and is defined by three groups of constructs: action control, goal related cognition, and volitional styles (Corno, 1993).

Action control refers to the strategies and knowledge used to manage cognitive and non-cognitive processes to achieve a particular goal. For example, how individuals allocate and control their attention, and how they use self-motivation techniques and handle intrusive emotions, such as anxiety, are part of the strategies considered by this concept of action control. *Goal-related cognition* refers to the adaptive use of learning strategies and the mindful effort invested. These include 1) timely application of deep or elaborative processing, and 2) the monitoring and appraisal processes that determine the extent to which effort is sustained. Here Corno (1993) seems to equate the "adaptive" character of the use of learning strategies to the "mindfulness" of the effort invested, remaining in the qualitative side of the idea of mindfulness. The quantitative dimension was elaborated by Salomon (1984) (see Section 2.5.1.1). *Volitional styles* refer to more or less stable individual differences that affect individuals' goal choice and action control processes: the management of both the task and the personal resources available.

Figure 14: Conation: Its Motivational and Volitional Dimensions (Corno, 1993)



Note: Diagram re-drawn from Corno (1993).

2.3.3. Engagement and Mindfulness

Another construct aiming at the same idea behind engagement is *mindfulness* (Salomon & Globerson, 1987). It was proposed as a common denominator behind cognitive, motivational and personality factors explaining why available skills and knowledge – and may be also resources – are often not used or underused. Salomon and Globerson (1987) defined mindfulness as the "*volitional, metacognitively guided employment of non-automatic, usually effort demanding processes"* (p. 625, emphasis in the original). The role of mindfulness in learning is based on empirical evidence showing that when mindfulness is evoked the learning outcomes are improved. This research is reviewed in section 2.5.1.1 under Salomon's model of mental effort and learning from T. Interestingly, the authors claimed that mindfulness is analogous to Berlyne's (1960) concept of "subjective uncertainty". This state of mind can be fostered by a variety of stimuli (e.g., incongruity and surprise), which more recently have been hypothesized to be central for explaining the motivational effects of games (see Section 2.2). This implies that games as a medium can bring about this state of mindfulness.

Different factors have seemed to affect the degree of mindfulness showed by individuals. Salomon and Globerson (1987) classified these factors into two groups: *proximal* and *distal* factors. By definition proximal factors can be experimentally manipulated (e.g., task demands), while distal factors represent enduring trait-like dispositions, such as individuals own tendencies to be mindful or our own need for cognition (Cacciopo, Petty, & Morris, 1983). Another distal source is represented by individuals' own experiences in a particular cultural context which are collected in the form of scripts representing common attributions, obvious solutions and socially shared intuitions and procedures.

Proximal factors proposed as more amenable to experimental and instructional manipulation are: situationally specific motivation, volition and the perceived demands and value of a task. The specific motivation refers to the desire to employ non-automatic effortful processes in light of individuals' own perception of efficacy. Without this desire it is unlikely that effortful or mindful cognition are employed at all. Volition, as already discussed in the previous section, provides the link between motivation and effortful behavior, by sustaining the doing in face of competing motivations or goals. Finally, the perceived demand of a task, situation, or source of information affects the level of mindfulness employed. When the situation is perceived as familiar, undemanding, or too demanding relative to individuals' perceived self-efficacy, little mindfulness can be expected to be displayed by the individuals. Conversely, when the demands of the task are perceived as within a reasonable range of effort expenditure, more mindfulness should occur. As a case in point, Salomon and Leigh (1984) showed that children preconceptions of televiewing as undemanding led them to process mindlessly a highly demanding TV program and therefore learning less than it would be the case if a more mindful approach had been employed. However, this culturally held preconception can be influenced by instructional or experimental manipulation. For example, these children were instructed to view a demanding TV show "to learn" or "for fun". When the show was viewed to learn, children showed more mindfulness. As this is central for the model of Salomon and for the purpose of this dissertation, Section 2.4.1.1 presents the details of Salomon's and others' research on mindfulness or Amount of Invested Mental Effort (AIME). What is important to highlight is the centrality of individuals and their willingness to engage appropriately in the “intellectual partnership” offered by a particular medium, such as an educational game.

2.3.4. Engagement and Flow

As pointed out at the beginning of the section on Engagement, this concept has also been considered from the perspective of Flow theory (Csikszentmihalyi, 1990) or optimal experience. Three alternative perspective on engagement and flow have been proposed: engagement as flow (Shernoff et al., 2003), flow as supporting engagement (Whitson & Consoli, 2009), and engagement as producing flow (Pavlas, 2010). Shernoff et al. (2003) considered engagement simply as a synonymous of flow and defined the latter as the simultaneous experience of *concentration*, *interest* and *enjoyment* during an activity. These three processes are supposed to give rise to the deep absorption in an intrinsically enjoyable activity (i.e., Flow). On the other hand, for Whitson and Consoli (2009) Flow theory may provide a way of understanding the qualitative experience of learning, which narrower approaches such as “time on task” have failed to provide. The authors reported a series of qualitatively oriented studies on the conditions of classrooms for producing flow experience as a way to foster sustained engagement. According to the authors, by affecting the conditions of flow it is possible to foster students’ engagement. However, without providing a definition and operationalization of what flow is supposed to foster, it remains difficult to test this assumption. One of these conditions is the balance between skill and challenge. However, some empirical data suggest that skill/challenge balance it is not enough to produce the optimal experience of flow (Schweinle, Turner, & Meyer, 2006; Engeser & Rheinberg, 2008). These authors suggested that more important than the skill/challenge balance is the relevance of the task and individuals’ perceived ability to solve it. In the field of educational games, engagement has been indirectly linked to flow through intrinsic motivation (Pavlas, 2010). According to the author, engagement works to “maintain the intrinsic motivation inherent to the game task” (Pavlas, 2010, p. 47) and as intrinsic motivation is central to flow, then engagement is important for supporting flow as well.

Finally, from the authors discussed above, flow represents a multidimensional construct. The followings are the dimensions highlighted by the different researchers interested in flow:

Kiili (2005)

- Balance challenge/skill
- Coherent action/feedback
- Activity inner logic
- Concentration
- Experience of time
- Self and activity together

Whitson & Consoli (2009)

- Challenging activities and the adequate skills
- Action and awareness are merged
- Concentration at the task
- Goals and feedback
- Paradox of control
- Loss of self-conciousness
- Transformation of time

Pavlas (2010)

- Challenge-skill balance
- Action-awareness merging
- Clear goals
- Unambiguous feedback
- Concentration
- Sense of control
- Transformation of time
- Autotelic experience
- Loss of self-conciousness

As can be seen from the list, the three authors seem to agree upon which characteristics together conform the experience of flow. When compared with the conceptualizations of engagement already described, it seems that one common idea shared by engagement and flow is the requirement of concentration. What is yet to be determined though is whether the object of this concentration is different in both concepts and whether or not flow's ideas of loss of self-consciousness and perception of time are also relevant for learning processes. For example, in the context of the main claims and issues concerning educational games, in Section 2.2.2 was described the study of Kiili (2005) on flow and learning from a business game. The study examined the role of flow state on learning and explorative behavior. Although no role could be established between these variables, the study might have been more useful if it had presented the correlations coefficients among the flow state dimensions (e.g., concentration, time distortion and loss of self-conciousness), exploratory behavior and learning. With this information it could be possible to explore more complex relationships between flow state dimensions and learning. The study might also have been more persuasive if it had used real learning measures instead of measures of "perceived learning".

Another aspect on which flow and engagement seem to show some overlap is in the relationship between individuals' action and individuals' awareness. Researchers have suggested that flow entailed a merge between actions and awareness. From the perspective of engagement, this seems to parallel the relationship between behavioral engagement and cognitive engagement. Similarly, in terms of attentional processes this merging of action and awareness reflect that humans usually can attend to some stimuli while ignoring others (see Section 2.5.2) and that an individual's action presupposes a previously selective attentional process. For example, for an individual to be able to execute an action (e.g., right click on a link on the internet), she needs first to have paid attention to that stimulus, that is, to have isolated the stimulus from other stimuli. These unattended stimuli can be other stimuli near the link on the Internet, it could be another person sitting next to our hypothetical individual or could be also the individual herself. In this last case, flow idea of merging action and awareness seems to fit to this basic attentional mechanism: individuals seem to be able to attend to one and only one stimulus at a time. On the other hand, the same might be true when the stimulus under consideration is time.

In summary, flow research still needs to explore how the different dimensions of flow affect specific cognitive processing and learning. As it will become apparent, the way in which cognitive engagement is defined and operationalized in this study seems to be more related to learning outcomes and its role on learning has been supported by a sufficient body of evidence.

2.3.5. Other conceptualizations of Engagement

This final section on engagement briefly describes other conceptualizations of engagement that can be found in the literature as more or less independent lines of research. Among the concepts found are "engaged learning" (Jones et al., 1994), "productive disciplinary engagement" (Engle & Conant, 2002), and "procedural, conceptual and consequential engagement" (Gresalfi, Barab, Siyahhan, & Christensen, 2009).

According to Jones et al. (1994) an individual is "engaged in learning" if she is 1) responsible for their own learning (i.e., self-regulated), 2) energized by learning (i.e., feelings of pleasure related to intrinsic motivation), 3) strategic (i.e., knowing how to learn by developing learning strategies), and 4) able to collaborate (i.e., skills to work with others). In Engle and Conant's (2002) notion of productive disciplinary engagement, the term engagement is not directly defined but some indicators are proposed:

1) More students make substantive contributions
2) The contributions were more often made in coordination with each other

3) Few students involved in “off-task” activities
4) Students attend to each other (e.g., assessed by eye gaze and body positioning)
5) Students make emotional displays
6) Students spontaneously get reengaged over long periods of time

More recently, Gresalfi et al. (2009) defined three types of engagement in mathematics. *Procedural* engagement involved using procedures accurately, but not necessarily with an understanding of why one performs the procedure (e.g., solving a mathematic problem correctly). *Conceptual* engagement involved the idea of “sense-making” or understanding (e.g., asking why an algorithm might be useful to solve a problem). *Consequential* engagement required “interrogating the usefulness and impact of the selection of particular tools on outcomes” (p. 22). For example, an individual should be able to explain how her choice of a particular statistical method supports her conclusions. They further suggested consequential engagement to be the highest level and to involve the interplay between the intentional choice of a tool in a particular situation, and reflecting on the consequences of that choice in terms of its impact on that situation. Similarly, Azevedo (2006) referred to engagement as pointing to the quality of the relationship between an individual and an activity, in a broad context of material and social infrastructures. For the author high engagement entailed an individual choosing an activity, persisting in it, investing personal resources such as effort, and showing positive affect toward the activity.

These last examples are based on theories coming from a socio-cultural approach to teaching and learning in which the unit of analysis is the “activity system” (Greeno, 2006), that is, the individual and the artifacts that surround her. The methodological implication is that discourse analysis is the main tool this approach have used to support its claims concerning learning and teaching. From this two issues arise for the present study. At a practical level, this approach does not lend itself to the analysis of “solo” activities such as learning online in a hypermedia environment, or as in this case, an educational game. At a theoretical level, this approach tends to reject basic assumptions from cognitive psychology, and therefore, it is not interested in building on this perspective. However, in this dissertation, the focus is mainly cognitive and based on previous research from an information processing perspective.

Finally, in the context of games for entertainment, engagement has been related to immersion, fun, flow or a mixture of any of these concepts. It has also been suggested to be either a pre-requisite to immersion or to come after immersion. It is clear that in the field of game studies the term engagement refers to different concepts. For example, Dow, Mehta, Harmon, MacIntyre, and Mateas (2007) referred to engagement as meaning Huizinga's (1949) "magic circle", in which individuals step away from real life. More generally they understood engagement as an individual's involvement and interest in a particular activity. Unfortunately, the broad concept of "involvement" makes difficult to determine what sort of involvement they were referring to. Others, for example Lankoski (2011), in trying to understand the relationship between players and their Player Characters (PCs) have distinguished between "goal-related engagement" and "empathetic engagement" using the notion of engagement in an aesthetic sense. According to the author these two types of "affects" influence each other. Goal-related engagement referred to the goals and affects that the game system creates in relation to the PCs and empathetic engagement referred to how players develop an emotional connection with their PCs. A seldom differentiation between motivation and engagement made in this field was explicitly suggested by Schoenau-Fog (2011). The author understood motivation as concerned mainly with the question "why people begin to play a game?" and engagement as concerned with the question of "what aspects of a game makes players want to keep playing". Therefore, engagement is conceptualized as a continuing desire to play. By means of a survey, the author examined the "triggers" of this engagement. Although this differentiation helped to better understand the concept of engagement with games, a more careful analysis of this concept in Schoenau-Fog shows a limited precision, which makes his definition easily interchangeable with others activities. The author asserted that engagement can be explained as a process where players perform a series of activities to achieve a particular goal and feel positive affect. However, not only players' engagement can be explained in this way but a whole set of daily activities as well. This limits the usefulness of Schoenau-Fog's ideas to precisely identify and define engagement as a construct.

So far this literature review has focused on games and educational games and discussed the experience they are supposed to produce in terms of the multidimensional construct of engagement. This raises the questions of which type of engagement or which dimension should be considered most relevant for learning. The following section will discuss the cognitive dimension or cognitive engagement as the most central for understanding and fostering learning processes and outcomes.

2.4. Cognitive Engagement

As pointed out in the introduction of this dissertation, cognitive engagement is regarded as a construct that might help understand the relationships between effort and thinking as suggested by Dewey's (1913) quotation. Cognitive engagement has also been proposed to be the most relevant dimension of engagement, and one of the most relevant factors affecting learning to the extent that represents what is usually meant by "task involvement" (2.2.2). Finally, it has been proposed to be one of the defining features of educational games (Section 2.2.1.2). What comes next is a detailed discussion of this construct and the description of two models that have been used in this dissertation to measure it in the context of playing *Genius Unternehmen Physik* (see Section 3).

Cognitive engagement has been considered to involve both the idea of investment (i.e., mental effort) in learning and the idea of self-regulation or being strategic. Cognitive engagement refers to the deliberate task-specific thinking undertaken by an individual while participating in a learning activity (Järvelä, Veermans, & Leinonen, 2008). Cognitively engaged individuals invest high amounts of effort and persistence to understand a topic (Rotgans & Schmidt, 2011). Newmann et al. (1992) described cognitive engagement as a "psychological investment in and effort directed towards learning, understanding, mastering the knowledge, skills or crafts that the academic work is intended to promote" (Newmann et al., 1992, p. 12). According to Fredricks et al. (2004), cognitive engagement referred also to the deep or surface cognitive strategies used to process information (cf. Craik & Lockhart, 1972; Marton & Säljö, 2005). According to Rotgans and Schmidt (2011), these definitions tended to consider cognitive engagement as an individual's stable trait. Therefore, they suggested that cognitive engagement should be considered as dependent of the task at hand, in particular in terms of the autonomy that the task may offer (e.g., self-study, working in groups, listen to a lecture, etc.). In this context, they developed a "context sensitive" measure of *situational* cognitive engagement (SCENG) to assess ongoing cognitive engagement with the different phases of problem-based learning (see Section 4.3.2).

An important body of literature has investigated the role of cognitive engagement on learning and achievement (e.g., Greene et al., 2004; Hannafin, 1989; Pintrich & Garcia, 1991). Conceptually, cognitive engagement has been conceptualized as the highest form of self-regulation (Corno & Mandinach, 1983) and closely related to concepts such as motivation to learn (Brophy, 2010), mindfulness (Salomon & Globerson, 1987), amount of invested mental effort-AIME (Salomon, 1984),germane (or generative) cognitive processing (DeLeeuw & Mayer, 2008), intentional learning (Bereiter & Scardamalia, 1989) and committed learning (diSessa, 2000), among others. For instance, Brophy asserted that motivation to learn entails "…a student's tendency to find academic activities meaningful…and to try to get the intended learning benefits from them." (p.249). For Salomon and Globerson (1987), mindfulness represented the "volitional, metacognitively guided employment of non-automatic, usually effort demanding processes" (p. 625), and included the more specific concept of Amount of Invested Mental Effort (AIME) discussed later in Section 2.4.1.1. Germane cognitive processing occurs when an individual engages in deep cognitive processing (e.g., mental organization of material while relating it to prior knowledge) (DeLeeuw & Mayer, 2008) that contributes to the construction and automation of schemas (Paas, Tuovinen, Tabbers, &Van Gerven, 2003). As mental effort is a central concept in this study, a more detailed discussion of its meaning is provided in section 5.1.1).

Cognitive engagement unifies two areas of research, self-regulated learning and motivation (Corno & Mandinach, 1983; Fredricks et al., 2004). Each of these areas has a well-established set of evidence that warrants the study of the role of cognitive engagement on learning. According to Corno and Mandinach (1983) motivation theories have hypothesized reasoning processes (e.g., expectations) with motivational effects on individuals. Furthermore, at least in classroom contexts, it has been suggested that students usually actively engage in cognitive interpretations of the environments and of themselves, which influence the amount and kind of effort they spend in a task. On the other hand, cognitive engagement has also been conceptualized from the learning literature in terms of strategic and self-regulated behavior. For Corno and Mandinach (1983), cognitive engagement represented the highest level of self-regulated learning encompassing the processes of acquisition and transformation. The details of Corno and Mandinach's model are further discussed in section 2.5.2.1. Others have considered cognitive engagement as entailing three components: knowledge, learning strategies and thinking strategies (Pintrich & Schrauben, 1992). Similarly, some researchers have equated cognitive engagement to strategy use and have made the classical distinction between surface and deep strategies (e.g., Greene et al., 2004). For example, Greene et al. defined cognitive engagement as the "amount and type of strategies that learners employ" (p. 4). This correlational approach has established important relationships between motivational processes, such as self-efficacy, intrinsic motivation and achievement goal orientation, and cognitive engagement, learning (i.e., course grades, achievement or academic performance), and effort. Through path analysis this line of research has provided relatively strong evidence concerning the role of cognitive engagement, and arguably warrants further research on this construct in specific learning environments such as educational games.

It is proposed in this dissertation that the connection within these two research areas (i.e., motivation and self-regulation) is through the qualitative distinction between deep and surface level of strategy use, which demand different amounts of invested mental effort. An individual more cognitively engaged will exert more mental effort, use deeper strategies, create more connections among ideas (schema acquisition) and achieve greater understanding. Similarly, the forms of engagement identified by Corno and Mandinach (1983) (cf. Howard, 1989), that is, self-regulation, task focus, resource management, and recipience, imply different amounts of invested mental effort. This dissertation attempts to establish a connection between the general psychological investment (i.e., mental effort) and the cognitive processes suggested by self-regulated learning from an information processing approach. Therefore, cognitive engagement will be understood in terms of *individuals' invested mental effort during a particular task, and its concomitant information processing strategies*.

Motivational processes lead ultimately to effort. Cognitive engagement as already suggested has clear relations to motivational constructs (e.g., motivation to learn) and has as its defining feature the idea of "effort toward learning" as proposed by Newmann et al. (1992). This raises questions as what effort entails, what is meant by "mental effort" and why is important, and finally how many types of effort could be distinguished and how they can be captured or measured. The next section discussed these issues and presents the two models – Salomon's 1984 Amount of Invested Mental Effort and Corno and Mandinach's (1983) model of Cognitive Engagement – that have been implemented in this dissertation to capture what is meant by mental effort and cognitive engagement.

2.4.1. Mental Effort

"…the educational significance of effort, its value for an educative growth, resides in its connection with a stimulation of greater *thoughtfulness* not in the greater strain it imposes." (Dewey, 1913, p.58).

Dewey distinguished between two forms of effort. One derived from what he called a *divided* activity and the other from a *unified* activity. The difference between the two was whether or not individuals' "growing" self was separated from the facts to be learned or whether or not there was an identity between the self and the facts to be learned. In the first case, effort implied a division of attention to the extent that the individual would calculate the exact amount of attention needed to complete a task, while saving for herself the rest of her thoughts towards things that really interested her. In the case of a unified activity, that is, an activity which emerges from the actions taken by the individual out of her own tendencies and needs, effort would represent a demand for *continuity* when facing difficulties, and it would only be significant within an "expanding" activity oriented toward the fulfillment of an end. In Dewey's words:

> …The demand for effort is a demand for continuity in the face of difficulties (…) Effort, like interest, is significant only in connection with a course of action, an action that takes time for its completion since it develops through a succession of stages (Dewey, 1913, p. 47).

For Dewey the "experience" of effort has two related components. On the one hand, effort entailed a degree of *mental stress*, in which the activity persisted in spite of temporal obstacles. This stress is experienced as an emotion born out of a mix of desire and aversion. In this sense, effort would consist of the peculiar combination of conflicting tendencies (i.e., "away from" and "towards to" tendencies – dislike and longing). On the other hand, effort entailed *thinking*, *inquiry* and *reflection*. It is born out of the mental stress experienced and its related emotion which turns into a *warning* to think: is the end worthwhile under the circumstances? Is there a better course of action? Thus, reflection can lead to either reconsidering the end or seeking new means (discovery and invention) to achieve the desired end. For the present dissertation, the most relevant aspect of

Dewey's ideas about effort is its connection with thought. For Dewey (1913) effort should not be "pure strain" but "the way in which the thought of an end persists in spite of difficulties, and induces a person to reflect upon the nature of the obstacles and the available resources by which they may be dealt with" (p.51).

This experience of resistance that the obstacles offer can have two effects on the individual. On the one hand, it can weaken the impetus in the forward direction toward the end desired, and thus causing discouragement. On the other hand, it can enhance the perception of the end by highlighting to the person how important is the end or how intense is the desire to reach that end. This, in turn, can provide additional energy to the individual's effort to achieve this end. In this manner, the obstacle should bring to the consciousness the end itself and its value, making the individual to think in what she is doing.

Arguably, Dewey's ideas of effort and its relation to thinking have parallel to more current ideas on motivation and learning strategies: 1) what for Dewey represented the core experience of effort as the mixed feeling of disliking and longing, later motivational scholar have called it approach/avoidance tendencies: 2) for Dewey, the experience of resistances as awareness of the end that energizes its pursue, represent the current definition of motivation as energized and sustained behavior towards a goal; finally, 3) Dewey's connection of effort with thinking in terms of obstacles versus available resources to tackle the former to achieve the end could correspond to the strategies described in the self-regulation literature.

More recently, Mulder (1982) reviewed the concept of mental effort and distinguished between three uses of the concept of effort. First, effort was suggested to be a factor explaining differences in intellectual performance. This line of reasoning can be found in Spearman explanation of variability on the general (G) factor of intelligence, in achievement motivation theories which use "effort" as a dependent variable and finally, in the theories of memory and their notion of a limited capacity processor (e.g., working memory). Second, it has been used as a metaphor of the physical effort in theories of mental workload. These theories attempt to explain the "workload" experienced by individuals (usually with the connotation of exhausting and stressful) when performing tasks using personal computers. Its link to physical activity can be seen in the emphasis of measuring, for example, oxygen consumption. Third, a sort of compensatory effort has been proposed in research areas concerned with the effects of stressors (e.g., noise) on performance and physiological states.

Concerning “psychological effort”, Mulder (1982) further proposed two main uses of the concept. One use is related to attention and the field of mental workload, while the other use is related to executive control processes and linked to research concerned with the effects of achievement motivation, incentives and knowledge of results, among others. In both cases attempts have been made to identify physiological indices of mental effort while performing a particular task. In the former, it is assumed the individual is motivated and the mental effort is directed by the intrinsic complexities of the task. The latter, corresponds to motivational, control process of mental effort and does not assume an already motivated individual. In brief, the former portraits effort as undergoing and the latter portraits effort as acting (i.e., passive versus active).

These two views have been reflected in the field of learning with technology. On the one hand, one group of researchers assumes motivated individuals and ask why is it that they cannot perform tasks appropriately. Their answer: too much mental effort *experienced* or *cognitive load*. On the contrary, another group of researchers (e.g., Salomon & Globerson, 1987) asked why some individuals even though they have the skills needed to perform a task successfully, do not employ them. Their answer: lack of mental effort *invested* or *Mindfulness*. This dissertation is concerned with this latter approach and portraits effort as an action, an investment on behalf of the individuals. Next section revises two models that represent this general approach to mental effort. In summary, mental effort entails two aspects. Mental effort as a quantity or amount as reflected in Salomon’s model and mental effort as a quality, reflected by Corno & Mandinach information processing model of cognitive engagement. Both models employed together can help understand how the amount of mental effort translates into appropriate information processing.

2.4.1.1. Amount of Invested Mental Effort: Concept, Model, Measures and Research

The work of Salomon is based on Bandura’s (1982) assumption of reciprocal determinism which stated that general human functioning involves a sustain interaction among environmental, cognitive and behavioral factors. This led Salomon to argue that television affected people’s perceptions of that medium, which in turn influenced the outcomes that determined people's subsequent actions with and attitudes toward the medium (Salomon, 1981, 1983, 1984; Salomon & Almog, 1998). Similarly, regardless of motivation and ability, it has been suggested that individuals' perception of a medium (e.g., TV, games, web tutorials, etc.) may affect the extent at which individuals invest mental effort in processing the information from that medium, which, in turn, affects learning outcomes (e.g., Glaser et al., 2012; Salomon 1984).

In the context of studies on televiewing (i.e., "literate viewing"), reciprocal determinism was useful to understand the difference in literate viewing of children from USA and Israel. Salomon (1981) reported that US children watched more TV than Israel, so he hypothesized that US children should have had a greater level of mastery of TV related mental skills. However, findings suggested that Israel children showed a greater level of literate viewing than US children. These results contradicted the central assumption which stated that the simple and direct exposure to the medium cannot affect literate viewing if not coupled with individuals' intentions to extract some knowledge from the TV. In the search of an explanation, Salomon explored the patterns of televiewing of US and Israel children and noticed that in the US children watched television as a solo activity, while children in Israel watched television together with their parents. In this context, Salomon asked whether Israel children took televiewing more seriously than their peers in the US, thereby investing more mental effort in processing the material of the TV. Therefore, Salomon further suggested that more "seriousness", should lead to more learning from television.

The concept. This "seriousness" was captured under the construct of Amount of Invested Mental Effort (AIME) – later theorized by Salomon and Globerson (1987) under the idea of mindfulness (Section 2.3.2.3). AIME was defined "as the number of non-automatic mental elaborations applied to a unit of material" (Salomon & Leigh, 1984, p. 120). This definition assumes that individuals have a "pool" of capacity that they can use to handle a particular task. Salomon directly referred here to Kahneman's (1973) model of attention (see Section 2.5.2). So defined, AIME was proposed to be a synthesis of similar constructs such as depth of processing, cognitive capacity, mental elaborations, and automatic versus non-automatic processing (Craik & Lockahrt, 1972; Rose & Craik, 2012; Chein & Schneider, 2012; Schneider & Shifrin, 1977). The central idea is that automatic processing is mainly controlled by the properties of the stimulus, while the non-automatic processing entails a voluntary and intentional desire to expend mental effort. As suggested in the section of mental effort (Section 2.4.1), this construct belongs to the mental effort as invested – Mulder "compensatory" models – rather than the effort experienced, echoing Dewey's core distinction between acting and undergoing. Furthermore, its motivational roots come from the idea that the AIME finally employed it is a decision of the individual based on a series of consideration concerning the task at hand, the assessed demands and the perceived ability to successfully engage in the task: "The control individuals exert over their deeper or more mindful processing of material implies at least a modicum of choice. This choice is based on what they perceive the material or the task to demand of them relative to their perceived self-efficacy in responding to these demands" (Salomon and Globerson, 1984, p.120).

Finally, AIME it is not equal to depth of processing because AIME is about the effort with which an individual elaborate information, which can be at surface or deep levels. In other words, it is possible to invest high amounts of effort in elaborating surface features, and automatically – effortless – to elaborate deep features. For example, when learning a new language, one can invest a lot of effort in understanding the syntactical structure of a paragraph (surface features) and not that much effort in try to understand the meaning of the paragraph and less the communicative intention of the paragraph (deep features). On the other hand, an expert in a research field can automatically detect the basic assumptions of a scientific paper and therefore the general structure of the argumentation (deep features).

The model. Salomon's model of learning from television consists of a series of specific relationship between self-efficacy (PSE), perceived demand characteristics (PDC), the amount of invested mental effort (AIME) and learning in the context of televiewing. In terms of the reciprocal determinism hypothesis, PSE and PDC correspond to the individual, AIME to the individual's behavior (i.e., effortful or effortless) and the television corresponds to the environment. Salomon (1984) expressed these relationships as follows:

> …when relevant skills of effortful elaboration are available to learners and in the absence of explicit, specific, and unambiguous task requirements, AIME is influenced by PDC and PSE. The stronger the perceptions of how much effort a well-known class of stimuli require, and the higher (or lower) one's PSE, the

> greater (or smaller) AIME is expended with the resultant differences in elaboration-based learning (Salomon, 1984, p. 650).

These relations can be summarized as follows:
1) The higher the PDC and the higher the PSE, a higher AIME can be expected.
2) The higher the PDC and the lower the PSE, a lower AIME can be expected.
3) The lower the PDC and the higher the PSE, a lower AIME can be expected.

In the first situation, an individual that perceives a situation or a medium as highly demanding and that also feels confident in meeting those demands, may be more likely to perceive the worth of investing mental effort. On the other hand, some individuals in front of these same demanding circumstances may feel less confident in meeting them and therefore may not invest the necessary amount of mental effort. In other words, individuals in this second situation, given their perceived low probabilities of success in meeting the processing demands of a medium, would tend to think that it is not worthwhile to even try. In general terms, the higher the perceived demands of a situation, the lower is individuals' confidence in meeting those demands; therefore, it may be possible that PDC and PSE have a more negative relationship. However, in everyday life where situations are not highly demanding it is reasonable to expect high self-efficacy together with high investment of mental effort. The third situation describes individuals' overconfidence on their abilities to meet the perceived low demanding situation. Here individuals find the situation to be "easy" therefore feeling a high self-confidence in addressing the situation successfully without much investment of mental effort.

These three situations reflect the complexity of the relationship between PSE and AIME. As Salomon (1984) suggested, the relationship between these two variables can be thought of as an U inverted. That is, in situations in which PSE is too high or too low, a low level of invested mental effort can be expected. On the other hand, under moderate levels of PSE the more likely it is that individuals invest higher amounts of mental effort.

The measurement. For methodological considerations (see Salomon & Leigh, 1984) self-reports have been proposed as the more suitable method to assess AIME. This approach has been followed later by the research described below. Essentially, what varies is the number of questions in the questionnaires (usually from 2 to 5) (see Table 15). In terms of content, the questions mainly ask for how hard individuals try to understand, how much they concentrate and think about a piece of material and how much effort they put while watching, reading, and searching information or solving a task. In previous studies, except for Heers' (2005) study, most stimuli have been comparatively simpler. Therefore, as detailed in the method section (see 4.3.2), for the stimulus used in this study, an educational game, the AIME questionnaire was divided into three main areas. One concerning a learning task, and the other two concerning the simulation/game itself. In this way, it will be possible to differentiate on what exactly the AIME was invested.

Table 15: Summary of Questions used to Measure AIME

Author	Questions	Reliability
Salomon (1984)	How hard did you try to understand the film (story) How hard did your friends in the room try to understand the film (story) How much did you concentrate while watching (reading)?" How easy to understand" was the TV or text story?	.81
Salomon & Leigh (1984)	Questions themselves not reported, but the topics asked: Students' effort expenditure, Peers' effort, grade-level the story (seen/read) is most appropriate, Students' concentration during watching/reading	.67
Cennamo, Savenye, & Smith (1991)	How hard did you concentrate while watching the lesson? How much did the lesson make you think? How hard did you try to understand the lesson? (the other three are not mentioned)	Cronbach Alfa .55
Bordeaux & Lange (1991)	How much do you concentrate or pay attention to the show?	Cronbach Alfa .80 & .73 (for child and

	How much do you try to figure out that is happening in the show? How much do you use your brain to understand what the characters are doing and why?	adult TV programs, respectively
Douglas (1994)	Not described. 33 items.	Cronbach Alfa .73
Supinsky (1995)	How hard did you concentrate while doing the interactive video lesson? How much did the lesson make you think? How hard did you try to understand the lesson? How much did you try to remember what you saw in the lesson? How much effort did you put into comprehending the interactive video lesson?	Cronbach Alfa .74
Brookhart & Durkin (2003)	How hard did you try on the classroom assessment? How much did you concentrate when you did the classroom assessment?	Average Cronbach Alfa .80 (across 12 assessment)
Heers (2005)	Wie stark mussten sie sich während der Aufgabenbearbeitung im Adventure konzentrieren? Wie stark haben Sie die bearbeiteten Aufgaben zum Denken angeregt? Wie sehr haben sie sich angestrengt, um die Chemie-Inhalte zu verstehen? Wie sehr haben sie sich angestrengt, um das Adventure-Spiel zu verstehen?	Cronbach Alfa .75
Brookhart, Walsh, & Zientarski (2006)	How hard did you try on the classroom assessment? How much did you concentrate when you did the classroom assessment?	Range and median alpha values (.68–.88, Mdn =.82)

Huang (2010)	I tried very hard when I did X I concentrated a lot while I did X I put a lot of effort into preparing for X	Cronbach Alfa .993
Rieh, Kim, & Markey (2012)	Effort invested into searching Thoughts put into searching Concentration during the search Usefulness of search results Importance of search	Not reported
Glaser, Garsoffky, & Schwan (2012)	How hard did you try to understand the film? How much did you concentrate during the film? How attentive were you in following the film? How much cognitive effort did you invest during the film?	.74

The research. In order to disentangle the issue of the primacy of the medium over the individual, the model also suggested that PDC can be influenced so at to bring about higher amounts of invested mental effort. This idea was supported empirically in studies of televiewing. For example, Kunkel and Kovaric (1983) explored the relationship between PDC - AIME and AIME - learning. They did not used two different media, but within the same medium they provided the participants with different instructions. Before showing the participant the television program, one group was told the program was for Public Broadcasting Service (PBS) and another group was told that the program was meant for commercial television. Within each group half the participants were told they would be tested on the content of the program (high PDC). The other half did not expect any test at the end of the session (low PDC). At the end, all participants were shown the *same* program. In a similar procedure, but comparing across media (i.e., TV and Books), Salomon and Leigh (1984) manipulated participants' PDC. One group had to read a story (Print-Group) and another group had to watch a television program (TV-Group). Researchers made explicit effort to maintain the two conditions similar in terms of content knowledge. Then, half of the group was told to watch/read for fun (PDC-Fun) and the other half was told to watch/read to learn the content of the TV/Book (PDC-Learn). Later, Heers (2005) in the context of immersive virtual environments to learn physics also divided participants into two groups according to an "elaboration instruction", which asked half the participant to seriously process the learning material embedded in the same virtual world (PDC-Learn). In the context of film viewing Glaser, Garsoffky and Schwan (2012) instructed one group of participants to watch the film for entertainment (Low PDC) and to another group of participants to watch the film in order to inform themselves about the central topic of the film (high PDC). Then the same film was presented to the two groups of participants.

The details of these studies and others are summarized in Table 16. Not all of them have tried to test Salomon's model directly, but most have used the notion of mental effort under the concept of AIME while exploring specific topics. At the beginning (e.g., Salomon, 1984) the goal was to test the relationships between PSE, AIME and individuals preconceptions concerning a medium (e.g., TV). Next, the issue of whether or not individuals' way of processing information from a medium can be influenced by manipulating the PDC was studied. Later, Salomon et al (1989) provided compelling evidence concerning transfer of knowledge through effortful internalization of a software's metacognitive support. Finally, a developmental approach pointed out that the mental effort while viewing TV depends on children's age (Bordeaux & Lange, 1991).

More into the 1990's the issue of AIME was applied in the context of learning from interactive video (e.g., Cennamo et al., 1991). The authors found no difference on AIME across interactive video, instructional television and television. However, positive relations between AIME and learning were found. Some contradictory results to those of Salomon, probably due to a developmental issue, had to do with the perception of what medium was the easiest. In Salomon's (1984) studies children found TV, but Cennamo et al. (1991) college students found interactive video to be easiest to learn from. Another contradictory finding of Cennamo and colleagues was that the more difficult it was to learn from a particular medium, the lower the recall scores. On the other hand, Salomon (1984) had suggested that if the medium is more demanding, the effort should be higher and therefore also the learning or recall of the material processed. Cennamo suggested that college students, as opposed to children, may have a better sense of their ability to learn from a medium.

Another group of studies has focused on exploring the influence of AIME on individuals' engagement with classroom or formative assessment (e.g., Brookhart & Durkin; Brookhart et al., 2006; Huang, 2010). This line of research uses a correlational approach to explore the relationships between AIME and particular motivational traits such as self-efficacy and goal orientation. Results are inconsistent concerning these relationships. While Brokhart et al. (2006) found a positive relationship between AIME and Active Learning and between AIME and performance assessment, Huang (2010) did not found any relationships between AIME and self-efficacy. Similarly, a study trying to understand the structure of relationships among different conative-related constructs found that AIME (mindfulness) had significant correlations with deep approach and action-control decision scale (Douglas, 1994).

A couple of studies have explored AIME in the context of more modern technology-enhanced environments. Heers (2005) manipulated PDC to explore the role of AIME in learning physics from a virtual environment or "learning adventure" software. The author did not find an effect of the "elaboration instruction" (i.e., high PDC) on the actual AIME invested. Rieh et al. (2012), in the context of internet-based activities such as web searches and the use of library system, found that the higher the PDC the more AIME was reported by the participants regardless of the system used. In summary, the studies that followed Salomon's pioneer work have examined whether or not AIME can be influenced by manipulating individuals PDC, how AIME relate to more trait-like motivational variables when trying to understand how conative-related construct affect each other, and whether or not AIME influence the learning outcome in contexts such as virtual environments, classrooms, and films.

So far it has been discussed the role of AIME, the quantitative dimension of mental effort, on learning and its relation to other constructs. However, the degree at which this mental effort translates to appropriate information processing has not been yet discussed. The next section discusses Corno and Mandinach information processing model of cognitive engagement and identify which processing strategies are more in line with the amount of invested mental effort as defined in this section.

Table 16: Summary of Research explicitly using Salomon's (1984) AIME Measure

Author & Subjects (Ss)	Treatment & Duration	AIME	Learning measure	Correlation/Comparison of AIME measures
Salomon (1984) 124 sixth-grade students	*Compared Ss' preconceptions of TV and Print, AIME and learning from these two sources. Findings showed individuals perceived TV as "easier" than Print. When no external instruction is available, these perceptions influence individuals AIME and, hence, learning.*			
	TV group $n = 62$ Print group $n = 62$ 13-25 min.	4 questions Cronbach Alpha = .81	Factual recognition Inference-making	*In TV group:* r= .04 & r = .69*, with recognition and inference, respectively. r=-.49** with self-efficacy r=-.28* with attributed realism. *In Print group:* r=.24 & r=.72* with recognition and inference, respectively. r=.37** with self-efficacy r=-.31* with attributed realism. *Between groups* t(122) = 2.21, p<.05 greater AIME in Print group F(1,122) = 7.42, p<.01. Significant media effect. Print group obtained higher achievement scores.
	Study 1: Ss in the Print group show higher levels of AIME. No differences on learning (recall & inference) were found.			

Salomon & Leigh (1984) Study 1: 64 sixth-graders Study 2: 87 sixth-graders	*Study 2: Ss in PDC-LEARN condition showed greater effort expenditure in TV, mobilized their abilities more, and generated more correct inferences. Therefore, TV material can be processed more "mindfully" when individuals' perceived demand characteristics are manipulated.*			
	Study 1: TV group *n* = 32 Print group *n* = 32	4 questions. No Cronbach Alpha reported.	Factual recall Inference-making	*In TV Group:* r=.11 recall r= .13 inference making *In Print Group*: r= .28 recall r= .42* inference making t(2.21) p<.05 AIME higher in Print Group
	Study 2: 2(TV vs. Print)x2(manipulated PDC) factorial design. TV Group: PDC-FUN *n* = 18 PDC-LEARN *n* = 22 Print Group: PDC-FUN n = 22			F (1, 83) = 3.93, p < .051 PDC-LEARN vs PDC-FUN F (1, 83) = 4.02, p < .05 Medium/PDC interaction r=.18 Recall (TV) r=.43 Recall (Print)

	PDC-LEARN n = 25			
Salomon, Globerson, & Guterman (1989) 74 seventh-grade students	*Ss using the version of the Reader Partner with metacognitive-like guidance reported higher mental effort in the process of reading and improved their later comprehension and quality of their essay than Ss in the other groups. The interaction with the Reader Partner affected the mastery of transferable skills through an effortful internalization of its guidance.*			
	Computer program “Reading Partner” Ver. 1: Metacognitve group (MC) *n* = 25 Ver. 2: Question group *n* = 25 Ver. 3: Control group *n* = 24	3 questions Cronbach Apha = .87	Metacognitive reconstruction Reading comprehension Quality of essays	F(2, 71) = 3.69, MSC = .63. Sheffe test AIME higher in MC group. r = .35 AIME & Metacognitive reconstruction in MC r = .20 AIME & Reading posttest in MC
Bordeaux & Lange (1991)	*Ss’ AIME during home viewing TV varied as a function of viewer age and type of program viewed. Child-program AIME scores were significantly higher four second grade-level children than for fourth and sixth grade level children. This developmental trend was not supported for adult-program watching. Findings warrant the use of mental effort measures in studies of processing of televised content.*			

116 second, fourth, & sixth-graders	Correlational study examining the developmental differences in AIME scores.	Four 4-point likert scale questions		Significant main grade level effect $F[2,110]=3.54$, $p= .03$, and significant grade level by program type interaction $F[2, 110]= 7.57$, $p=.001$.
Cennamo, Savenye, & Smith (1991)	*Ss were presented with three different media (interactive video, instructional television, and television) and were evaluated on their preconceptions of these media and the perceived mental effort as a function of responding to practice questions*.			
71 college students				No significant differences on AIME among the three media. Correlations (total sample) between AIME and recall $r = .33$, $p < .01$, and between AIME and inference, $r = .27$, $p = .012$.
Douglas (1994) Sixty college students	*Ss responded to different instruments measuring conative-related constructs. Among the measures were mindfulness, approach to learning, action versus state orientation, and mastery versus performance orientation.*			
	Correlational study among conative variables	33 items for Mindfulness	No	Mindfulness showed significant correlations between Deep Approach (.83), Strategic Approach (.65), and Action Control-Decision scale. A significant negative correlation with Surface Approach (-.40) and Apathetic approach (-.27) and not significant correlation with neither Mastery Orientation (maybe due to low reliability) nor Academic Self-confidence.

Supinsky (1995)	*Ss in two instructional strategies showed no significant differences on measures of achievement, amount of invested mental effort and attitudes.*			
	Group 1 *n*=45 Individualized video/cooperative learning Group 2 *n*=44 Cooperative/interactive video 50 minutes session each group	Five 5-points likert questions	Achievement	No significant differences between the groups was found t(87)=.53, p= .59. Group 1 M=3.87, SD=.498. Group 2 M=3.82, SD=.501. No correlations between AIME and achievement were reported.
Brookhart & Durkin (2003) 96 students	*Individual differences in motivational variables measured while engaged in different classroom assessment format yielded a consistent positive correlation between active learning and AIME scores. In general, performance assessments showed positive correlations among mastery goal orientation, performance goal orientation and AIME.*			
	Correlational study of a teacher's classroom assessments	Two 5-point likert scale		Significant positive correlations between AIME and Active Learning across assessment tasks and subject (from .18 to .67) and with Mastery Goal Orientation (.25 and .75) and Performance Goal Orientation (from .23 to .60). Negative correlation with Superficial Learning (from -.13 to -.66). Mixed results for AIME and achievement.

Heers (2005) 60 college students	*Ss were instructed to play a learn adventure software for learning physics under condition of high-low Immersion and with or without an "elaboration instruction" for investing mental effort with the physic content. Results showed that the elaboration instruction failed to produce the corresponding levels of invested mental effort. Higher mental effort led to higher exploration behavior, but no higher knowledge acquisition.*			
	Experimental design. Four groups high and low in immersion and with or without elaboration instruction. *n* = 15 per group.	Four 7-point likert scale	Knowledge acquisition test	Two ways ANOVA yielded no main effect for the "elaboration" instruction. No interaction effects were found.
Brookhart, Walsh, & Zientarski (2006) 223 eight graders	*Motivation and effort correlational patterns with achievement on classroom assessments were examined. Motivational variables (e.g., self-efficacy) predicted achievement after controlling for prior achievement. After controlling for these variables, effort variables (i.e., AIME and active learning strategies) did not predict additional variance in classroom achievement.*			
	Correlational study of 4 teachers'	Two 5-point likert scale	Classroom assessment	AIME across teachers and assessments: M=4.12, SD=.75.

	classroom assessment			
Huang (2010)	*Ss participated in either Divergent or Convergent assessment activities. Trait characteristics, such as self-efficacy, were correlated with event-related motivational variables, such as AIME and perceived task characteristics, and strategy use variables, such as commitment control. AIME did not show a significant relationship with the trait motivational variables.*			
105 college students	Correlational students under Divergent and Convergent assessment situations.	Three questions, no anchor points mentioned	Not reported	AIME was not a variable significantly related to traits motivational variables in neither the Divergent or Convergent assessment activities.
Rieh, Kim, & Markey (2012) 15 college students	*Ss were required to perform two types of tasks (Product Task and Research Task) using two information retrieval systems (Library system and web search). Findings showed that people perceiving the task as demanding reported more AIME, regardless of the retrieval system used.*			
		Five 10-points likert scale	Not reported	The difference in AIME for the more demanding task was smaller than the difference in the less demanding task. For example: Concentration during search under the Product Task (easy): Web search: M=5.27; SD=2.54

				Library search: M=7.4 SD=1.68 Concentration during search under the Research Task (demanding): Web search: M=6.73; SD=1.75 Library search: M=7.13; SD=1.68
Glaser, Garsoffky, & Schwan (2012)	*Ss viewed two films either for entertainment (entertainment condition) or to inform themselves about the topic of the film (information condition). These two different goals for viewing a film did not influence knowledge acquisition. However, participants with an information goal reported a greater AIME than participants with an entertainment goal.*			
60 college students	Study 2. A 2x2x2 factorial design, with Narrative distance (close vs. distant), instruction (entertainment vs. information), and film content (Hattusa vs. Tucume)	Four 5-point likert scale	Knowledge acquisition	Significant main effect of instruction on AIME, $F(1,66)=4.85$, $p<.05$. No learning effect between groups.

2.4.2. Self-regulation

According to Schunk and Zimmerman (1994) Self-regulation refers to "students' self-generated thoughts, feelings, and actions, which are systematically oriented toward attainment of their goals" (p. ix). Regardless of the specific definition, some key components seem to be part of any definition of self-regulation (Pintrich & De Groot, 1990): metacognitive strategies for planning and monitoring, management and control of effort, and actual cognitive strategies used to learn (e.g., elaboration). What follows is a description of one model of self-regulation from an information processing approach which is central for the empirical study conducted in this dissertation.

2.4.2.1. Corno & Mandinach's Model

Corno and Madinach (1983) considered cognitive engagement as one form – and the highest one– of self-regulation. For the authors, self-regulated learning is "an effort to deepen and manipulate the associative network in a particular area (which is not necessarily limited to academic content), and to monitor and improve that deepening process" (Corno & Mandinach, 1983, p. 95). The authors attempted to link normally separated lines of research, that is, motivation, learning and instruction. They suggested that motivation theories hypothesize reasoning processes, such as expectations, that have motivational effects on individuals. More importantly, they suggested that, at least in classroom contexts, students usually engage actively in cognitive interpretations of the environments and of themselves, which influences the amount and kind of effort they spend in a task. However, these interpretative processes described by motivation theories are disconnected from what the students in actuality do when "engaged". Secondly, these interpretative processes are seen both as influences and consequences of cognitive engagement. A similar argument made Pintrich and Schrauben (1992). They suggested that students' cognitions and perceptions about themselves and their environment mediate their behavior. The degree of adaptation of the behavior to the environment would depend on individuals self-regulatory processes.

The model distinguished five component processes that corresponded to two different classes of information processing: *acquisition* and *transformation*. The acquisition process is composed of *Alertness* and *Monitoring*. Alertness is about attending and receiving incoming information (e.g., reading line-by-line). Monitoring involves continuously tracking and gathering information, rehearsing information and self-checking progress on a task (e.g., double-checking answers to a test). The transformation process is composed of *selectivity*, *connecting* and *planning*. Selectivity is discriminating stimuli and distinguishing relevant from irrelevant information (e.g., forming a picture of the goals of the task). Connecting is about searching and linking familiar knowledge to new information (e.g., using prior knowledge or experiences). Planning involves organizing a sequence or routine for solving the task.

According to Corno and Mandinach (1983) the acquisition process is expected to control the transformation process (i.e., selectivity, connecting new information with what is stored in memory, and planning specific performance routines). The model assumed that individuals' activities are focused on manipulating the associative network they possess. It also assumed that self-regulated processes are related to ability and that they can also be acquired through training. Corno and Mandinach (1983) considered that for a given task an individual may use more or less of either the acquisition or transformation process. According to Howard (1989), implicit in the model of Corno and Mandinach (1983) was the assumption that the forms of cognitive engagement proposed should vary in terms of the mental effort they require:

> Self-regulated learning, then can be seen as an adaptive combination of cognitive skill and effort. Not only does a self-regulated student have the ability to engage in the appropriate kinds of cognitive processes in optimal amounts for a given task, but this student also adapts the effort applied in this work to the task at hand (Howard, 1989, p.13).

Consequently, a self-regulated student is one that is able to 1) engage flexible in the four forms of engagement, 2) to be motivated to use the strategies of these forms of engagement, and 3) to employ action control or volitional strategies. The four forms of engagement differentiate themselves for the degree of use of either acquisition or transformation processes. For instance, cognitive engagement involves high levels of use of both acquisition and transformation process. Table 17 summarizes these four forms of engagement:

Table 17: The Four Forms of Cognitive Engagement by Corno and Mandinach (1983)

		Use of Acquisition processes	
		High	Low
Use of transformation processes	High	Self-regulated learning	Task Focus
	Low	Resource management	Recipience

Note. Acquisition entails Attending and Monitoring strategies; Transformation entails Selectivity, Connecting, and Planning.

Concerning the other three forms of engagement, Corno and Mandinach (1983) characterized *Resource management* as a form of cognitive engagement in which the individual intentionally avoid the mental effort of carrying out information transformation processes. By gathering help from other sources these individuals seldom use information acquisition and meta-level planning. Some initial research conducted by the authors showed that success expectation could not correspond with the increase of self-regulated learning. *Recipience* was presented as an even more passive response in which the learning environment would accomplish most of the transformation and low-level monitoring. Only some acquisition processes, such as rehearsal, could be executed. It is considered less demanding than resource management and represents a general "passivity" with little mental investment. The ways in which the learning environment, a teacher or a particular technology may lead to Recipience is by "short-circuiting" techniques (e.g., advanced organizers, diagrams and summaries, etc.). The authors argued that high ability students tend to engage in self-regulated learning even when the instruction has short-circuited some processes. For example, a student presented a diagram might decide to construct her own one if that is helpful to better understand the topic at hand. Short-circuiting might be useful for lower level tasks and for increasing task-specific self-efficacy in low achievers. Finally, *Task Focus* is expected to occur when individuals intentionally activate more transformation processes (i.e., selectivity, connecting new information to existing one, and task-specific planning) than acquisition processes and should be a useful strategy for tasks requiring quick analytic responses, little self-checking and use of external resources. It is considered to be a form of intelligent investment of mental effort and highly likely to occur in, for example, test taking situations. Corno and Mandinach (1983) identified the effective and ineffective strategies of students when responding to a spatial ability test (see Table 18). The effective strategies used by Task Focus are mainly selectivity, connecting and planning. Concerning the ineffective strategies, the authors interpreted them as a high concern or sensitivity to extra-task elements that would reflect, they suggested, alertness and monitoring, hence a dominant acquisition function.

Table 18: Examples of Task Focus and Alerting/Monitoring Strategies

Effective strategies	Ineffective strategies
Careful attention to specific information	Skips the selection of information
Analytic comparison of item features	Tendency to add additional information
Mental figure rotation	Hesitancy reflected in double checking answers
Winnowing out likely incorrect response options	Unwillingness to guess when unsure

Howard (1989) provided an example of the different forms of engagement and their respective cognitive processes involved (i.e., alertness, selectivity, etc.) in the context of a reading task. The author hypothesized how each of the four forms of engagement should look like. In Recipience approach, student should use Alertness (e.g., read the chapter quick), some monitoring (e.g., a superficial consideration whether the text was understood), and rehearsal (e.g., rehearse the outline of the chapter).
A Resource Management approach would employ Alertness and Rehearsal (e.g., read chapter and outline and rehearse the information), Planning (e.g., overview chapter before reading, look for possible questions) and a general tendency to find out from others what the main discussion questions could be. In the case of Task Focus the student would use Alertness, Selectivity and Connecting (e.g., reading the chapter, highlighting, writing down questions that come to mind, adding some ideas to the given outline), and Planning (e.g., go back and forth the chapter to identify sources to answer a discussion question). Finally, self-regulated learning would include Alertness, Selectivity, Connecting (e.g., read chapter, write down questions that come to mind, highlighting, adding information to the outline), Planning (e.g., go to other resources to get additional information, overview chapter before reading, look for discussion questions), Monitoring (e.g., go back to parts of the chapter to check on understanding and/or clarify confusions).

As depicted by the example and the theoretical perspective, the relationships between four forms of engagement and the information processing strategies are illustrated below:

Table 19: Cognitive Processes in each Form of Cognitive Engagement

	Acquisition		Transformation		
	Alertness	Monitoring	Selectivity	Connecting	Planning
RE	Yes				
RM	Yes	Yes			Yes
TF	Yes		Yes	Yes	Yes
SRL	Yes	Yes	Yes	Yes	Yes

Note. RE = Recipience; RM = Resource Management; TF= Task Focus; SRL = Self-Regulated Learning.

An example of an empirical study trying to identify the four forms of engagement and their respective processing was conducted by Rogers and Swan (2004). They applied this model in the context of an Internet search task. Through the observation of overt behaviors and a questionnaire of perceived strategy use of 80 undergraduates, the authors explored whether the strategies assessed through direct observation of overt behavior and questionnaires of perceived use of strategies would reveal the four forms of cognitive engagement consistent with Corno and Mandinach's (1983) model. Behaviors such as clicking, typing a keyword, and taking notes were collected from a pilot study. A frequency score was used for each of the strategies proposed by the model and a sum of the frequencies of the strategies for each of the two information processes (i.e., acquisition and transformation) was obtained. Over 8000 behaviors were coded. From them around 20% were discarded. From the remaining behaviors coded 3292 pertained to Selectivity, 1840 to Connecting, 923 to Alertness, 526 to Planning, and 225 to Monitoring. Cluster analysis on the strategy-use-per-minute yielded four groups with different levels (High or Low) of acquisition and transformation processes: The High Acquisition/High Transformation, corresponding to the self-regulated learning; the Low Acquisition/High Transformation, corresponding to Task Focus; the High Acquisition/Low Transformation, corresponding to Resource Management; and, finally, the Low Acquisition/Low Transformation, corresponding to Recipience. This study suggests that it is possible to empirically distinguish these processes, although not without some difficulties and inconsistencies in the data (Howard, 1989).

This model of cognitive engagement captures the qualitative dimensions of the concept not considered by Salomon' (1984) model of AIME. Together they provide a useful framework to start exploring this important construct in the context of educational games.

2.5. A Conceptual Framework for Research on Engagement in Educational Games

As discussed in section 2.2 researches on educational games have focused mainly on the effects or outcomes rather than on the process of learning from games. Likewise, several studies have postulated mechanisms for explaining the effectiveness of the games studied, but with almost no empirical data to support the claims concerning such mechanisms. This leads to a sort of "circular" argument (Eisenhart, 1991). For Eisenhart the difficulty of researchers is to explain empirical results in a non-circular manner. That implies the need to show that the data occur in a particular way *because of* the processes described by the explanation and not accidentally. What may happen is that researchers simply describe or identify data in terms of a predetermined framework. In doing this, researchers assume but do not *demonstrate* that the explanation derived from the framework is appropriate. In this sense, a conceptual framework is defined as an *argument* of what concepts (i.e., variables) are important and why they are needed in order to explain the empirical results obtained when attempting to understand a phenomenon. More formally, according to Eisenhart (1991), a conceptual framework is a

> …skeletal structure *of justification*, rather than a skeletal structure of explanation based on a formal logic (i.e., formal theory) or accumulated experience (i.e., practitioner knowledge). A conceptual framework is an argument including different points of view and culminating in a series of reasons for adopting some points–i.e., some ideas or concepts– and not others. (Eisenhart , 1991, p. 209).

This section summarizes different perspectives about games, learning and engagement in order to build a conceptual framework along the lines suggested by Eisenhart (1991). The present framework attempts to synthesis the concept of engagement as defined in Section 2.3, the gaming model proposed by Garris et al., (2002), Kolb's (1984) model of experiential learning, and Corno's (1993) volitional processes.

From the section 2.3 on engagement, it was established that this construct entailed three dimensions: behaviors, emotions and cognitions (See Table 20). As already mentioned behavioral engagement referred to active participation and includes effort, concentration, attendance, following the rules, and avoiding trouble. Emotional engagement referred to the extent individuals experience positive and negative reactions to teachers, peers, and activities in general, and included emotions such as interest, enjoyment, enthusiasm, feelings of belonging and valuing of learning. Cognitive engagement is defined as investment in learning and includes self-regulation, thoughtfulness, and willingness to go beyond the basic requirements to master difficult skills (Fredricks et al., 2004). Therefore, an individual showing a high engagement is characterized by high attention, interest and enjoyment, and effort to master new skills. Meanwhile, a low engaged individual is characterized by boredom, inattentiveness, and passivity (Bohnert, Fredricks & Randall, 2010).

Garris et al. (2002) gaming model is a highly influential one. However, as suggested before (2.2.2), it mixes motivational and volitional constructs together, when in fact a separation and distinction between them can help understand better the effects of games in general and educational games in particular. First, as can be seen in Table 20, Garris et al. (2002) model considered under their "user judgment" category aspects of human functioning, such as task involvement, interest and enjoyment. On the contrary, according to the conceptual framework proposed here, the general notion of "task involvement" is operationalized in terms of *cognitive engagement* as reflected by 1) individuals use of transformation processes of selecting relevant information, connecting incoming information with prior/familiar information, and planning a set of actions to achieve a learning goal, and 2) self-reported mental effort invested or mindfulness. From the perspective of volition this cognitive engagement corresponds to the volitional process of *goal-directed cognition*, which entails adaptive strategy use, mindful persistence and mental effort. On the other hand, Garris et al. (2002) notion of interest and enjoyment are considered to belong to the emotional engagement dimension. Second, Garris et al. (2002) model considered also "task involvement" as related to attentional processes. On the contrary, the present framework locates the attentional component of human functioning as behavioral engagement. From the perspective of volition, this engagement would correspond to the process of action control, mainly related to its resource allocation component. In summary, the process model of Garris et al (2002) and its initially fuzzy concept of "task involvement", has been differentiated into two distinct dimensions of engagement which correspond to two distinct volitional processes. In this way, the present framework explicitly understand engagement as a multidimensional and volitional construct different from motivational constructs such as, for example, self-efficacy.

From the perspective of experiential learning theory (Kolb, 1984) learning corresponds to “the process whereby knowledge is created through the transformation of experience. Knowledge results from the combination of grasping and transforming experience” (Kolb, 1984, p. 41). Arguably, this grasping and transformation of experience may well correspond to Corno and Mandinach’s (1983) acquisition and transformation processes, that is, within the qualitative aspects of cognitive engagement. Furthermore, if the way of grasping the experience is through what Kolb (1984) called the “concrete experience”, then this aspect could be reflected by the behavioral dimension of engagement. On the contrary, if the grasping of experience is made through abstract conceptualization, then this grasping could be represented as the acquisition process described by Corno & Mandinach (1983). Finally, if transforming the experience is related to reflective observation, then the transformation processes proposed by Corno and Mandinach (1983) could capture this aspect of Kolb’s model (1984). In summary, it is suggested that Kolb’s experiential theory might be operationalized as both cognitive and behavioral engagement.

Having established the main dimensions of engagement and how they seem to relate to central models of learning from games (Garris et al., 2002) and volitional processes relevant for learning (Corno, 1993), the next step is to depict the general *conceptual framework* for research on engagement and educational games. This conceptual framework has been inspired by Garris et al. (2002), which also seemed to have inspired other models of motivation and games (Mattheiss et al., 2009; Wainess, 2010), and also by Ennemoser (2009). The present conceptual framework presents two main characteristics. On the one hand, it borrowed the classical input–process–outcome structure from Garris et al (2002) to organize the variables that relate to learning. On the other hand, it borrowed from the mediational analysis approach from McKinnon (2008) and Ennemoser (2009) the distinction between mediators and moderators and between theory of action and conceptual theory. In this way, the present framework provides specific routes to understand how an intervention (i.e., an educational game) produces the expected outcomes and under what conditions and for *whom* such outcomes can be expected. Below these two characteristics are separately discussed and the contributions of the present conceptual framework are highlighted.

Input–Process–Outcome (IPO). Garris et al. (2002) organized the variables related to learning from games in terms of an IPO model. The input consisted of the game characteristics (e.g., fantasy, rules, challenge, etc.) and the instructional content. The process considered a cycle of "re-engagement" including the users' judgments and behaviors together with the game's feedback. The outcome referred simply to learning outcomes. Finally, "debriefing" was suggested as a bridge between the process and the outcomes and considered to be a type of instructional support. Influenced by this gaming model Wainess (2010) and Mattheiss et al. (2009) proposed a model of motivation in game-based learning by organizing the variables in a similar ways. Wainess (2010) grouped the variables into three categories: independent variables (e.g., games features), moderator and mediator variables (e.g., cognitive load), and dependent variables (e.g., learning outcomes). It improved Garris et al. (2002) model by expanding the number of game characteristics and by establishing specific links between these features and motivational variables such as expectancies. The model of Wainess (2010) represents also an improvement by having classified some variables in terms of "mediating" and "moderating" variables (e.g., cognitive load). However, he did not distinguished between the two types of variables. On the other hand, the model of Mattheiss et al. (2009) built directly upon Garris et al. (2002), by changing the labeling of the groups of variables: conditions, activities and outcomes. As an improvement, the authors added as part of the input the role of the individual in terms of the expectations they might have.

Mediational perspective. The present conceptual framework proposed the use of the mediational perspective in order to develop an empirically based theory of learning from games. This perspective is important if the interest is to understand how something works, for whom and under what conditions. Ennemoser (2009) suggested that mediation and moderation could help understand how interactive processes – central to games – produce particular effects. As Ennemoser writes: "Serious games researchers have to pay attention to what exactly happens between cause and effect and have to theorize about the involved processes in terms of mediation and moderation (Ennemoser, 2009, p. 356). According to the author, the mediation question is "how does it work?" and the moderator question is "does it work for all in all conditions?"

The starting point is to explore the processes by which an educational game (input) produces changes in learning, motivation or attitudes (outcome). Shedding light into these processes entails the identification of variables that explain how or why an educational game affects a specific outcome. These variables have been called *mediation* variables or *mediators* and explain how one variable transmits its effect to another one (MacKinnon, 2008).

A mediational analysis attempts to determine the extent to which a third variable translates the effect of one variable to another. This third variable could be a *mediator variable*, a variable that occupied an intermediate place in the cause chain between the independent and the dependent variable. As Mackinnon writes: "I*n a mediational analysis, the independent variable causes the mediator which then causes the dependent variable*." (Mackinnon, 2008, p. 8, emphasis in original). For example, the *deliberate use* can be a mediator of the causal chain between the interactivity of the game and learning (Ennemoser, 2009). The third variable could also be a *moderator variable*, a variable that interacts and change the relation between the independent and dependent variables, so that this relationship changes at different levels of the moderator. As Mackinnon explains: "*A moderator is a variable that changes the sign or strength of the effect of an independent variable on a dependent variable*." (Mackinnon, 2008, p. 8, emphasis in original). Following the example between interactivity and learning, a moderator could be individuals' computer literacy. So that the effect of interactivity on learning would occurred for individuals high on computer literacy.

A mediational analysis entailed three general steps: 1) the identification of possible mediators that may affect the outcome of interest; 2) to determine whether or not a particular mediator is causally related to the outcome of interest; and 3) to manipulate this causally related mediator so as to change the outcome. Points 1) and 2) provide a *conceptual theory* of learning from educational games, that is, how the mediator variable affects the outcome. Point 3) provides a *theory of action*, that is, how a particular educational game or intervention affects the mediators identified. For example, in the context of this dissertation, the conceptual theory suggests that learning from games is affected by cognitive engagement (i.e., the main mediator). As for the theory of action, it is assumed that cognitive engagement is affected by the educational game as designed (see Section 3.1.1.1) and by individuals' previous perception of the medium and the task demands.

Table 20: Comparison of Frameworks for the Study of Engagement in Educational Games

Input–Process–Outcome Game Model (Garris et al., 2002)	**Engagement (Fredericks et al., 2004; Corno & Mandinach, 1983)**	**Experiential model (Kolb, 1984)**	**Volitional processes (Corno, 1993; Salomon & Globerson, 1987)**
User judgment	**Cognitive engagement**		**Goal-related cognition**
	Acquisition		*Adaptive strategy use*
Task involvement	Alerting (receiving incoming stimuli, tracking/gathering information)	*Grasping the experience*	
(e.g., depth of involvement with learning)		Concrete experience (sensory cortex)	
	Monitoring (continuous tracking of information, rehearsing, self-checking)	Abstract conceptualization	
	Transformation		
	Selecting (discriminating/distinguishing relevant from irrelevant information)	*Transforming the experience*	
		Reflective Observation	
	Connecting (searching and linking familiar knowledge to incoming information)	Active Experimentation	
	Planning (Organizing a task approach sequence or performance routine).		
	Learning strategies (process interview)		

	Mindfulness/Mental effort (Self-report)		*Mindful persistence and effort*
User judgment Interest Enjoyment	**Emotional engagement** Interest, enjoyment and/or intrinsic motivation and/or *flow* (self-report)		**Outcomes** *Flow and other affects*
Task involvement (level of attention) **User behavior** Direction, intensity, quality "persistent re-engagement"	**Behavioral engagement** *Attention allocation* (e.g., Fixation Durations, Dwell Time, Reading depth) *Mental effort* (e.g., pupil diameter)	*Concrete experience* (Focused attention).	**Action control** *Resource allocation* *Protective actions* *Copying with distractors*
System feedback (Assessment of progress towards goal) *Confidence* (self-efficacy)			

What follows is a brief description of the conceptual framework depicted in Figure 15. The dashed lines mean that the list of variables is not exhaustive but they can be easily expandable to other relevant factors that might be suggested by other lines of research. First, for the framework, the variables on the left belong to the *input*. From this, two groups of variables can be distinguished: independent variables and moderator variables. Independent variables are the game itself in terms of its characteristics, the game surroundings, the instructional support and the content layout. The game characteristics refer to the highly cited features that make games motivational. However, the present framework subsumed them under the broader notion of game design patterns, suggesting that these patterns should be identified and studied, instead of discrete features (see Section 2.1.1 and 2.2.5). The technical game surrounding refers to both the type of platform and graphics (e.g., 2D or 3D). The social surrounding refers to whether or not the game is played as an individual or collective activity. The instructional support and content layout refer to all the features – internal or external to the game itself – that have been designed in order to support individuals' knowledge acquisition. On the other hand, the moderators are variables related to the individual or person characteristics. It can be distinguished between cognitive and conative variables, together with other organismic variables such as gender. The variables selected in Figure 15 correspond to the variables identified as relevant for learning purposes from the literature review conducted here (Section 2.2) and from the previous frameworks mentioned above (e.g., Wainess, 2010; Mattheiss et al., 2009) . For example, it is reasonable to expect that high and low prior knowledge or high and low working memory capacity could have an effect on individuals' process of learning. From the mediational perspective, this means that either prior knowledge or working memory capacity can change (i.e., moderate) the relationship between an independent variables (e.g., game played alone versus game played in pairs) on individuals engagement.

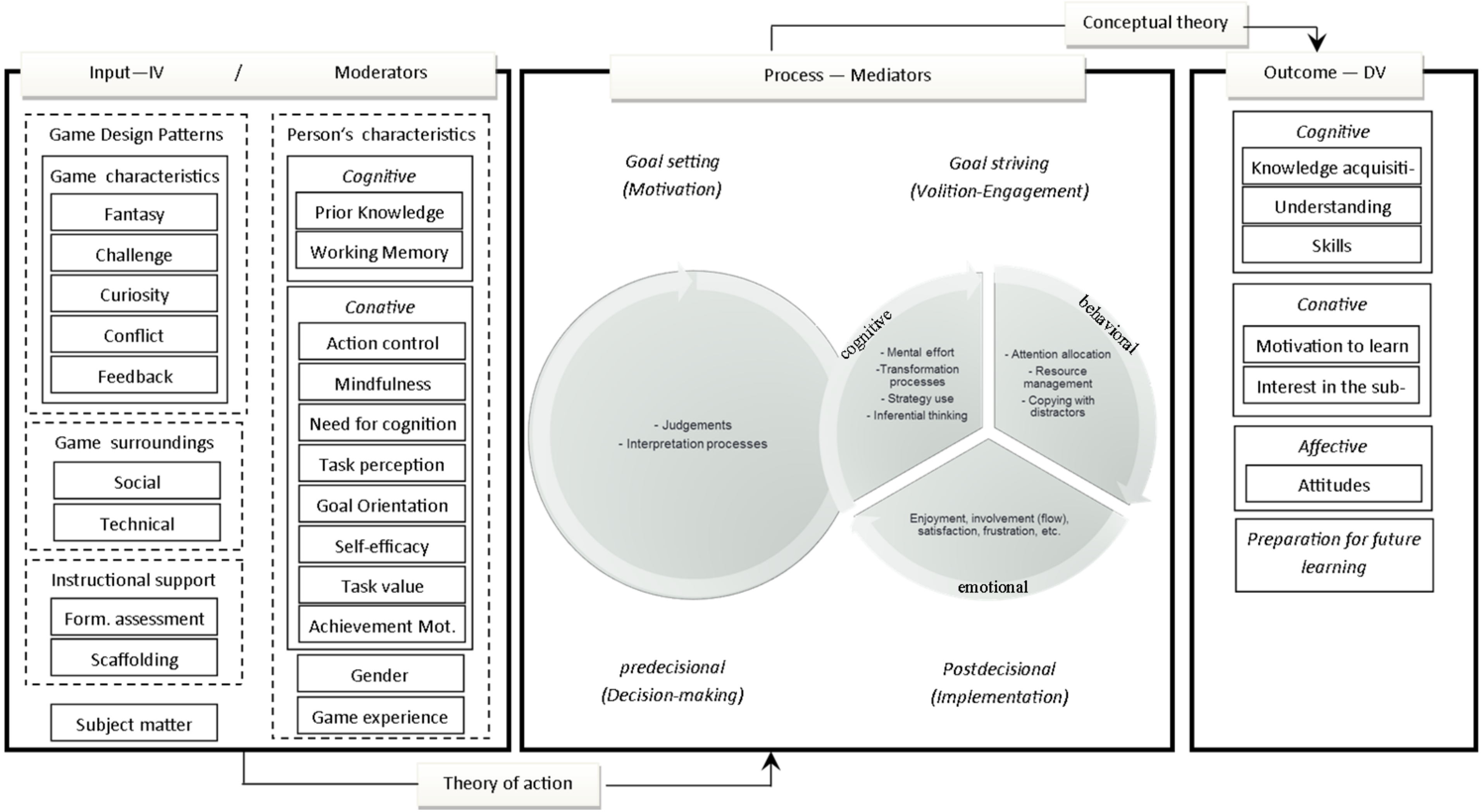
Second, for this framework, the *process* entails both the goal setting and goal striving processes. The former considered from the perspective of motivation and the latter considered from the perspective of volition and engagement. Both perspectives are treated as mediators of the process of learning from games. This framework does not adhere to any particular theory of motivation as, for example, Wainess (2010) did in his proposal. On the contrary, following Corno & Mandinach (1983), the framework assumes that any motivational theory or construct (e.g., expectancy value theory) ultimately generates a specific set of judgments and interpretations. What these interpretations are will defined the specific goal set by the individuals and, which she is supposed to strive for during the post-decisional or implementation phase of this motivational-volitional process. This idea is depicted through the two concentric circles in the middle of Figure 15. These processes are assumed to occur based on individuals' characteristics (i.e., moderators) and the specific gameplay that an educational game affords, and are assumed to occur in an "on-going" manner, and therefore highly situationally bounded. Following the example of prior knowledge, individuals with lower prior knowledge might be less likely to engage with the content knowledge embedded in the game. On the contrary, individuals with higher prior knowledge might well engage appropriately with such content knowledge.

Finally, the *outcome* in the framework brings together for sets of possible educational outcomes. It distinguished between cognitive, conative and affective outcomes, together with the notion of preparation for future learning (Bransford and Schwartz, 1999). Among the cognitive outcomes, it is possible to distinguish between content knowledge, understanding and skills. The conative outcomes include motivation to learn and interest in the subject matter. The affective dimension considers the positive attitudes toward technology and the subject matter embedded in the game. These conative and affective outcomes have been proposed as the central advantages of educational games.

The present dissertation has considered a few of the possible variables relevant to learning, mainly prior knowledge, task perception and self-efficacy. As detailed in section 4, prior knowledge took the form of a pre and posttest of content knowledge. The task perception included both the perception of the investment of effort that game warrants and individuals' perceived task demands (i.e., play to learn versus play for fun). How these variables affected individuals' engagement, in particular cognitive engagement, and therefore learning was the main purpose of this dissertation.

Previous sections have defined and operationalized the central construct of cognitive engagement. The next sections (2.5.1 and 2.5.2) describes how the other dimensions of engagement (i.e., emotional and behavioral) are considered and operationalized.

Figure 15: Conceptual Framework for the Study of Learning and Engagement in Educational Games from a Mediational Analysis Perspective



2.5.1. Emotional and Behavioral Engagement

Emotional engagement. Emotional engagement is related to individuals' affective reactions, which include interest, boredom, anxiety and happiness (Skinner & Belmont, 1993). In the context of virtual environments, some authors defined engagement entirely as an emotional state (Schuurink & Toet, 2010). In general, the qualitative distinction between positive emotions and deeper involvement or investment has not been made, with exception of the concept of flow (Fredricks et al., 2004). In their conceptual framework, Fredricks et al. positioned flow as the key concept for this dimension of engagement. Following the suggestions of Bohnert et al. (2010), the ideas of liking, fun, enjoyment and feelings of involvement are regarded as belonging to this category of emotional engagement.

Behavioral engagement. This engagement can be defined in three ways (Fredricks et al., 2004). Firstly, as a set of positive conducts such as following the rules and classroom norms. Secondly, as participation in extra-curricular activities. Thirdly, as involvement in learning and academic tasks, including behaviors such as concentration, effort and attention. For this dissertation, the last definition of behavioral engagement is used. Normally, behavioral engagement has been measured either through direct observation in order to determine the amount of on-task versus off-task behavior or through self-reports. Studies in the context of educational games have used log-files analysis to trace individuals' explorative behavior and how they interact with the different element of the game. For instance, Quest Atlantis (Barab, Thomas, Dodge, Carteaux, & Tuzun, 2005) and River City (Nelson, 2007), both Multiuser Learning Environments (MUVE), have used log files to study how individuals use the learning resources, the hints and the academic feedback embedded in the games. These analyses have discovered that few students actually use or access (that means clicking on a particular feature of the game) the resources and hints designed (e.g., Hickey et al., 2009). Cordova and Lepper (1996), operationalized the behavior of participants during gameplay by collecting a series of indicators such as the use of hints, number of problems solved correctly, strategic play, use of complex operations, spaces per turn and number of times a more challenging version of the program was selected. The authors only found differences in strategic play, use of complex operations and the use of more challenging versions of the game, but no different were found in terms of hint use and problems solved correctly, which are more learning related behaviors. This pattern of results illustrates how individuals engaged qualitatively different in the game in terms of gameplay, and apparently qualitatively similar in terms of processing the information related to the learning tasks. The question is what they actually did while solving the problems and using the hints. In the case of the MUVE, the question is what individuals actually do with the resources embedded, if accessed at all? Do they spend enough time processing the information existent in these resources? Do they read or just skim the information available? In order to respond to these questions regarding how individuals actually process the information in educational games, this dissertation takes a step forward and attempts to measure directly a central aspect of individuals' "on-task" behavior, that is, their patterns of attentional involvement and resource allocation. For that, this study attempts to capitalize in the advantages of current non-immersive eye tracking technologies for recording individuals' eye movements (Duchowski, 2007). What follows is a discussion of theoretical links between attention, processing and eye movements.

2.5.2. Attention and Eye Movements Research

Before getting into the technicalities and assumptions concerning eye tracking research, a brief description of the effortful model of attention developed by Kahneman (1975) is presented. The goal is to use this model as a theoretical background to justify the focus on attention and resource allocation as operationalized by the eye tracking measures used in this dissertation.

An effortful model of attention. Kahneman's (1973) considered two central aspects of attention: *selectivity* and *intensity*. From everyday experience it is clear that humans can attend to some stimuli while ignoring others, which implies the existence of mechanisms that govern the significance of the stimuli in the environment. After having selected the stimuli, individuals apply or invest certain amount or intensity in their attention process. In this sense, to attend is to "apply oneself". Kahneman mentioned the work of Berlyne (1960) as one of the first researchers that studied this intensive aspect of attention and listed three "collative properties" of stimuli that control individuals' attention: novelty, complexity and incongruity. These properties, under the category of Curiosity, were proposed by Malone (1981) as partly explaining the "fun" in games (see Section 2.2.3) and have been since then highly mentioned as a central fundament for the engagement power of games. The study of the stimuli properties implies a certain degree of passivity or involuntary involvement of individuals. However, as Kahneman suggested, it seems also interesting to study the phenomenon of *voluntary selective attention* to stimuli, because these stimuli are relevant to the task an individual has *chosen* to pursue. This voluntary attention is “an exertion of effort in activities which are selected by current plans and intentions” (Kahneman, 1973, p. 4). When the current plans and intentions are related to learning a subject or processing information related to that subject, it follows that these learning activities and/or information processing should consume most of an individuals’ effort.

The main assumption of the model is that the level of effort that an individual exerts while performing a task is not a function of the individual’s intentions, but a function of the task demands. Therefore, the individual’s voluntary control of the effort exerted is very limited. However, as Kahneman (1975) acknowledged, there are some evidence that motivation may well play a role. For example, he mentioned how sleep deprived individuals performed at a normal level in a task when highly motivated. The fatigue that came with the sleep deprivation increased the difficulty of continuing performing the task, and the motivated individual *compensated* this difficulty by increasing the level of effort exerted. Although motivational factors are mentioned in the model, they are slightly integrated as influencing the momentary intentions of the individuals. Furthermore, the general idea that the “mobilization of effort in a task is controlled by the demands of the task, rather than by the performer’s intention” (Kahneman, 1973) seems insufficient to explain individuals’ mental effort (cf. Sarter, Gehring, & Kozak, 2006). This focus on the task demands is only one part of the complex phenomenon of mental effort (see Section 2.4.1). As Salomon (1984) suggested, the degree of mental effort investment depends also on factors such as individuals self-efficacy and individuals perception of the task – not only how demanding it is, but also includes other aspects such as the medium involved (TV in Salomon’s studies). Finally, Kahneman’s model does not explicitly address what factors might influence the components of the model.

Eye movements. After this brief review of Kahneman’s (1973) model of attention, the discussion centers on the eye movements’ research. The basic assumption in this area is that there is a connection between the eye and the mind. The eye-mind hypothesis states that where a person is looking at and what is she thinking on or interested in are highly correlated events (Just & Carpenter, 1980). The eye tracking methodology allows for the recording of moment-to-moment processing behavior and momentary changes of effort on individuals in the context of simulation and games (De Rivecourt, Kuperus, Post, & Mulder, 2008; Marshall, 2005), but it needs auxiliary data in order to interpret the eye movements measures (Holmqvist, Nyström, Andersson, Dewhurst, Jarodzka, & Van de Weijer, 2011). Holmqvist et al. divided eye tracking data into *events* and *representations*. The former are any countable entity, and can be simple such as fixations, saccades, and blinks or complex such as regression scanpaths. The latter are recalculation of data and are not countable (e.g., Transitions).

Saccades and fixations represent the two fundamental eye movements. Saccades are the eye movement themselves and fixations represent the period of time that the eyes remain fairly still. Saccades are quick, ballistic movements and last between 30 and 120 milliseconds. Fixations are relatively stable periods that typically range between 150 and 600 milliseconds (Olsson, 2007). According to Rayner (2009), three key points seem to be accepted by researchers concerning eye movements. First, the amount of information processed on any fixation depends on the task (e.g., reading or scene perception). For instance, fixations and saccades tend to be longer on scene perception because more information is being processed (mean of 300msc. and 4,5 degrees, respectively). Second, the difficulty of the stimulus makes fixations longer and saccades shorter. Third, the difficulty of the task itself (e.g., searching for an object in a scene versus looking at the scene for a memory test) influence eye movements. Finally, viewers integrate and process visual information and the efficient processing of information occurs on each fixation. Normally, studies use the number of fixations, mean fixation duration and total inspection time as relevant indicators of information processing and learning (Rayner, 1998). However, another important indicator is pupil diameter given its relationship with mental effort and workload (e.g., Marshall, 2005).

Another central concept in eye tracking research is the *Areas of Interest (AOI).* According to Holmqvist et al. (2011) an AOI defines a region in the stimulus in which the researcher has a particular interest. Thus, AOI, as opposed to fixation and saccades, are defined in relation to the content of the stimulus. Holmqvist et al., suggested that when using AOIs several points should be considered:

1. The research hypothesis should decide what AOI will be defined
2. Each AOI should cover an area with an homogeneous semantic
3. Do not put objects too close together so that there is no margin between AOIs

4. Avoid overlapping AOIs
5. The minimal AOI size is limited by the accuracy of the recorded data

Once an AOI has been defined, it allows the definition of three basic events: *dwells*, *transitions* and *AOI hits*. The dwell (i.e., "gaze" or "glance") is defined as one visit in an AOI from entry to exit. A transition (i.e., "gaze shift") corresponds to the movement from one AOI to another. Finally, a hit corresponds to a fixation whose coordinates fall inside the AOI. The fixation-based hit is very important and used in many measures such as Dwell time and Reading depth, measures described later below and used in this dissertation.

Holmqvist et al. (2011) distinguished also four research traditions in eye tracking: visual search, reading research, scene perception, and usability. All these traditions, except maybe for the case of usability, offer agreed upon techniques and experimental procedure that can be adapted to tackle a specific research question that may well be outside any specific tradition. According to Holmqvist et al. two things differentiate usability from the other three research traditions. First, usability does not yet count with experimental conventions, measures and procedures as the other traditions do. Second, usability focuses on higher-level cognitive processing than the other traditions. Below a brief description of some applied fields that focus on this "higher-level" cognitive processing is presented as they relate closer to the purposes of this dissertation.

From a cognitive load perspective, eye tracking indicates individuals' shifts of attention (van Gog & Scheiter, 2010) and therefore is suitable to study how presentation formats affects individuals' eye movements behavior and, hence learning. Similarly, multimedia research on learning has shown that the more time individuals spend looking at "relevant areas" or Areas of Interest (AOI) the higher the comprehension scores and transfer scores (e.g., Bouchex & Lowe, 2010). As summarized by Meyer (2010), eye tracking can shed light onto questions related to "how" an instructional method works by providing measures of "perceptual processing" during learning. The link between this perceptual processing and learning outcome was found in four of the six studies conducted in the special issue of eye tracking and multimedia learning (Meyer, 2010). For example, Boucheix and Lowe (2010) used eye tracking to examine the impact of different cueing strategies for directing individuals' attention to low salience but relevant features of a complex animation. The animation was divided into several AOIs representing functional parts of a piano system. The authors collected the number of fixations and the total duration of fixations in the different AOIs. Results showed that individuals spent more time looking at the relevant areas when relevant features were highlighted (cued). This pattern of eye movement had a positive impact on learning.

From a game studies perspective, research using eye tracking measures have studied the learning experience of players, the sense of presence and immersion of players (Alkan & Cagiltay, 2007; Jennett et al., 2008; Kallinen, Salminen, Ravaja, Kedzior, & Sääksjärvi, 2007), although the main current interest in this field is the development of eye tracker devices for human-computer interaction (i.e., gaze interaction) for persons with disabilities (Nacke et al., 2011). For instance, Alkan & Cagiltay (2007) used eye tracking to study the learning experience of players interacting with the game *Return of the Incredible Machine Contraptions*. Combined with interview, eye tracking data allowed the authors to identify where players shifted their attention during the game, as defined by specific areas of interest. Results showed that individuals mean fixation duration occurred in the contraptions area where participants are supposed to think about the possible solutions to the puzzles presented. Kallinen et al. (2007) investigated the sense of presence defined as involvement and immersion into a stimulus. They hypothesized that 1st and 3rd person shooter game, representing high and low presence respectively, would lead to differential involvement, reflected on the number of fixations and gaze time outside the screen of the game, self-reports and physiological measures. However, they found significant differences only on self-reports and not on the eye movement's data. They assert that many different moderating factors could have affected the relationship between involvement and eye movement. In a similarly effort, Jennett et al. (2008), investigated whether individuals' eye movements changed during an immersive task (i.e., playing *Half-life*) compared to a non-immersive task (i.e., clicking at squares). Results showed that individuals in the non-immersive task showed an increased on the number of fixations, while individuals on the immersive condition showed a decreased on the number of fixations. The immersiveness of the two conditions was confirmed by individuals self-report of immersiveness. Significantly higher level of immersion was reported by individuals playing *Half-life*. Authors concluded that in the immersive condition, individuals' decreased fixations reflected a more focused attention on the visual components relevant to the game. Taken together, these studies represent an attempt to examine the experiences of individuals while playing games. Even though they do not focus on the learning experience during the interaction with an educational game, they provide some insights about the methods already used in similar fields of research.

Finally, from an educational game perspective, the use of eye tracking technology is in its infancy, although promising research has already been conducted (Kickmeier-Rust et al., 2011; Law, Mattheiss, Kickmeier-Rust, & Albert, 2010). Kickmeier-Rust et al. (2011) compared the visual patterns of low and high performers in the educational game prototype *80Days*. Results showed that good learners scanned visual field evenly with longer saccades, attending to relevant areas of the screen more frequently. There were also different visual patterns when navigating in 3D or 2D spaces. Muir and Conati (2012) analyzed hint usage in *Prime Climb*, an educational game for learning factorization skills. The authors used total fixation time spent on previously defined AOI on the hints displayed. It was also calculated a ratio of fixations per word in order to get a closer examination of "how carefully a student scans a hint's text" (Muir & Conati, 2012, p. 115). Results showed no learning gains and a very modest use of the system generated hints.

Law et al. (2010) conducted a particularly interesting study in the context of the already mentioned educational game *80Days*. The interesting feature is the use of eye tracking with auxiliary data such as interviews, mental effort and game quality questionnaires. The purpose of the study was to examine the potential of eye tracking for evaluating educational games in general and to study vicarious learning from an educational game. The authors used the number of fixation per second as the eye tracking measure of interest and interpreted it as an indicator of information processing. This measure was computed by "averaging the total number of fixations over the total viewing time for both micro-missions…" (Law et al., 2010, p. 479). Participants (n=24) viewed a recording of the game *80Days*. The study examined the impact of the position of an NPC (non player character) and the presence/absence of an adaptive hint on individuals visual attention, learning, perceived quality of the game and cognitive load. Results showed no effect of the manipulation on visual attention and learning. However, pre-post learning gains for the whole sample did yield statistical differences. The study did not find an effect of the manipulation on the perceived qualities of the game nor on the perceived cognitive load. Finally, a set of correlations among the measures show that the eye tracking measure of number of fixations per second had a negative correlation coefficient with the perceived effort invested. The authors, based on Jennet et al. (2008), explained these findings in terms of higher immersion: "engaged learners tended to show a lower number of fixations per second, implying a higher degree of immersion" (Law et al., 2010, p. 483). Likewise, the lack of correlation between the eye tracking measure and the learning gains was suggested to be related to either a limitation of the number of fixation to capture deeper levels of information processing or to ceiling effects on the learning test.

Although the eye tracking measures used in this dissertation are based on the previous studies discussed above, in the present study a measure of reading depth was used to capture "how carefully" individuals read/scan text within the educational game used here (cf. Muir & Conati, 2012). The next paragraphs provide a description of the measures used in the current study: fixation durations, dwell time, total dwell time and reading depth.

Fixation durations provide information concerning how long the eyes have been still in a particular position. It is the most used measure in eye tracking research and has received several names such as "fixation time" and also has been confused with "dwell time" (Holmqvist, et al., 2011). The interest of this measure lies, as already mentioned, in the assumption that fixations and attention are two synchronous events and what is fixated by the individuals is what it is processed (Just & Carpenter, 1980). In other words, fixation durations reflect perceptual intake and processing. However, two points are important to keep in mind. One is that fixating something does not necessarily entail attentive processing. Second, the intake period should be equal to the stillness as defined by the algorithm used, and it is known that different algorithms produce different fixation durations.

With these considerations in mind, there is still sufficient support to interpret longer fixation duration with deeper and more effortful cognitive processing (Holmqvist et al., 2011). Among the exceptions to this general "rule of thumb", Holmqvist et al. mentioned some situations in which the length of the fixation duration may involve other factors, which for the purpose of this dissertation are considered to be of minimal influence (see Table 21).

Table 21: Exceptional Situations were Longer Fixation Duration = Not Deeper Processing

Situations	Fixation Duration
“Daydreaming”	Longer
Expertise	Longer
Neurological impairments	Longer
Quick moving of an stimulus	Longer
High stress	Shorter

Dwell time is an event related to the Area of Interest (AOI) and is concerned with how long, measured from entry to exit, the eye/gaze remained inside an AOI. It is defined as “one visit in an AOI, from entry to exit” (Holmqvist et al., 2011, p. 386). In comparison with fixation durations, dwell time are longer, more dispersed and dependent on an AOI. It has also shown a pronounce skewness. This measure is affected mainly by the semantics of the object and by the task that is being accomplished. High values of dwell time can have diverse meanings such as interest, informativeness, and difficulty for extracting information. A related measure of interest for this study is the *Total Dwell Time*, which is supposed to be more sensitive to slow and long-term cognitive processing. It represents the sum of all the dwell time in the same AOI. Finally, another measure related to dwell time is *Reading Depth*. This measure captures how deeply a piece of text is read and has been used (Holsanova, Rahm, & Holmqvist, 2006) in newspaper reading, showing that the reading depth is significantly lower for longer newspaper articles. The operationalization requires a defined AOI with text in it as input and the size of the AOI (e.g., pixels). The output is a measure of “depth” in terms of time by squared pixels or squared centimeters. A practical advantage of this measure is that it works for any combination of stimuli, such as text and pictures.

2.6. Overview of Research Questions and Hypotheses linking Cognitive Engagement and Learning

From the discussion conducted in the literature review, several key points relevant for learning from educational games emerged. First, what makes for an effective educational game from a design perspective. Here the idea of integrating the content and the fantasy of the games is a central design goal for most educational game designers (i.e., “intrinsic” design goal). This type of design will protect the quality of individuals’ game experience usually understood under the concept of deep engagement in terms of flow state.

However, the concept of engagement has had different meanings and, more importantly, different dimensions that are important to distinguish while examining the process of learning from educational games and the role these dimensions might have on individuals' learning. The distinction and examination of these dimensions can provide with a richer characterization of individuals' experience while playing an educational game. Second, the role of mental effort and its connection to deep processing of information and to reflective thinking are central for learning in general and for learning from media such as TV, interactive video, internet, film, in particular. Third, as mental effort constructs deal mainly with the amount or quantity, it is needed to understand how this intensity translate into appropriate information processing, the qualitative dimension of mental effort. These two aspects are the key components of the concept of *cognitive engagement*, which the literature has shown to have a consistent relation to learning outcomes and motivational variables central to learning in different contexts. Finally, these ideas have been formalized in Salomon's (1984) and Corno and Mandinach's (1983) models of mental effort and cognitive engagement, respectively, and integrated into a broader framework that attempts to operationalize models of learning that have remained theoretical (e.g., Garris et al., 2002). Therefore, how might individuals' cognitive engagement be stimulated while playing an educational game and how this engagement relates to behavioral and emotional engagement are important issues to address in order to better understand learning from educational games and beyond. Similarly, understanding which factors affect individuals' engagement might also contribute to the general understanding of learning from this type of technology. Hopefully, in time research such this one can contribute with useful knowledge to improve the design of games and overcoming the the perception of educational games as "bad" game design. These core ideas are reflected in the following research questions and hypotheses.

2.6.1. Expected Effects of *Genius Unternehmen Physik* and the Perceived Demand Characteristic (PDC) of the Task on Learning

The research question concerning this educational game is:

Research Question 1: What effects does the educational computer game *Genius Unternehmen Physik* have on individuals' recall of content knowledge?

Games in general are considered to offer higher levels of “interactivity” than previous media such as TV and interactive video. The more interactive the medium, the more likely is to be perceived as more demanding, overcoming individuals preconception concerning how much effort is warranted to learn from a particular medium. For both Kahneman (1973) and Salomon (1984), the perceived demands of the task represented the core factor influencing the amount of effort required to actively process information from a particular source. Likewise, Tobias and Fletcher (2011) pointed out that in order to be able to have evidence of learning and transfer, the game and the transfer task should overlap the cognitive processes required. In the case of the present game, there is plenty of information that needs to be attended, selected and elaborated. This exact information is asked or explicitly presented in the recall test. On the other hand, educational games, such as *DimensionM*, have shown learning effects though it has been categorized as an “extrinsic” game. As already described, intrinsic games are supposed to protect flow (Habgood & Ainsworth, 2011; Kiili, 2005) and to make more likely the connections concerning the content knowledge when the content and the fantasy are coupled together (Malone, 1981). However, the scant empirical evidence available does not seem to support completely these claims (cf. Habgood, 2007; Habgood & Ainsworth, 2011; Kiili, 2005; Pavlas, 2010). Therefore, it is reasonable to expect learning gains for each group, as compared between pre and post recall test:

Hypothesis 1: Individuals will exhibit significantly greater recall of content knowledge on the posttest than in the pretest.

Concerning the effect of manipulating the task demands on participants’ learning, the second research question is:
Research Question 2: Is there a mean difference in recall of content knowledge between individuals instructed to play to learn physics and individuals instructed to play for fun?

In the context of Salomon’s (1984) model of mental effort, several studies have instructed participant to engage in an activity either for “fun” or “to learn”, with some variances in the message (e.g., Heers, 2005; Rieh et al., 2012; Salomon & Leigh, 1984). This message should evoke differential patterns of cognitive engagement (i.e., mental effort invested and cognitive processing). Similarly, following Kahneman’s (1973) effortful model of attention and the eye-mind hypothesis (Just & Carpenter, 1980) behavioral engagement (i.e., attention allocation) with the material embedded in the educational game should also be affected. Therefore, an impact on the recall of content knowledge is to be expected. Although Salomon and Leigh (1984) could not find differences concerning recognition, they did find differences on individuals' inferences. However, the measure of recall used here are in some way in-between recognition and inferences. That is, they are assumed to require more mental effort than simply recognition of items. Secondly, the nature of the media used here, an educational game, as compared with the original studies in TV is by definition a richer one that may foster more readily individuals' "mindfulness" through increasing their perceived demand characteristics (e.g., Hannafin, 1989; Salomon & Globerson, 1987):

Hypothesis 2: Individuals instructed to play to learn physics will exhibit greater recall of content knowledge on the posttest than individuals instructed to play for fun.

2.6.2. Expected Effects of the Perceived Demand Characteristic (PDC) of the Task on Cognitive Engagement

As for the effect of manipulating the task demands on participants' cognitive engagement, the third research question is:

Research Question 3: Are there mean differences in cognitive engagement – as measured by AIME Task, SCENG and Transformation Processes – between individuals instructed to play to learn physics and individuals instructed to play for fun?

Salomon's (1984) model predicted that, in the absence of a clear instruction, the amount of invested mental effort or AIME should be influenced by individuals' self-efficacy and perceived demand characteristics of the tasks (PDC). As individuals differences on self-efficacy and general AIME were controlled between the two conditions, it is reasonable to expect that changing individuals PDC towards a more demanding task (i.e., playing to learn Physics instead of playing for fun), should increase their AIME. Therefore, participants reported AIME with the Tasks (AIME Task) and situational cognitive engagement with the tasks (SCENG) should be higher for those instructed to play to learn.

As AIME represents only the quantitative aspect of cognitive engagement, the qualitative aspect should also be influenced by the experimental manipulation. According to Corno and Mandinach's (1983) model, different forms of cognitive engagement imply different amounts of mental effort (Howard, 1989). As discussed in 2.4.2.1, the degree of use of the acquisition and transformation processes define four forms of cognitive engagement, being self-regulated Learning (Corno & Mandinach, 1983) or Comprehensive Engagement (Howard, 1989) together with Task Focus the forms that more intensively demand the effortful use of transformation processes. On the other hand, the form of engagement Recipience, is the most passive and effortless one of the four forms. Therefore, it is plausible to expect that individuals instructed to learn should show to a greater extent forms of cognitive processing that are more effortful, that is, more task focused, while individuals instructed to play for fun should show a more Recipience approach to the tasks in the game, that is, a minimum use of the effortful transformation processes. Therefore, altogether, individuals' higher PDC (i.e., the ones instructed to learn) should show greater cognitive engagement:

Hypothesis 3a: Individuals instructed to play to learn physics will exhibit a greater Amount of Invested Mental Effort (AIME Task) than individuals instructed to play for fun.
Hypothesis 3b: Individuals instructed to play to learn physics will exhibit greater Situational Cognitive Engagement (SCENG) across tasks than individuals instructed to play for fun.
Hypothesis 3c: Individuals instructed to play to learn physics will exhibit a higher level of transformation processes than individuals instructed to play for fun.

2.6.3. Expected Effects of the Perceived Demand Characteristic (PDC) of the Task on Behavioral Engagement

Concerning the effect of manipulating the task demands on participants' behavioral engagement, the fourth research question is:

Research Question 4: Are there mean differences in behavioral engagement – as measured by Fixation Duration, Total Dwell Time, and Reading Depth – between individuals instructed to play to learn physics and individuals instructed to play for fun?

As already suggested it is expected that the perception of the task demands have an influence on individuals' cognitive engagement in terms of mental effort invested and degree of use of transformation processes while engaged with the learning material of the game. As discussed in Section 2.5.2 attention refers to the effortful application of oneself in activities selected according to one's specific goals and intentions. As Kahneman (1973) and Salomon (1984) suggested, individuals choose an activity to pursue and then allocate the effort necessary to perform such activity. Therefore, if individuals attempt to learn physics during the play of the educational game used in this study, their intentional selective attention should be focused mainly on the pages related to the content to be learned. On the other hand, eye tracking research has provided some indicators of focused attention, visual attention or attention allocation that are appropriate for the purpose of the study to use them as indicators of individuals' behavioral engagement. Therefore the following effects are expected:

Hypothesis 4a: Individuals instructed to play to learn physics will exhibit longer Fixation Duration than individuals instructed to play for fun.
Hypothesis 4b: Individuals instructed to play to learn physics will exhibit longer Total Dwell Time than individuals instructed to play for fun.
Hypothesis 4c: Individuals instructed to play to learn physics will exhibit higher Reading Depth than individuals instructed to play for fun.

2.6.4. Expected Relationships among Cognitive Engagement, Behavioral Engagement and Recall of Content Knowledge

As for the relationships among the dimensions of engagement, the fifth research question is:
Research question 5: How do cognitive and behavioral engagement relate to each other and to learning?

The multidimensional concept of engagement (i.e., cognitive, behavioral and emotional) reflects the general multivariate nature of human functioning. In particular, models of mental effort and focused attention connect these two construct closely together (Kahneman, 1973; Salomon, 1984). Attention is considered to be the starting point in the chain of information processing and is under the control of an executive component in memory that is effort intensive. Likewise, it is expected that higher cognitive and behavioral engagement should correlate positively with the learning outcomes as has been suggested by previous research on AIME (e.g., Salomon, 1984; Cennamo, 1993) and by the already discussed eye-mind hypothesis (Just & Carpenter, 1980). Therefore, the following relationships are expected:

Hypothesis 5a: There will be a statistical positive correlation among AIME Tasks, SCENG and Transformation Processes.
Hypothesis 5b: There will be a statistical positive correlation between AIME Task and Learning.
Hypothesis 5c: There will be a statistical positive correlation between SCENG and Learning.
Hypothesis 5d: There will be a statistical positive correlation between Reading Depth and Learning.

2.6.5. Expected Relationships among Cognitive Engagement and Emotional Engagement

Concerning the relationships between AIME measures and emotional engagement, the sixth research question is:
Research question 6: How do individuals' AIME measures relate to each other and to individuals' Emotional Engagement?

The review on educational game design (see sections 2.2.2 through 2.2.4) suggested that there is modest evidence on the effectiveness of intrinsic fantasy/integration designs over more extrinsic ones (i.e., edutainment-like, Egenfeldt-Nielsen, 2005). Furthermore, Malone's (1981) evidence seemed to point to the opposite direction. Likewise, the assumption that more extrinsic-oriented games would more probably break the flow of the game has not been directly examined (Habgood & Ainsworth, 2011). Kerres et al. (2009) also suggested that the learning experience in educational games is by design one of disruption and tension between the game and the learning content. Apparently, no empirical evidence of these claims has been provided yet. In consequence, it is hypothesized that *Genius Unternehmen Physik*, when including the learning tasks, should not necessarily hinder the overall playing experience of the users and users not necessarily perceive this educational game as two “separated” or “disruptive” experiences. Therefore, it is plausible to expect a positive correlation between individuals’ emotional (i.e., “having fun” and “feeling involved” in the game) and cognitive engagement. Likewise, a positive correlation between cognitive engagement measures (i.e., AIME Task, AIME Simulation, SCENG) with the tasks and with the simulation/game is expected:

Hypothesis 6a: There will be a statistical positive correlation between AIME Tasks and AIME Simulation.
Hypothesis 6b: There will be a statistical positive correlation between AIME Tasks and Emotional Engagement.
Hypothesis 6c: There will be a statistical positive correlation between AIME Simulation and Emotional Engagement.

2.6.6. Expected Relationships among Self-efficacy, General AIME and Cognitive Engagement

Concerning the role of participants’ self-efficacy and general AIME on their cognitive engagement, the seventh research question is:

Research question 7: Do individuals’ initial Self-efficacy and general AIME relate to their actual cognitive engagement under the unspecific instruction to play for fun as compared with the instruction to play to learn physics?

Hypothesis 7a: Individuals instructed to play for fun will exhibit a statistically greater correlation coefficient between Self-efficacy and AIME Task than individuals instructed to play to learn physics.
Hypothesis 7b: Individuals instructed to play for fun will exhibit a statistically greater correlation coefficient between Self-efficacy and SCENG than individuals instructed to play to learn physics.
Hypothesis 7c: Individuals instructed to play for fun will exhibit a statistically greater correlation coefficient between Self-efficacy and Transformation processes than individuals instructed to play to learn physics.

Hypothesis 7d: Individuals instructed to play for fun will exhibit a statistically greater correlation coefficient between General AIME and AIME Task than individuals instructed to play to learn physics.

Hypothesis 7e: Individuals instructed to play for fun will exhibit a statistically greater correlation coefficient between General AIME and SCENG than individuals instructed to play to learn physics.

Hypothesis 7f: Individuals instructed to play for fun will exhibit a statistically greater correlation coefficient between General AIME and Transformation processes than individuals instructed to play to learn physics.

3. Plan and Implementation of the Study

3.1.1. Selecting the Educational Game: *Genius Unternehmen Physik*

Ritterfeld et al. (2009) built a database with 612 serious games. From this 48% corresponded to games emphasizing some type of practicing skill and only 16% were targeted to the adult population. This empirically suggests the difficulties of finding games specific for population older than 14 years old and in particular in German language (Petko, 2009). In appendix D there is a list of the games consulted, reviewed, played by the researcher or tested by a participant. There was no previous preference for a particular genre (e.g., action genre, strategy games, simulation games, etc.). Table 22 shows the strategies used for the searching of educational games as candidates for the present study.

Table 22: Type of Sources Used to Locate Educational Games

Type of Source
Journals
British Journal of Educational Technology
Simulation & Gaming
Computers & Education
Computers and Human Behavior
Journal of Educational Computing Research
Game Studio, International journal of game research
Instructional Science
Journal of the Learning Sciences
Academic Databases
ERIC, Educational Resources Information Center (ERIC)
EBSCO HOST, Research Database (http://search.epnet.com/)
ProQuest Dissertation and Thesis Database
Social Networks
Academia.edu
Xing.com

Researchgate.net
LinkedIn
Technology and game organizations
Futurelab (http://www.futurelab.org.uk/)
Gamusutra (http://www.gamasutra.com/)
Serious Game Initiative
Games for Change
DiGRA, Digital Games Research Association
Federation of American Scientists
Academic Websites
The Education Arcade, an MIT-University of Wisconsin Partnership
Listserver
Gamesnetwork

The decision criteria for choosing an educational game were:

- The game should fulfill at least two of the following characteristics: present a conflict, an explicit goal structure, and feedback, plus the integration of specific learning material.
- The game allow a single player mode
- The game's learning goals are clear and not too simplistic and trivial (e.g., 2+2=4)
- The game's design allow for active gameplay in a relatively short periods of time (30–45 min.).
- The game is in German
- The game includes the information or content knowledge in some text or graphic manner, i.e., a particular type of representation.
- The game's subject matter is science

Screenshots of a game reviewed before the main study was conducted is provided in Table 23. Even though the games has some interesting features, the fact that they were in English and the time needed to play them represented serious limitations that did not fit the time and population constraints of the study.

Table 23: An Example of Educational Games Reviewed

Source: http://www.spielbar.de/neu/wp-content/uploads/2009/11/remission_scr3.jpg

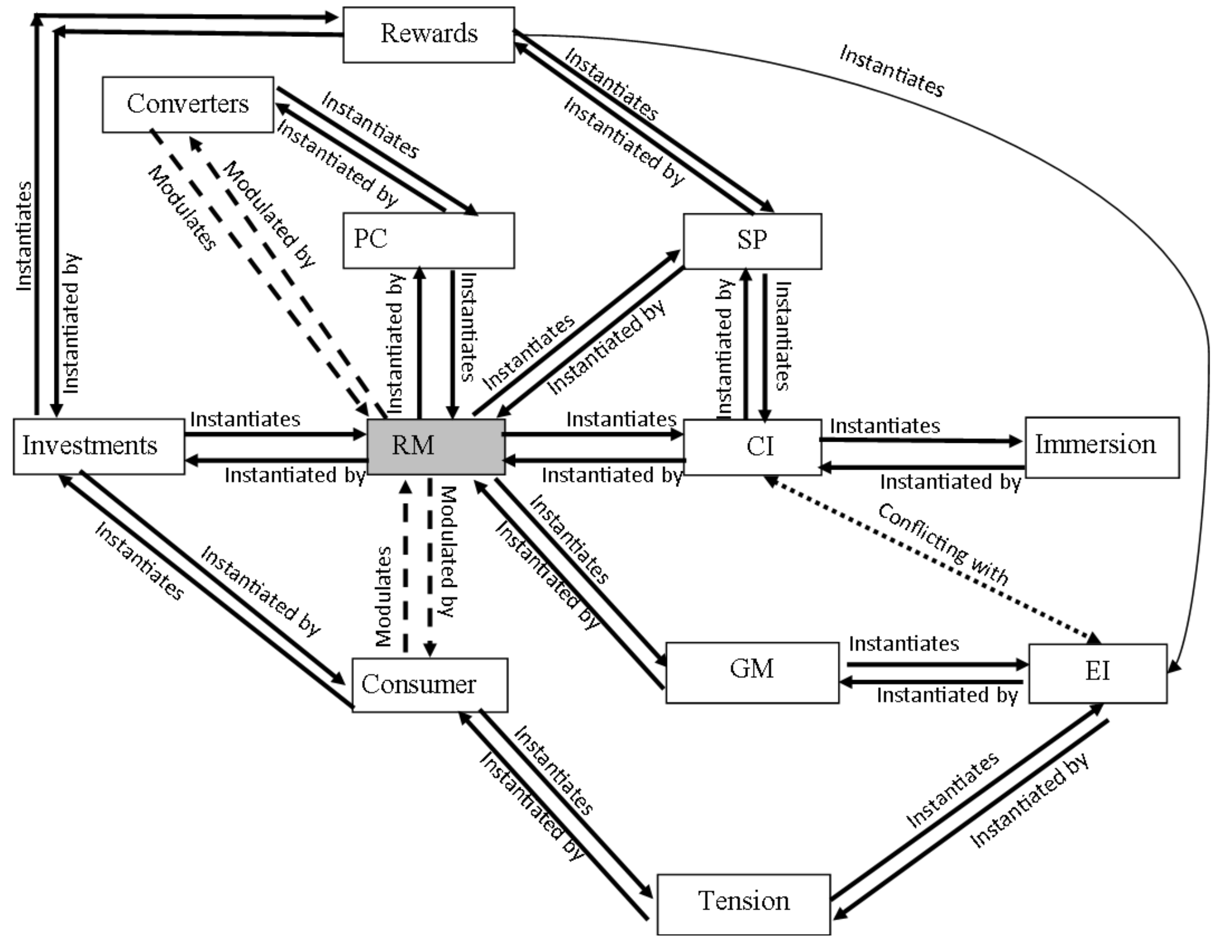
After reviewing several educational games, *Genius Unternehmen Physik* (Figure 17) fulfilled the basic requirements in the context of the experimental constraints. *Genius Unternehmen Physik* was designed to teach physics and belongs to the category of edutainment. It is a game intended for pupils in high school and contains advanced learning materials conforming to German school curricula. The player is positioned as a “entrepreneur” who needs to deal with an environment that simulates an economic world where the player must produce bicycles, hire workers, build fabric and houses, pay taxes, make investment and profit from the selling of the bicycles produced. Coupled with these activities are a set of tasks and exercises about physics, for example, the laws of pulleys levers. In summary, the game offered an explicit goal structure, feedback and integrates learning content in texts and graphics.

3.1.1.1. Game Description

The structural analysis of *Genius Unternehmen Physik* is based on Björk and Holopainen (2005) model described in section 2.1.1.1. The aim of this analysis is to understand what patterns exist in a particular game design (cf. Filsecker & Kerres, 2013). Central to *Genius Unternehmen Physik* is the resource and resource management pattern. Figure 16 summarizes the main patterns in *Genius Unternehmen Physik* and their theoretical interaction. The solid arrows mean that the pattern instantiates or cause the other pattern in the direction of the arrow. *Investment*, for example, instantiates *Resource Management*. That is, if *Genius Unternehmen Physik* presents the pattern Investment (Committing Resources to something to get the rewards later) this provokes a Resource Management situation (Players plan, manage, and control resource flows in order to reach the game's goals). In *Genius Unternehmen Physik* players have an initial amount of money (i.e., resources) to build different units. Players need to plan how many of each type of unit to build in order to reach the goal of the game. These goals can be defined by the players (e.g., get more money, not to lose money or expand the business, among others). The precondition of the Resource Management patterns is the existence of limited resources influencing players' abilities to reach a specific goal within the game through the use of *Producers* and *Consumers*. In strategy games, for example, resources are used to determine the actions a player is able to execute and are usually implemented in the form of Producer-Consumers patterns (the pattern determines the lifetime of game elements, governing the flow of gameplay.). This pattern is usually chained with *Converters* (they produce different game elements from other game elements). The main converters in this game are the fabrics which produce bicycles. Resource Management is modulated by *Consumers* (a sort of resource is consumed as a consequence of a player action). *Genius Unternehmen Physik* presents as main resource a certain amount of money that is consumed by players' actions. This resource is used as investments through which players build factories, cole mines, houses for workers, etc., which support the production and future selling of the bicycles. These investments should promote *Stimulated Planning* (to encourage players to plan aspects of the game) as the player has to plan the use of these resources and assess the possible later *Rewards* (to receive something perceived as positive, or the relief of something perceived as negative, for completing goals in the game). Within this gameplay, the tasks players need to solve are a sort of Producers (some kind of a resource is produced as a consequence of a player action), and Converters which turn the result of the answers to the task into money. Likewise, solving the tasks produce patterns of Rewards within the game.

Finally, five modes of play can be distinguished in *Genius Unternehmen Physik*. A mode of play limits players' activities by allowing certain actions and making others more rewarding (Björg & Holopainen, 2005). In particular, a mode of play represents different phases of a game. A mode of play is defined by different actions, goals, interface, and game time. In *Genius Unternehmen Physik* the *Main Mode* allows players to build and destroy units. Its interface is the whole territory of the "city" plus the menu on the right of the screen, while the game time is running (see Figure 17). The *Office Mode* allows players to open letters and the main Journal where the content knowledge of physics is embedded. According to the fictional world, this mode represent the link of the player to the "outside" world and "scientific community". During this mode the time of the game is paused. The *Statistics Mode* allows players to read the basic stats of the business and also to change the value of the workers' wages. Game time is also paused. The *Journal Mode* allows players to browse through a 2D journal and read the content related to topics of physics. The fictional world presents this journal as part of a regular publication from the scientific society of physics. The content is helpful for answering the learning tasks. Game time is also paused. Finally, the *Task Mode* allows players to enter numbers and written responses as well as selecting the correct alternative among a set while the game time is paused. The Journal and Task Modes are closely related to the learning goals of the game, while the first three are closer related to the game itself.

Figure 16: Key Game Design Patterns Affecting *Genius Unternehmen Physik*'s Gameplay



Note. PC= Producer-Consumer; RM = Resource Management; SP= Stimulated Planning; CE= Cognitive Immersion; GM= Game Mastery; EI= Emotional Immersion.

Figure 17: The Main Mode of Play in *Genius Unternehmen Physik*

3.1.1.2. The Learning Material

In this section an overview of the tasks is provided. For each task the main instruction, the number of screens, the correct solution and the main section of the journal that contained the correct answer are provided. The tasks are described in the order in which they appear in the game. Each task was introduced with reference to some aspect of the game storyline or fictional world. For example, the first task about insulation material was presented as a problem of insulation of the outside toilets used by the workers of the company. Another feature of all the tasks of the game is that after providing with the correct answer the game gives feedback in relation to the story of the game. If the answer is incorrect the feedback is either showing the right solution or providing with the right procedure. Finally, correct and incorrect answers have monetary consequences in the game. That is, when correctly answered, most of the tasks provide a certain amount of money, and when incorrect, a certain amount of money is discounted from the individual's account.

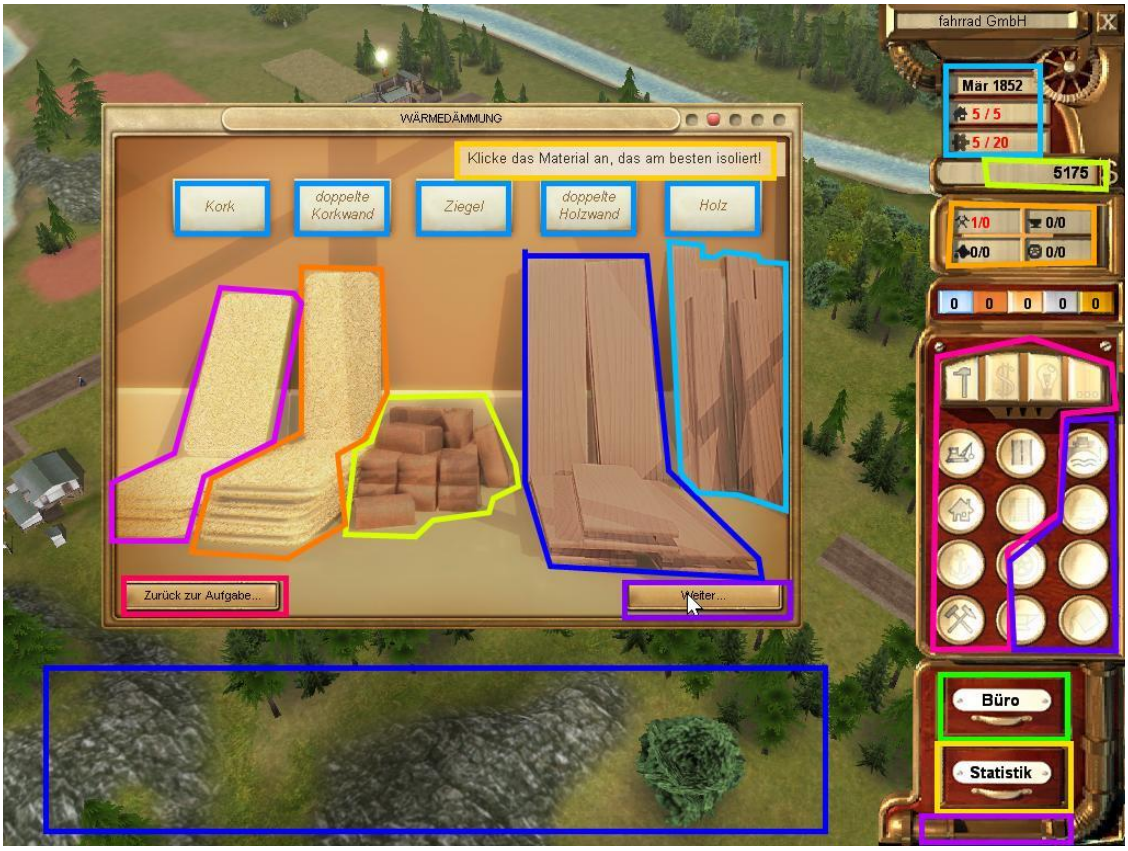
Task 0. Heat Insulation (Wärmedämmung). This task requires that the individuals recognize which material isolated best a particular construction. Individuals are provided with a set of four pictures with the different materials that might isolate better a particular building and from which only one was correct.

Task 1. Physical states (Aggregatzustände). In this task individuals have to recognize and choose which images represent the three physical states of matter (i.e., solid, liquid, and gas). Individuals have to click on three images among several distractors portraying the three physical states of matter. In this case, the right images are a cube with water, the boiling vapor from a teapot, and a painting of a white mountain.

Task 2. Density (Dichte). This task has in actuality two stages. In the first stage, individuals have to identify the weight of a coin by adding pieces of weights in a balance until both the coin and the weights added are equal. In the second stage, individuals are asked to calculate the density of the coin and are provided with, again, the weight of the coin in grams and its volume. After having mentally calculated the density, individuals had to locate the correct density from a list of five alternatives.

Task 3. Perpetuum Mobile. In this task individuals need to decide whether or not a new incoming employee that claims to have discovered the machine of the perpetum movement should be hired or not based on the scientific perspective on the issue of perpetum movement.

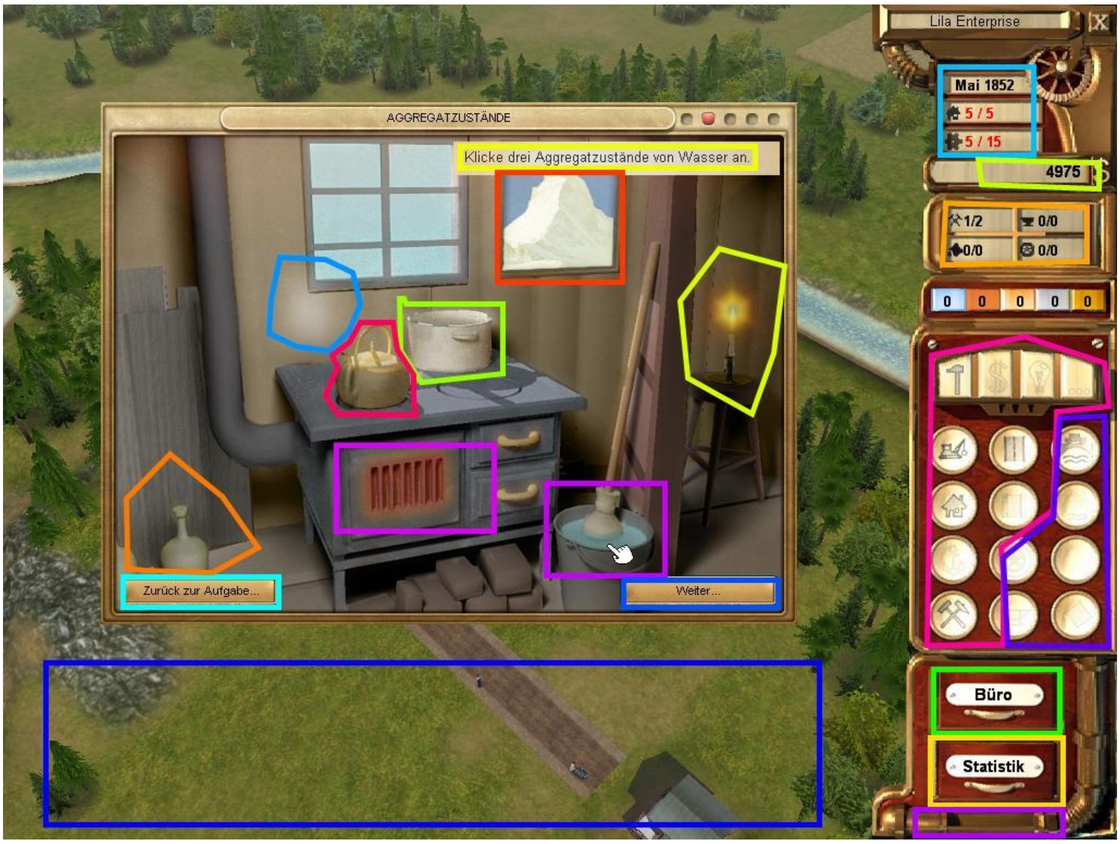
Figure 18: Screenshot Insulation Task in *Genius Unternehmen Physik*



Task 4. Planets (Planeten). In this tasks individuals are asked to arrange in the right order the planets of the solar system, starting with Mercur and finishing with Neptun.

Task 5. Pulley I (Flaschenzug I). Similar to the task on density, this one also have two stages. In stage one, individuals need to choose with which pulley, among four possibilities, more load could be pull up. Once they have chosen the correct pulley, they are asked to calculate the maximum load of the pulley given the number of ropes and the force applied in Newton. Then individuals have to write down the answer using the keyboard.

Figure 19: Screenshot Physik State of Matter Task in *Genius Unternehmen Physik*



Task 6. Pulley II (Flaschenzug II). This task uses the same context of the first one, but asks a different question. In this case individuals had to determine the number of ropes needed in order to pull up the load a few meters. The data provided were the number of ropes and the distance that the load had to be pull up.

Task 7. Pressure cooker (Schnellkochstopf). This task presented individuals with one normal cooking pot and a pressure cooker, each equipped with a thermometer. Once clicked on each of the thermometers, these showed that the two cooking pots had different temperatures. Then from four alternatives, individual had to select the one with the right explanation for the different temperatures found.

Figure 20: Screenshot Density Task in *Genius Unternehmen Physik*

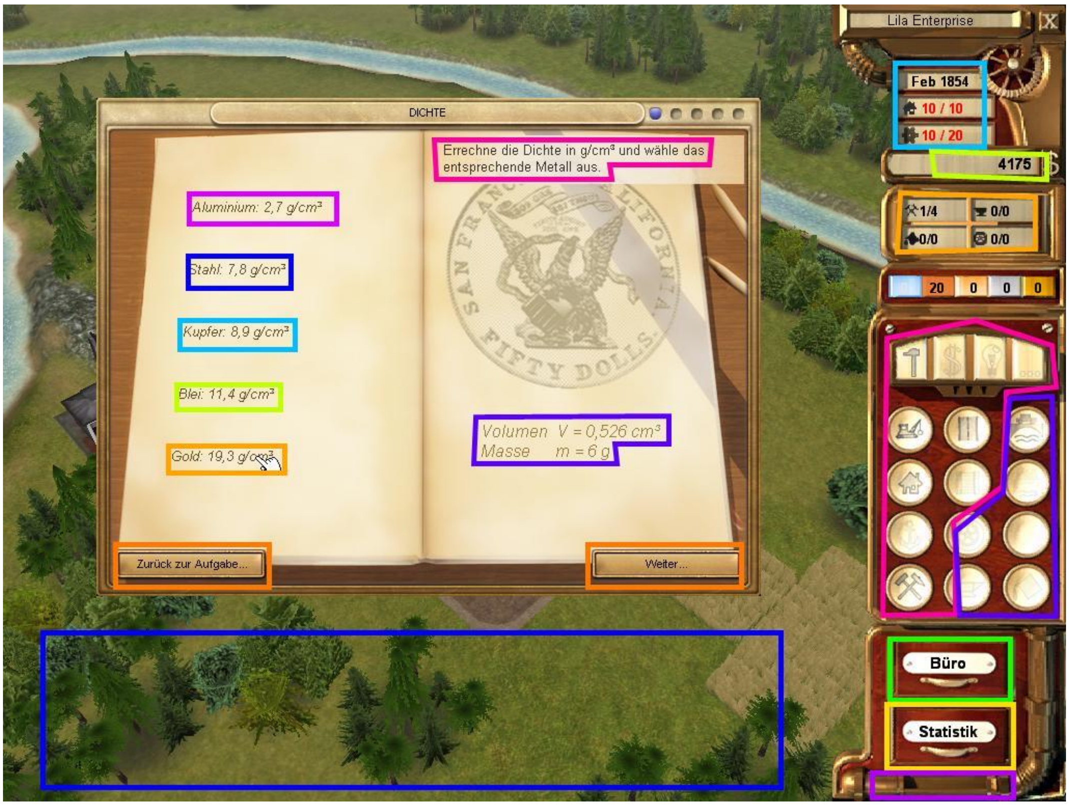


Figure 21: Screenshot Perpetuum Mobile Task in *Genius Unternehmen Physik*

Figure 22: Screenshot Planets Task in *Genius Unternehmen Physik*

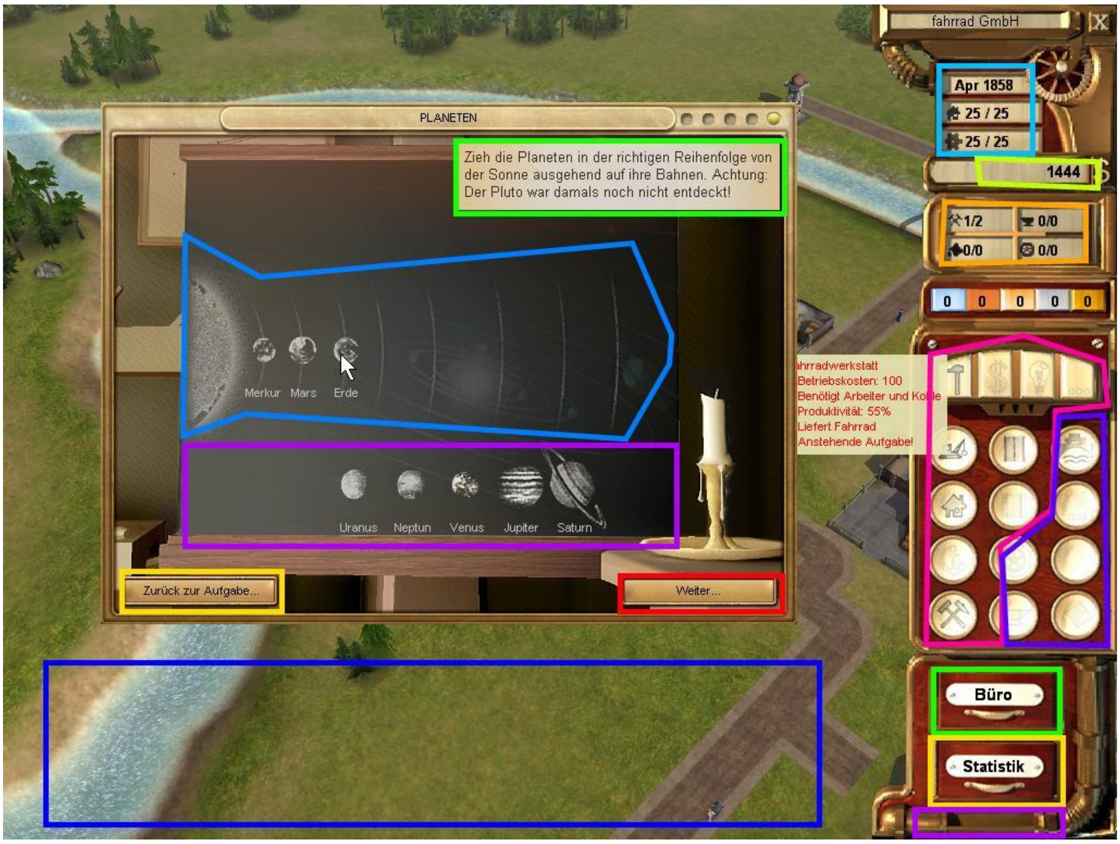
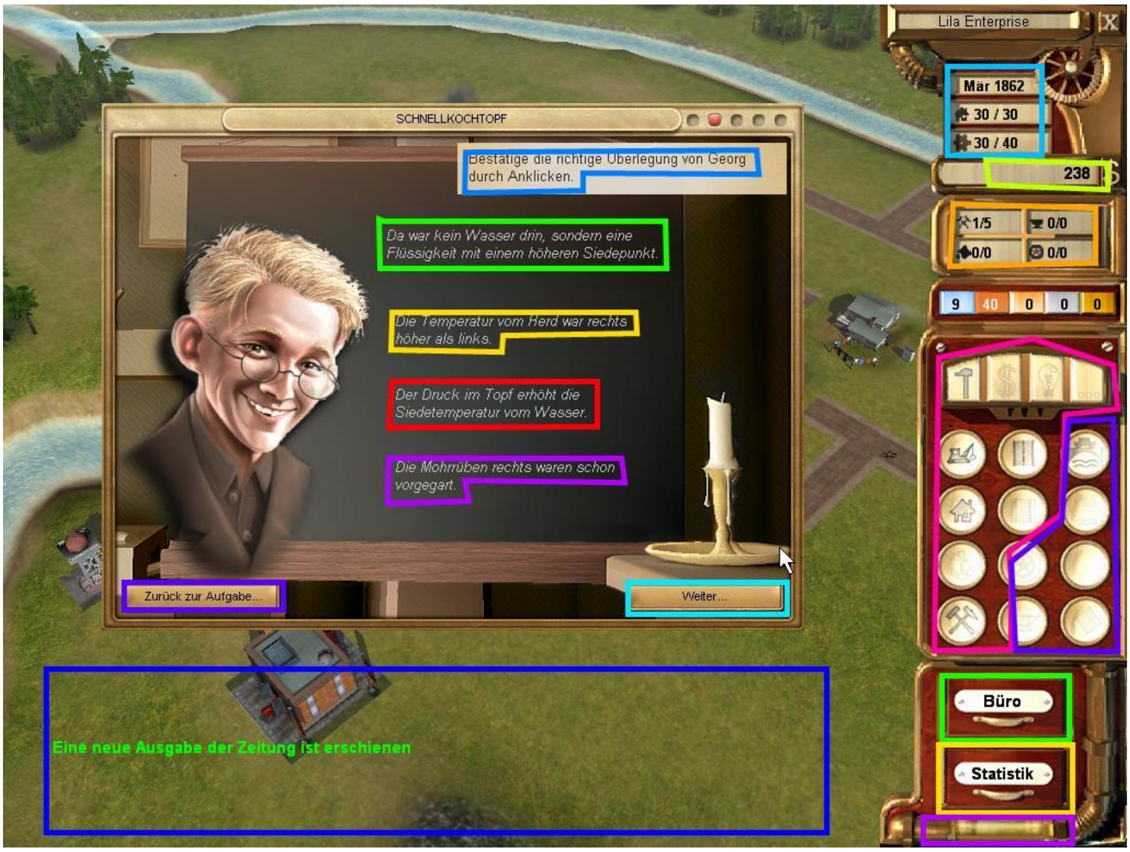


Figure 23: Screenshot Pulley I Task in *Genius Unternehmen Physik*

Figure 24: Screenshot Pulley II Task in *Genius Unternehmen Physik*

Figure 25: Screenshot Pressure Cooker Task in *Genius Unternehmen Physik*



3.1.1.3. The Pilot Study

The purpose of the pilot study was to gather information about the experimental design, so that its deficiencies can be identified and solved. In particular, this pilot study searched for evaluating the general instruction that operationalized the independent variable, the experimental setting in terms of time for gameplay, and individuals self-reports. Sixteen participants played the game *Genius Unternehmen Physik* during 20 minutes and answered to a set of questionnaire and interviews in English. Participants were randomly assigned to two conditions depending on the instruction received. Half of the participants received the instruction “play for fun” and the other half the instruction “play to learn”.

The results of this pilot study (see Filsecker et al., 2011) highlighted several points to be corrected in the main study conducted for this dissertation:

- The English level was not homogeneous among the participant
- The instructions were somehow too broad and lacked specificity
- The session seemed to be too brief for the majority of the participants

Finally, another pilot study was conducted to examine the type of technique to assess the qualitative aspect of the dissertation related to the information processing strategies employed while solving the tasks. Five participants were instructed to play a similar game *Task Force Biologie* and were asked to think aloud while playing. Results clearly showed how difficult was for the participants to play the game and think aloud at the same time. This was reflected in long periods of silence despite the experimenter instruction and reminding to keep talking while playing. Therefore, for the main study, the think aloud protocol was discarded and in its place a retrospective interview was designed and implemented (Section 4.4.2).

4. Method

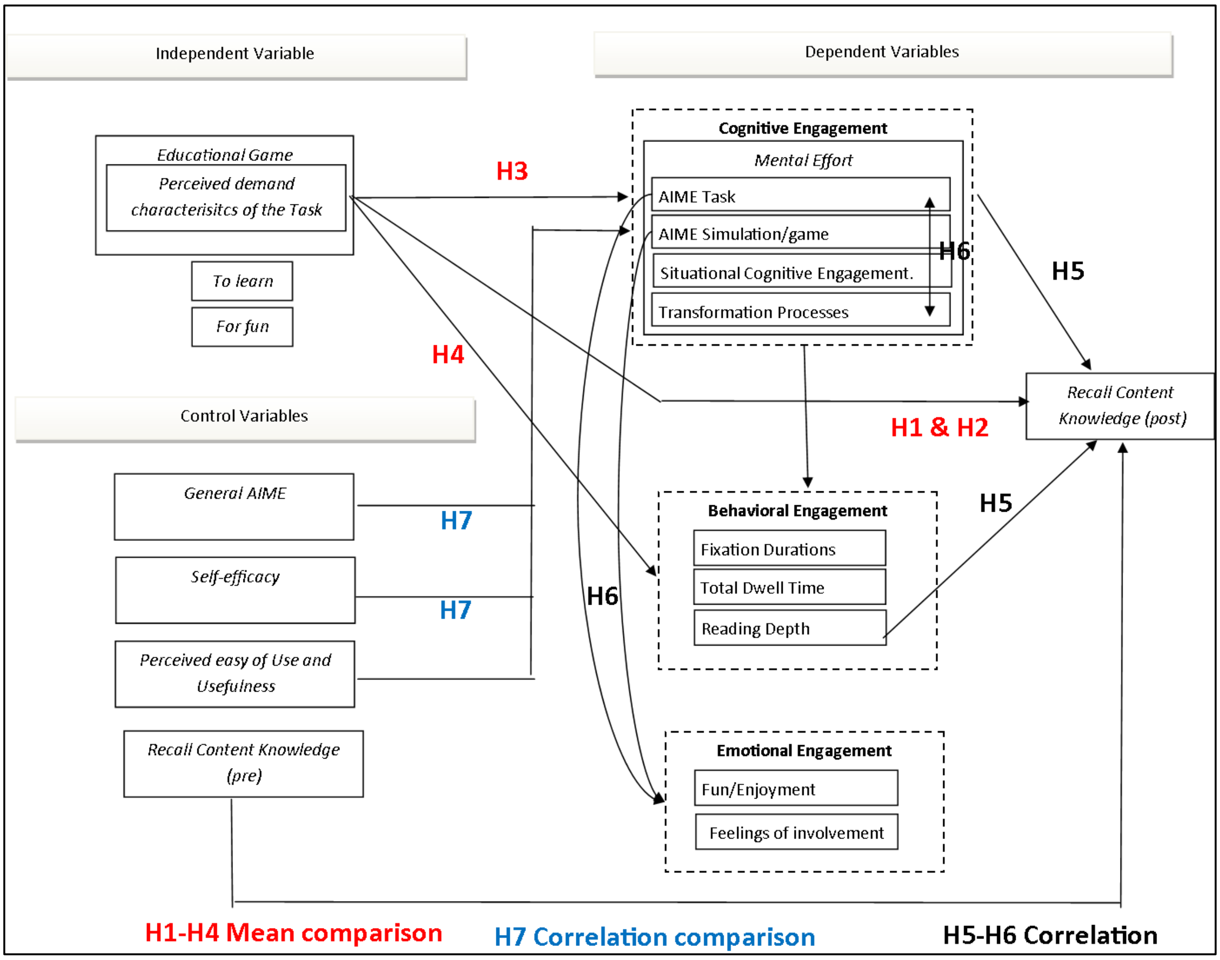
4.1. Research Design

This study employed a randomized experimental pretest-posttest control group design (Shadish, Cook, & Campbell, 2002). It involved two groups or conditions: one experimental group, which followed the instruction to play to learn, and a control group, which followed the instruction to play for fun. Participants were randomly assigned to one of the two conditions. The sessions run for each participant individually. At the end of the approximately 60 minutes participants were thanked and debriefed.

The session consisted of five phases. In the first one, participants were welcomed and explained the different activities they were expected to perform. This was followed by a five minutes explanation of the basic controls and purposes of the educational game *Genius Unternehmen Physik* and a time for questions on behalf of the participant was allowed. In the second phase participants were presented the stimulus (i.e., the game) through the Tobii Eye Tracker system. This system allows to record different types of media. For this study the Screen Recording option was employed. This option is appropriate for presenting external software application such as *Genius Unternehmen Physik*. The third phase entailed answering the questionnaires containing the dependent variables. In the fourth phase, the records of participants' gameplay were replayed from the beginning showing to the participants their fixations and gaze data as red balls and lines, respectively. At this point the interview started with general questions concerning the game and the participants' main goals and representation of the game. Then each of the tasks that participants attempted to solve was located on the video and participants were asked to answer a questionnaire of situational cognitive engagement (SCENG) together with the specific questions of the interview (see Section 4.4.2). Finally, in the fifth phase, participants were administered the recall posttest and then debriefed.

Figure 26 presents a general overview with the independent, dependent and control variables and the hypotheses that link the variables together. Hypotheses 1 to 4 are about comparing means in the two conditions produced by the independent variable. Hypothesis 5 and 6 are correlational and Hypothesis 7 compares the correlation coefficients of specific control variables and cognitive engagement variables between the two conditions. Sections 4.1.1 and 4.3 provide the detail for each variable. In the diagram, arrows pointing to the whole box with the discontinuing frame mean that the hypotheses expect differences in all the variables within that box. For example, H3 expects to have differences in terms of all the cognitive engagement measures (exception for AIME Simulation/game), which breaks it into H3a, H3b, and H3c (see Section 5.3.3), which for simplicity are not further depicted in the diagram. The next sections provide conceptual and operational definitions of each of the variables shown in this diagram.

Figure 26: Overview of the Variables and Hypothesis of the Study



4.1.1. Independent variable

According to the AIME model (see Section 2.4.1.1), AIME is significantly affected by individuals' perceived demand characteristic (PDC) of the medium, a material, an instruction or a task. The instruction received by the participants referred to the manipulation of the perceived demands characteristics of the task as used by Salomon (1984) and others in studies involving televiewing, film, interactive video and internet use (see Section 2.4.1.1). One group of participants was instructed to play to learn physics, while another group was instructed to learn for fun. The participants instructed to play the game with the intention to learn the content related to physics embedded in the game received the following instruction:

Your goal during this game is to learn as much as you can about physics. At the end of the game you should respond a test measuring how much physics you learned.

After this instruction, participants in this group were asked whether or not they understood the instruction and were provided with the opportunity to ask any question they might have. In the following sections, this experimental condition is called the condition *to learn*. On the other hand, the other group of participants was instructed to just have fun while playing the game. The concrete instruction was:

Your goal during this game is to have fun.

Participants in this group were also asked whether or not they understood the instruction and were provided with the opportunity to ask any question they might have. In what follows, this experimental condition is called the condition *for fun*.

The condition to learn was considered to represent an optimal approach to learn from an educational game, while the condition for fun was considered to represent a more typical approach to games to the extent that they call for a more relaxed processing of the material embedded in the game.

4.2. Participants

Between October 10, 2011 and December 12, 2011 forty-two German speaking adults, ranging in age from 19 years to 33 years, participated in the study (Figure 27).

As a reward participants received either a signed document in exchange of credit for taking part in a study or €10. The rest of the participants took part in the study as part of the course activities in the context of a seminar taught by the researcher. Table 24 shows basic information concerning the sample. Most of them were women in their final semesters of the Erziehungswissenschaft (EW) bachelor program at Duisburg-Essen University. One third corresponded to students in their initial semester of the Angewandte Kognitions- und Medienwissenschaft (Komedia) bachelor program at Duisburg Essen University. Most of them seldom played any sort of computer game.

Figure 27: Flow of Participants through the Experimental Phases

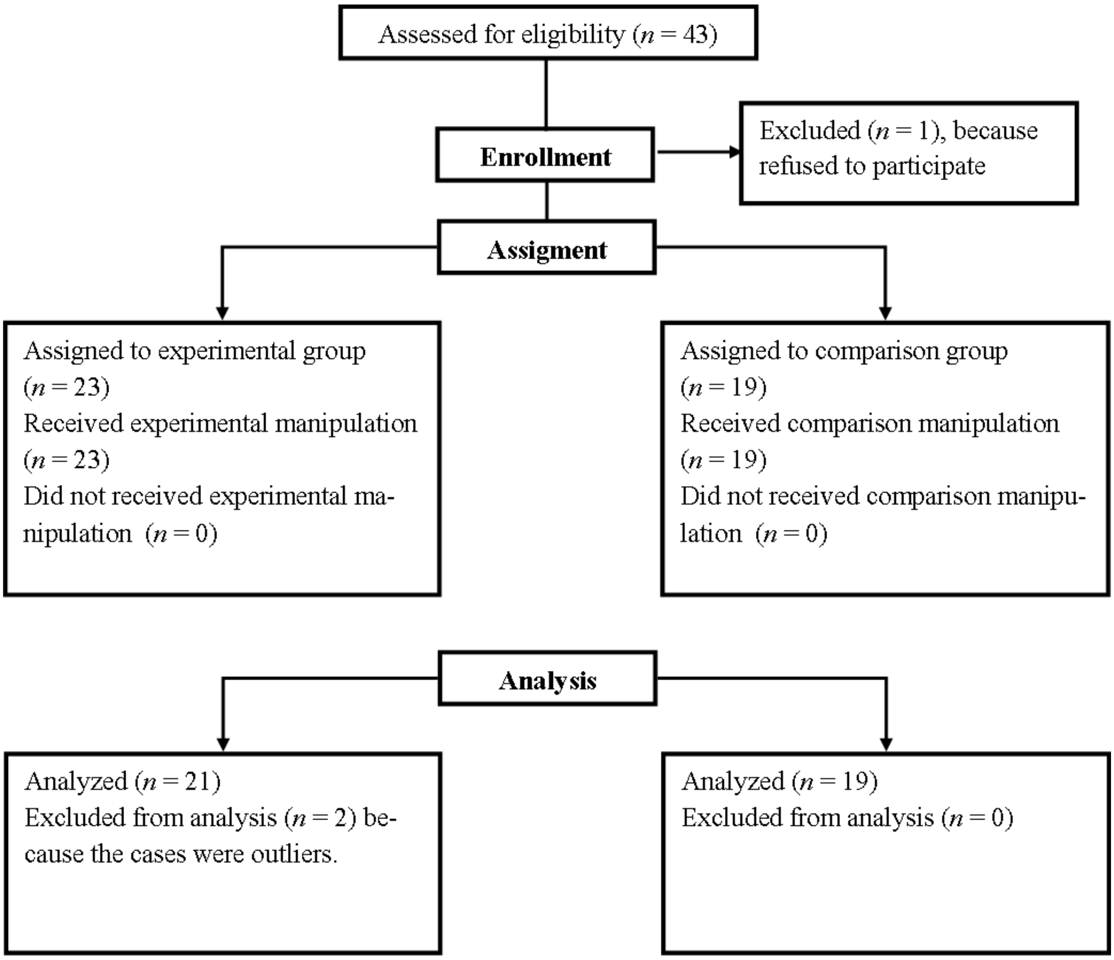


Table 24: Participants Demographics (*n* = 42)

Characteristics	*N*	%
Gender		
Female	36	85.7
Male	6	14.3
Time spent playing computer games		
Never	16	38.1
Less than once a week	13	31.0
Once a week	4	9.5
Twice a week	2	4.8
Almost every day	3	7.1
Study program		
Erziehungswissenschaft	28	66.7
Angewandte Kognitions- und Medienwissenschaft	14	33.3
Current semester enrolled		
First	5	11.9
Third	5	11.9
Fourth	1	2.4
Fifth	13	31.0
Sixth	10	23.8
Seventh	2	4.8
Ninth	1	2.4

Note: Totals of percentages are not always 100 due to missing data.

4.3. Measurement Procedures

This section presents the different instruments and strategies used to measure and assess each of the control and dependent variables of the study. Appendix I shows the original items in English and their final version in German.

4.3.1. Control Variables

The following is a list with the instruments and control variables measured so that it is possible to assess the homogeneity of the groups in aspects relevant for the hypothesis tested. All of them were administered through an online survey system[7] one week before the study.

Game use. A survey was used to collect information about individuals' experience with games. The present survey is based on Buchanan's dissertation (2006). A sample item is: "How often do you play computer games?"

Perception of the media. This questionnaire is based on the work of Salomon (1984) and Cennamo (1993). The survey asked participants "How much computer games makes you think?" and "How much effort do you invest when playing a computer game?" Through these types of questions a general scale of mental effort, the General Amount of Invested Mental Effort (or General AIME), and perceived difficulty was developed.

Self-efficacy. The Self-efficacy Subscale for Computer Gaming (five items) from the *Self-efficacy in Technology and Science* –SETS– (Ketelhut, 2004) was used. The subscale was translated into German for the purpose of this study. A sample item is: "No matter how hard I try, I do not do well when playing computer games."

Recall pretest. This test represented the knowledge acquisition that occurs during the gameplay session. The test was administered one week before the experimental session. The content knowledge was assessed by subject-matter specific items addressed explicitly by the educational game. The items require constructed responses (Salomon, 1984; Cennamo, 1993). The items consist of 12 open-ended questions referring to verbal information explicitly presented during the game. A scoring rubric was developed to score the answers to the items. An example of question was „Berechnen Sie die maximale Last und Trage das Ergebnis ein. Zugkraft F2=2 Newton tragende Seilstücke n=4. Maximale

[7] The website used was www.surveymonkey.com

Last F1=“ or „Beschreiben Sie die drei Planetengesetze von Johannes Kepler.“ The questions are provided in Appendix H. The highest theoretical score was 25.9 points.

4.3.2. Dependent Variables

The dependent variables of the study referred to the different self-reported measures of cognitive and emotional engagement obtained from individuals self-reports, plus the measures of behavioral engagement obtained from participants' eye movements. The following description is organized in the order in which the data were collected.

Behavioral engagement. Participants' behaviors were recorded during the game play using Tobii 1750 eye tracking system. The Tobii Studio screen records individuals' eye movements and mouse clicks while interacting with the computer (e.g., in the context of a webpage or a computer game). The monitor used is a 17" TFT monitor with a resolution of 1280x1024 pixels. It allows also lower resolutions such as 1024x768, the one used in this study. While interacting with the computer participants corneal reflections are illuminated by five near-infrared diodes. The sampling frequency of the eye tracker is 50 Hz, which correspond to an interval of 0.02 seconds. This means that every 0.02 seconds an eye gaze is recorded (Olsson, 2007). As any signal processing system, eye tracking needs to remove the noise in the data and distinguish between two fixations and a saccade within the variance of measurement noise. This task is achieved by developing algorithms to filter the data. Tobii Studio provides two of such filters: Tobii Studio and Clear view. For this study, the fixation filter used was the default Tobii Studio developed by Olsson (2007).

Among the more than hundreds of measures from eye tracking available (Homlqvist et al., 2011), the study focused on key measures assumed to reflect deeper cognitive information processing and more likely represent indicators of effort, persistence and choice (see Table 25). Below is a description of the procedure conducted to obtain the measures of fixation durations, dwell time, total dwell time, and reading depth for each learning page for the Journal Mode and the Task Mode. Although the measures as described below were taken from the Journal and the Tasks, for the analysis of these measure a composite measure using both the journal and the tasks pages was produced.

Table 25: Eye Tracking Measures for Behavioral Engagement

Category	General Target Question	Output	Type of processing
1. Fixation Durations	For how long was the eye still in a position?	Fixation durations in seconds.	Eye-mind hypothesis: what is fixated equals to what is processed.
2. Dwell time	For how long, measuring from entry to exit, did gaze remain inside the AOI?	The duration of the dwell in seconds.	Depending on the object and the task Dwell time can represent interest and/or difficulty in extracting information.
3. Total dwell Time	For how long did gaze remain inside the AOI in total?	Sum of all the dwell time in the same AOI	Total Dwell Time represents slow and long-term cognitive processing
4. Reading depth	How "deeply" is the text read?	A depth (in pixels) or proportion of the text looked at.	How much of the information delivered in text form is actually read.

Note. AOI = Area of Interest.

Fixation duration (FD). This variable was obtained from Tobii studio software. For each of the pages selected as relevant for learning from the game a group of AOIs (see Section 2.5.2) was defined and a particular name was given for each of them. The learning related pages referred to two of the main modes of play identified in the game (see Section 3.1.1.1), that is Journal Mode (JM) and Task Mode (TM). As already described the JM contains information and content related to physics and represents the main knowledge resource that participants can use to answer the knowledge questions addressed during the TM and later in the recall test.

The JM is composed of 11 pages that distribute the necessary information to be able to respond to the questions during the game and also to be able to respond to the posttets designed for this study. For each of these pages a group of AOI was defined. There are some AOIs that are common to all Modes of Play because they are always visible to the players even though they are inactive. Figure 28 illustrates this. To the right of the image, the Menu with different tools is always visible but it is not functional unless the participant closes the main window in the middle. Within this middle window different portion of texts and images are displayed. In this case, the window corresponds to one of the pages displayed during the JM. The first three AOIs (i.e., in yellow, blue, and orange) are called, for example, “Q1.2_1”, “Q1.2_2”, and “Q1.2_3” referring to the number question in the final recall test (i.e., Q1.2). This question asks about the order of the planets of the solar system. Within these three AOIs is the information needed for being able to answer the question after the game and accomplishing the task during the game. Therefore, for this page a mean fixation duration for all the fixations within these three AOIs is computed.

Figure 28: An Example of AOIs Definition within a Learning Page during Journal Mode of Play (JM)

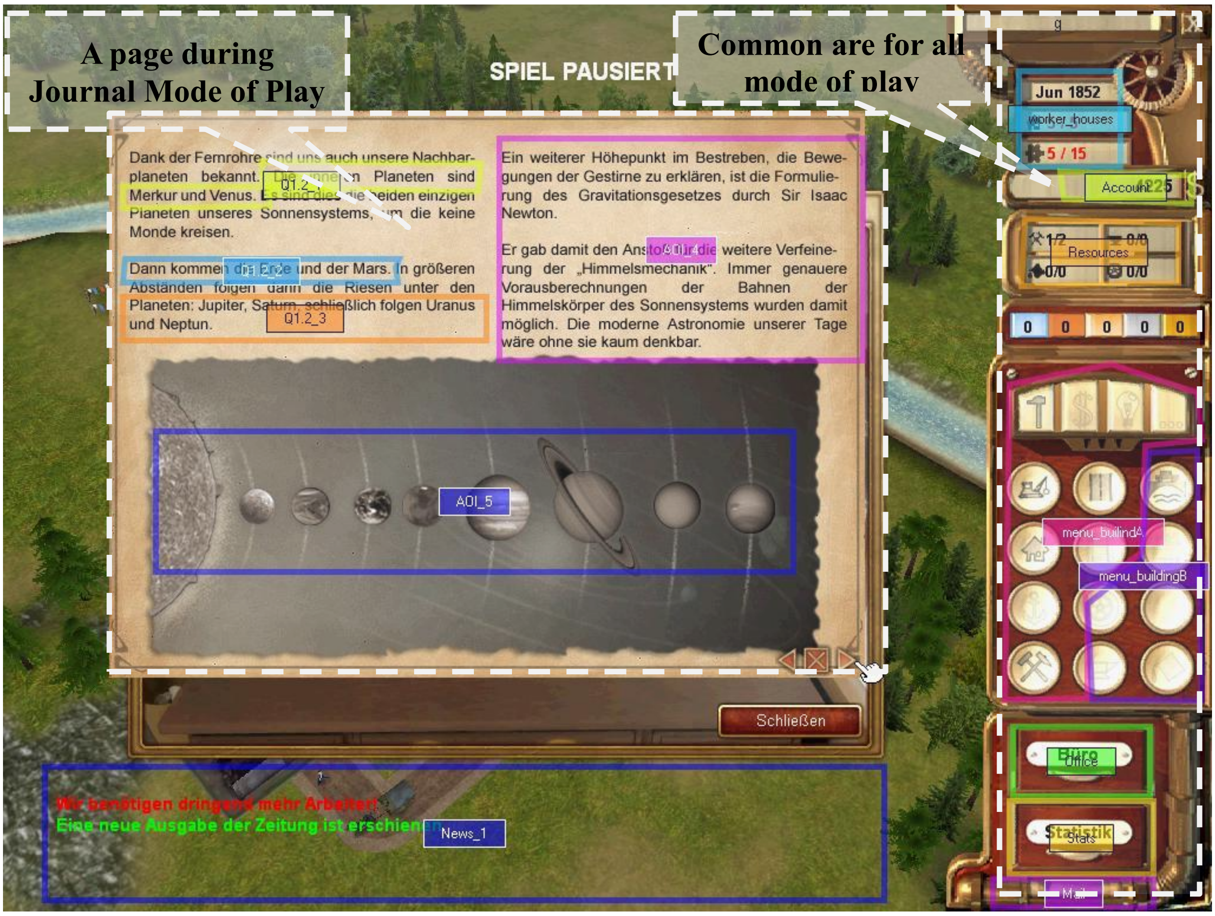


Figure 29 shows the fixations of one participant on the AOIs described above and Figure 30 shows a graphic of the mean fixation durations for each AOI, including the non-AOI mean fixation durations (i.e., fixations that fell out of any defined AOI).

Figure 29: A Participant's Scanpath across and within AOIs during the Journal Mode of Play (JM) – Example 1

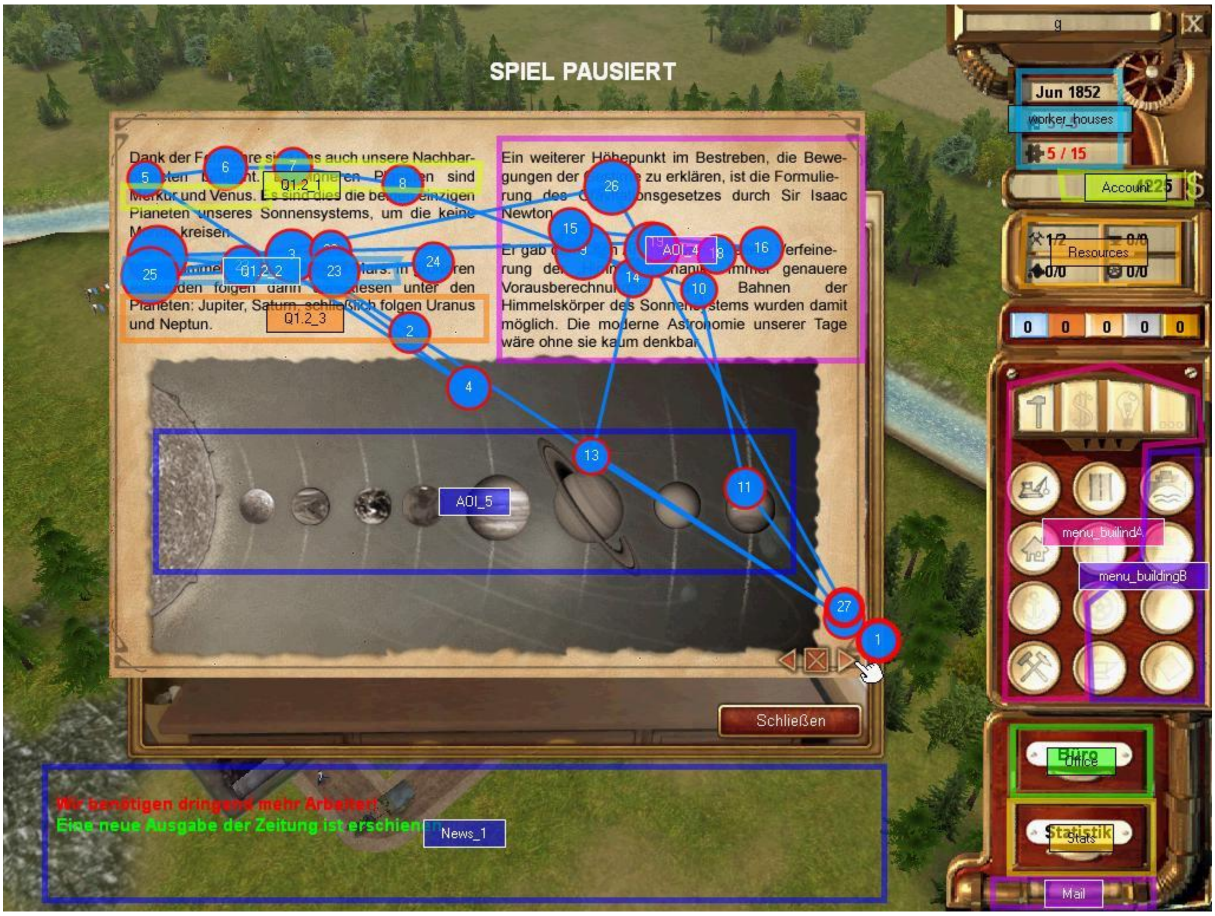
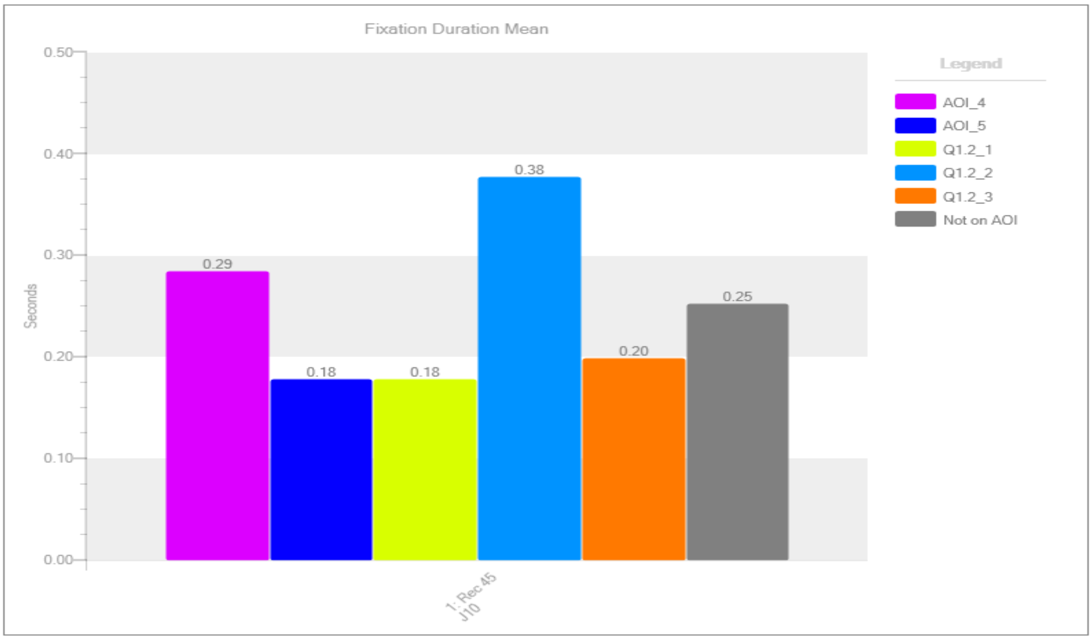


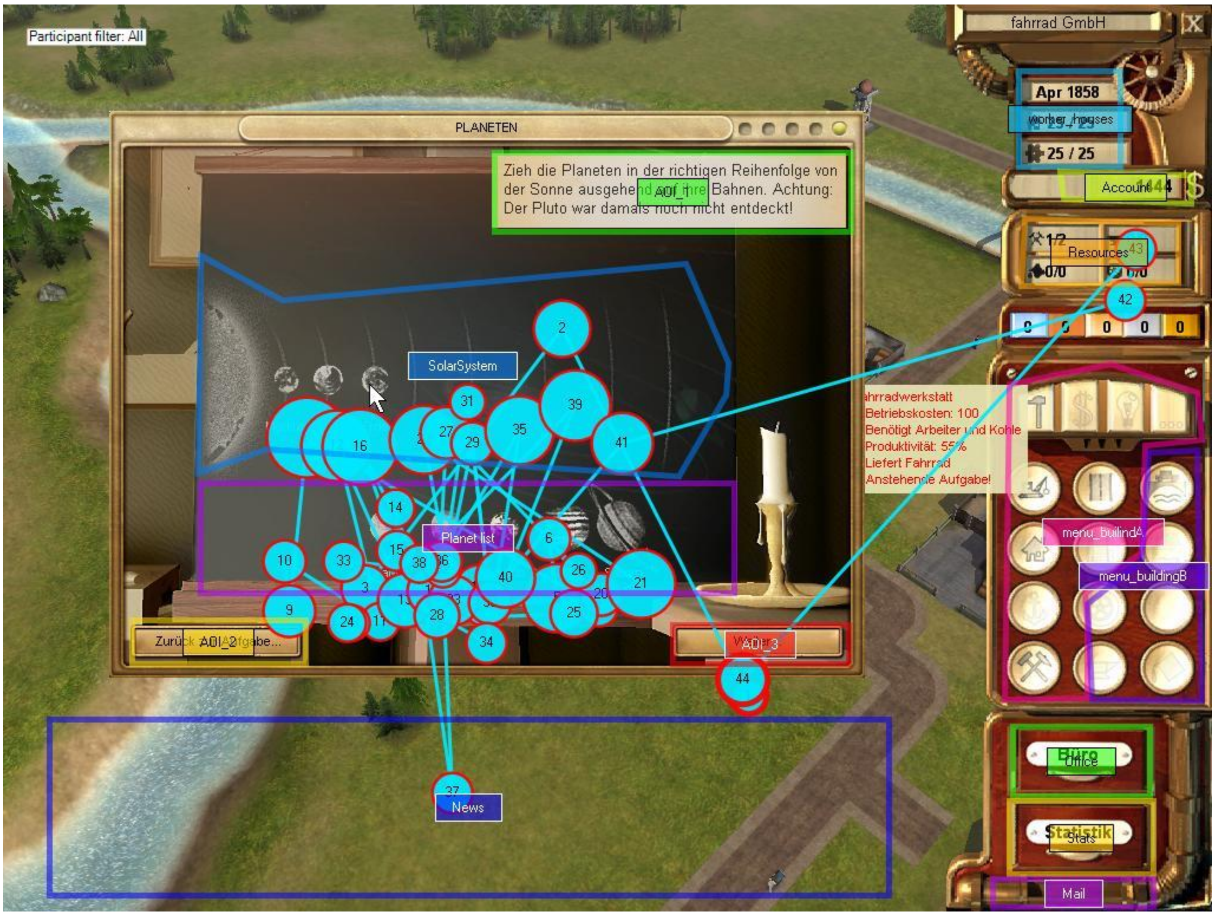
Figure 30: An Example of a Participant's Fixation Duration Means in each AOI Identified during the Journal Mode of Play (JM)



As the data show the mean fixation durations for the three AOIs are 0.18, 0.38, and 0.20. Therefore, for these three AOIs the mean fixation durations would be (0.18 + 0.38 + 0.20)/3=0.25. This means that the participant's mean fixation duration is 0.25 seconds.

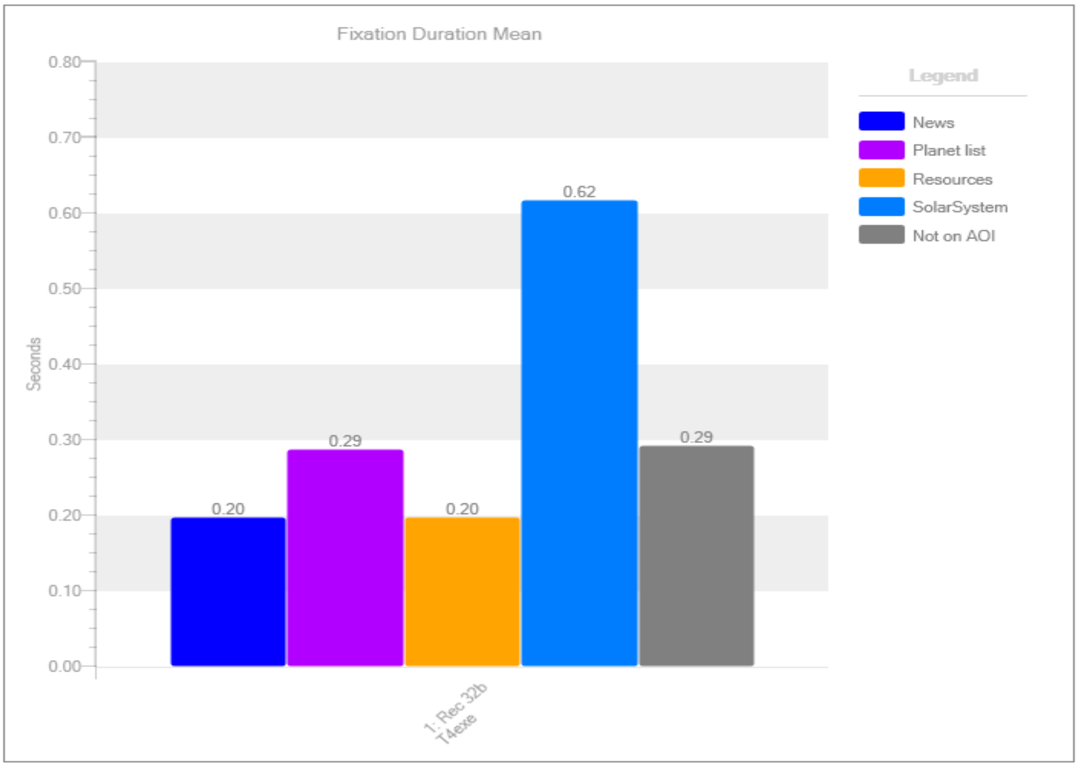
In the Task Mode (TM) individuals are expected to answer the different questions asked during the game. For each question there is an introduction that situates the question in the context of the game story, followed by the question itself, that is, the instructions and the basic data needed to answer it. For each of the pages containing a question, a group of AOIs was defined. Again, there are some AOIs that are common to all Modes of Play. Figure 31 illustrates this. To the right of the image, as mentioned above, the Menu and its tools is displayed. The middle window contains different portions of texts and images.

Figure 31: A Participant's Scanpath across and within AOIs during the Task Mode of Play (TM) – Example 2



In this case the central AOIs (i.e., in blue, purple and light green) are called "SolarSystem", "Planetlist" and "AOI_1". This question asks individuals to order each of the planets from the closest to the farthest to the sun. Figure 32 displays the values of the mean fixation durations for each of the AOIs already showed in Figure 31.

Figure 32: An Example of a Participant's Fixation Duration Means in each AOI Identified during the Journal Mode of Play (JM)



The same process described above for the calculation of the mean fixation durations during the JM is used here. For the AOIs already mentioned the mean fixation durations are 0.62, 0.29, and 0.00, for the Solarsystem, Planetlist and AOI_1, respectively (AOI_1 is 0 because it was not fixated at all as can be seen in Figure 31). Therefore, the final value for this individual in this page would be (0.62 + 0.29 + 0.00)/3=0.33. This means that the participant's mean fixation duration is 0.33 seconds. This example further illustrates the value of identifying and selecting particular AOIs within a window and timeframe. If all the AOIs are included (i.e., Resources and News) the final mean fixation duration for the same participant would have been 0.28. This represents more time, but unrelated to the physic content.

Dwell time (DT). This variable is based on fixation durations and represents the mean duration of one visit to an AOI (see Section 2.5.2). In order to get the dwell time for a learning page, the same procedure described above was followed.

Concerning the Journal Mode (JM), the dwell time for a specific page is calculated by averaging the dwell times for each relevant AOI in the page. For instance, Figure 33 shows a participant's scanpath entailing fixation durations and transitions within a window of 3.1 seconds. Here the individual visited the AOI depicting the planet four times. The first visit corresponded to fixations number 1 and 2 (fixation 3 falls out of the AOI), the second visit to fixations 4 and 5 (fixation 6 falls out of the AOI), the third visit to fixation 7 and the last visit started with fixation 13. In the case of the upper AOI, it shows one visit corresponding to fixations 8, 9, 10, 11, and 12. Each of these groups of fixation durations is averaged and the result represent the dwell time for each AOI.

Figure 33: A Participant's Scanpath across and within AOIs during the Journal Mode of Play (JM) – Example 3

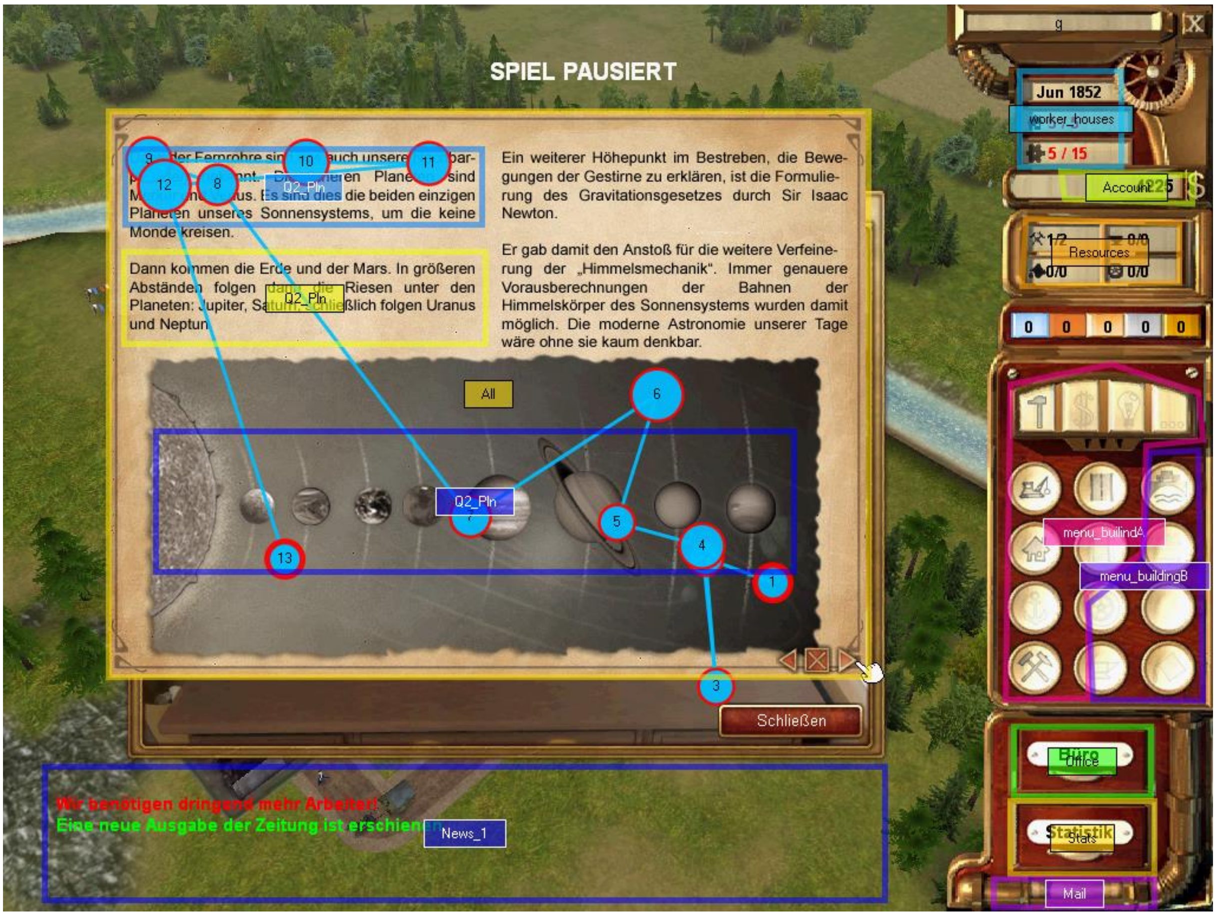
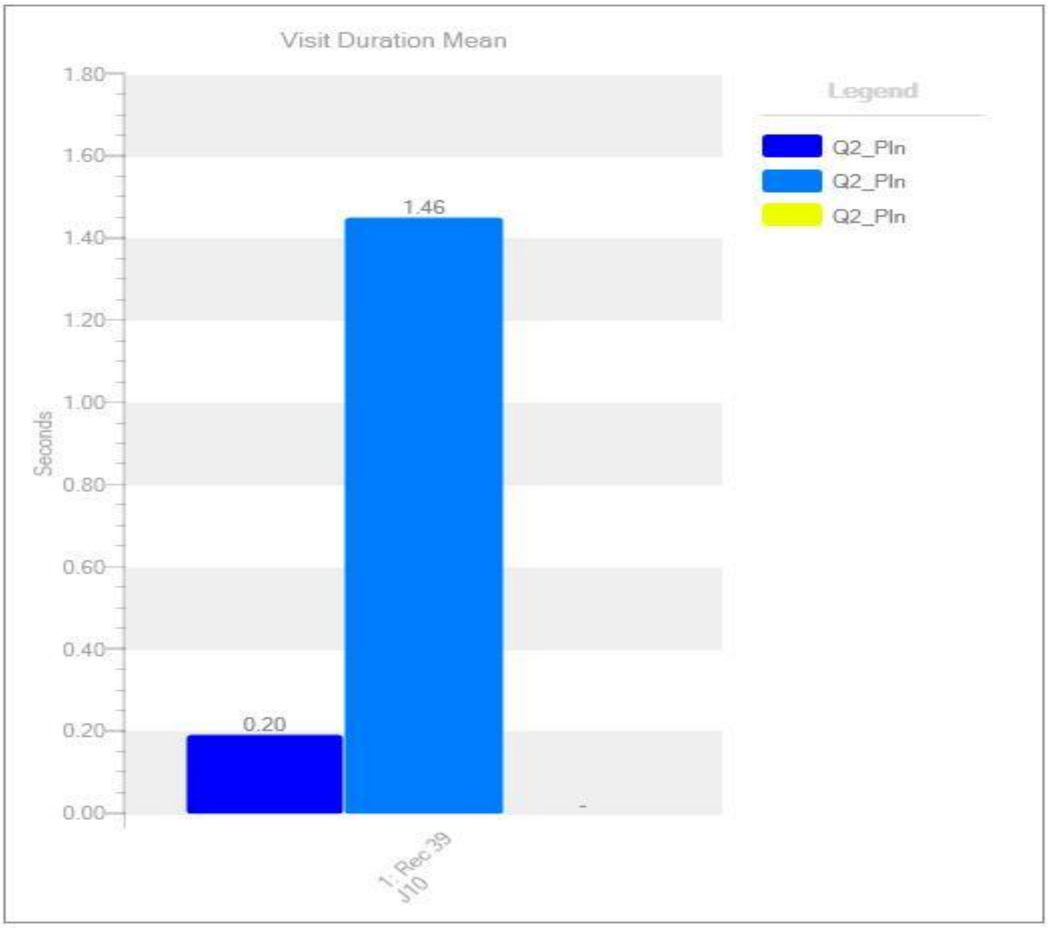


Figure 34 below shows the mean values for the dwell time of the visited AOIs. It can be seen that the graph does not contain any values for the AOI in yellow, which means it was not visited at all within the window of 3.1 seconds. For instance, the dwell time for the upper AOI is equal to 1.46 seconds. This value is obtained by multiplying the number of fixations in the AOI by the mean fixation durations in the AOI. In this case 5*0.29 = 1.46. For the three AOI of this example, the mean dwell time would be (0.20 + 1.46 + 0)/3 = 0.55 seconds.

Figure 34: An Example of a Participant's Mean Dwell Time in each AOI Identified during the Journal Mode of Play



Concerning the Task Mode, the same procedure is employed. For instance, Figure 35 shows a participant's scanpath entailing fixation durations and transitions within a window of 5.1 seconds. Here the participant visited the upper AOI with the instructions one time (even though it can be seen that fixations 1 and 12 are within, fixation 1 is actually considered to fall out of the AOI), the AOI Solarsystem was visited one time and the AOI Planet list one time. For the upper AOI, the visit corresponded to fixations 2 through 6 (fixation 1 falls out of the AOI). For the lower AOI, the visit corresponded to fixations 7 through 11. For the AOI Solarsystem, the visit corresponded to fixation number 12. For each of these groups of fixation durations an averaged was calculated to represent the dwell time for each AOI.

Figure 35: A Participant's Scanpath across and within AOIs during the Task Mode of Play (TM) – Example 4

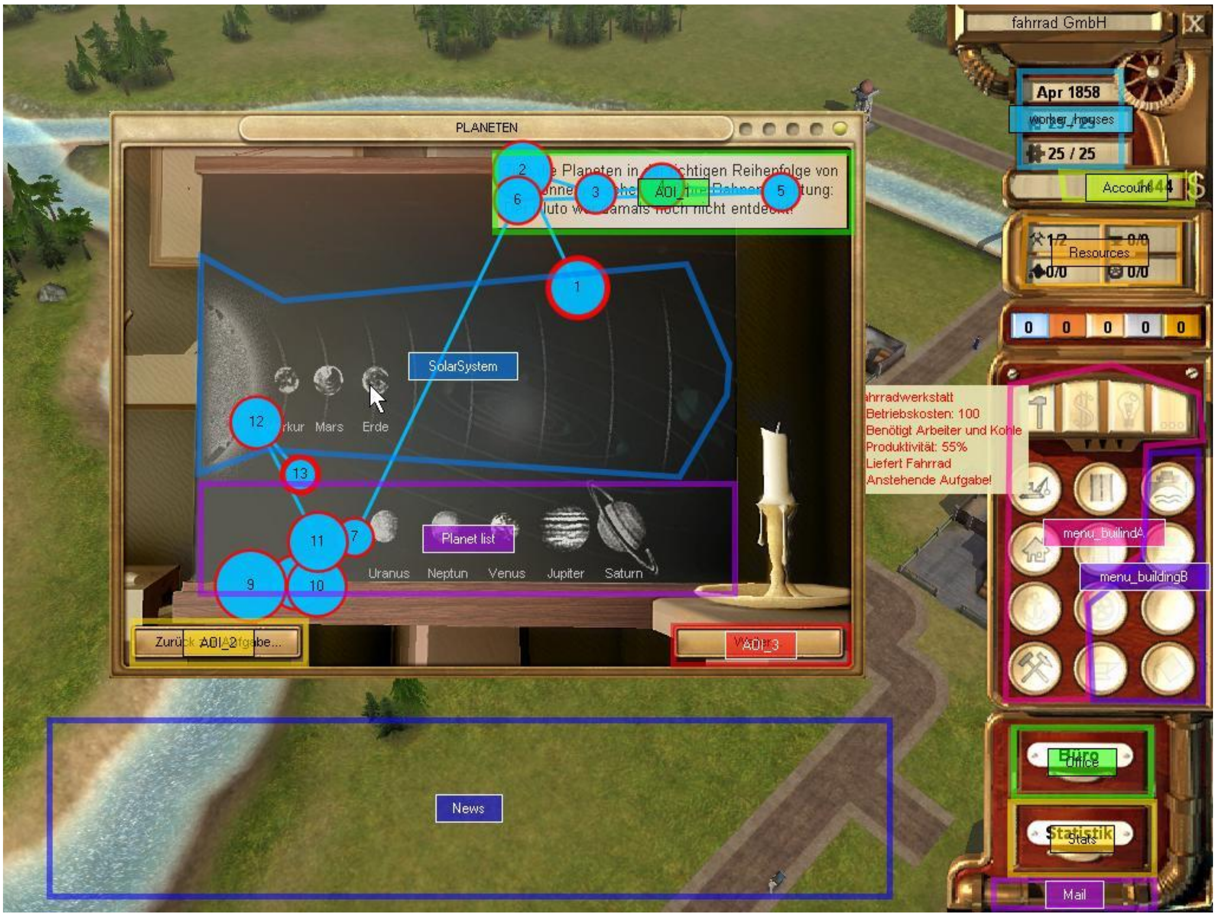
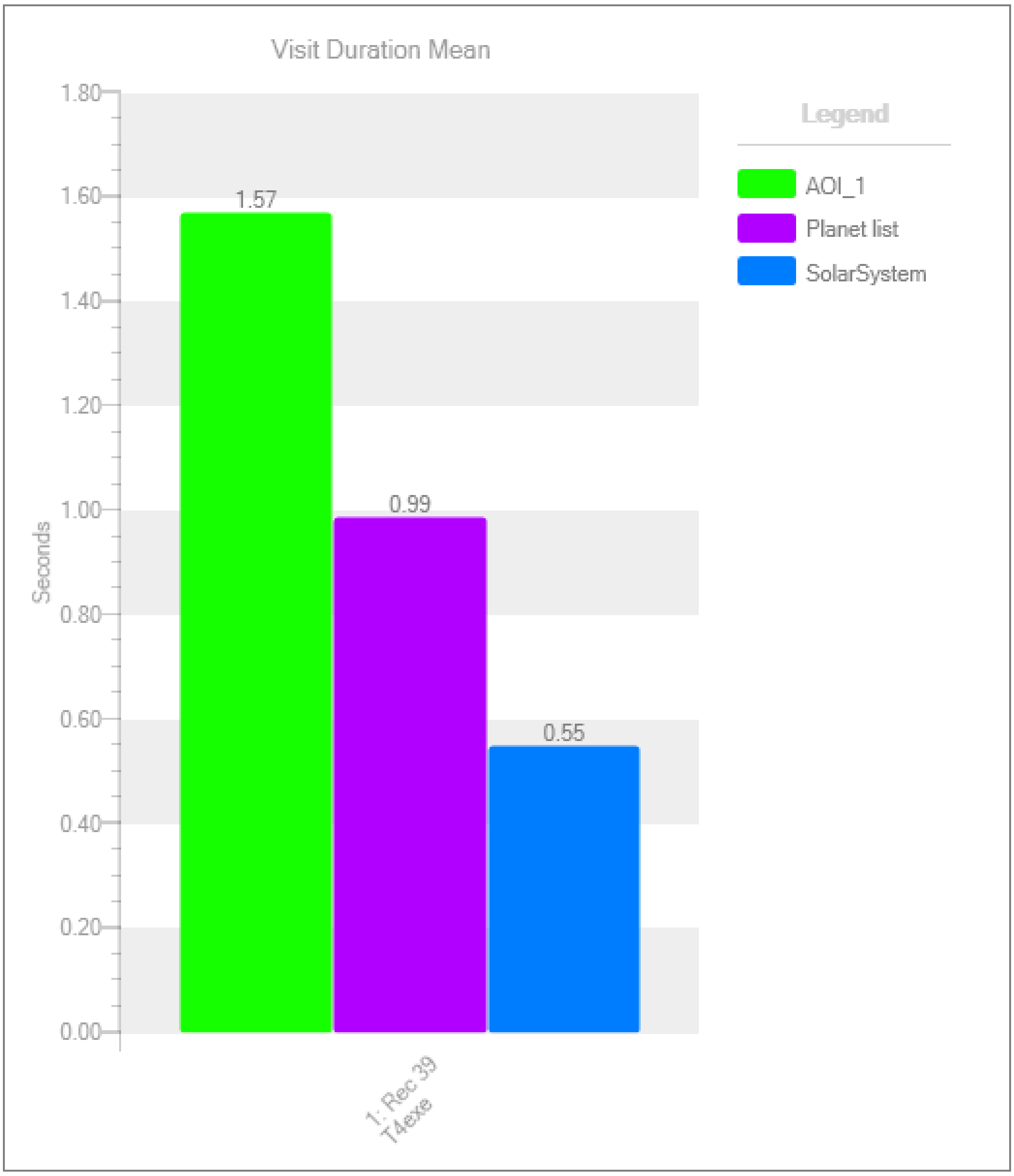


Figure 36 below shows the mean values for the dwell time of the visited AOIs. During the window of 5.1 seconds, the higher mean dwell time was 1.57 seconds for the AOI containing the instructions for the task. The second was the AOI showing the pictures of each planet, with a mean of 0.99 seconds. The last AOI received a visit of only 0.55 seconds. For the three AOIs of this example, the mean dwell time would be (0.57 + 0.99 + 0.55)/3 = 0.70 seconds.

A closely related measure of dwell time is total dwell time. This variable represents the mean duration of all visits to an AOI. It is calculated by adding the mean dwell time of each visit to a specific AOI. For example, let's assume that AOI "Solarsystem" was visited two times. Let's also assume that the dwell time of the first visit was .45 seconds and the dwell time of the second visit was .35 seconds. A total dwell time for the AOI "Solarsystem" would be .45 + .35= .80 seconds. Finally, once the rest of the AOIs' total dwell times are calculated (e.g., .30 and .15), the mean Total Dwell Time for the page is .80 + .30 + .15=1.25 seconds.

Figure 36: An Example of a Participant's Mean Dwell Time in each AOI Identified during the Task Mode of Play



Reading depth. As described in Section 2.5.2 this measure requires the area of the AOI and the dwell time spent in the AOI. Using the example during the Journal Mode below is a comparison of the same individual's reading with a control subject instructed only to read carefully the text. Figure 37 shows that the participant instructed to read carefully (i.e., the "reader") had longer and more fixation durations and therefore a longer dwell time than the participant playing the game (i.e., the "player"). The reader obtained a mean fixation duration of 0.51 second against the 0.27 seconds from the player. Likewise, the reader obtained a mean dwell time of 2.32 against a dwell time of 2.12.

Figure 37: "Reader" versus "Player" Scanpaths within an AOI during the Journal Mode of Play (JM)

(a) (b)

Next a measure of the area is needed to compute the reading depth measure. First, the AOI area needs to be converted into pixels. As the presentation of the stimulus was done through Tobii Studio Screen Recording option at a 1024x798 resolution, the area of the stimuli corresponded to the area of the screen, which is 1024x798 = 786432 square pixels. Second, these pixels need to be converted into centimeters. This conversion needs to consider the size of the screen and the DPI (dots per inch) used. In this case, an online pixel calculation (www.pxcalc.com) was used and the data about the screen resolution and the diagonal of the screen used during the experimental session was introduced. The output provided the horizontal and vertical dimension of the screen in centimeters. In this case those values were 34.54cm * 25.91 cm. which gives an area in centimeters of 894.93. Third, the initial AOI area provided by Tobii as a percentage of the screen is converted into squared centimeters. For example, if an AOI represent an area of 2.47% (ca. 0.025) the area in squared centimeters would be 0.025*894.9 = 22.1. Finally, using the Mean Dwell Time in seconds provided by Tobii, it is possible to calculate the Reading depth measure. Following the example, if the Mean Dwell Time was 2.5 seconds for this AOI, then the reading depth would be 2.5 seconds divided by 22.1 squared centimeters, that is, 0.11 mean dwell time per square centimeter. For the reader in Figure 377, the Mean Dwell Time is 2.32 seconds and the AOI area in squared centimeters is 22.01. Therefore, her reading depth is 2.32/22.01 = 0.10 mean dwell time per square centimeter. On the other hand, for the player this value is 2.12/22.01 = 0.09. This means that the reader engaged in a slightly deeper reading of the text contained in this AOI than the player during the Journal Mode. Appendix J shows the areas in % and squared centimeters of each AOI used to calculate the reading depth measure.

Cognitive engagement. This variable addressed the quantity and quality of individuals' investment of mental effort when engaged with academic content. For collecting information about the amount of individuals' mental effort, the following measures were used: Amount of Invested Mental Effort (AIME) (Salomon, 1984; Cennamo et al., 1991), and 4-item Situational Cognitive Engagement scale (SCENG) (Rotgans & Schmidt, 2011). The AIME measure was tailored to three aspects of the game, that is, the Simulation, the learning Tasks, and the Statistics of the game. Therefore, three scales were developed asking for the amount of invested mental effort in these three activities within the game. Appendix E presents the actual German version used for the study. Therefore, three scales named AIME Simulation, AIME Task and AIME Statistics were used in this study. The technical details of AIME are summarized in Table 15 , Section 2.4.1.1.

On the other hand, the SCENG measure was used only for the Tasks embedded in the game (see Appendix F for the version in German). The questions were asked for each learning task participants were engaged in. The selection of this measure was motivated by the following reasons: 1) its "sensitivity" to the task demands; 2) its short scale quick to administrate; 3) its psychometrical properties; and, 4) the learning context in which it was developed and validated (i.e., a learning task). The SCENG questionnaire was developed by Rotgans and Schmidt (2011) to overcome the limitations of current measures of cognitive engagement, which according to the authors were too general and unable to capture small contextual variations such as the type of task students might engage in the classroom (e.g., self-study, group work, listen to a lecture, etc.). SCENG measure is composed of three dimensions:

- Students' perceptions of their present engagement with the task
- Students' rating of their effort and persistence with task
- Students' feeling of being absorbed by the task

Considering the notion of cognitive engagement as a latent variable, Rotgans and Schmidt (2011) engaged in the validation of the measure under the logic of structural equation modeling or SEM (Hancock and Mueller, 2001). Hancock's coefficient H (i.e., the reliability coefficient of SCENG), for the exploratory and confirmatory samples was .93 and .73, respectively. Both scales (AIME and SCENG) used a 5 point-likert scale. Below are the items used in the current study:

Table 26: The Scales' Items used to Measure Mental Effort.

AIME	SCENG
How hard did you try to understand the (simulation/learning task/statistics)?	I was engaged with the topic at hand
How difficult was the (simulation/learning task/statistics) to understand?	I put in a lot of effort
How much did you concentrate while (playing, reading learning task, reading the statistics)	I wish I could still continue with the work for a while
How much did the (simulation/learning task/Statistics) make you think?	I was so involved that I forgot everything around me.
How much did you try to remember what you saw in the (simulation/learning task/statistics)?	
How much effort did you put into comprehending the (simulation/learning task/statistics)?	

Note. AIME = Amount of Invested Mental Effort; SCENG = Situational Cognitive Engagement

For the qualitative aspect of the effort invested, an interview was conducted based on Corno and Mandinach's (1983) model of cognitive engagement. The interview represented a compromise between the retrospective think-aloud analysis (RTA) (van den Haak & de Jong, 2003) and the process-oriented interview approach (POI) (Järvelä & Salovaara, 2004). It has been designed specifically to fulfill one of the purposes of this study, that is, explore the goal-related cognitive processes during gameplay. Therefore, it shares an open and more unstructured section similar to the RTA and a more focused section similar to the POI. Several reasons led to create this interview process instead of using either the common concurrent think-aloud analysis (CTA) technique or POI only. The CTA was discarded for the following reasons: 1) The extra effort required by the CTA can distort important measures of attention allocation and concentration (Guan, Lee, Cuddihy, & Ramey, 2006), 2) RTA has shown better performance due maybe to the workload associated to the CTA (van den Haak & de Jong, 2003), phenomenon highly probable in complex stimuli such as games, 3) RTA is more focused on explanations than CTA (Bowers & Snyder, 1990), especially if cued (van Gog, Paas, van Merrienboer, & Witte, 2005). On the other hand, the POI departures from traditional RTA in that it looks to explore individuals' motivational goals and cognitive strategies (Järvelä & Salovaara, 2004), as well as the acquisition and transformation processes related to cognitive engagement (Corno & Mandinach, 1983). The cued part of the interview process refers to the fact that participants watched the recording of their gameplay including their eye movements. Therefore, the technique used here is called Cued-Retrospective Process Oriented Interview or CPOI. Section 4.4.2 describes the process of analysis of the interviews.

Emotional engagement. Chen, Kolko, Cuddihy, and Medina's (2011) Game Engagement Questionnaire was translated and adapted for the purposes of the present study. According to the authors, the questionnaire sought to integrate different concepts relevant in the experience of gameplay. In this way, the questionnaire included items assessing usability aspects, flow, and "fun". For this study, the questionnaire comprehended 11 items (see Appendix K). Some examples of the items were "I felt like I was inside the game world" and "I enjoyed a lot the theme and content of this game". By means of a factor analysis, the items that represented more closely the emotional engagement dimension were identified and related to the degree of involvement, enjoyment and fun aspects usually explored in the literature on flow state and games.

Recall posttest. This test is identical to the one used as pretest. For details see description in section 4.3.1.

4.4. Scoring Procedures

The following section describes how the recall test and the interview were scored. For the test, a rubric was used based on the content knowledge embedded in the game. In the case of the interview a coding scheme reflecting the model of cognitive engagement (Corno & Mandinach, 1983; Howard, 1989) was applied.

4.4.1. Recall Test

The test was scored based on a rubric (see Appendix L) for assessing the level of recall for each of the questions. Each question was assigned a slightly different value based on the assumption that recalling some information may be more difficult than other. The maximum possible score was 25.4 points. The scoring process entailed three steps. In step one, the researcher and an assistant independently scored a sample of the responses on the pretest and posttest. Then, the scores obtained for each participant were compared and the discrepancies were discussed until an agreement was reached. Next, independently and blind to the experimental condition, the entire pretest and posttest were scored by the researcher and the assistant.

4.4.2. Cued-Retrospective Process Oriented Interview (CPOI)

The content of the interview was quantified according to a coding scheme developed to reflect the cognitive engagement of the participant with the tasks embedded in the game (Corno & Mandinach, 1983). The interviews were transcribed and analyzed with MAXQDA version 10. With this software a process of inter-rater reliability was conducted with a total of 8 selected interviews.

The process of development and application of the coding scheme followed the general steps that traditional quantitative content analysis entails (Schreier, 2012). Quantitative content analysis is focused on the manifest meaning of the material and therefore little use of the broader context is needed. It is more concept-driven than data driven, which makes central the handling and checking of the reliability of the coding frame employed (Chi, 1997; Schreier, 2012). The process of building the coding frame followed 4 steps. In the first step, only the relevant information from the transcription was selected. In the second step, the structure of the coding scheme was specified, that is, the main dimensions and subcategories of the coding frame. For this it was used a concept-driven strategy which borrowed from the model of cognitive engagement of Corno & Mandinach (1983) and similar research (e.g., Howard, 1989). Next, the definition and specification of each dimension of cognitive engagement was established. Finally, the researcher and an assistant applied the coding frame to analyze 5 interviews. Both codes were compared and disagreements were discussed until the final coding frame was generated. Then a new set of 3 interviews were coded with the revised coding frame (see Appendix M) and an inter-rater reliability was established (see Section 5). What follows is the definition of each dimension of the coding frame employed.

As described in section 2.4.2.1, Corno and Mandinach's (1983) model entails the processes of acquisition and transformation, with the subsequent cognitive processes of Attention, Monitoring, Selectivity, Connecting, and Planning. However, Howard (1989) based on Corno & Mandinach's model, introduced some modifications, part of which were employed in this study. Basically, the modification had to do with defining two levels (i.e., general and specific) to the main dimensions of the model. For the present study, the operationalization of these categories was as follows.

Acquisition processes. For *Attention*, it was first identified an Attention "minus" or negative attention (see Howard, 1989). Each statement reflecting an explicit superficial tracking and /or reception of information was coded as Attention (-). An example of an Attention (-) statement is "…also ich hab' das überflogen" (i.e., I just skimmed it). All other statements reflecting the incoming of information were coded as Attention. For example, the statement "Nee, also ich hab' das ja öfter nachgelesen" (i.e., no, I often read it again) was coded as Attention. The codes for *Monitoring* were divided into General Monitoring and Specific Monitoring. *General Monitoring* reflected the number of statements referring to a level of awareness or general understanding of the task itself or the progress toward achieving the goal of the task. "Also ich wusste das jetzt nicht ganz genau." is an example of General Monitoring statements (i.e., "I did not really know that"). *Specific Monitoring* reflected statements showing individuals going back to existent sources of information in order to understand it, check, rehearse or clarify a specific aspect of the task or state the understanding or lack of understanding of a particular object or process of the task at hand. An example of this type of statement is "Also ich hab' das extra nochmal nachgelesen, damit ich weiß wie ich darauf komme, ob das jetzt Gold ist oder nicht Gold ist" (i.e., "I read that again, so that I can determine whether or not that is gold").

Transformation processes. The first dimension introduced here was *Planning* "minus". Similar to Attention minus, this dimension attempted to identify statements showing clearly the lack of any further consideration of the actions made and where the use of guessing and trial and error were the main strategies used. An example of this type of statement is “Und das war dann eher so ... vermutet, geraten, sag ich mal so…” (i.e., “I supposed, I guessed, let’s put it in that way”). *General Planning* was indicated by statements showing an assessment of the task’s requirements, decision-making concerning the use of resources or possible strategies. It is reflected also in statement of short-term goals or chain of actions directed toward a goal. An example of this type of statement is “Ja, ich hab' die Formel gesucht für die ähm für die Seillänge.” (i.e., Yes, I searched the formula for the length of the rope). *Specific Planning* reflected statements describing the steps employed to solve the task or a particular sequence of procedures together with the goal that inspired them “So zu sagen: Strategie eins: Erinner' dich an die Formel, Strategie zwei: guck mal was an Zahlen da steht und wie man die verbinden kann und ähm dann einfach eintippen” (i.e., “so to speak: strategy one, remember the formula, strategy two see which numbers are there and how they can be related and then just type in”). *Selectivity* was indicated by statements showing the search for particular pieces of information, the discrimination among stimuli and the distinction between relevant and irrelevant information. An example of this type of statement is “Ähm hauptsächlich hab' ich halt ... die Errechnung angeguckt.” (i.e., “I watched mainly at the computation”). *Connecting* referred to the search and linking of information with familiar/prior knowledge. Following Howard (1989) this dimension was further divided into general and specific. *General Connecting* referred to statements showing general awareness of everyday knowledge. An example of this statement is “ja, ich hatte Physik in der Schule” (i.e., “yes, I had physics in school”). *Connecting specific* referred to statements showing the connection of two or more elements of a task or the explicit use of specific prior knowledge obtained from an informative source. An example of this type of statement is “Ich habe in der Zeitung gelesen, deswegen wußte ich schon (i.e., “I read it in the journal, that’s why I knew it”). Here the “Journal” refers to a resource embedded in the educational game used in the study (see Section 3.1.1.1).

In order to distinguish between acquisition and transformation processes, two measures were developed. For acquisition, the frequencies for General Monitoring, Specific Monitoring, and Attention were added up. Frequencies of Attention minus were subtracted to this sum. The final frequencies were divided by the number of tasks each participant attempted to solve. Similarly, the process of transformation entails the dimensions of Planning, Selectivity and General Connecting and Connecting Specific.

Following the same logic already described, the frequency of statements related to transformation processes was calculated as follows:

Table 27: Acquisition and Transformation Processes Index

Acquisition	Transformation
	Dimensions
Attention (+)	General Planning
Attention (-)	Specific Planning
General Monitoring	Planning (-)
Specific Monitoring	General Connecting
	Connecting Specific
	Selectivity
	Index
Acquisition=[(General Monitoring + Specific Monitoring + Attention)-(Attention minus)]/Number of tasks	Transformation = [(Planning + Selectivity + General Connecting + Connecting Specific)-(Planning minus)]/Number of tasks

4.5. Statistical Analysis

This section provides a description of the main measures of location and variation relevant for this study, together with a description of the type of graphical summaries of data employed. Next, the main features of the test of group differences and association are described. The description and observations made here are based on the work of Wilcox (2012).

4.5.1. Descriptive Statistics

The goal of descriptive statistics is to provide a useful summary that convey parsimoniously the key properties of a determined set of data within the context of an empirical study. This summary is usually achieved by measures of location and variation calculated from the sample data, together with information about the shape of the data distribution.

Measures of location. A measure of location is a number that usually represents the typical individual or object under study. In more technical terms, a summary of data that fulfill the following two properties is a measure of location: 1) the value must lie between the smallest and largest value from a list of values, and 2) if all the values are all multiplied by a constant *b*, then the value of the measure of location is multiplied by *b* as well. The central measures of location are the *sample mean*, the *sample median,* the *trimmed mean*, and the *winsorized mean*.

The *sample mean* can be calculated by averaging a list of *n* observed values or by multiplying each value for its frequency and adding up the results. The mean is intended to represent or estimate the population mean from which the sample was drawn. However, the sample mean is highly sensitive to unusually large or small values called *outliers*. To quantify this sensitivity it is used the *finite sample breakdown point* of the sample mean, which refers to "the smallest proportion of observations that can make it arbitrarily large or small" (Wilcox, 2012, p.21). In this case a single observation (i.e., one large or small value) can make the sample mean goes up or down, regardless of the other values. This means that the breakdown point for the sample mean is $1/n$. Therefore, in the presence of outliers, the mean may poorly reflect the typical value or response.

Another measure of location that overcomes the limitation of the mean is the *sample median*. The sample median represents the extreme form of trimming, that is, the removing of a proportion of the smallest and highest values and averaging the remaining ones. The sample median, when an odd ascending ordered sample size is used, corresponds to the middle value. When the sample size is even, the sample median corresponds to the average of the two middle values. The finite sample breakdown point of the sample median is .5 – the highest possible value. Therefore, while the mean is highly sensitive to outliers, the median represent the opposite end of this sensitivity. That is, the sample median is resistant to the presence of outliers in the data.

Finally, the *trimmed mean* represents a compromise between the sample mean and the sample median in terms of their sensitivity to outliers. The trimmed mean removes a particular proportion of the lowest and highest values in the data (e.g., 10%). This proportion represents its finite sample breakdown point. Even though it is an open question how much to trim, as a rule of thumb trimming a proportion of 20% is a good choice. A particular variation of this procedure occurs when calculating the *winsorized mean.* Instead of removing the smallest and largest values trimmed, when winsorizing these values are set equal to the smallest and largest values *not* trimmed.

Measures of variation. Individuals respond different to conditions. This variation is reflected in the data and because of this the sample mean rarely is equal to the population mean. The goal is to take into account the sources of this variation when estimating the population mean with the sample mean. The central measures of variation are the *sample variance,* the *standard deviation*, the *interquartile range*, *median absolute deviation* (MAD), and the *winsorized variance*.

The *sample variance* is calculated by subtracting the sample mean from each of the n observation in the sample and then squaring the difference. Then, these results must be added up and divided by $n - 1$. The square root of the sample variance is called *standard deviation*. As with the sample mean, the sample variance is not *resistant* to outliers, which means that any unusual value can inflate the sample variance. In technical terms, this means that the finite sample breakdown point of the sample variance is only $1/n$.

The *interquartile range* is usually employed in the detection of outliers. Roughly stated, the calculation of the interquartile range involves removing the smallest and largest 25% values of the data and then taking the difference between the smallest and largest value remaining. The particular methods of calculation depend on the purpose of using the interquartile range.

The *median absolute deviation* (MAD) is relevant when trying to detect the presence of outliers and is calculated by subtracting the sample median from each observed value and then taking absolute values. Its finite sample breakdown point is .5.

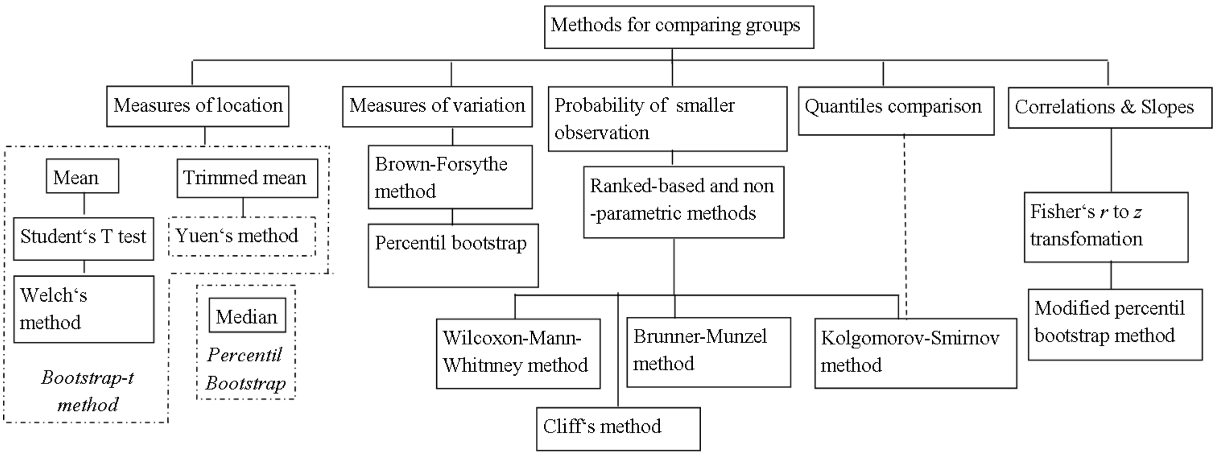
Finally, the *winsorized variance* is useful when working with the trimmed mean. The winsorized variance is the sample variance of the winsorized values. The procedure is identical to the winsorized mean and the finite sample breakdown is also equal to the amount winsorized.

Data distributions. In order to describe the distribution of the data Histograms and Box-plots represent two commonly used graphs employed for examining whether or not the distribution is normal and whether or not there are outliers in the data, respectively. As this study compared mainly means between groups (see Section 4.5.2) a focus on the skewness of the data was considered given that it tends to impact tests of means whereas kurtosis tends to impact test of covariance structures not implemented in this study (Byrne, 2012). Additionally, tests of univariate normality were conducted in order to have a complete picture of the shape of the data.

4.5.2. Test of Group differences and Test of Association

In order to test the hypotheses this study used parametric and robust procedures. For comparing two groups there are four general approaches: comparing measures of location (e.g., media or median), comparing measures of variation, focusing on the probability that an observation in one group is smaller than an observation in a second group, and comparing simultaneously all the quantiles in the two distributions (Wilcox, 2012). In turn, for each of these approaches there are several methods that provide alternatives perspectives on how groups differ (Figure 38). This study is interested in the typical behavior between the two conditions (i.e., for fun or to learn) and the degree at which the independent variable affects the correlation of specific variables. Therefore, the mean, the 20% trimmed mean and the correlation comparison methods are used to compare participants under the two experimental conditions.

Figure 38: Methods for Comparing Two Groups of Observations



Univariate versus Multivariate. The present study explored the effect of an experimental manipulation on different dependent variables. Having different variables normally entails different tests which can inflate the Type Error I. For dealing with this issue lowering the alpha level through a Bonferroni adjustment is a familiar practice for researchers. On the other hand, there is also the possibility of using multivariate approaches that deal with the problem of multiple comparisons or tests. Given the characteristics of this study concerning the relatively small sample size, the univariate approach using Student's T and Bootstrap methods without the Bonferroni adjustment were preferred to multivariate alternatives or the univariate plus Bonferroni alternative. The following considerations led to this decision.

Concerning the Bonferroni adjustment, it is not clear whether or not is a good solution for the issue of controlling the Error Type I (Bender & Lange, 2001; Perneger, 1998; Sankoh, Huque, & Dubey, 1997). There seems to be different scenarios where using the Bonferroni approach might be appropriate and some scenarios when it might not. Roughly speaking, using the Bonferroni adjustment seems to be appropriate when more than two groups are compared as in the case of a "one-way" analysis of variance. It is also appropriate to use it when a single null hypothesis is tested using more than one test, as in the case of a repeated measure design. A more complex scenario occurs when there are multiple hypotheses being tested with a single test each, as in the case of the present study. On the other hand, using Bonferroni adjustment depends on the interest and research questions under examination. If a researcher is interested in the general question of whether or not a set of background (e.g., independent) variables are related to a set of other (e.g., dependent) variables, and for addressing this question a number of comparisons are made, the Bonferroni adjustment might be appropriate. On the contrary, if the researcher's interest is in specific relationships among the variables supported by theoretical or technical considerations, and for that each relation is expressed by a specific hypothesis, then Bonferroni adjustment should not be used. This latter situation is the one that better represent the purposes and research questions of the present study. In addition, Bonferroni adjustments, by lowering the Alpha level, decrease the power of the test. Similarly, it seems to work well only with a few numbers of comparisons (Bender & Lange, 2001). Finally, Bonferroni adjustments refer to the *comparisonwise error rate*, which does not seem to be that useful and is concerned with controlling the error of the global null hypothesis – all individual null hypotheses simultaneously true –, which "…is of limited interest to the researcher." (Bender & Lange, 2001, p.343). In a similar vein, it has been claimed that Bonferroni adjustment control the Type I Error of testing an irrelevant null hypothesis: "Bonferroni adjustments provide a correct answer to a largely irrelevant question." (Perneger, 1998, p. 1236). Therefore, no Bonferroni adjustments were used for testing the hypotheses of this study.

Finally, in the case of having several dependent variables of interest as in the present study, the use of multivariate analysis of variance (MANOVA) seems to be an appropriate alternative. Even though this approach can deal with a set of dependent variables, it has also some requirements and assumptions, hard to meet with the data of this study. First, the MANOVA is used as an alternative to several ANOVAs, that is, when more than two groups are examined. However, it is theoretically possible to use it for two groups. Second, MANOVA's sample size requirement is high when power, effect size, number of groups and dependent variables are considered. For example, for three groups, a power of .7, moderate effect size and only 3 dependent variables, the number of individuals per cell is about 40. Third, for MANOVA to be robust concerning the assumption of normality, the cell should have the same number of individuals or cases (Läuter, 1978). Finally, MANOVA has its own limitations such as the several assumptions that need to be met, the ambiguity in interpreting the effect of an independent variable on a dependent variable, and the limited situations in which MANOVA offers a more powerful solution than common ANOVAs or in the case of the present study, than Student's T (Tabachnick & Fidell, 1996). In light of the above considerations, the present study used the Student's T test and the Bootstrap-t method as complementary approaches (Wilcox, 2012) without any Bonferroni adjustments.

Student's T test. This test seems to be sensitive to small departures of normality and homoscedasticity. A rule of thumb is to assume both when the sample size is relatively large. However, it is not clear what this number should be (commonly, 30 cases per group is considered to be large). Furthermore, there are several tests to check for both assumptions (e.g., Levene's test for equal variances and Kolgomorov-Smirnov and Shapiro-Wilk for normality). However their power is not always enough to detect situations in which these assumptions should be discarded. To address this problem modern statistical procedures, such as bootstrapping and trimmed mean have been developed. As Student's T requires normality and because distributions are never exactly normal, small departures from normality can bias Student's T results. On the one hand, small departures form normality, through the inflation of the standard error of the mean, can reduce drastically the power of the test. And as outliers tend to affect the distribution by inflating the standard error, its detection is central when testing hypothesis. On the other hand, non-normality limits the ability either to control Type I error or the probability coverage when computing confidence intervals. For example, when sampling from a mixed normal distribution (i.e., a distribution "normal" in the middle but with heavier tails than the normal distribution), the current Type I error is 0.022 (n=20), a much lower value than the nominal 0.05. This reduces power. In other words, the researcher has a lower probability of rejecting the null hypothesis when it is false. Similarly, when sampling from a skewed distribution, the actual Type I error can be as high as 0.42 (n=12). This means that now the researcher has an extremely high probability of rejecting the null hypothesis.

In summary, when sampling from normal distribution with unequal variances or with unequal sample sizes, Students T test starts to show problems. Even though Student's T test was designed for being sensitive to mean differences between two groups, in actuality is sensitive to a *set of other ways* in which distributions can vary, such as different amounts of skewness. However, modern bootstrap methods have been developed to counteract these issues highly likely to occur in everyday social research.

Bootstrap-t method. In general terms a bootstrap method provides a procedure to test hypotheses and compute confidence intervals without assuming normality. As in everyday practice the samples are drawn from population with unknown distributions, the bootstrap-t method addresses this issue by re-sampling with replacement from the observed values in the current data set. The assumption is that this procedure yields a more accurate approximation to the true distribution than simply assuming normality of an unknown distribution. The advantage in terms of controlling Type I error probabilities with different bootstrap methods over Student's T is different depending on the type of distribution and sample sizes. When sample sizes are small (n = 20) as is the case of this study, and the data shows slight departures from normality, the problem arises of controlling the Type I error probability and keeping the power of the test high enough. For example, for Lognormal distributions the Type I error even though higher than the nominal value (0.05) it is less with bootstrap methods than with Student's T. Similarly, for skew distribution the Alpha level is in actuality greater than the nominal level when using Student's T as opposed to bootstrap methods.

Correlation. When conducting correlational analysis, it is important to keep in mind the factors that might affect the magnitude of Pearson correlation found. First, the distance of the points to the regression line between the two variables. The further the points are to the line, the smaller the magnitude of the correlation. Second, the magnitude of the slope around which the data points are centered. The lower the magnitude of the slope, the lower the magnitude of the correlation. Third, the presence of outliers. One outlier suffices to bring the magnitude of the correlation to zero or to a fairly large value even though the remaining values show an opposite tendency. Fourth, the restriction in range in any of the variables. The restriction of the range can either increase or decrease the correlation magnitude. Fifth, the curvature of the data, and finally, the reliability of the measure used.

In order to overcome the influence of outliers, when comparing correlations, the bootstrap method for calculating confidence intervals was preferred to the Fisher r-to-z Transformation (Wilcox, 2012).

5. Results

In this section the results are presented. PASW Statistics 18 was used for descriptive analysis, the Student T test and the calculation of correlations. For the bootstrap confidence interval for trimmed means and correlations the software R was employed. MAXQDA was used to analyze the interviews conducted.

First, the results of the scale construction and their reliability are presented. Second, a descriptive analysis of the control and dependent variables and their distribution is provided. Third, descriptive statistics of the eye tracking data and interview data is presented. Fourth, the results of the hypothesis testing are presented. Finally, a re-analysis of the data based on the median-split method explores the behavior of those individuals high and low on Amount of Invested Mental Effort (AIME). When reporting the results of statistical test of differences (e.g., Student' T), the *statistical* is used to describe t values with a probability of .05 or less. The term *significant* is used to describe the effect size of the differences found (Larson-Hall, 2010). Similarly, when reporting the correlation coefficients, the term "statistical positive correlation" is used to reflect that the coefficient has a probability of .05 or less. Its degree of significance is assessed by the strength of the coefficient (i.e., weak, moderate or strong) (Sheskin, 2007).

5.1. Scales

What follows is a description of the reliability of the data collected through the online survey and at the end of the experimental session. In particular, a description of the Cronbach's Alpha (α) values of the scales used for the control and dependent variables is provided (Table 28).

The lowest Alpha occurred in the *Perception of Computer Games Difficulty* scale (α=.237) and was not included for further analysis. The highest Alpha (α=.773) occurred in the *Self-efficacy in Computer Gaming* scale. From the scales *General Amount of Invested Mental Effort* and *Self-efficacy in Computer Gaming* one item was removed in order to increase the overall Alpha of the scales.

Concerning the dependent variable *Amount of Invested Mental Effort (AIME)*, the lowest Alpha (α=.486) occurred in the scale asking about the Simulation (see Table 29). The highest Alpha (α=.894) occurred in the scale asking about the learning tasks. In both of these scales one item was removed to improve the overall Cronbach's Alpha of the scales.

Table 28: Cronbach's Alpha and Number of Items for the Control Variables Scales

Scale	Initial Alpha	Initial Number of Items	Final Alpha if one item removed
Self-efficacy in Computer Gaming	.773	5	.789[1]
General Amount of Invested Mental Effort	.689	6	.772[1]
Perception of Computer Game Difficulty	.237	3	.295[2]

[1] The final scales were calculated as the mean of the items after removing the item causing the Cronbach's Alpha to be lower.
[2] This scale was not use in any further analysis.

For the dependent variable *Situational Cognitive Engagement*, the lowest Alpha (α=.447) occurred in the scale concerning the learning Task0 (see Table 30). The highest Alpha (α=.852) occurred in the scale concerning the learning Taks2. In the learning Task0, Task6 and Task7, one item was removed to improve the overall Cronbach's Alpha of the scales.

Table 29: Cronbach's Alpha for Amount of Invested Mental Effort (AIME)

Scale	Initial Alpha	Initial Number of Items	Final Alpha if item removed
Amount of Invested Mental Effort: Tasks	.772	6	.894[1]
Amount of Invested Mental Effort: Simulation	.486	6	.698[1]
Amount of Invested Mental Effort: Statistics	.806	6	.874

[1] The final scales were calculated as the mean of the items after removing the item causing the Cronbach's Alpha to be lower.

Table 30: Cronbach's Alpha for Situational Cognitive Engagement (SCENG)

Scale	Initial Alpha	Initial Number of Items	Final Alpha if item removed
Situational Cognitive Engagement: Task0	.447	4	.606[1]
Situational Cognitive Engagement: Task1	.729	4	.757
Situational Cognitive Engagement: Task2	.823	4	.852
Situational Cognitive Engagement: Task3	.798	4	.800
Situational Cognitive Engagement: Task4	.681	4	.669
Situational Cognitive Engagement: Task5	.768	4	.793
Situational Cognitive Engagement: Task6	.593	4	.640[1]
Situational Cognitive Engagement: Task7	.661	4	.796[1]

Table 31 summarizes the items removed in each scale together with the consequent improvement on the Alpha values.

Table 31: Items Removed from the Control and Dependent Variables Scales and Cronbach's Alpha Improvement

Scale	Item removed	Alpha's improvement
Self-efficacy in Computer Gaming	"Ich kann bei Computerspielen standing für eine lange Zeit spielen"	+.016
General Amount of Invested Mental Effort	"Wie sehr, denken Sie, kann Inhalt von Videospielen nützlich sein, um eine tiefere Lernerfahrung zu haben?"	+.083

Amount of Invested Mental Effort: Simulation	"Wie schwierig war es, die Wirtschaftssimulation zu verstehen?"	+.212
Amount of Invested Mental Effort: Tasks	"Wie schwierig war es, die physikalischen Wissensaufgaben zu verstehen?"	+.122
Situational Cognitive Engagement: Task0	"Ich habe mich sehr stark, in dieser Wissensaufgabe eingearbeitet"	+.159
Situational Cognitive Engagement: Task6	"Ich habe mich sehr bemüht, diese Wissensaufgabe zu lösen"	+.047
Situational Cognitive Engagement: Task7	"Ich habe mich sehr bemüht, diese Wissensaufgabe zu lösen"	+.135

Concerning the reliability of the *recall* pre and posttests, two raters (i.e., the researcher and an assistant), blind to the experimental condition, scored all the pretests and posttests of the 42 participants of the study based on a scoring rubric (see Appendix L). This produced two separate matrices with 42 rows representing each participant and 12 columns representing each of the questions. This yielded 504 (i.e., 42x12) scoring comparisons in the pretest and 504 scoring comparison in the posttest, for a total of 1008 comparisons. A value of 1 was given when the two raters agreed on the score obtained of a participant in a particular item. The process was repeated for each participant in all the items. The two raters agreed on 949 of the 1008 scoring comparisons, which was 94% agreement (i.e., 949/1008=.94).

The Game Engagement Questionnaire, developed by Chen et al. (2011) was adapted and translated into German. The questionnaire consisted of 11 items and showed a high Alpha (α=.816). As the scale was derived by using items reflecting different aspects of engagement, the interest here was on those items reflecting the emotional experiences of being involved and having fun while playing the game. In order to accomplish that, a factor analysis was conducted on the 11 items. Table 32 shows the items according to their factor loading (from highest to lowest) and grouped in the factor they belong according to the factorial solution.

Table 32: Summary of Items and Factor Loadings for Varimax Orthogonal Three-Factor Solution for the Game Engagement Questionnaire (N = 42)

	Factor loading			
Item	1	2	3	Communality
GEQ_Invol_Das Spiel hat mich voll und ganz eingenommen	**.832**	.406	-.011	.857
GEQ_Invol_Ich fühlte so stark in das Spielerlebnis hineingezogen. dass ich die Zeit vergessen habe	**.827**	.247	-.228	.797
GEQ_Invol_Ich fühlte mich, als ob ich innerhalb der Spielwelt war	**.827**	-.009	.075	.690
GEQ_Fun_Inhaltlich und thematisch hat mir das Spiel viel Spaß gemacht	**.697**	.514	-.069	.755
GEQ_Invol_Ich fühlte mich total in das Spielerlebnis hineingezogen	**.692**	.452	-.110	.696
GEQ_game_Die Interfacegestaltung (Graphik und Toneffekte) des Spiels fand ich sehr gut	**.684**	-.053	.313	.569
GEQ_game_Ich habe das Gefühl. gut darin zu sein. das Spiel zu steuern	.312	**.766**	.045	.686
GEQ_Ich konnte stets antizipieren. was. in Reaktion auf die von mir gestartete Aktionen als nächsten passieren würde	.143	**.723**	-.061	.546
GEQ_game_Die Wissensaufgabe bereichern die Geschichten/den Plot ("storyline") erheblich	.030	**.603**	.497	.611
GEQ_game_Die Figuren bereichern die Geschichte/den Plot ("storyline") erheblich	.010	.331	**.860**	.849
GEQ_game_Ich habe innerhalb des Spiels eine kurze Verzörgerung zwischen meinen Aktionnen und den erwarteten Ergebnisse bemerkt	-.010	-.298	**.614**	.466

Note. Boldface indicates highest factor loadings.

The first factor is composed of items reflecting the experience of involvement and enjoyment, representing for this study the emotional component of the concept of engagement. On the other hand, the item referring to the quality of the interface seems conceptually different from the idea of emotional engagement. The subsequent Alpha showed that removing this item improved the value of alpha from α=.892 to α=.902. Therefore the scale was built with the first five items with the highest load on factor number one. The second factor referred to the degree of perceived control while playing the game and showed a modest Alpha of .612 and the third factor an even lower one (α=.442). These results are to be expected given the sensitivity of the Cronbach's Alpha to the number of items. Therefore, these two factors were not included in any further analysis.

Finally, in order to assess the reliability of the coding frame used to analyze the interviews for capturing participants' cognitive processing, the *intercoder agreement* function of MAXQDA was employed (Maietta, 2008). Thirty five interviews with a total of 75.467 words were analyzed. The coding frame (see Section 4.4.2) resulted in a total of 3831 coded sentences (see Appendix M). From these codes 36% were related to the cognitive processes while solving the learning tasks, 9.79% represented further reflections from the participants, 10.28% corresponded to participants goals for the game, and finally, 43.93% represented not coded sentences. Therefore, the 3831 codes were the input for the MAXQDA function *segment agreement in percentage.* This function checks the extension of a code (i.e., how much text was coded in a particular code) and then compares it with the extension of the same code as coded by the second coder. To achieve this comparison, the function offers an option that allows the researcher to determine the percentage of text agreement to be considered as correlating. Normally, this percentage is set to 90%. This percentage was used in this case for estimating the inter-rater reliability. The average reliability obtained was 69%.

5.2. General Descriptive Statistics

Descriptive statistics concerning the variables measures of location and shape of their distribution is provided. For the case of the control variables, test of group differences are also conducted here to evaluate the initial comparability of the two groups studied. The dependent variables are divided into questionnaire data (Section 5.2.2), next the data obtained from the interviews conducted (Section 5.2.2.1).and, finally, the data obtained through the eye tracking (Section 5.2.3).

5.2.1. Control Variables

The control variables consisted of General Amount of Mental Effort, Self-efficacy in Computer Gaming Scale, and prior knowledge (Recall pretest). Descriptive statistics together with graphic and non-graphic methods to examine the distribution of the variables are provided. Finally, the analysis for the control variables were conducted using a two-tailed test of significance and the effect size was reported using Cohen's d calculated with Murphy and Myors' formula:

Equation 2: Murphy and Myors' (2004) Effect Size Formula

$$d = \frac{2t}{\sqrt{df\ err}}$$

Table 33 shows the measures of location and variation together with their distributions for the total sample. Table 34 provides descriptive statistics and test of normality separated for each experimental condition.

Table 33: Media, Standard Deviation and Sample Size for the Control Variables

	Total Sample		
Control variables	M	SD	*n*
General Amount of Invested Mental Effort	2.94	.66	40
Self-efficacy in Computer Gaming	3.67	.82	40
Recall Pretest	3.29	3.27	42

Some of the control variables present slightly departures from normality as shown by their skewness and normality test. A graphical approach to understand the distribution of these variables comes next.

Table 34: Descriptive Statistics and Tests of Normal Distribution for the Control Variables by Experimental Condition

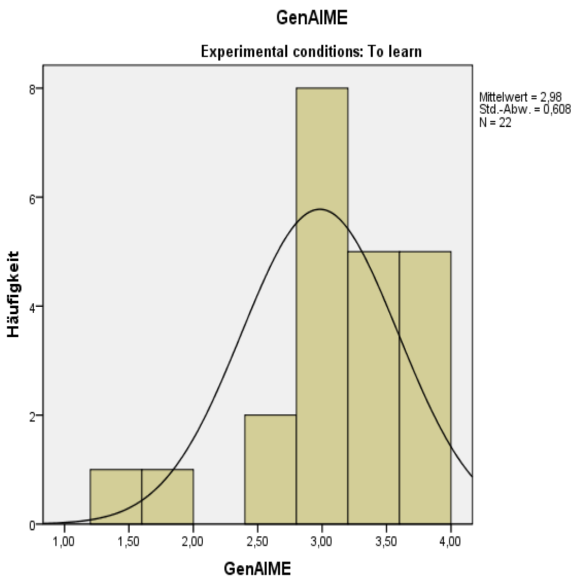
Experimental Condition	M (SD)	Median	Skewness (SE)	Kolmogorov-Smirnov	Shapiro-Wilk
To Learn					
General Amount of Invested Mental Effort	2.97 (.63)	3	-1.07 (.512)	.195*	.895*
Self-efficacy in Computer Gaming	3.67 (.87)	4	-.63 (.512)	.194*	.949
Recall Pretest	3.66 (2.61)	3.7	.279 (.512)	.089	.944
For Fun					
General Amount of Invested Mental Effort	2.86 (.78)	2.9	-.58 (.564)	.156	.953
Self-efficacy in Computer Gaming	3.78 (.77)	4	-.18 (.564)	.174	.946
Recall Pretest	3.50 (2.84)	3.85	.75 (.564)	.133	.949

Note. * Statistically significant (Alpha < .05).

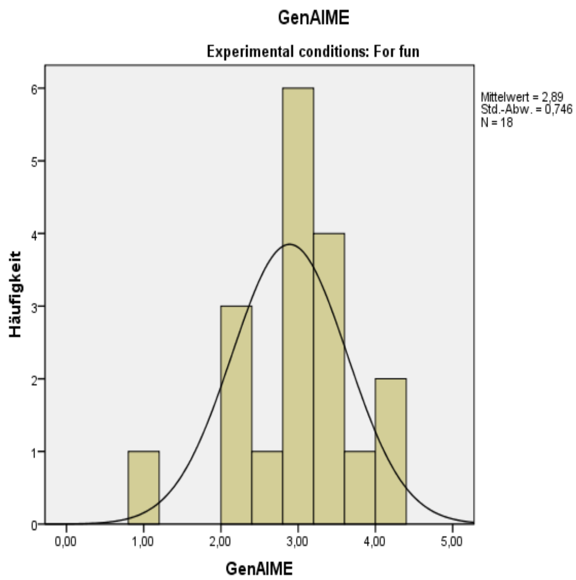
General Amount of Invested Mental Effort. Frequencies suggested a slightly departure from normality in the condition to learn (Figure 39a) and normal distribution in the condition for fun (Figure 39b) with presence of outliers (Figure 39c). The mean score for the total sample was 2.94 (SD = .66) out of 5 points. Participants in the condition to learn obtained a score of 2.97 (SD = .63) and participants in the condition for fun a score of 2.86 (SD = .78). No statistical difference on the General AIME was found, $t(38) = .434$, $p = .66$, Cohen's $d = 0.14$. Bootstrap-t method yielded a 95% CI [-0.22, .62].

Figure 39: Frequencies (a, b) and Boxplots (c) of General Amount of Invested Mental Effort by Experimental Condition

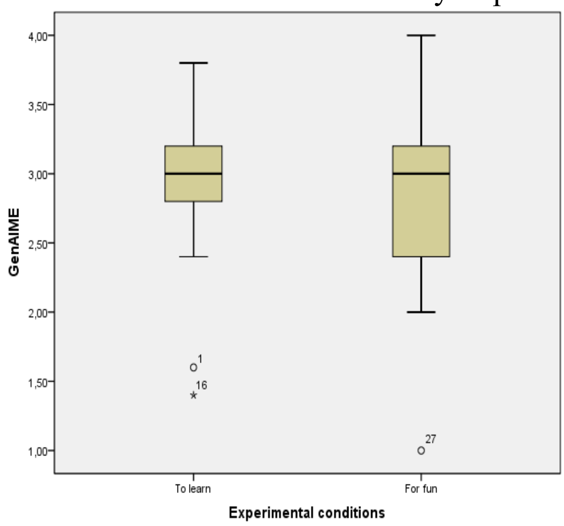
(a) Frequencies General Amount of Invested Mental in Condition to Learn



(b) Frequencies General Amount of Invested Mental in Condition for Fun



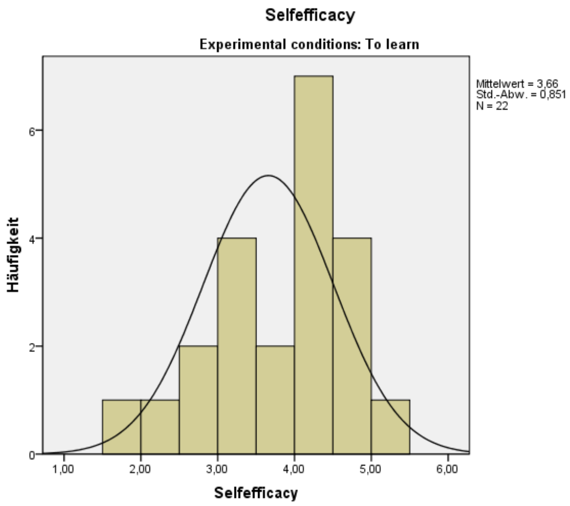
(c) Boxplot General Amount of Invested Mental by Experimental Condition



Self-efficacy in Computer Gaming. Frequencies suggested a normal distribution in both conditions (Figure 40ab) without presence of outliers (Figure 40c). However, Table 34 showed a significant KS test. The mean score for the total sample was 3.67 (SD = .82) out of 5 points. Participants in the condition to learn obtained a score of 3.67 (SD = .87) and participants in the condition for fun a score of 3.78 (SD = .77). No statistical difference on the Self-efficacy in Computer Gaming was found, $t(38) = .133$, $p = .895$, Cohen's $d = .04$. Likewise the bootstrap-t method yielded a 95% CI [-0.59, .73].

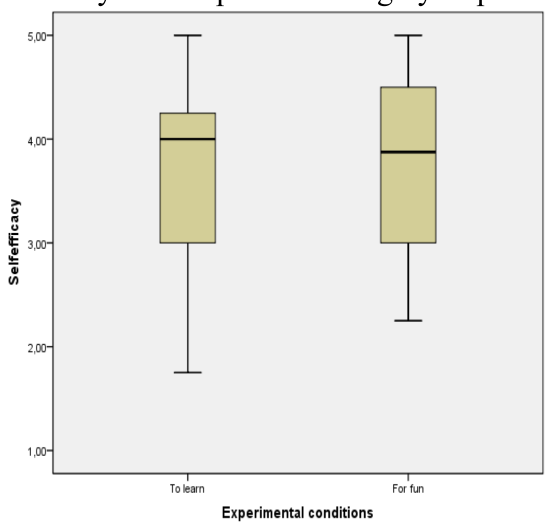
Figure 40: Frequencies (a, b) and Boxplots (c) of Self-efficacy in Computer Gaming by Experimental Condition

(a) Frequencies Self-efficacy in Computer Gaming in Condition to Learn



(b) Frequencies Self-efficacy in Computer Gaming in Condition for Fun

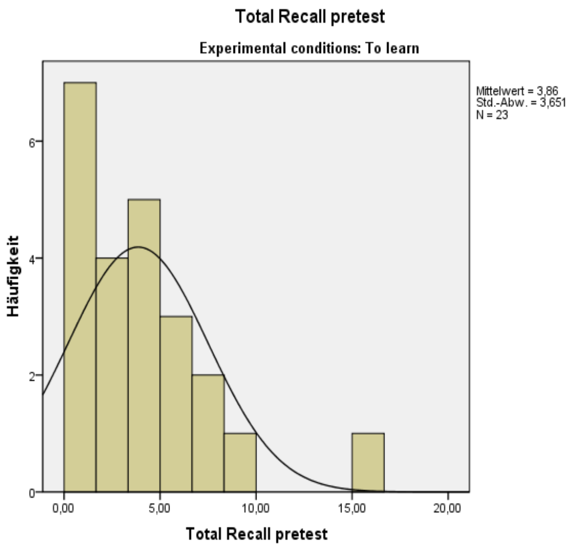
(c) Boxplot Self-efficacy in Computer Gaming by Experimental Condition



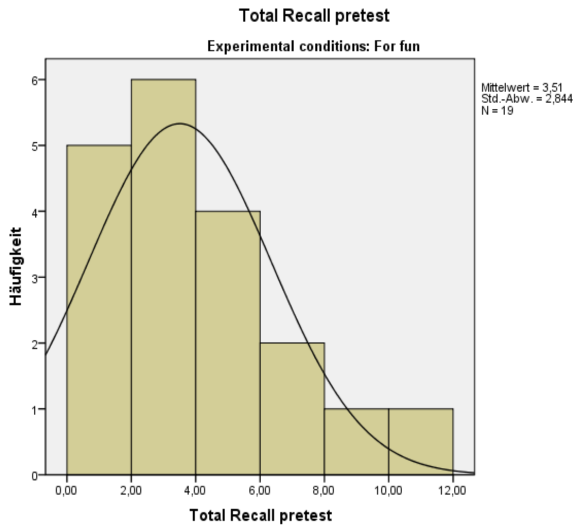
Recall Pretest. Frequencies suggested a slightly departure from normality in both conditions (Figure 41ab) with the presence of one outlier (Figure 411c). The mean score in the Recall Pretest for the total sample was 3.29 (SD = 2.61) out of 25.4 points. Participants in the condition to learn obtained a score of 3.66 (SD = 2.61) and participants in the condition for fun a score of 3.50 (SD = 2.84). The frequencies show a small skewness in the data in both conditions (see Figure 411). No statistical difference on prior knowledge was found, $t(40) = .342$, $p = .946$, Cohen's $d = .11$. Likewise the bootstrap-t method yielded a 95% CI [-2.09, 1.9].

Figure 41: Frequencies (a, b) and Boxplots (c) of Recall Pretest by Experimental Condition

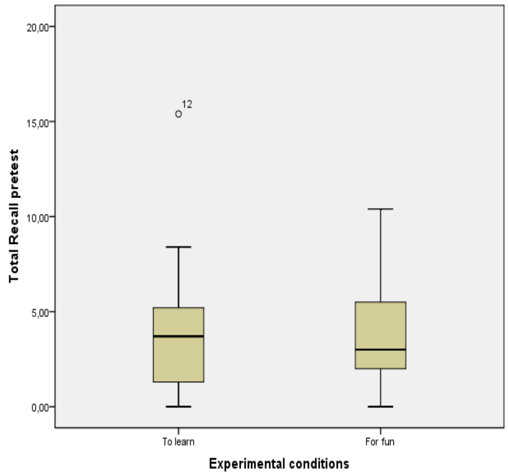
(a) Frequencies Recall Pretest in Condition to Learn



(b) Frequencies Recall Pretest in Condition for Fun



(c) Boxplot Recall Pretest by Experimental Condition



In summary, the distributions of the control variables concerning Recall Pretest showed a normal distribution in both groups. In the condition for fun all variables showed a normal distribution. On the other hand, in the condition to learn General Amount of Invested Mental Effort, and Self-efficacy showed slightly departures from normality. Concerning the presence of outliers, data showed the presence of outliers only on General Amount of Invested Mental Effort and Recall Pretest. These results warranted the use of complementary bootstrap methods for the hypothesis testing.

As for the test of difference between conditions, the control variables - General Amount of Invested Mental Effort, Self-efficacy in Computer Gaming, and Recall Pretest – did not show a statistical difference between the condition to learn and the condition for fun. Therefore, in terms of these variables, both groups are comparable.

5.2.2. Dependent Variables

The dependent variables of interest consisted of Amount of Invested Mental Effort with the Simulation, the Tasks and the Statistics of the game (i.e., AIME Simulation, AIME Tasks, and AIME Statistics), Situational Cognitive Engagement (SCENG), and Emotional Engagement. Behavioral Engagement' descriptive statistics are provided in section 5.2.3. Inferential statistics are reported in the section 5.3 Hypothesis Testing.

Table 35 shows the measures of location and variation of the dependent variables together with their distributions for the total sample. Table 36 provides descriptive statistics and test of normality separated for each experimental condition.

Table 35: Media, Standard Deviation and Sample Size for the Dependent Variables

	Total Sample		
Control variables	M	SD	*n*
Amount of Invested Mental Effort			
Simulation	3.58	.57	42
Tasks	3.46	.85	42
Statistics	2.63	.64	42
Situational Cognitive Engagement	2.84	.69	42
Emotional Engagement	4.05	1.36	38
Recall Posttest	7.80	4.51	42

Table 36: Descriptive Statistics and Tests of Normal Distribution for the Dependent Variables (Questionnaire Data) by Experimental Condition

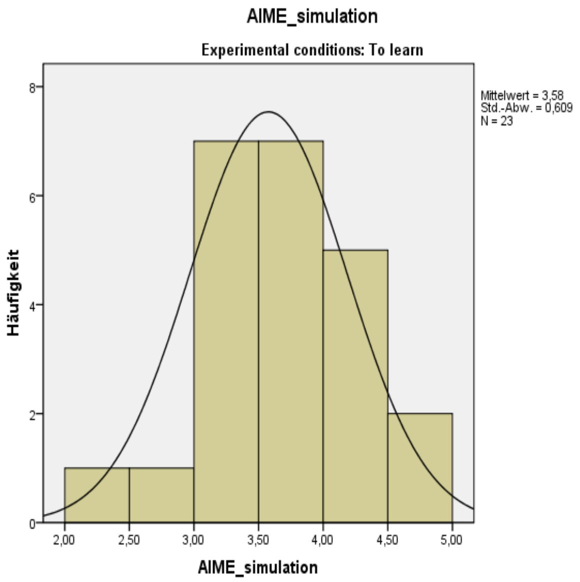
Experimental Condition	M (SD)	Median	Skewness (SE)	Kolmogorov-Smirnov	Shapiro-Wilk
To Learn					
Amount of Invested Mental Effort					
Simulation	3.61 (.48)	3.6	-.19 (.512)	.14	.97
Tasks	3.63 (.60)	3.6	-.006 (.512)	.07	.97
Statistics	2.59 (.70)	2.58	.95 (.512)	*.18*	*.91*
Situational Cognitive Engagement	2.97 (.55)	2.96	-.79 (.512)	*.18*	*.91*
Emotional Engagement	4.26 (1.45)	4.9	-.80 (.512)	.19*	.90*
Recall Posttest	8.01 (3.92)	8	-.008 (.512)	.11	.95
For Fun					
Amount of Invested Mental Effort					
Simulation	3.5 (.52)	3.6	-.55 (.564)	.14	.93
Tasks	3.2 (.96)	3.1	-.27 (.564)	.11	.96
Statistics	2.71 (.56)	2.55	.30 (.564)	.15	.94
Situational Cognitive Engagement	2.59 (.77)	2.4	-.07 (.584)	.23*	*.90*
Emotional Engagement	3.91 (1.05)	3.7	1.25 (.564)	.20	.88*
Recall Posttest	9.09 (5.05)	9.9	-.10 (.564)	.13	.94

Some of the dependent variables presented slightly departures from normality as shown by their normality test. A graphical approach to understand the distribution of these variables comes next.

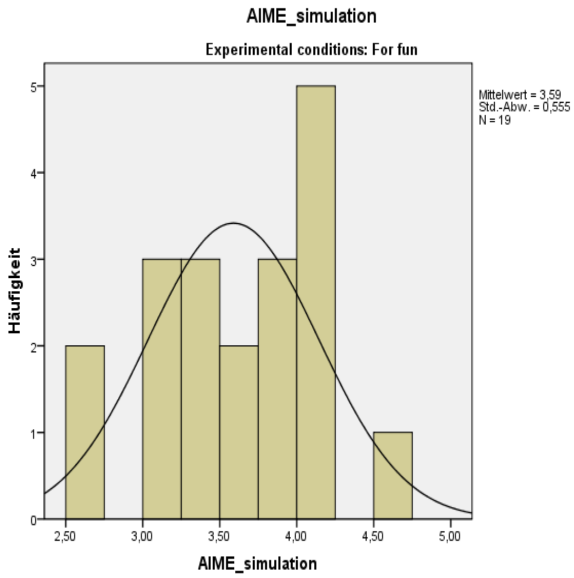
Amount of Invested Mental Effort on the Simulation. Frequencies suggested a normal distribution in both conditions (Figure 42ab) without presence of outliers (Figure 42c). The mean score for the total sample was 3.58 (SD = .57) out of 5 points. Participants in the condition to learn obtained a score of 3.61 (SD = .48) and participants in the condition for fun a score of 3.5 (SD = .55).

Figure 42: Frequencies (a, b) and Boxplots (c) of Amount of Invested Mental Effort on the Simulation by Experimental Condition

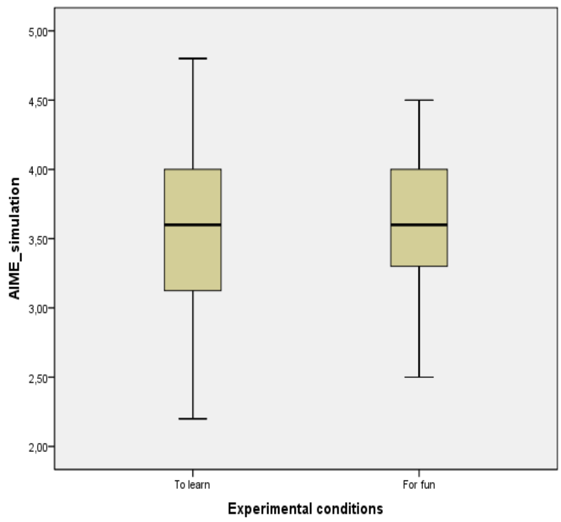
(a) Frequencies Amount of Invested Mental Effort on Simulation in Condition to Learn



(b) Frequencies Amount of Invested Mental Effort on Simulation in Condition for Fun



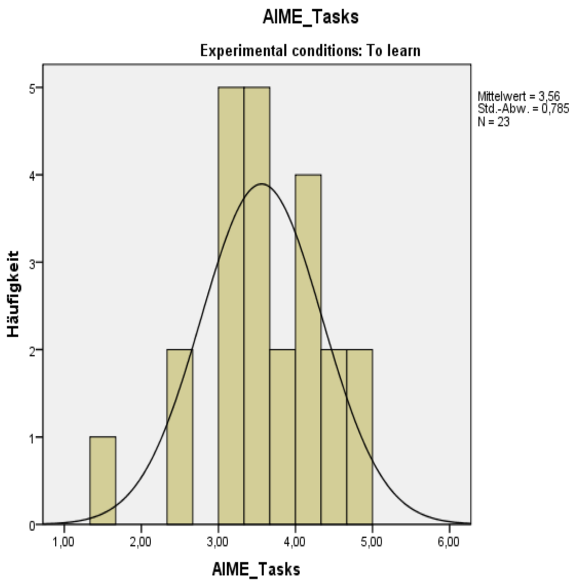
(c) Boxplot Amount of Invested Mental Effort on Simulation by Experimental Condition



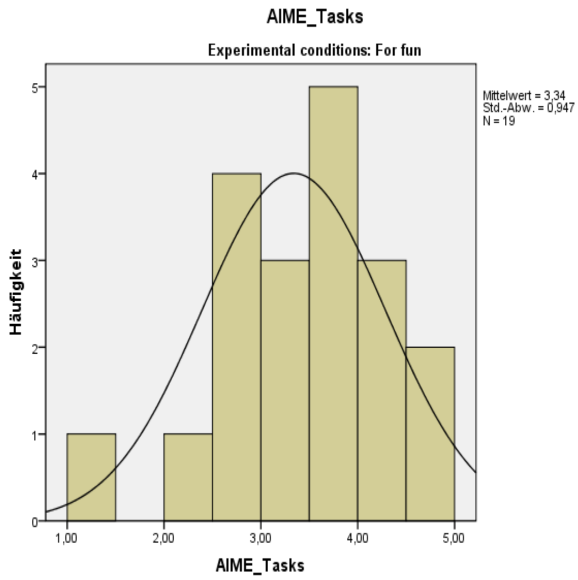
Amount of Invested Mental Effort on the Tasks. Frequencies suggested a normal distribution in both conditions (Figure 43ab) without presence of outliers (Figure 43c). The mean score for the total sample was 3.46 (SD = .85) out of 5 points. Participants in the condition to learn obtained a score of 3.63 (SD = .60) and participants in the condition for fun a score of 3.2 (SD = .96).

Figure 43: Frequencies (a, b) and Boxplots (c) of Amount of Invested Mental Effort on the Tasks by Experimental Condition

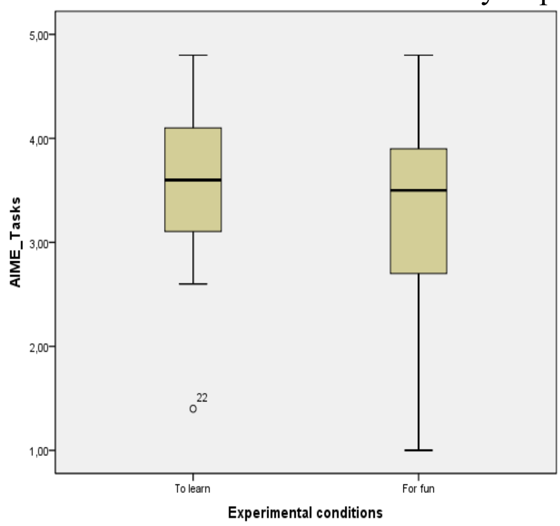
(a) Frequencies Amount of Invested Mental Effort on Tasks in Condition to Learn



(b) Frequencies Amount of Invested Mental Effort on Tasks in Condition for Fun



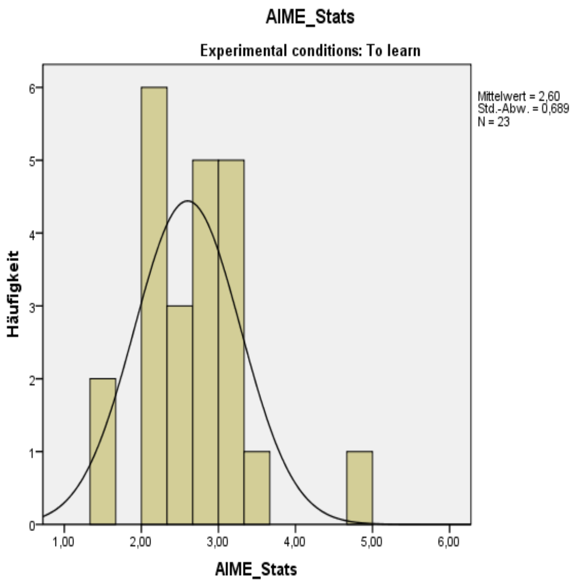
(c) Boxplot Amount of Invested Mental Effort on Tasks by Experimental Condition



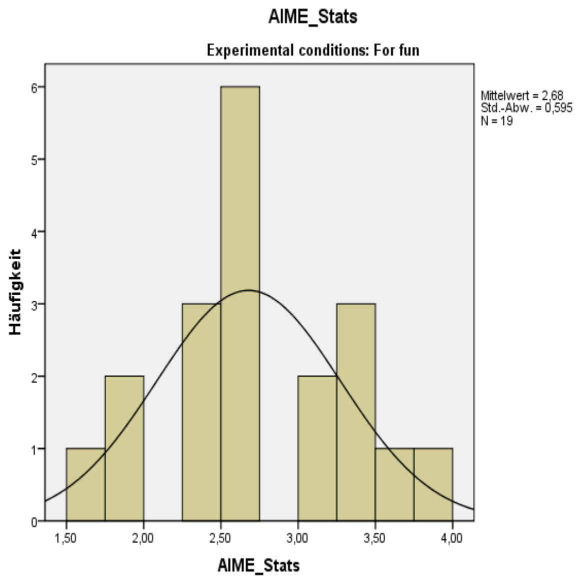
Amount of Invested Mental Effort on the Statistics. Frequencies suggested a normal distribution in both conditions (Figure 44ab) with the presence of one outlier (Figure 44c). The mean score for the total sample was 2.63 (SD = .64) out of 5 points. Participants in the condition to learn obtained a score of 2.59 (SD = .70) and participants in the condition for fun a score of 2.71 (SD = .56).

Figure 44: Frequencies (a, b) and Boxplots (c) of Amount of Invested Mental Effort on the Statistics by Experimental Condition

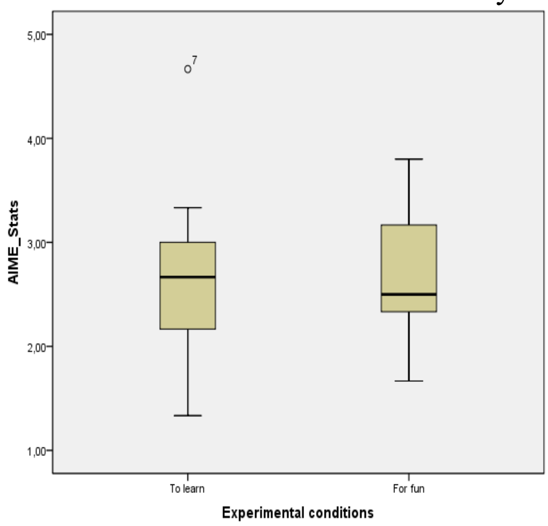
(a) Frequencies Amount of Invested Mental Effort on Statistics in Condition to Learn



(b) Frequencies Amount of Invested Mental Effort on Statistics in Condition for Fun



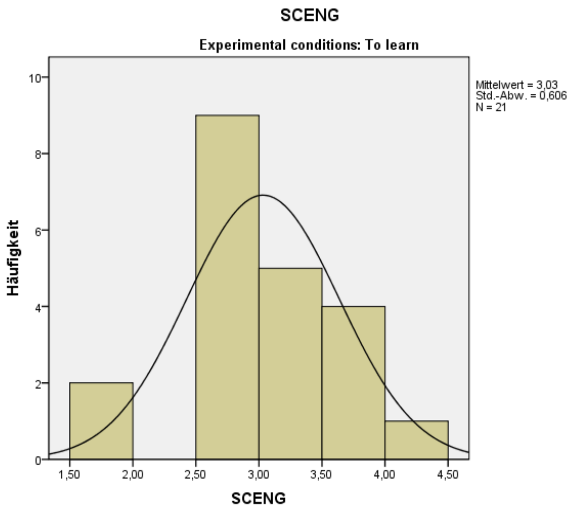
(c) Boxplot Amount of Invested Mental Effort on Statistics by Experimental Condition



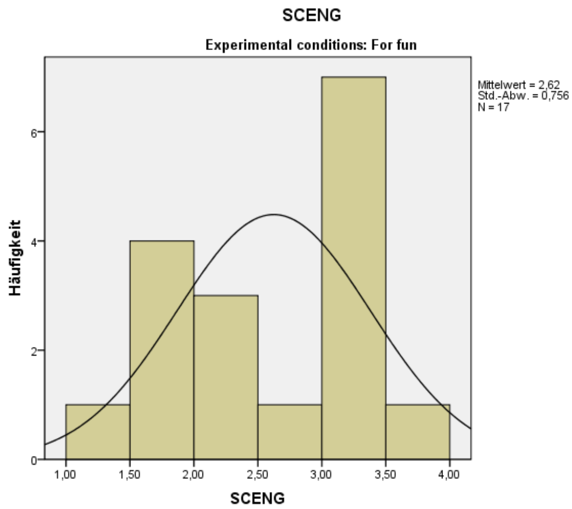
Situational Cognitive Engagement (SCENG). Frequencies suggested a slight departure from normality in the condition to learn (Figure 45ab) with the presence of outliers (Figure 45c). The mean score for the total sample was 2.84 (SD = .69) out 5 points. Participants in the condition to learn obtained a score of 2.97 (SD = .55) and participants in the condition for fun a score of 2.59 (SD = .77). (Appendix N shows the SCENG values for each task).

Figure 45: Frequencies (a, b) and Boxplots (c) of Situational Cognitive Engagement by Experimental Condition

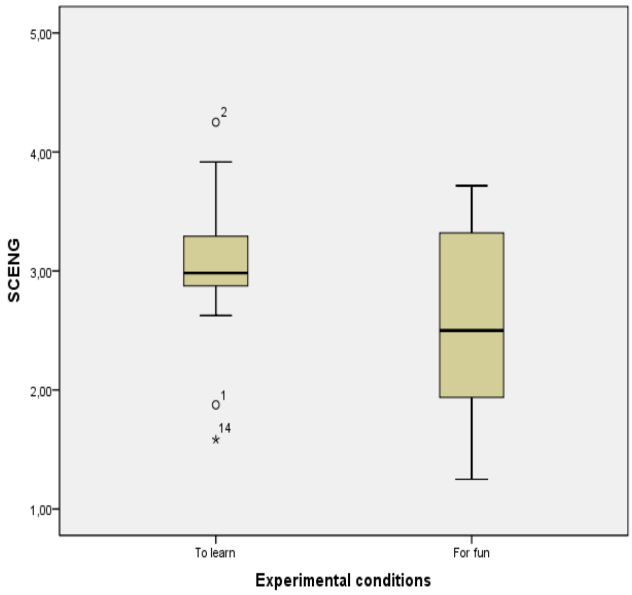
(a) Frequencies Situational Cognitive Engagement in Condition to Learn



(b) Frequencies Situational Cognitive Engagement in Condition for Fun



(c) Boxplot Situational Cognitive Engagement by Experimental Condition

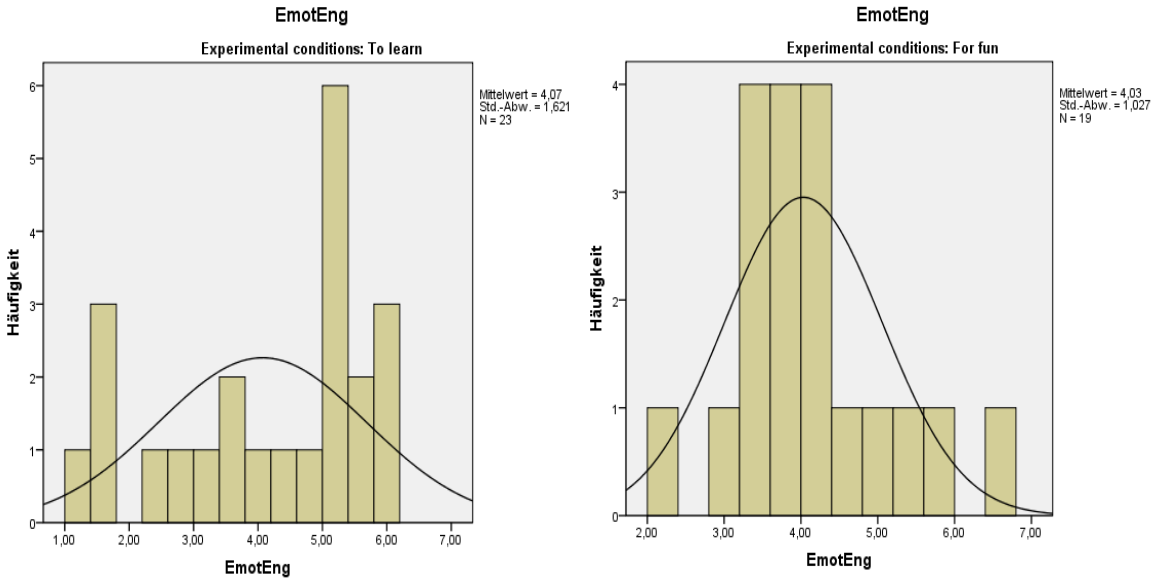


Emotional Engagement. Frequencies suggested slight departures from normality in both conditions (Figure 46ab) with the presence of outliers (Figure 46c). The mean score for the total sample was 4.05 (SD = 1.36) out of 7 points. Participants in the condition to learn obtained a score of 4.26 (SD = 1.45) and participants in the condition for fun a score of 3.91 (SD = 1.05).

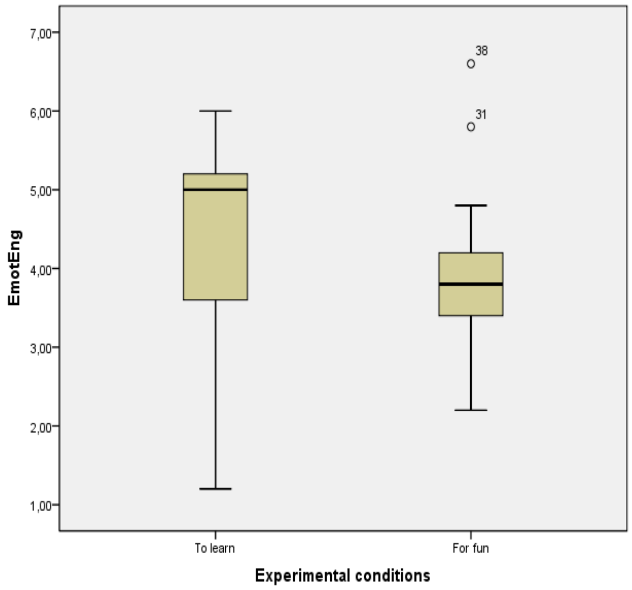
Figure 46: Frequencies (a, b) and Boxplots (c) of Emotional Engagement by Experimental Condition

(a) Frequencies Emotional Engagement in Condition to Learn

(b) Frequencies Emotional Engagement in Condition for Fun



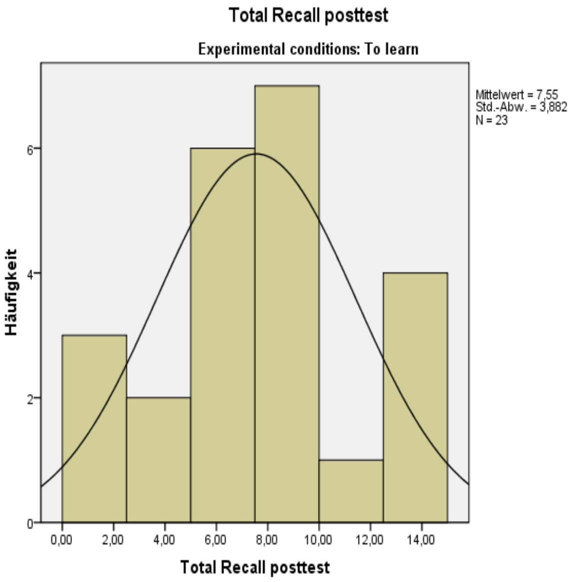
(c) Boxplot Emotional Engagement by Experimental Condition



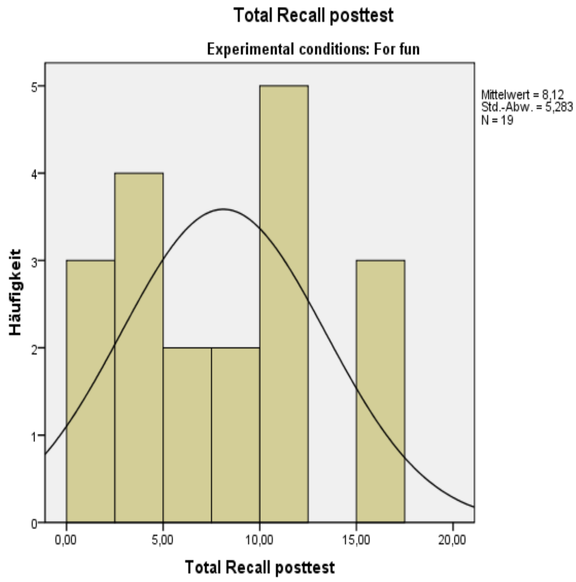
Recall posttest. Frequencies suggested a normal distribution in both conditions (Figure 47ab) without the presence of outliers (Figure 47c). The mean score for the total sample was 7.80 (SD = 4.51) out of 25.4 points. Participants in the condition to learn obtained a score of 8.01 (SD = 3.92) and participants in the condition for fun a score of 9.09 (SD = 5.05).

Figure 47: Frequencies (a, b) and Boxplots (c) of Recall Posttest by Experimental Condition

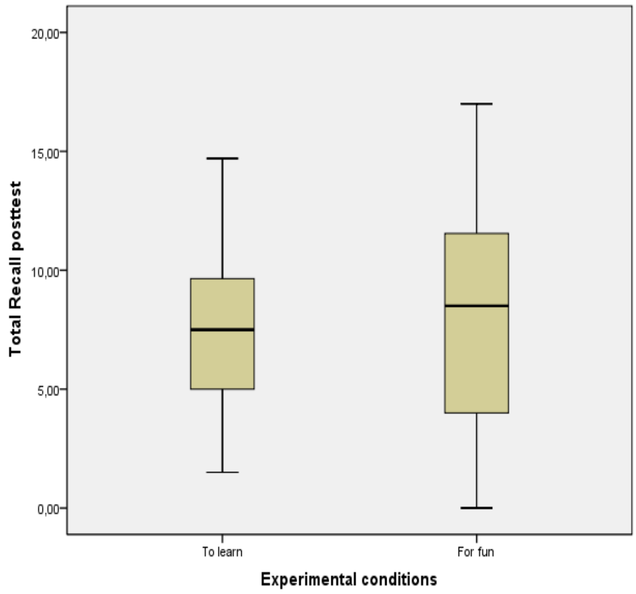
(a) Frequencies Recall Posttest in Condition to Learn



(b) Frequencies Recall Posttest in Condition for Fun



(c) Boxplot Recall Posttest by Experimental Condition



For the measures calculated from the eye tracking data, the following descriptive statistics were obtained.

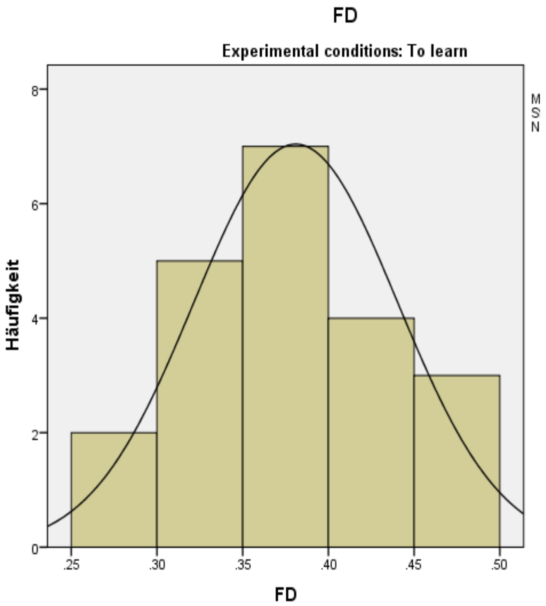
Table 37: Descriptive Statistics and Tests of Normal Distribution for the Dependent Variables (Eye Tracking Data) by Experimental Condition

Experimental Condition	M (SD)	Median	Skewness (SE)	Kolmogorov-Smirnov	Shapiro-Wilk
To Learn					
Fixation Durations	.38(.01)	.39	-.380 (.501)	.165	.961
Total Dwell Time	9.24(.99)	8.86	.656 (.501)	.124	.929
Reading Depth	.07(.006)	.07	.127 (.501)	.164	.956
For Fun					
Fixation Durations	.34(.01)	.34	-.101 (.524)	.109	.889
Total Dwell Time	8.52(1.12)	7.35	.768 (.524)	.170	.416
Reading Depth	.06(.006)	.064	-.205 (.524)	.117	.830

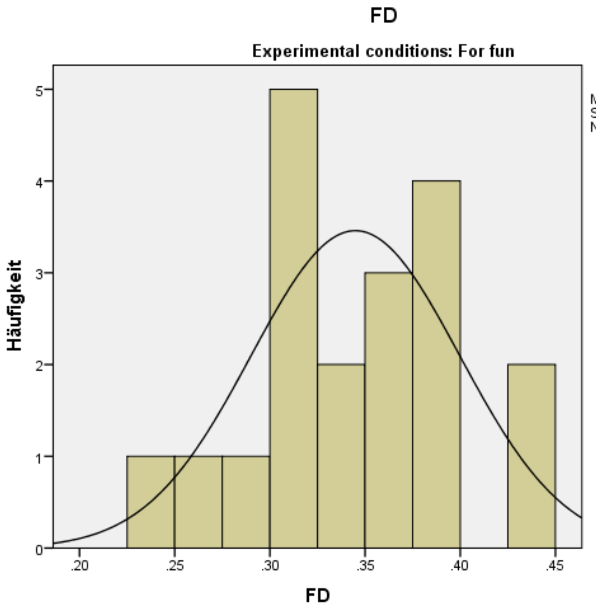
Fixation Durations. Frequencies suggested a normal distribution in both conditions (Figure 48ab) without the presence of outliers (Figure 48c). Participants in the condition to learn obtained a score of .38 (SD = .01), while participants in the condition for fun obtained a score of .38 (SD = .06).

Figure 48: Frequencies (a,b) and Boxplots (c) of Fixation Durations by Experimental Condition

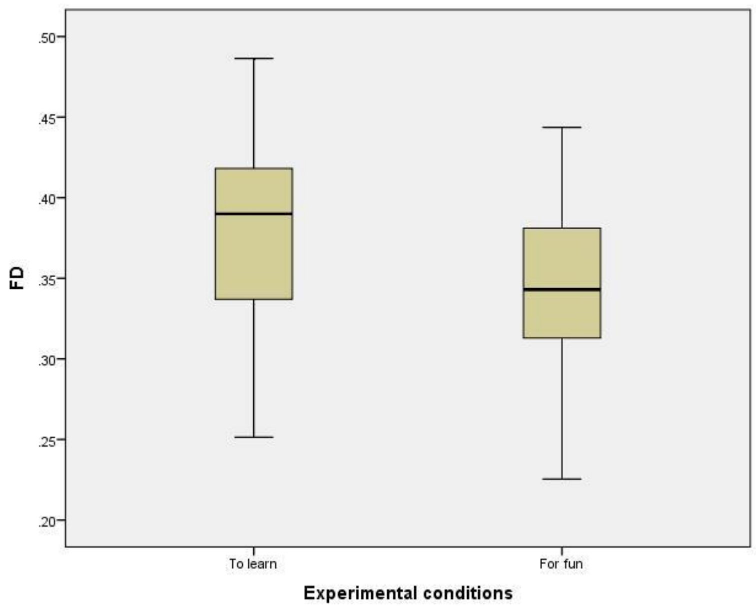
(a) Frequencies FD Cond. to Learn



(b) Frequencies FD Cond. For Fun



(c) Boxplot FD by Exp. Condition

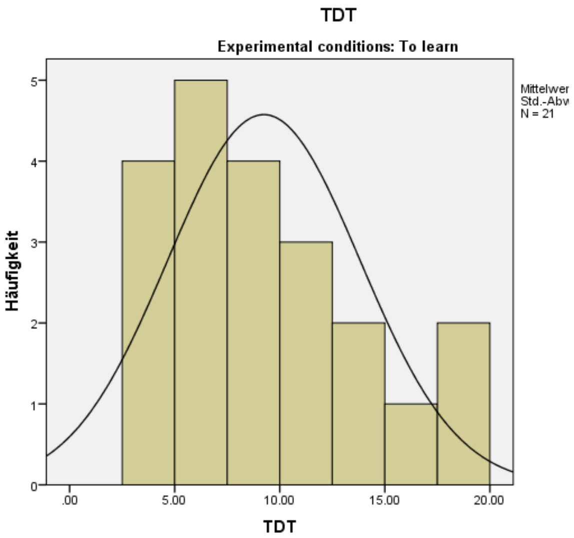


Note. FD= Fixation Durations; Cond.= Condition; Exp. = Experimental.

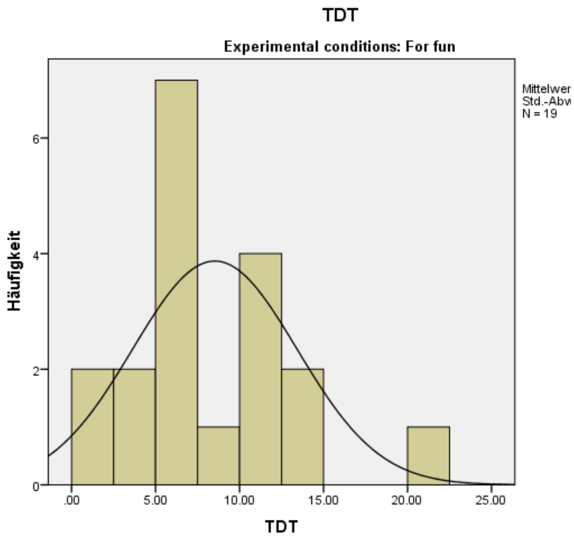
Total Dwell Time. Frequencies suggested a normal distribution in both conditions (Figure 49ab), and the presence of outliers (Figure 49c). Participants in the condition to learn obtained a score of 9.24 (SD = .99), while participants in the condition for fun obtained a score of 8.52 (SD = 1.12).

Figure 49: Frequencies (a, b) and Boxplots (c) of Total Dwell Time by Experimental Condition

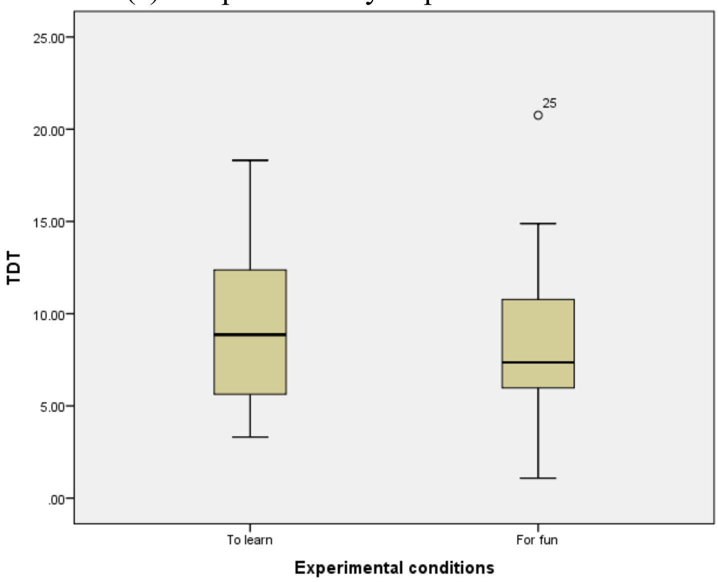
(a) Frequencies TDT Cond. to Learn



(b) Frequencies TDT Cond. For Fun



(e) Boxplot TDT by Exp. Condition

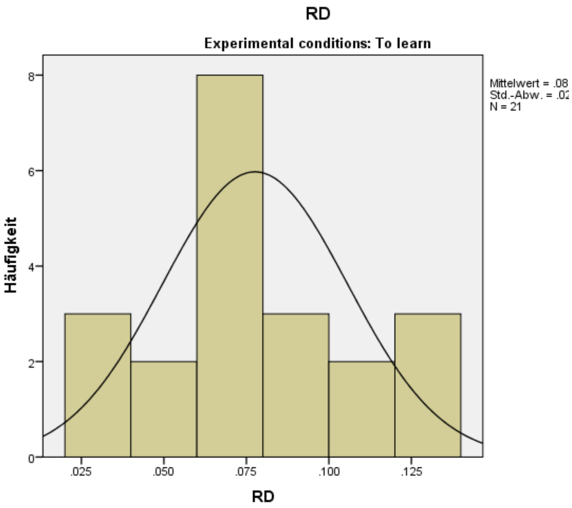


Note. TDT= Total Dwell Time; Cond.= Condition; Exp. = Experimental

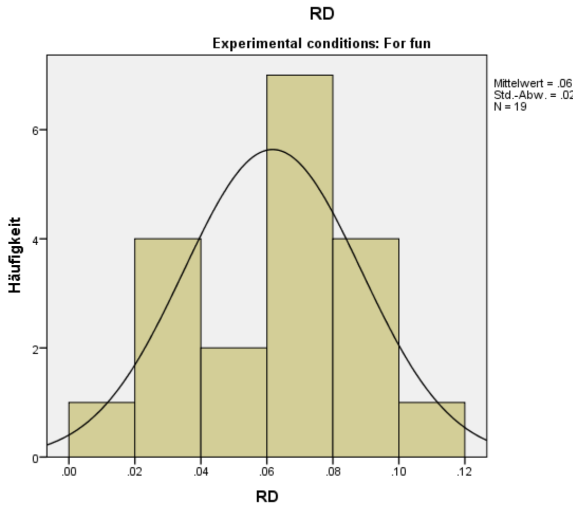
Reading Depth. Frequencies suggested a normal distribution in both conditions (Figure 50ab). No presence of outliers (Figure 50c). Participants in the condition to learn obtained a score of .07 (SD = .006), while participants in the condition for fun obtained a score of .06 (SD = .006).

Figure 50: Frequencies (a, b) and Boxplots (c) of Reading Depth by Experimental Condition

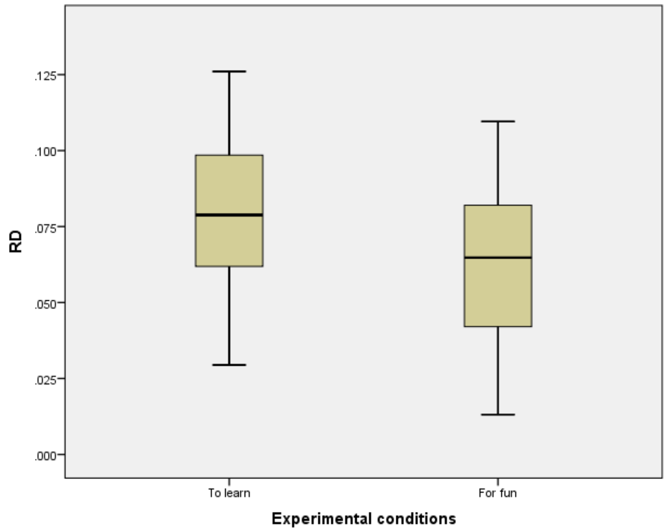
(a) Frequencies RD Cond. to Learn



(b) Frequencies RD Cond. For Fun



(e) Boxplot RD by Exp. Condition



Note. RD= Reading Depth; JM= Journal Mode; TM= Task Mode; Cond.= Condition; Exp. = Experimental

In summary, the distributions of the dependent variables concerning mental effort – Amount of Invested Mental Effort on the Simulation (AIME Simulation), Amount of Invested Mental Effort on the Tasks (AIME Tasks), Amount of Invested Mental Effort on the Statistics (AIME Statistics), and Situational Cognitive Engagement (SCENG) –

showed a normal distribution. Likewise, AIME Statistics and SCENG presented outliers. The Emotional Engagement and Recall Posttest showed slightly departures from normality with the presence of outliers. Concerning the dependent variables from the eye tracking data, all of them – Fixation Durations, Total Dwell Time and Reading Depth – showed a normal distribution for both groups. Total Dwell Time showed one outlier.

5.2.2.1. Interview Data

A total of 41 interviews were conducted. However due to an unfortunate technical problem, 6 interviews could not be retrieved from the recorder. The remaining 35 interviews were transcribed and imported to MAXQDA version 10. The average duration of the interview process – which included several questionnaire administrations – was 24 minutes (approx.). In the condition to learn 19 interviews were transcribed (M= 24 minutes, Range= 6 to 45 minutes). Similarly, in the condition for fun 14 interviews were transcribed (M= 24 minutes, Range= 7 to 40 minutes). The range depended upon several factors such as the number of tasks the participant solved and therefore the number of questionnaires and questions she had to answer about them. It also depended to some extent on individual difference in terms of how talkative participants were during the interview. Below the descriptive statistics for each processing strategy is provided:

Table 38: Media and Standard Deviation for the Acquisition and Transformation Processes by Experimental Condition

	Learn		Fun	
Processes	M	SD	M	SD
1. Acquisition				
Attention				
Attention (+)	.44	.23	.55	.59
Attention (-)	.26	.25	.39	.38
Monitoring				
General	.46	.56	.38	.36
Specific	.77	.44	1.03	1.03
2. Transformation				
Selectivity	.78	.57	.75	.56
Connecting				
General	.37	.33	.63	.34
Specific	.92	.59	.62	.49
Planning				
General	.3	.29	.29	.18
Specific	.05	.07	.03	.07
Planning (-)	.47	.49	.3	.38

After coding the interviews, the data was summarized as the frequency of statements representing acquisition and transformation processes. The procedure of these calculations was described earlier in section 4.4.2. Table 39 provides descriptive statistics and test of normality separated for each experimental condition.

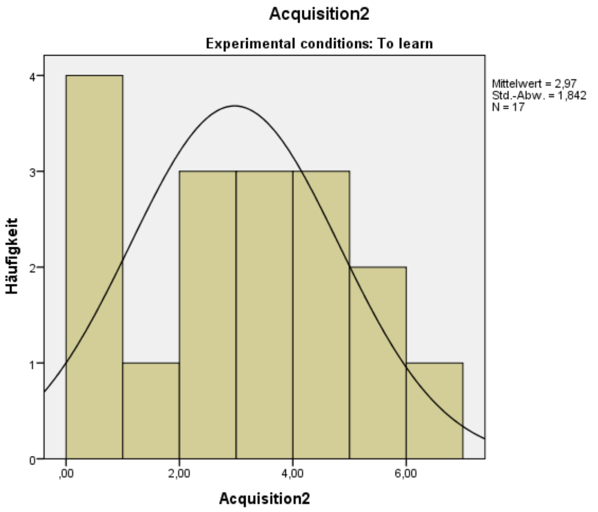
Table 39: Descriptive Statistics and Tests of Normal Distribution for Acquisition and Transformation Processes by Experimental Condition

Experimental Condition	M (SD)	Median	Skewness (SE)	Kolmogorov-Smirnov	Shapiro-Wilk
To Learn					
Acquisition	2.97 (1.84)	3	.2 (.550)	.123	.960
Transformation	4.21 (2.97)	4	.759 (.550)	.167	.911
For Fun					
Acquisition	2.72 (2.17)	2.2	.926 (.637)	.155	.923
Transformation	3.4 (1.47)	3.7	-.810 (.637)	.140	.937

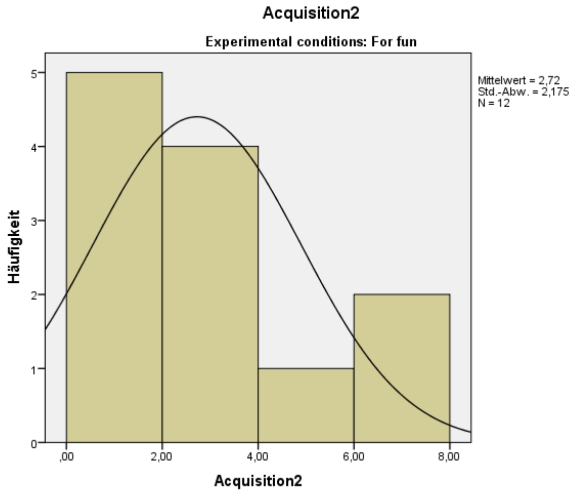
Acquisition. Frequencies suggested a slight departure from normality in the condition for fun (Figure 51b) without the presence of outliers (Figure 51c). Participants in the condition to learn obtained a score of 11.52 (SD = 1.73) and participants in the condition for fun a score of 12.75 (SD = 3.31).

Figure 51: Frequencies (a, b) and Boxplots (c) of Acquisition Processes by Experimental Condition

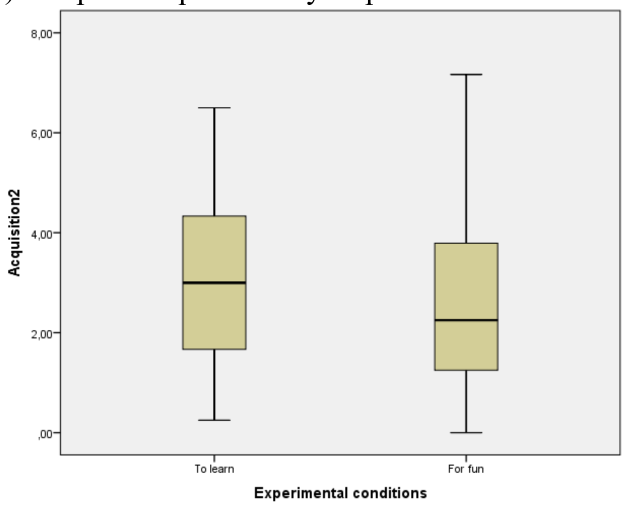
(a) Frequencies Acquisition in Condition to Learn



(b) Frequencies Acquisition in Condition for Fun



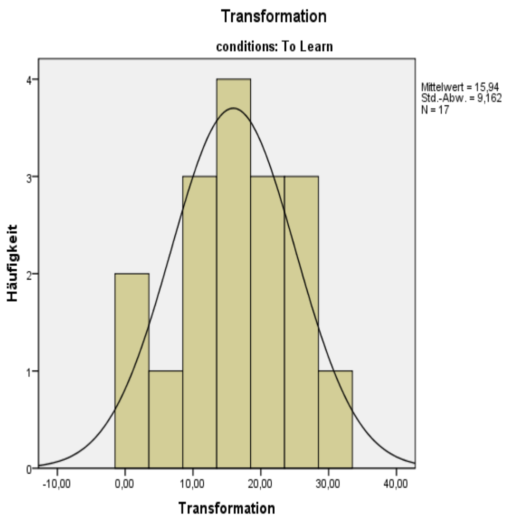
(c) Boxplot Acquisition by Experimental Condition



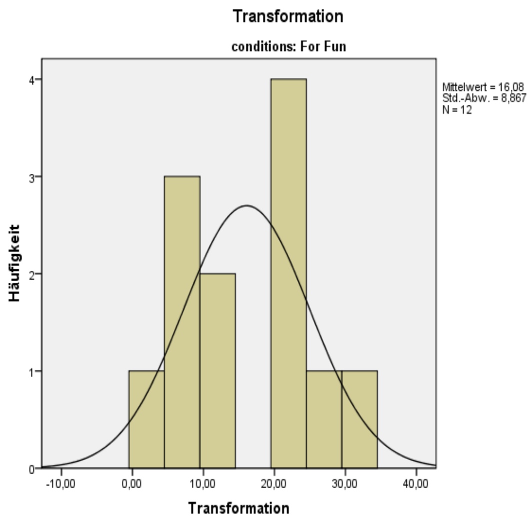
Transformation. Frequencies suggested a normal distribution in both conditions (Figure 52ab) with the presence of one outlier (Figure 52c). Participants in the condition to learn obtained a score of 15.94 (SD = 2.22) and participants in the condition for fun a score of 16.08 (2.55).

Figure 52: Frequencies (a, b) and Boxplots (c) of Transformation Processes by Experimental Condition

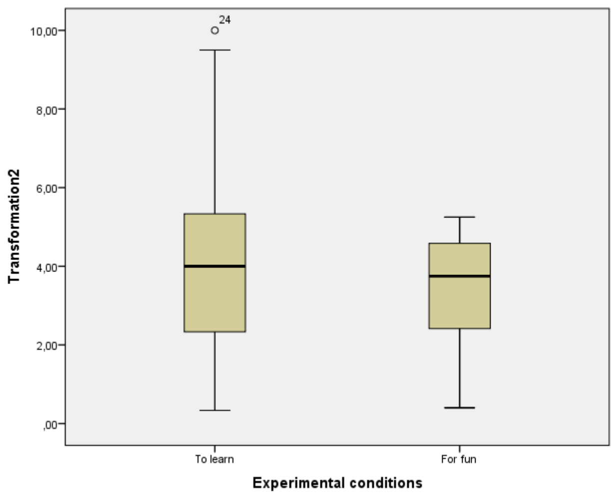
(a) Frequencies Transformation in Condition to Learn



(b) Frequencies Transformation in Condition for Fun



(c) Boxplot Transformation by Experimental Condition



In summary, the distribution of the dependent variable Acquisition showed a slightly departure from normality, while Transformation showed a normal distribution, though with the presence of one outlier. These results warranted the use of complementary bootstrap methods for hypothesis testing (see Section 5.3).

5.2.3. Eye Tracking Data

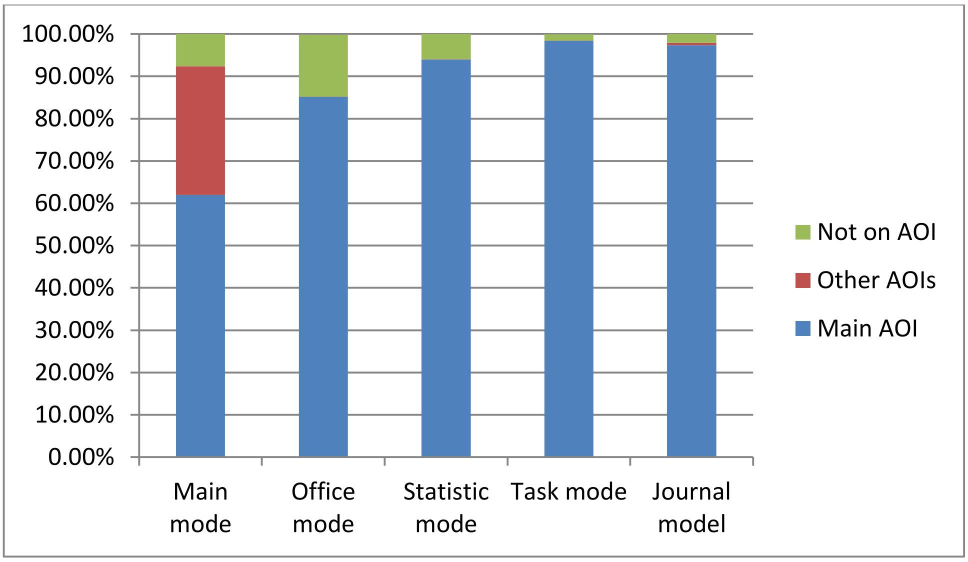
This section provides an overview of participants' general behavior during the play of the game by identifying the average amount of time and the proportion of time that participants spent looking at a particular area of interest (AOI) in terms of the different modes of play defined. As indicated in section 3.1.1.1, the *Main Mode* of play (MM) refers to the activities in which participants were allowed to build, explore the game landscape and the menu options available as well. The main AOI for the MM was defined as the central area of the screen. The *Others* AOI refers to the menu area to the right of the screen and to the message area below the screen. All AOIs correspond to the sum of the main AOI with the Others AOIs. Finally, *Not on AOI* refers to the time spent in areas not defined in any particular AOI and corresponded to small areas such as the border of the screen and the areas in-between two or more AOIs. The mean total time across modes of play was 1946.1 seconds or 33 minutes approximately. Table 40 and Figure 53 show how the time spent during the playing of the game was distributed across the modes of play in terms of mean time and proportion of time spent in each AOI.

Table 40: Descriptive Statistics for Total Dwell Time Spent on AOIs by Modes of Play

	Main AOI		Other AOIs		All AOIs		Not on AOIs	
	M	SD	M	SD	M	SD	M	SD
Total Dwell Time								
Main mode	645.9	192.5	317.5	117.5	963.5	260.5	79	36.5
Office mode	35	26.3	.02	.15	35.1	26.3	6	5.2
Statistic mode	48.4	33.1	.02	.15	48.4	33.1	3.1	3.3
Task mode	318.3	.13	.13	.67	318.5	121.6	5	4.2
Journal mode	474.5	297.5	2.9	3.1	477.4	298.1	10.1	10.9

The data show that participants' time spent looking at an AOI was mainly on the main AOI, that is, the AOI that defined the specific mode of play. For example, participants in the Journal Mode had to read through this journal the themes and topics related to physics. While this mode was "on" the rest of the game was in pause, but participants could still cast glances at the menu on the right of the screen and other areas outside the main window displaying the journal content. On the other hand, when participants were in the main mode of play (MM), where they could build houses, bridges, streets, and other elements offered in the menu, although the main AOI here corresponded to the central space on the screen, the menu was also relevant for this mode of play. And so it is shown in Figure 533 with the red column indicating that participants did spend time in the menu area and other AOI besides the central one. This is not the case for the other modes of play. This might me due to the fact that the rest of the modes caused the game to change to a "pause" state.

Figure 53: Proportion for Total Dwell Time Spent on AOIs by Modes of Play



When analyzing this behavior between conditions, it seems that the participants in the condition to learn spent more time on the modes corresponding to the learning of physics (i.e., Task Mode and Journal Mode) than participants in the condition for fun. Likewise, the latter spent more time in modes of play more related to the game than the participants in the condition to learn. These values were not statistically different (see Table 41).

Table 41: Descriptive and Inferential Statistics for Total Dwell Time in the Main AOI by Mode of Play and Experimental Condition

	Learn		Fun					
Total Dwell Time	M	SD	M	SD	*df*	*t*	*p*	Cohen's *d*
Main mode	639.2	230.8	653.7	141.8	39	.238	.813	.03
Office mode	32.3	21.4	38.2	31.3	31.15	.698	.490	.12
Statistic mode	39.2	21.2	59	41.1	26.04	1.886	.071	.36
Task mode	342.9	140	289.4	90.7	35	1.351	.185	.22
Journal model	502.9	325.6	443	268.3	38	.631	.532	.1

5.3. Hypothesis Testing

This section presents the results of the inferential tests conducted to see whether or not the hypotheses proposed received empirical support. For the group differences the Student T (one tailed) was used together with the bootstrap method, so that for variables that presented either some slight departures from normality or outliers (see Section 5.2) Type I error could be controlled through this method. Correlational analyses were conducted using Pearson's correlation and winsorized Pearson's correlation coefficient. Finally, for comparing Pearson correlations the bootstrap method was used (see Appendix O for the complete Pearson's correlation matrix and scatterplots). When the observed T value of a Student T or a Pearson correlation is greater than the T value at the Alpha level of .05, the result is described in terms of *statistical* differences. Its *significance* is evaluated in terms of its corresponding effect size (Larson-Hall, 2010). According to Cohen (1992) the magnitude of the differences between two means are: Small=.20, Medium=.50 and Large=.80. Likewise, for the difference between two Pearson correlation coefficients, Cohen (1992) suggest the following magnitudes: Small=.10, Medium=.30, and Large=.50. For the case of evaluating the strength of a Pearson correlation coefficient (r), Sheskin (2007) suggests the following magnitudes: Strong= if r is greater than .70, Moderate= if r is between .30 and .70, and Weak= if r is lower than .30. As the assumption of normality for Student T is fragile with a relatively small sample size, results are complemented by the bootstrap method as described in the previous section. Concerning the assumption of equal variance, when is not met a welch tests is reported instead. As for the assumptions related to Pearson correlation, the normal distribution of variables is described in Section 5.2. When either normality has not been met or the presence of outliers has been detected the winsorized correlation coefficient is calculated instead of the Pearson coefficient.

5.3.1. Hypothesis 1

The research question addressed here is: What effects does the educational computer game (*Genius Unternehmen Physik*) have on the individuals' learning of physics related content? (RQ1)

To address this question the following hypothesis was tested:

Hypothesis 1: Individuals in both conditions will exhibit significantly greater recall of content knowledge on the posttest than in the pretest.

As shown in Table 42 both conditions obtained a significant gain from the pre to the posttest. In the condition to learn, a statistical gain of 4.2 points was found, $t(22) = 3.816$, $p = .001$, Cohen's d = .39. The effect size of .39 for the gain differences indicated a medium effect. Likewise the bootstrap-t method for comparing the trimmed means between the two occasions (pre-post) conditions yielded a 95% CI [1.88, 6.58]. As the interval did not include zero it is possible to reject the null hypothesis of no differences between the pre and the posttest. In the condition for fun, statistical gain of 4.5 points was found, $t(18) = 4.627$, $p = .000$, Cohen's d = .54. The effect size of .54 for the gain differences indicated a medium effect. Likewise the bootstrap-t method for comparing the trimmed means between the two occasions (pre-post) yielded a 95% CI [1.3, 7.79]. As the interval did not include zero it is possible to reject the null hypothesis of no differences between the pre and the posttest. Therefore Hypothesis 1 was supported by the data.

Table 42: Descriptive and Inferential Statistics for the Recall Test by Experimental Condition

Condition		Pretest M	Pretest SD	Posttest M	Posttest SD	*df*	*t*	*p*	Cohen's *d*
	Learn	3.85	3.65	7.55	3.88	22	3.816	.001	.39
	Fun	3.5	2.84	8.11	5.28	18	4.627	.000	.54

5.3.2. Hypothesis 2

Concerning the effects on learning this hypothesis answers the question: Is there a mean difference in recall of content knowledge between individuals instructed to play to learn physics and individuals instructed to play for fun? (RQ2).

Hypothesis 2: Individuals instructed to play to learn physics will exhibit greater recall of content knowledge on the posttest than individuals instructed to play for fun.

Table 43 shows that there was no statistical difference in recall between the two conditions, $t(39) = -.2$, p = .843, Cohen's d = .07. The effect size of 0.07 for the differences in recall indicated a negligible effect. Likewise the bootstrap-t method for comparing the trimmed means between the two conditions yielded a 95% CI [-3.66, 2.35]. As the interval did include zero it is not possible to reject the null hypothesis of no differences between conditions on the Recall posttest. Therefore, Hypothesis 2 was not supported.

Table 43: Descriptive and Inferential Statistics for Recall Posttest by Experimental Condition

	Learn		Fun					
Dependent variables	M	SD	M	SD	*df*	*t*	*p*	Cohen's *d*
Recall Posttest	7.55	3.88	8.11	5.28	39	.2	.843	.07

5.3.3. Hypothesis 3a-c

The following hypotheses (Hypothesis 3a-c) operationalized the central question of this study concerning the amount and quality of the participants' cognitive engagement: Are there mean differences in cognitive engagement – as measured by AIME Task, SCENG and Transformation Processes – between individuals instructed to play to learn and individuals instructed to play for fun? (RQ3).

Hypothesis 3a: Individuals instructed to play to learn will exhibit a greater Amount of Invested Mental Effort (AIME Task) than individuals instructed to play for fun.

Amount of Invested Mental Effort (AIME). This variable was anchored to three different aspects of the game. One aspect concerned the tasks, that is, the solving of the physical exercises embedded in the game (i.e., AIME Task). The other two had to do with aspects of the game itself (AIME Simulation and AIME Statistics). The important one here is AIME Task. Participants mean score (see Table 44) for the total sample was 2.96 (SD =

.65) out of 5 points. Participants in the condition to learn obtained a score of 3.65 (SD = .64) and participants in the condition for fun a score of 3.33 (SD = .94). There was no statistical difference on AIME Task between conditions, $t(36) = 1.285$, $p = .206$, Cohen's d = .4. The effect size of .4 for the differences in AIME Task indicated a medium effect. Likewise the bootstrap-t method for comparing the trimmed means between the two conditions yielded a 95% CI [-2.26, .82]. As the interval did include zero it is not possible to reject the null hypothesis of no differences between conditions. Therefore Hypothesis 3a was not supported by the data.

Table 44: Descriptive and Inferential Statistics for the Amount of Invested Mental Effort (AIME Task) by Experimental Condition

	Learn		Fun					
Dependent variables	M	SD	M	SD	*df*	*t*	*p*	Cohen's *d*
Amount of Invested Mental Effort (AIME Task)	3.65	.64	3.33	.94	39	1.285	.206	.4

Note. The p values on the table should be divided by two for a one tailed test.

The following hypothesis is concerned with the difference between the experimental conditions as to the level of Situational Cognitive Engagement (SCENG), a measure taken for each of the learning tasks and averaged across tasks (see Section 4.3.2).

Hypothesis 3b: Individuals instructed to play to learn will exhibit greater Situational Cognitive Engagement (SCENG) across tasks than individuals instructed to play for fun.

Situational Cognitive Engagement across tasks was 2.97 (SD = .55) in the condition to learn and 2.59 (SD = .77) in the condition for fun (Table 45). There was a statistical difference on the situational cognitive engagement between groups, $t(30.38) = 1.801$, $p = .04$ ($p/2$ for a one tailed test), Cohen's d = .32. The effect size of .32 for the differences on situational cognitive engagement indicated a small effect. However, the bootstrap-t method for comparing the trimmed means between the two experimental conditions yielded a 95% CI [-0.2, 1]. As the interval did include zero it is not possible to reject the null hypothesis of no differences between conditions, but in light of the results of the Student's T, the Hypothesis 3b is considered partially supported.

Table 45: Descriptive and Inferential Statistics for Situational Cognitive Engagement (SCENG) by Experimental Condition

	Learn		Fun					
Dependent variable	M	SD	M	SD	*df*	*t*	*p*	Cohen's *d*
Situational Cognitive Engagement (SCENG)	2.97	.55	2.59	.77	30.38	1.801	.08	.32

Note. Due to unequal variance, Welch test was conducted for this variable. The p values on the table should be divided by two for a one tailed test.

The next hypothesis is concerned with the cognitive processes that occurred during each of the learning tasks and represents the cognitive strategies, in terms of type of processing (qualitative dimension) of the cognitive engagement construct.

Hypothesis 3c: Individuals instructed to play to learn physics will exhibit a higher level of transformation processes than individuals instructed to play for fun.

Participants in the condition to learn reported an average of 4.21 (SD = 2.97) statements related to transformation processes, while participants in the condition for fun reported an average of 3.45 (SD = 1.47) (Table 46). There was no statistical difference on the level of transformation processes reported between conditions, $t(27) = .809$, $p = .967$, Cohen's d = .31. The effect size of .31 for the differences in the level of transformation processes indicated a small effect. Likewise, the bootstrap-t method for comparing the trimmed means between the two experimental conditions yielded a 95% CI [-5.35, 4.76]. As the interval included zero it is not possible to reject the null hypothesis of no differences between conditions. This result did not support Hypothesis 3c.

Table 46: Descriptive and Inferential Statistics for Transformation Processes by Experimental Condition

	Learn		Fun					
Dependent variables	M	SD	M	SD	*df*	*t*	*P*	Cohen's *d*
Transformation	4.21	2.97	3.45	1.47	27	.809	.967	.31

Note. The p values on the table should be divided by two for a one tailed test.
Transformation= Transformation Processes (Planning, Selectivity, Connecting).

Summary. Cognitive engagement – as measured by SCENG – showed a statistical difference between the two conditions. Participants in the condition to learn reported higher levels of situational cognitive engagement during the learning tasks embedded in the *Genius Unternehmen Physik*. The significance of this result is reflected in its small effect size. On the other hand, the cognitive engagement measures of AIME Task and transformation processes did not yield statistical differences between conditions. Participants reported the same level of mental effort invested and the same frequency of transformation processes used such as planning, connecting and selectivity. However, the significance of AIME Task, and transformation processes is reflected in their small to medium effect sizes. Therefore, Hypothesis 3b was supported by the data, while 3a and 3c were not supported by the data.

5.3.4. Hypothesis 4a-c

The following hypotheses operationalized the question: Are there mean differences in behavioral engagement – as measured by Fixation Duration, Total Dwell Time, and Reading Depth – between individuals instructed to play to learn and individuals instructed to play for fun? (RQ4).

Hypothesis 4a: Individuals instructed to play to learn physics will exhibit longer Fixation Duration than individuals instructed to play for fun.
Hypothesis 4b: Individuals instructed to play to learn physics will exhibit longer Total Dwell Time than individuals instructed to play for fun.
Hypothesis 4c: Individuals instructed to play to learn physics will exhibit higher Reading Depth than individuals instructed to play for fun.

The results of the fixation durations, total dwell time and reading depth are presented in Table 47 .

Table 47: Descriptive and Inferential Statistics for Behavioral Engagement by Experimental Condition

	Learn		Fun					
Dependent variable	M	SD	M	SD	*df*	*t*	*p*	Cohen's *d*
Fixation Duration	.38	.05	.34	.05	37.98	1.994	.053	.65
Total Dwell Time	9.24	4.57	8.52	4.89	36.94	.476	.637	.16
Reading Depth	.07	.02	.06	.02	37.86	1.826	.076	.59

Note. Due to unequal variance, Welch test was conducted for the three variables; p values should be divided by two for one tailed test.

Mean Fixation Durations. Participants in the condition to learn showed an average of .38 seconds (SD = .05) and participants in the condition for fun showed an average of .34 seconds (SD = .05). There was a statistical difference on the fixation duration means showed between conditions, $t(37.98) = 1.994$, $p = .025$, Cohen's d = .65. The effect size of .65 for the differences in fixation duration indicated a medium effect. Likewise, the bootstrap-t method for comparing the trimmed means between the two experimental conditions yielded a 95% CI [.003, .07]. As the interval did not include zero it is possible to reject the null hypothesis of no differences between conditions. Therefore, Hypothesis 4a was supported by the data.

Total Dwell Time. Participants in the condition to learn showed an average of 9.24 seconds (SD = 4.57) and participants in the condition for fun showed an average of 8.52 seconds (SD = 4.89). There was no statistical difference on the total dwell time between conditions, $t(36.94) = .476$, $p = .637$, Cohen's d = .16. The effect size of .16 for the differences in total dwell time indicated a negligible effect. Likewise, the bootstrap-t method for comparing the trimmed means between the two experimental conditions yielded a 95% CI [-2.49, 4.23]. As the interval included zero it is not possible to reject the null hypothesis of no differences between conditions. Therefore, Hypotesis 4b was not supported by the data.

Reading depth. Participants in the condition to learn showed an average depth of .07 (SD = .02) seconds per square centimeter and participants in the condition for fun showed an average depth of .06(SD = .02) seconds per square centimeter. There was a statistical difference on the reading depth between condition, $t(37,86) = 1.826$, $p = .038$, Cohen's d = .59. The effect size of .59 for the differences in reading depth indicated a medium effect. Likewise, the bootstrap-t method for comparing the trimmed means between the two experimental conditions yielded a 95% CI [-.004, .04]. As the interval (barely) included zero it is possible to reject the null hypothesis of no differences between conditions. Therefore, Hypothesis 4c was partially supported by the data.

Summary. Fixation durations and reading depth showed a statistical difference. Participants in the condition to learn showed longer fixation durations and deeper reading of the specific Areas of Interest (AOIs) related to the content knowledge embedded in the game. The significance of this result was reflected in its medium effect size. On the other hand, total dwell time did not yield a statistical difference. Therefore, Hypothesis 4a was supported by the data, Hypothesis 4c was partially supported by the data, and Hypothesis 4b was not supported by the data.

5.3.5. Hypothesis 5a-d

These hypotheses address the question: How do cognitive and behavioral engagement relate to each other and to learning? (RQ5).

Hypothesis 5a: There will be a statistical positive correlation among AIME Tasks, SCENG and Transformation Processes.
Hypothesis 5b: There will be a statistical positive correlation between AIME Task and Learning.
Hypothesis 5c: There will be a statistical positive correlation between SCENG and Learning.
Hypothesis 5d: There will be a statistical positive correlation between Reading Depth and Learning.

As shown in Table 48, there is a statistical positive correlation among the mental effort indicators. The amount of invested mental effort on the Task (AIME Tasks) and the situational cognitive engagement (SCENG) showed a statistical positive correlation, r=.71, p<.01. Both variables showed a strong correlation coefficient. Similarly, transformation processes showed a statistical positive correlation with SCENG, r=.31, p<.05. These two variables showed a moderate correlation coefficient. Finally, transformation processes also showed a statistical positive correlation with AIME Task. These two variables showed a moderate correlation coefficient. Therefore, Hypothesis 5a was supported by the data.

Table 48: Pearson Correlations for the Measures of Cognitive Engagement

Measures	1	2	3
1. AIME Task			
2. SCENG	.71**		
3. Transformation	.41*	.31*	

Note. *p <.05; **p<.01. AIME Task=Amount of Invested Mental Effort with the Task; SCENG= Situational Cognitive Engagement; Transformation = Transformation Processes (Planning, Selectivity, Connecting).

Table 49 shows the winsorized correlations between cognitive engagement measures and learning. Of the two cognitive engagement measures, only SCENG showed a positive winsorized correlation with learning, r=.37, p<.05, with a moderate correlation coefficient. AIME Task did not yield a statistical positive correlation. As for the measure of behavioral engagement, Reading Depth showed a statistical positive correlation with learning, r=.32, p<.05, with a moderate correlation coefficient. Therefore, Hypothesis 5c and 5d were supported by the data, while Hypothesis 5b was not supported by the data.

Table 49: Winsorized Correlations between AIME Task, SCENG, Reading Depth and Learning

Measures	Learning
AIME Task	.25
SCENG	.37*
Reading depth	.32*

Note. p <.05. AIME Task=Amount of Invested Mental Effort with the Task; SCENG= Situational Cognitive Engagement.

Summary. The Pearson correlations among the cognitive engagement measures showed a statistical positive relation. Participants who reported higher levels of mental effort invested with the learning tasks embedded in the game also reported higher levels of situational cognitive engagement and higher frequencies of the use of transformation processes such as planning, connecting and selectivity. On the other hand, participants who reported higher levels of situational cognitive engagement during the learning tasks showed a greater recall of the content knowledge embedded *in Genius Unternehmen Physik*. Similarly, participants who showed deeper levels of reading of the content knowledge embedded in the game showed a greater recall. On the contrary, AIME Task did not yield a statistical positive correlation with learning. Participants who reported higher mental effort with the learning tasks did not show greater recall of the content knowledge embedded in the game. Therefore, Hypothesis 5a, 5c, and 5d were supported by the data, while Hypothesis 5b was not supported by the data.

5.3.6. Hypothesis 6a-c

The hypotheses below address the question: How do individuals' AIME measures relate to each other and to individuals' Emotional Engagement? (RQ6)

Hypothesis 6a: There will be a statistical positive correlation between AIME Tasks and AIME Simulation.
Hypothesis 6b: There will be a statistical positive correlation between AIME Tasks and Emotional Engagement.
Hypothesis 6c: There will be a statistical positive correlation between AIME Simulation and Emotional Engagement.

As shown in Table 50, there was a statistical positive correlation among the two mental effort indicators and emotional engagement. The amount of invested mental effort with the Task (AIME Tasks) and amount of invested mental effort with the simulation (AIME Simulation) showed a statistical positive correlation, $r=.69$, $p<.01$, with a moderate correlation coefficient. Similarly, emotional engagement showed a statistical positive correlation with AIME Task, $r=.43$, $p<.01$, with a moderate correlation coefficient, and with AIME Simulation, $r=.45$, $p<.01$, with also a moderate correlation coefficient. Therefore, Hypothesis 6a, 6b, and 6c were supported by the data.

Table 50: Pearson and Winsorized Correlations between AIME Task, AIME Sim, and Emotional Engagement

Measures	1	2	3
1. AIME Task			
2. AIME Sim	.69**[2]		
3. Emotional Eng.	.43**[1]	.45**[1]	

Note. ** p<.01; [1]Winsorized correlations; [2]Pearson correlation. AIME Task=Amount of Invested Mental Effort with the Task; AIME Sim= Amount of Invested Mental Effort with the Simulation; Emotional Eng.= Emotional Engagement.

Summary. The statistical positive correlations showed that participants who invested higher levels of mental effort with the learning tasks also tend to invest similar levels of mental effort with the simulation/game itself. The significance of this relation is reflected in the almost strong correlation coefficient. Similarly, participants who invested mental effort either in the task or in the simulation tended to report higher levels of fun and feelings of involvement in the game (i.e., emotional engagement). The significance of this relation is reflected in the moderate correlation coefficient. Therefore, Hypothesis 6a, 6b, and 6c were supported by the data.

5.3.7. Hypothesis 7a-f

These Hypotheses address the question: Do individuals' initial self-efficacy and general AIME relate to their actual cognitive engagement under the unspecific instruction to play for fun as compared with the instruction to play to learn physics? (RQ7).

Hypothesis 7a: Individuals instructed to play for fun will exhibit a statistically greater correlation coefficient between Self-efficacy and AIME Task than individuals instructed to play to learn physics.

Hypothesis 7b: Individuals instructed to play for fun will exhibit a statistically greater correlation coefficient between Self-efficacy and SCENG than individuals instructed to play to learn physics.

Hypothesis 7c: Individuals instructed to play for fun will exhibit a statistically greater correlation coefficient between Self-efficacy and Transformation processes than individuals instructed to play to learn physics.

Hypothesis 7d: Individuals instructed to play for fun will exhibit a statistically greater correlation coefficient between General AIME and AIME Task than individuals instructed to play to learn physics.

Hypothesis 7e: Individuals instructed to play for fun will exhibit a statistically greater correlation coefficient between General AIME and SCENG than individuals instructed to play to learn physics.

Hypothesis 7f: Individuals instructed to play for fun will exhibit a statistically greater correlation coefficient between General AIME and Transformation processes than individuals instructed to play to learn physics.

Concerning Self-efficacy, Table 51 shows that there was no statistical difference between self-efficacy and AIME Tasks, CI [-.19,1.27], between self-efficacy and SCENG, CI[-.11,1.29], and between self-efficacy and transformation processes, CI[-.97,.76]. However, Self-efficacy/AIME Task correlation and Self-efficacy/SCENG correlation showed a moderate and large coefficient, respectively. Therefore, Hypothesis 7a, 7b, and 7c were not supported by the data.

Table 51: Pearson Correlation Coefficients and Bootstrap 95% Confidence Intervals for Self-efficacy and Cognitive Engagement Measures by Experimental Condition

	Self-efficacy		
Dependent variables	Learn	Fun	Bootstrap 95% IC
AIME Task	-.218	.399	[-.19,1.27]
SCENG	.012	.631	[-.24,1.27]
Transformation	.401	.223	[-.97,.76]

Note. *p<.05; **p<.01. AIME_Tasks= Amount of Invested Mental Effort on the Tasks; SCENG=Situational Cognitive Engagement; Transformation = Transformation Processes (Planning, Selectivity, Connecting).

On the other hand, Table 52 shows that a statistical difference was found between conditions on the correlation between General AIME and AIME Task, CI [.04,.09]. The correlations of General AIME with SCENG and with Transformation did not yield statistical differences, CI [-.78,.50], CI [-.32,1.13], respectively. However, General AIME and Transformation showed a moderate correlation coefficient in the condition for fun, while only a weak coefficient in the condition to learn. Likewise, General AIME and

AIME Task showed a strong correlation coefficient in the condition for fun. Therefore, Hypothesis 7d was supported by the data, while Hypothesis 7e and 7f were not supported by the data.

Table 52: Pearson Correlation Coefficients between General AIME and Cognitive Engagement Measures by Experimental Condition and for the Total Sample

	General AIME		
Dependent variables	Learn	Fun	Bootstrap 95% IC
AIME Task	.102	.616*	[.04,.09]
SCENG	.595	.400	[-.78,.50]
Transformation	.130	.409	[-.32,1.13]

Note. *p<.05; AIME_Tasks= Amount of Invested Mental Effort on the Tasks; SCENG=Situational Cognitive Engagement; Transformation = Transformation Processes (Planning, Selectivity, Connecting).

Summary. The correlation coefficients between self-efficacy and cognitive engagement measures did not yield statistical difference although at face value they seem very different[8]. However, the correlation coefficients between self-efficacy and AIME Tasks and SCENG were moderate in the condition for fun, compare to the weak coefficient obtained in the condition to learn. On the other hand, the correlation coefficients between General AIME and cognitive engagement measures were moderate to strong, with General AIME and AIME Task showing statistical difference. For participants in the condition for fun, the higher the general level of invested mental effort with games, the higher mental effort invested in the learning tasks embedded in *Genius Unternehmen Physik*. The correlation coefficient between transformation processes such as planning, connecting and selectivity, and General AIME although not statistically different, was moderate in the condition for fun and only weak in the condition to learn. Participants in the condition for fun who reported a higher General AIME tended to some extent to use more frequently the transformation processes during the learning tasks. Therefore, Hypothesis 7d was supported by the data, while Hypotheses 7a, 7b, 7c, 7e, and 7f were not supported by the data.

[8] One reason for this could have been the amount of variance that there exist in the variables, which prevents the data to reject the null hypothesis of no difference (Wilcox, personal communication, April 16, 2013).

5.3.8. Summary of Hypotheses

Table 53: Summary of the Results of the Hypothesis Testing

Hypotheses	Result
RQ1: What effects does the educational computer game (*Genius Unternehmen Physik*) have on the individuals' learning of physics related content?	
H1: Individuals will exhibit higher posttest scores than pretest scores	Supported
RQ2: Is there a mean difference in recall of content knowledge between individuals instructed to play to learn and individuals instructed to play for fun?	
H2: Individuals instructed to play to learn will exhibit greater recall of content knowledge on the posttest than individuals instructed to play for fun.	Not supported
RQ 3: Are there mean differences in cognitive engagement – as measured by AIME Task, SCENG and Transformation Processes – between individuals instructed to play to learn and individuals instructed to play for fun?	
H3a: Individuals instructed to play to learn will exhibit a greater Amount of Invested Mental Effort (AIME Task) than individuals instructed to play for fun.	Not supported
H3b: Individuals instructed to play to learn will exhibit a greater Situational Cognitive Engagement (SCENG) than individuals instructed to play for fun.	Partially supported
H3c: Individuals instructed to play to learn will exhibit a higher level of transformation processes than individuals instructed to play for fun.	Not supported
RQ4: Are there mean differences in behavioral engagement – as measured by Fixation Duration, Total Dwell Time, and Reading Depth – between individuals instructed to play to learn and individuals instructed to play for fun?	
H4a: Individuals instructed to play to learn will exhibit longer Fixation Duration than individuals instructed to play for fun.	Supported
H4b: Individuals instructed to play to learn will exhibit longer Total Dwell Time than individuals instructed to play for fun.	Not supported
H4c: Individuals instructed to play to learn will exhibit higher Reading Depth than individuals instructed to play for fun.	Partially supported
RQ5: How do cognitive and behavioral engagement relate to each other and to learning?	
H5a: There will be a statistical positive correlation among AIME Tasks, SCENG and Transformation Processes.	Supported
H5b: There will be a statistical positive correlation between AIME Task and Learning.	Not supported Supported

H5c: There will be a statistical positive correlation between SCENG and Learning.	Supported
H5d: There will be a statistical positive correlation between Reading Depth and Learning.	
RQ6: How do individuals’ AIME measures relate to each other and to individuals’ Emotional Engagement?	
H6a: There will be a statistical positive correlation between AIME Tasks and AIME Simulation.	Supported
H6b: There will be a statistical positive correlation between AIME Tasks and Emotional Engagement.	Supported
H6c: There will be a statistical positive correlation between AIME Simulation and Emotional Engagement.	Supported
RQ7: Do individuals’ initial Self-efficacy and General AIME relate to their actual cognitive engagement under the unspecific instruction to play for fun as compared with the instruction to play to learn?	
H7a: Individuals instructed to play for fun will exhibit a statistically greater correlation coefficient between Self-efficacy and AIME Task than individuals instructed to play to learn physics.	Not supported
H7b: Individuals instructed to play for fun will exhibit a statistically greater correlation coefficient between Self-efficacy and SCENG than individuals instructed to play to learn physics.	Not Supported
H7c: Individuals instructed to play for fun will exhibit a statistically greater correlation coefficient between Self-efficacy and Transformation processes than individuals instructed to play to learn physics.	Not supported
H7d: Individuals instructed to play for fun will exhibit a statistically greater correlation coefficient between General AIME and AIME Task than individuals instructed to play to learn physics.	Supported
H7e: Individuals instructed to play for fun will exhibit a statistically greater correlation coefficient between General AIME and SCENG than individuals instructed to play to learn physics.	Not supported
H7f: Individuals instructed to play for fun will exhibit a statistically greater correlation coefficient between General AIME and Transformation processes than individuals instructed to play to learn physics.	Not supported

5.4. Further Analysis of Participants' Experience

In principle, the experimental manipulation of the perception of the task and the actual subjective perception can be different. It can be argued that individuals might not have been aware, at least during the entire experimental session, of the initial instruction and that initial individual differences might have influenced the amount of mental effort invested while playing *Genius Unternehmen Physik*. Therefore, the purposes of the following analysis were: to examine whether or not there are statistical differences on the control variables between individuals reporting high and low AIME Task; to examine whether or not there are statistical differences on specific dependent variables related to mental effort (i.e., Amount of Invested Mental Effort with the Simulation or AIME Simulation, Amount of Invested Mental Effort with the Statistics or AIME Statistics, situational Cognitive Engagement or SCENG, Transformation Processes, and learning); finally, to present qualitative information on participants' expectations and goals while playing the game. To achieve this purposes the total sample was divided using the median-split procedure on the AIME Task variable (Heers, 2005). A Student T (one tailed test) was conducted for examining whether or not there were statistical differences and the effect sizes were reported to assess the significance of such differences. No Bonferroni adjustments were used for reasons presented in Section 4. For the qualitative data, exemplars were selected and their frequencies were presented when describing participants' reports on their experiences with the game.

5.4.1. High versus Low AIME Task

AIME Task and control variables. Table 54 shows the results for each variable. It can be seen that the only variable that showed a statistical differences between the participants reporting high and low AIME Task was General AIME. Participants high on AIME Task also reported higher levels of General AIME, that is, reported that they invested higher levels of mental effort when processing material from a game than participants that reported lower AIME Task. This suggests that the actual amount of mental effort invested by the participants might have been determined by their initial perception of games and how much mental effort this new media warrant. The high significance of this result is reflected in its large effect size. Finally, prior knowledge and self-efficacy did not show a statistical difference between the two groups.

Table 54: Descriptive and Inferential Statistics for the Control Variables by High and Low AIME

	AIME Task High		AIME Task Low					
Control variables	M	SD	M	SD	*df*	t	*p*	Cohen's *d*
Pretest	2.91	2.23	3.56	2.81	38	.815	.420	.26
General Amount of Invested Mental Effort	3.24	.40	2.75	.69	37	2.634	.012	.86
Self-efficacy in Computer Gaming	3.72	.90	3.63	.75	38	.352	.727	.11

Note. p values should be divided by two for one tailed test.

AIME Task and dependent variables. Table 55 shows the results for each of the variable considered. Both cognitive engagement variables (i.e., SCENG and Transformation Processes) showed statistical differences together with moderate effect sizes. The participants who reported higher AIME Task, also reported higher SCENG and a higher frequency of using transformation processes such as Planning, Connecting and Selectivity. On the other hand, the AIME Simulation also showed a statistical difference between the groups. Participants reporting higher levels of AIME Tasks also reported higher levels of AIME with the simulation/game than the participants who reported lower levels of AIME Task. The significance of this result is reflected in its large effect size. Likewise, participants reporting higher levels of AIME Task also reported higher levels of Emotional Engagement than participants who reported lower levels of AIME Task. The group of participants who invested high levels of mental effort also reported positive feelings of fun and of being involved with the game. The significance of this result is reflected in its large effect size. Concerning the recall of content knowledge, the posttest showed an almost statistical difference with a moderate effect size in favor of the participants who reported higher levels of AIME Task. This suggests that the AIME Task tended to go together with higher levels of recall of the content knowledge embedded in the game. Finally, the behavioral engagement measures of Fixation Durations, Total Dwell Time and Reading Depth did not statistically differ between the two groups.

Table 55: Descriptive and Inferential Statistics for the Dependent Variables by High and Low AIME

	AIME Task High		AIME Task Low					
Dependent variables	M	SD	M	SD	*df*	*t*	*p*	Cohen's *d*
Posttest	8.76	5.02	6.85	3.84	40	1.284	.174	.40
AIME Simulation	3.90	.46	3.26	.50	40	4.239	.000	1.34
AIME Statistics	2.77	.74	2.50	.50	40	1.376	.177	.44
Situational Cognitive Engagement (SCENG)	3.2	.53	2.4	.66	36	3.65	.001	.6
Transformation	4.71	2.76	3.02	1.78	27	1.926	.065	.74
Emotional Engagement	4.60	.96	3.49	1.50	40	2.860	.007	.90
Fixation Durations	.35	.05	.37	.06	35.45	1.205	.236	.40
Total Dwell Time	8.22	4.47	9.57	4.90	37.67	.910	.368	.30
Reading Depth	.07	.03	.07	.02	36.59	1.09	.28	.36

Note. p values should be divided by two for one tailed test.

Summary. Participants who reported higher levels of mental effort invested during the learning tasks embedded in the game (AIME Task High) show also higher cognitive engagement during the learning task together with a higher level of emotional engagement understood here as feelings of involvement and fun. These engagements (cognitive and emotional) go in hand with the mental effort invested in the simulation part of Genius Unternehmen Physik. Likewise, these engagements seem to support to some extent the recall of the content knowledge embedded in the game. On the other hand, the cognitive engagement actually showed by participants seems to be influenced to an important extent by their initial perceptions of games and the amount of effort they warrant in order to process information and learn from them.

5.4.2. Qualitative data

The interview conducted, together with exploring participants' information processing, it also asked for general topics related to the experience of playing *Genius Unternehmen Physik.* Aspects of such experience were the main goals individuals set for themselves and the reasons why they decided to engage with the instructional content of the game. Through these questions it was also explored the degree at which participants followed the experimental instruction (i.e., to play learn physics or to play for fun) and other issues that emerged from the interviews: the issues of the role of the pretest in the game, the effect of the experimental manipulation of the task demands and motivations or reasons participants provided for solving or not a learning task. This information help characterized the subjective experience relevant for the research questions of this study.

The pretest. One of the issues that spontaneously emerged from the participants was their explicit expectation of the items that they had to answer during the pretest. To the simple question of "why did you read this" or "why did you open this" (referring to a page with instructional content) 14 participants referred to as having being "waiting" for the questions to arise (8 in the condition to learn and 6 in the condition for fun). Some of them appeared a bit concerned with their "poor" performance on the pre-test and used the experimental session and the game so as to "catch up" with their knowledge of physics. For example, a participant when asked why s/he decided to solve a particular learning task (i.e., "question"), s/he said:

"(…) weil ja vorher in dem Test die Fragen genau die Fragen dazu kamen. Und ähm ich konnte die meisten leider nicht beantworten (…), wie gesagt, vor ähm vor im Vorfeld von diesen physikalischen Aufgaben nicht beantworten konnte." (Transcript 39). [*Translation.* (…) because the questions in the test were the same. And I could not answer almost any of them (…) again, previously I could not answer these questions about physics].

Similarly, when asked why someone decided to read the "Journal" with all the content knowledge about physics, one participant responded:

"(…) ich wusste viele Teile davon nicht und die standen in der Zeitung drin und deshalb hab' ich das zum Nachbessern- Ich hab' das jetzt gestern Abend nicht mehr gemacht nachzugucken, was jetzt die richtigen Antworten waren und das hab' ich jetzt da gemacht." (Transcript 40). [*Translation.* I did not know any of them and they were inside the Journal, therefore to correct them- yesterday I have not looked at the correct answers but I did so now].

It is clear from these statements that the pretest might have cued some participants so as to make them use the learning resources in order to check for the right answers. As participants usually did not have only one purpose for taking actions within a game, here also it was reflected the several elements behind someone's decision to engage with the learning aspect of the game. The following dialog illustrates this:

Participant: Weil ich äh davon ausgegangen bin, dass mir der Inhalt für den weiteren Spielverlauf natürlich helfen kann. [*Translation.* because I assumed that the content could be useful for the game progress]
Interviewer: Mhm.
Participant: Und ... äh aus Interesse natürlich auch. [*Translation.* and because of interest too]
Interviewer: Ah, aus Interesse. [*Translation.* because of interest]
Participant: Ja, zumal ich ja auch den Fragebogen auch letzte Woche noch beantwortet hatte. Das waren ja die gleichen- Das waren ja die Antworten auf diese Fragen und ich wusste da auch nicht alle Fragen so heraus. [*Translation.* Yes, precisely because I have responded last week the survey. They were the same– they were the same answers to these questions and I did not know the answers to all of them]
Interviewer: Mhm.
Participant: Und von daher war‘s schon interessant die Antworten, dann auch nachzulesen.
Interviewer: Alles klar. [*Translation.* And thus it was interesting to read the answers]. (Transcript 41).

The participant in transcript 41 above mentioned two main sources or reasons why s/he decided to read the learning content embedded in the Journal. On the one hand, because s/he assumed that answering the questions could have an impact on the later progress on the game. On the other hand, s/he mentioned interest too, but an interest based on the experience of not having been able to respond to the pretest, so that having the answers in the Journal was an opportunity to read and know what the correct answers were.

The experimental instruction. Concerning the extent at which participants followed the experimental instruction, the following examples from a total of 13 cases (10 in the condition to learn and 3 in the condition for fun) illustrate how different the instructions seem to have been taken up by the participants.

“Ich hab die ganze Zeit so: Ach, du musst bauen du musst bauen und hier noch Häuser bauen (…) Also ich hab' das zu spät dran gedacht, dass es ja eigentlich um die physikalischen Aufgaben geht.” (Transcript 52).
[*Translation.* The whole time I did this: ok, you must build, build and here you must build houses (…) So, it was too late until I realized it really was about physics tasks.]

“ehrlich gesagt, im Laufe des Spiels hab' ich son– Weiß ich nicht, so eigene Ziele ... mehr verfolgt. Ich weiß nicht…” (Transcript 39).
[*Translation.* To be honest, during the game I followed, I do not know, I followed my own goals…]

“Ja, also zwischendurch war‘s weg. Äh...aber es kam durch die Aufgaben ja auch immer wieder. Und immer wenn Aufgaben waren, da waren dann dachte ich, ja du hast eben nicht nur...äh... nicht nur dieses Strategiespiel sondern da ist ja noch 'was hinter. ”(Transcript 63).
[*Translation.* Yes, in the meantime it was gone….but it returned through the learning tasks over again. And each time that there was a learning task I thought, yes you should not only this strategy game but what is behind it…]

“Ja, ich weiß nicht. Das hatte ich gar nicht mehr so im Hinterkopf. Das war für mich jetzt nicht der Hauptgrund das Spiel zu spielen, sondern ich wollte ja diese Fabrik endlich an 's laufen kriegen und war total abgelenkt dadurch, dass ich keine blöden Arbeiter da rein gekriegt hatte. ”(Transcript 34).
[*Translation.* Yes, I do not know. I did not keep it in mind. It was not the main reason for playing the game, but to get up and running the business and I was really distracted by the fact that I did not get any stupid employee.]

It is also important to notice that the instruction to play for fun could have also had some idiosyncratic manifestations which might have had an impact on participants’ in-game behavior. As the following extract of one interview shows, the participant originally in the condition for fun, found the learning tasks more interesting and “fun” than the simulation:

Participant: Spaß zu haben? [*Translation.* Have fun?]
Interviewer: Spaß zu haben. Und? [*Translation.* Have fun. And?]
Participant: Ja, deshalb hab' ich mich mit dem Wirtschaftsteil, mit den Statistiken nicht beschäftigt, weil ich das immer, weiß ich nicht, nicht spaßig finde. [*Translation.* Yes, that is why I did not get involved with the business part, with the statistics, because I do not find them fun.]
Interviewer: Okay. Aber die Aufgabe? [*Translation.* Ok, but the learning task?]
Participant: Was? [*Translation.* What?]

Interviewer: Die kleine Aufgabe, die du gelöst hast. [*Translation.* The small learning task that you solved.]

Participant: Ja. [*Translation.* Yes.]

Interviewer: Diese machten Spaß? [*Translation.* Are they fun?]

Participant: Ja, zum Teil schon. [Yes, in part.]

Interviewer: Könntest du ein bisschen mehr-? [*Translation.* Could you a little bit more…]

Participant: Ja, also wie gesagt, also ich finde so Textaufgaben eigentlich nicht so gut, aber selbst die letzte Aufgabe, die ich nicht geschafft habe. So Formeln und so was so was zu lösen macht mir eigentlich Spaß. [*Translation.* Yes, as I told you, I find the learning tasks not really good, but even the last learning task which I could not solve. So, I have fun solving formulae and similar.]

Interviewer: Ach so. Okay. Interessierst du besonders an der Physik? [*Translation.* Ok, do you have a special interest in physics?]

Participant: Ja, eigentlich schon. [*Translation.* Yes, I do.] (Transcript 23)

Motivation. The following examples illustrate the different reasons why participants engaged with the learning tasks embedded in *Genius Unternehmen Physik*. Understanding the reasons of the participants' action might help distinguish between actions related to individual differences and actions related to the design of the game. Likewise, reasons help understand goal setting processes that are the building blocks of goal striving (i.e., engagement) processes. The examples below were selected from 27 participants. As can be seen, the reasons vary. Most participants seem to have considered the consequences of solving or not a learning task. Others mentioned curiosity or interest as a main reason for attempting to solve the task. For other participants, the features of the tasks (i.e., how easy or meaningful they are) seemed to be the main reason for solving them. Finally, other mentioned elements of the fictional aspects of the game such as helping a non-player character and momentary/situational interests related to the task and/or the game. Examples of these alternatives reasons for attempting to solve a learning task are presented below.

Table 56: Examples of Participants Reasons to Engage with the Learning Tasks on *Genius Unternehemen Physik*

General consequences of the game

"...also ich hab's einfach gemacht, weil ähm mir jetzt nicht klar war, was das für Auswirkungen hat, wenn ich das nicht tue." (Transcript 27). [*Translation.* I just did it because it was not clear to me what could be the outcome if I do it.]

"…mir war in dem Moment auch noch nicht klar was für Konsequenzen das hat, ehrlich gesagt." (Transcript 38). [*Translation.* It was not clear to me what were the consequences of it, honestly.]

"Weil ich mir dachte, dass ich dadurch höchstwahrscheinlich einen Nutzen habe, um weiter spielen zu können." (Transcript 33). [*Translation*. Because I thought to myself that it is highly possible that I can use it in order to keep playing.]

Fictional world & mechanics of the game

"Und dann hab' ich gesehen, dass ich dafür extra Geld bekommen hab' und dann hab' ich mir gedacht: Mach ich die anderen doch auch. " (Transcript 59). [*Translation*. And then I saw that I got extra money for it, and then I thought to myself: I will do the others too.]

"Ja, weil ich ja unbedingt das Geld wollte. " (Transcript 51). [*Translation*. Yes, because I wanted the money in any case.]

"Ja, weil da der ähm Professor da gesagt hat, dass man dem helfen soll (…)Ich dachte, das gehört zum Spiel. Man hat ja dann auch Geld bekommen, wenn man ihm weiter hilft. " (Transcript 37). [*Translation*. Yes, because the Professoer said one should help him (…) I thought that is part of the game. It can be money received if one helps him.]

"Ja, deshalb hab' ich einfach angeklickt, weil ich dachte es passiert nichts. Aber dann hab' ich gemerkt, dass Geld abgezogen wird. " (Transcript 51) [*Translation*. Therefor I just clicked, because I thought nothing would happen. But later I noticed that money is substracted.] "Ja. Ich hab' halt die ganze Zeit gedacht, da würde eine Konsequenz raus kommen, wie dann bei dem, wie bei den Investoren, ne. Dass wenn ich das jetzt falsch mache, ich total loose. ” (Trasncript 38). [*Translation*. Yes, I thought some consequences should come, as in the case of the investors. If I do it wrong I loose.]

Features of the tasks

“Also ich hab' jetzt nicht so viel Wert darauf gelegt, dass die Antwort wirklich richtig ist, weil ja, weil wie gesagt ich mir nicht vorstellen kann, (lacht) dass das so wichtig für die ist, dass die sa so ein tolles Klohäuschen haben. ” (Transcript 49). [*Translation*. So I did not care that much whether the answer was right or not, because I could not imagine (laugh) how important it could be to have a nice badroom.]

“Also ich hab' ja dadurch kein kein äh keinen ähm finanziellen Vorteil gehabt, oder so. Weiß ich nicht. Einfach, weil mir die Aufgabe gestellt wurde und weil mir auch sofort ‘ne Lösung eingefallen ist wahrscheinlich. ” (Transcript 39). [*Translation*. I did not get any financial advantages. I do not know. Simply beacause the task was there and I immediately found a way to solve it.]

Momentary preferences of participants

“Aber ich wollte in dieser Sache jetzt nur die Aufgabe lösen, weil ich ja spielen möchte.” (Transcript 35). [*Translation*. I just wanted to solve the task, because I wanted to play.]

“Ich war da nicht so- hab mich da nicht so mit beschäftigt, weil mich das Spiel mehr interessiert hat, in dem Moment.” (Transcript 45) [*Translation*. I did not get involved with it, because at that moment the game interested me more]

Interest & curiosity

“Ja, aber wenn ich dann aber mal vor dem Computer bin, möchte ich das auch lösen. Also nur aus Neugier. ” (Transcript 50). [*Translation*. Yes, when I am in front of the computer, I want to solve it too.]

“Okay, hab' ich denn überhaupt 'nen Wissensstand oder- (lacht) Halt auch die Neugier: Wie viel weiß ich denn überhaupt? ” (Transcript 32).

[*Translation*. Ok, do I have any knowledge (laugh) or also the curiosity: how much do I know about it at all?]
"ich wollte das dann auch nicht nur blöd spielen, sondern auch ein bisschen was verstehen und äh auch einfach aus Interesse, einfach mal 'n bisschen mehr Informationen zu bekommen, über Sachen, die man noch nicht weiß." (Transcript 60). [*Translation*. I did not want to play stupidly, but also to understand a little bit and also because of interest, to get a bit of information about stuff I do not know.]
"Weil ich ja nicht wusste was noch kommt. Ich hab' einfach mal geguckt was passiert äh was ich machen soll und hab' das einfach gemacht, äh weil ich neugierig war, was noch weiter passiert." (Transcript 31). [*Translation*. I did not what was to come. I just looked what happens, what I should do and simply did it, because I was curious about what will happen next.]

Finally, the dialog below shows how a participant seems to have developed two reasons for doing the learning tasks. At the beginning, it was about money, but later the participant mentioned that for him/her it was important the goal of the game as instructed, that is, to learn physics.

Interviewer: Und äh allgemein: Warum hast du diese Aufgabe versuchen zu lösen?
Participant: ...Ich hab Geld gekriegt. [*Translation.* Yes, as I told you, I find the learning tasks not really good, but even the last learning task which I could not solve. So, I have fun solving formulae and similar.]
Interviewer: Ach so-
Participant: Hinterher. Das ist mir beim ersten Mal aufgefallen und... das war ja Ziel des Spiels, dass ich etwas über Physik lerne und dann.
Interviewer: Ach so beides? Ein bisschen beides?
Participant: Ja während- genau. Einmal hab ich Geld gekriegt zwischendurch, das ist natürlich nett wenn man ähm (lacht) so baut wie ich- ähm und ähm... ja, eben auch was über Physik zu lernen. (Transcript 63).

In summary, the qualitative data showed the highly differentiated experiences in terms of the extent at which participants were influenced by the pretest and by the experimental instruction. It seems reasonable to suggest that the pretest had an impact on participants' expectation and their behavior in the game and that the experimental manipulation seems sometimes to have been forgotten or replaced by participants' own goals. Finally, the reasons why participants decided to engage with the learning tasks present a high variation. These are related to the curiosity of the consequences the task might have had on the game, the role of the game fictional and mechanics features, and more personal interest on the content of the task or situational interest on the events of the game.

6. Discussion: Engagement and Learning from Educational Games

The purpose of this study was to examine the effect of manipulating individuals' perception of the demand characteristics of playing an educational computer game either for fun or to learn on individuals' cognitive engagement and learning when their perceptions of games and their self-efficacy to learn from them are controlled. More generally, the goal of this dissertation was to examine the construct of cognitive engagement on its quantitative and qualitative dimensions and how it related to other types of engagement – behavioral and emotional – to support learning processes in the context of an educational game (i.e., *Genius Unternehmen Physik*).

In order to achieve these purposes, Salomon's (1984) model of Amount of Invested Mental Effort and Corno and Mandinach's (1983) model of cognitive engagement were used to understand the role of cognitive engagement in terms of amount of invested mental effort and information processing in learning from a medium such as *Genius Unternehmen Physik*. At the same time, these models were subsumed under the broader concept of engagement (e.g., Fredricks et al., 2004) and its three central dimensions (i.e., cognitive, behavioral and emotional engagement), providing a comprehensive framework for the study of engagement and cognitive engagement in educational games and beyond.

The present study involved an eye tracking lab experiment combined with the administration of questionnaires and an interview. The data analysis examined the effect of a specific instruction (i.e., either to play to learn physics or to play for fun) – intended to influence participants' perception of the demands of the tasks of playing a game – on participants' cognitive and behavioral engagement and learning. It also explored the influence of individual differences in terms of initial perceptions of games and self-efficacy on participants' actual cognitive engagement. In order to examine whether or not high AIME Task had an impact on learning independently of the experimental manipulation, a median-split on AIME Task created the high AIME and low AIME groups on which was possible to compare the differences on the control and dependent variables.

Hypotheses about learning. It was expected that the game session would produce a gain in recall of content knowledge in both conditions (H1) and a higher score on the posttest for the participants in the condition to learn (H2). Results showed that participants in both groups had a statistical greater score between the pre and posttest with a significance reflected in a medium effect size. Therefore, H1 was supported. The difference on the recall posttest between the condition to learn and the condition for fun did not show a statistical difference, so that H2 could not be supported.

Hypotheses about effects of Perceived Demands Characteristics (PDC) of the task on Cognitive Engagement. The effect of the PDC should have been reflected on higher self-reported amoun of invested mental effort (AIME) (H3a) and situational cognitive engagement (SCENG) (H3b) with the learning tasks in the condition to learn when compared to the condition for fun. It was also expected that the manipulated PDC would show higher reported frequency of transformation processes (H3c) used when engaged in solving the learning tasks in the condition to learn when compared to the condition for fun. Results showed that the expected effects of PDC on participants' cognitive engagement could be only partially established. Situational cognitive engagement was statistically higher in the condition to learn according to the one tailed Student's T, but given that the bootstrap confidence interval included zero, the Hypothesis H3b is considered to be only partially supported. On the other hand, Hypotheses 3a and 3c were not supported by the data.

Hypotheses about effects of Perceived Demands Characteristics (PDC) of the task on Behavioral Engagement. The effects of PDC were expected to be reflected on participants' behavioral engagement or attentional behavior. The effect of the PDC should have been reflected on higher Fixation Durations (H4a) and higher Total Dwell Time (H4b) with the learning content in the condition to learn when compared to the condition for fun. Likewise, it was expected a deeper reading of the learning content in the condition to learn (H4c) when compared to the condition for fun. Results showed that participants in the condition to learn had higher fixation durations while engaged in the learning content. Therefore, Hypothesis H4a was supported. Participants' reading depth was higher in the condition to learn, but the bootstrap confidence interval included zero. Therefore, H4c is considered to be partially supported with some caution. Hypothesis H4b was not supported.

Hypotheses about the relationship among variables. Cognitive engagement measures should be related, so that invested mental effort, situational cognitive engagement and transformation processes should show positive correlations (H5a). The recall of content knowledge should be positively related to the invested mental effort (H5b), the situational cognitive engagement (H5c), and the reading depth (H5d). Likewise, it was also expected positive relationships between the mental effort measures (AIME Task and AIME

Simulation) and emotional engagement in terms of fun and feelings of being involved in the game. It was expected that a higher invested mental effort with the learning tasks should lead to higher invested mental effort with the simulation (H6a). Higher emotional engagement should lead to higher invested mental effort with the learning tasks (H6b) and to higher invested mental effort with the simulation (H6c).

Results showed that the three measures of cognitive engagement are highly correlated, which supported Hypothesis H5a. The expected relationships between cognitive engagement measures and recall of content knowledge could be only partially established. The invested mental effort did not show a positive correlation with the recall of content knowledge, so that Hypothesis H5b could not be supported. Situational cognitive engagement had a positive correlation with recall of content knowledge; therefore Hypothesis H5c was supported by the data. A positive correlation was established between reading depth and recall of content knowledge, by which Hypothesis H5d is considered to be supported by the data.

In the case of the mental effort measures, a positive correlation between invested mental effort with the learning tasks and the invested mental effort with the simulation was established. Therefore, Hypothesis H6a was supported. For the case of emotional engagement, it showed a positive correlation with the invested mental effort with the learning tasks and also with the invested mental effort with the simulation. Therefore, Hypothesis H6b and Hypothesis H6c were supported by the data.

Hypotheses about the role of PDC, Self-efficacy and General AIME on Cognitive Engagement. It was expected that the PDC would have had a differential effect on the correlations between self-efficacy and cognitive engagement measures, and on the correlations between General AIME and cognitive engagement measures. A higher correlation coefficient between self-efficacy and the cognitive engagement measures in favor of the condition for fun was expected. In particular, the Pearson correlation coefficients between self-efficacy and AIME Task (H7a), self-efficacy and SCENG (H7b), and self-efficacy and transformation processes (H7c) should be stronger in the condition for fun. Likewise, the Pearson correlation coefficients between General AIME and AIME Task (H7d), General AIME and SCENG (H7e), and General AIME and transformation processes (H7f) should be stronger in the condition for fun. Results showed that no effects of PDC could be established on the correlations coefficients between self-efficacy and AIME Task, self-efficacy and SCENG, and self-efficacy and transformation processes.

Therefore, Hypothesis H7a, Hypothesis H7b, and Hypothesis H7c were not supported by the data. On the other hand, the effect of the PDC could only be established on the correlation coefficients between General AIME and AIME Tasks, while General AIME and SCENG, and General AIME and transformation processes correlation coefficients did not show a statistical difference. Therefore, only Hypothesis H7d was supported by the data.

Results of the median-split on AIME Task. An explorative analysis was conducted to examine the role of the actual invested mental effort with the learning task, given that the experimental manipulation did not yield the expected effects on this central variable. By comparing the control variables between the groups high and low in AIME Task it could be established that participants in the high AIME group also reported higher General AIME or the general perception that games warrant the investment of an important amount of mental effort. On the other hand, participants who reported higher levels of mental effort invested during the learning tasks embedded in the game (AIME Task High) showed also higher cognitive engagement during the learning task together with a higher level of emotional engagement (i.e., feelings of fun and involvement).

In summary, the results described above show that only a portion of the expected effects of the manipulation of participants' perceived demands characteristics (PDC) was established. The manipulation of participants' PDC showed a limited effect on the measures of cognitive engagement. On the dependent variable recall of content knowledge, the manipulation of PDC did not have an effect. On the other hand, the PDC had an effect on two of three eye tracking measures related to behavioral engagement (i.e., fixation durations and reading depth). Likewise, greater recall of content knowledge was associated with higher situational cognitive engagement and reading depth. A close relationship among the mental effort measures and participants' emotional engagement was established. Individual differences in terms of the perception of games were related to the actual amount of mental effort invested, which in turn seem to have an almost statistical effect on recall of content knowledge and emotional engagement. The following bullet points provide a general perspective of the present study and will be discussed in the next section:

1. Participants increased their recall of content knowledge while playing *Genius Unternehment Physik*. However, the manipulated PDC did not influence the recall of content knowledge between the two conditions (i.e., to learn and for fun).

2. The manipulation of the PDC had a partial effect on participant's cognitive engagement. However, it showed an effect on participant's behavioral engagement.
3. There was a positive correlation of recall of content knowledge with situational cognitive engagement (SCENG), as well as with reading depth. No positive correlation was found between recall of content knowledge and invested mental effort with the task (AIME Task).
4. There was a positive correlation among invested mental effort with the learning tasks (AIME Task), invested mental effort with the simulation (AIME Simulation), and Emotional Engagement (i.e., feelings of fun and involvement).
5. General AIME showed a statistically stronger correlation with AIME Task in the participants under the condition for fun.
6. The actual invested mental effort with the learning tasks (AIME Task) goes together with higher levels of situational cognitive engagement, transformation processes, emotional engagement, and almost with a greater recall of content knowledge. The actual invested mental effort (AIME Task) seems to have been mainly influenced by participants initial General AIME, that is, by their initial perceptions of games and the amount of effort they warrant in order to process information and learn from them.

6.1. Interpreting the Main Results

The present dissertation shares the core assumption of Salomon's model (1984) concerning human functioning: performance, and in this case learning, is the resultant of a sustained interaction among individuals' cognitive and behavioral factors situated in a particular environment (Bandura, 1982; Dewey, 1938). In the context of this study, cognitive and behavioral factors correspond to participants' cognitive and behavioral engagement, respectively. The environment corresponds to the educational game *Genius Unternehmen Physik*. Salomon's model (1984) predicts 1) higher AIME in the presence of a clear and specific instruction to process a material in order to learn it (i.e., increasing the perceived demand characteristic of a task) (see Kunkel & Kovaric, 1983; Salomon & Leigh, 1984; Glaser et al., 2012), and, therefore, 2) higher learning, while in the absence of such type of instruction 3) AIME should be influenced by individuals initial perceptions of the demands of the medium and their self-efficacy to deal with such demands.

The results of the previous studies that have examined AIME across different media can be summarized as follows: 1) studies that have provided empirical support to both the

PDC-AIME relationship and the AIME-Learning relationship (Salomon et al., 1989), 2) studies that have only provided empirical support for the relationship between PDC and AIME (Salomon & Leigh, 1984; Salomon, 1984; Beentjes, 1989; Glaser et al., 2012), and 3) studies that have provided empirical support to neither of the two relationships, that is, PDC-AIME and AIME-Learning (Cennamo et al., 1991; Supinsky, 1995; Heers, 2005). The studies in this last category have suggested issues related to AIME sensitiveness (Cennamo et al., 1991), competition for cognitive resources (Supinsky, 1995) and the complexity of the virtual environment used (Heers, 2005). Concerning this last point, it is significant that in these last three studies the complexity of the stimuli used was much higher than in the studies that found a PDC-Learning relationship. The present study, which also used a complex environment (cf. Heers, 2005), falls also in this last category. The possible explanations of the present results are discussed below.

6.1.1. Perceived Demands Characteristics (PDC) and Cognitive Engagement: Measure Sensitiveness, Overload and Relinquishment

Concerning the first prediction of Salomon's (1984) model mentioned above, it could not be replicated in the present study. Participants in the condition to learn did not show higher AIME in comparison to the participants in the condition for fun. However, participants in the condition to learn did show higher SCENG. Two possible explanations based on the idea of measure sensitiveness and overload of participants' cognitive resources is provided below. A third explanation considers the possibility of participants' general relinquishment to engage with the learning requirement of physic content. It is proposed that SCENG was differentially more sensitive than AIME to the PDC manipulation. On the other hand, competition for participants' cognitive resources coming from both the game as cognitive overload and from the participants as volitional judgments are suggested to have hindered the power of the experimental instruction. Finally, a general relinquishment to cognitively engage may also explain the lack of effects of PDC on cognitive engagement measures.

The first possibility is that SCENG was more sensitive than AIME to the experimental manipulation. Research on cognitive load (e.g., DeLeeuw & Mayer, 2008) suggests that different measures (e.g., dual task, rating scales and posttest) might tap on to different types of loads (i.e., intrinsic, extrinsic and germane). More importantly, this research employs these measures several times across stimuli with different properties (e.g., stimuli with high and low complexity), turning the measures sensitive to the properties of the stimuli. In the case of the present study, SCENG was administered to each participant as many times as learning tasks were solved by them. In other words, if a participant tried to solve 5 learning tasks (see Section 3.1.1.2 for a description of the learning tasks), then 5 times the SCENG questionnaire was administered. Therefore, it may well have been more sensitive to individuals' cognitive engagement and later recall. However, the authors who developed the measure of SCENG, to be originally sensitive to the tasks (Rotgans & Schmidt, 2011), concluded that the measure seems to be more sensitive to the degree of individuals' knowledge construction and not changes in the task demands. Their explanation makes sense in their study of the phases of problem-based learning, which assumes a progressive construction of knowledge, but in the case of the learning tasks embedded in the *Genius Unternehmen Physik* such a progression is unlikely given the diversity and relative independency of the topics covered by these tasks. Furthermore, the fluctuations on SCENG across tasks may support the case for the role that the task demands (see Appendix N) might have played in the case of the present study.

It is also possible that during the game session participants cognitive resources were occupied by an overload coming from the game and by volitional judgments in the form of ruminations after failure or in front of the learning material perceived as difficult to tackle. In any or both situations the experimental manipulation of participants' PDC might have lost its original power (cf. Heers, 2005). Concerning cognitive overload, the complexity of the educational game used can be theorized in terms of the cognitive load imposed on participants' cognitive architecture (e.g., Nelson & Erlnadson, 2008; Kalyuga & Plass, 2009; Schrader & Bastiaens, 2012). The main argument of cognitive load theory in the context of games is that games pose heavy processing requirements on human's limited working memory capacity (Kalyuga & Plass, 2009). For example, Schrader and Bastiaens (2012) found that participants in a simply hypertext environment experienced less cognitive load and performed better in retention and transfer tests of physics than participants using an educational game. Some of the features of games that might overload participants' working memory are, for example, navigation tasks, searching hidden cues, processing narratives and contextual information, separated representations requiring search and match process, excessive information presented at once, and limited guidance to balance the lack of prior knowledge (Kalyuga & Plass, 2009; Nelson & Erlandson, 2008). In *Genius Unternehmen Physik*, the game's buildings, trees and landscape, together with the menu and the different messages that participants become from the game represent extraneous material that may not have a direct influence on a specific learning goal (cf. Nelson & Erlandson, 2008). The plethora of visual and textual information that appear in multiple locations at once may overload participants working memory. Furthermore, in the game many messages are presented at the same time in the form of text and voice, which might overload participants working memory by providing redundant information (i.e., redundancy effect, Mayers & Moreno, 2003; Nelson & Erlandson, 2008). In the context of these general sources of cognitive load, some of which are inherent in environments such as games (Kalyuga & Plass, 2009), the instruction at the beginning of the game session to play the game in order to learn physics it is likely to have found little space in participants working memory, diminishing the effectiveness of the manipulated PDC (cf. Heers, 2005). These authors (Kalyuga & Plass, 2009; Nelson & Erlandson, 2008) also acknowledge the tension of applying the cognitive load framework to educational games and virtual environments. For example, Kalyuga and Plass (2009) recognize that the sources of extraneous load are deliberately broken in games as a mean to engage players cognitively. Similarly, Nelson & Erlandson (2008) analysis of multiuser learning environment (MUVE) wondered whether or not some multimedia principles such as the use of *segmenting* (i.e., information in segments controlled by the learner) and *coherence* (i.e., excluding interesting but extraneous material) (Mayer & Moreno, 2003) could be counterproductive to the overall purpose of MUVEs, that is, to mimic real-world inquiry processes. Therefore, cognitive load is relevant to educational games when the learning material poses a high intrinsic load (i.e., high level of complexity of the material relative to the learner) so that minimizing the extraneous load becomes central to avoid overloading individuals' working memory capacity. A discussion of these cases is provided below in Section 6.2.1.

The other sources that might have used participants' cognitive resources are volition related thoughts. Intrusive thoughts and ruminations related to preoccupation and hesitation states could have occupied participants' cognitive resources (Koole, Kuhl, Jostmann, and Vohs, 2005). According to Koole et al., individuals' tendency to ruminate after possible threats or frustrations (i.e., inhibition of positive affects) is labeled state orientation, as opposed to action orientation where individuals are ready to engage in *task-relevant* cognitions (Corno, 1993). Participants during the game session could have experienced certain negative affects such as frustration together with intrusive ruminations that might have occupied part of their cognitive resources, leaving less space for engaging in learning or cognitive related goals. In this scenario, the instruction to play to learn physics could have been either relegated to a second priority given the participants' hesitation to engage with such content as physics or because such rumination represented an "overload" of participants working memory. In the former case, the analytical thinking needed to implement the intention to learn physics could have been hindered. In the latter, no space to represent the goal they want to achieve is left in participants' cognitive resources (Koole et al., 2005). Some evidence that the issues raised by action control theory might have played a role come from the informal observation and reporting of participants during the interviews. For example, some participants showed frustration and the consequent hesitation in keep pursuing the solving of the learning tasks and the diminishing of the analytical thinking that supports the implementation of goals. Others also showed ruminations in terms of not being able to do the task, having no interest, feeling the subject was unrelated to their skills and remembering of past failures with the subject matter. However, this alternative although speculative is reasonable and has implications for the design of affective and conative feedback to support complex learning processes together with cognitive ones as suggested by the use of prior what-why questions before reading texts about content knowledge and the implementation of the so called "conative" feedback (see Section 6.3.3).

Interest or an automatic self-assessment concerning ones competencies in the area of physics could have been a factor that represented the relinquishment of participants to engage with the learning content. According to Salomon (1984) and Corno and Mandinach (1983), individuals in general interpret the situation and their abilities to deal with it in ways that either encourage them to invest effort or discourage them to even try. In other words, participants low confidence in being able to solve tasks related to physics and lack of interests in the topic precluded them to "cross the Rubicon" (Corno, 1993; Kuhl & Beckmann, 1985) and start the goal striving, volitional state, that is, to commit to learn physics and play the game with a learning goal in mind. Participants seem to spend most of the time assessing what are the tasks about, whether they understand it or not and how they can solve it. In Corno and Mandinach's (1983) terms, participants remained at the acquisition level of information processing without engaging more significantly in transformation processes. This contention is to some extent supported by the data, which showed similar frequencies of acquisition and transformation in both groups (see Table

39). Furthermore, when the way to solve the task implied reading content related to physics, participants seem to prefer to solve the tasks by themselves using logic or simply guessing. Some evidence of this can be found in the positive correlations between reading depth when reading the journal and the no correlation between reading depth while solving the learning task (Appendix O).

Finally, concerning the transformation processes as coded from the interviews conducted, Corno & Mandinach (1983) proposed that a higher use of transformation processes lead to a particular form of engagement called Task Focus. This form of engagement occurs when individuals intentionally activate more transformation processes (i.e., selectivity, connecting new information to existing one, and task-specific planning) and should be useful for tasks requiring quick analytic responses, little self-checking and use of external resources. It is considered to be a form of intelligent investment of mental effort and highly likely to occur in, for example, test taking situations. The tasks embedded in *Genius Unternehmen Physik* are similar to tasks that usually appear in test taking situation. On the other hand, the tasks also corresponded to "quick" analytic responses, such as calculating the force of a pulley in Newtons. In front of such tasks and the lack of statistical differences on SCENG and AIME, the lack of statistical differences in transformation processes should be related more to issues of overload and relinquishment discussed above than to the type of tasks presented to the participants.

6.1.2. Cognitive Engagement and Learning: The Transfer Appropriate Processing Account

SCENG showed a positive relationship with recall of content knowledge, while AIME did not. Although this goes in line with some of the previous studies (Salomon et al., 1989; Salomon & Leigh, 1984), it also shows an inconsistency with others that did found positive AIME-Learning correlations (e.g., Cennamo et al., 1991). Most of these relationships though are reflected by weak correlation coefficients (see Table 16). The possible explanations provided below are based on the theory of transfer appropriate processing (McCrudden, 2011; Morris, Bransford, & Franks, 1997; Rose & Craik, 2012). This theory suggests that retention on memory depends not only on the level of processing, but also on the match between the learning and the retrieval activities or on the degree at which the test requirements match the processes used for encoding the information (Rose & Craik, 2012). A similar argument has been proposed by Tobias and

Fletcher (2011) in the context of transfer to curricula from educational games. They suggested that in order for such a transfer to occur, there must be an overlap between the cognitive processes engaged during the game and the ones required by an external task. Under this general theory of appropriate cognitive processing, the differential relationship between AIME and SCENG with the recall of content knowledge is discussed in terms of AIME and SCENG tapping different constructs and therefore processing, the allocation of the adequate strategies, and the role of information processing versus inferential activity.

The first possibility is that AIME and SCENG measure different constructs, and therefore different processing required for different learning outcomes. Results showed that the recall of content knowledge, although with a moderate effect size from pre to posttest within subjects, did not show a statistical difference between the two conditions (see Table 42 and Table 43). Overall, the highest theoretical score on the recall test was 25.9, while the mean score obtained was not higher than 8 points (only a 33% of the total score). SCENG might have aimed at more "superficial" processes enough for achieving a certain level of recall, while AIME could be actually aiming at more fundamental processes not needed for recall or inferences, but for higher level of reasoning and thinking. As shown in Table 15, AIME questions are about how hard individuals try to *understand*, how much they *concentrate* and *think* about a piece of material and how much *effort* they put while watching, reading, and searching information or solving a task. On the other hand, SCENG is about the *perception* of being engaged with the task, the *effort* and *persistence* with the task and a *general feeling* of being absorbed by the task. AIME and SCENG only tap on the idea of effort, but SCENG does not include any "higher order" cognitive process such as understanding or thinking. It can be argued that questions such as "describe the three laws of planets' movements" or "Write down the equation for the pulley's law" (Appendix H) might not require deep thinking or understanding, and maybe it is enough to have had a sense of "engagement", "absorption" and effort to address them successfully.

Another possibility important to discuss is the degree at which participants that reported higher AIME were unable to allocate the appropriate processing strategies. From the framework on cognitive engagement (see Section 2.4), Salomon's AIME is conceptualized as the quantitative aspect of mental effort, while Corno and Mandinach's acquisition and transformation processes represent the qualitative aspects of mental effort. The main assumption is that a higher amount of mental effort should go in hand with more effortful processing strategies, such as selecting information, connecting new information with prior information, planning a set of steps to achieve a solution, are more appropriate to learning. For example, in their study of film viewing, Glaser et al. (2012) gave participants two instructions while watching a film: watch to learn (i.e., learning goal) or watch for entertainment (i.e., entertainment goal). Participants with a learning goal showed higher AIME, but no higher knowledge acquisition. The authors hypothesized that inappropriate processing strategies could have explained these results. As they did not provide measures of such processing strategies, the hypothesis was left open to future research. In the present study, Acquisition and Transformation processes were quantified from the interviews conducted at the end of the experimental session. Table 48 shows positive correlation coefficients among Transformation processes, AIME and SCENG. Furthermore, AIME and Transformation processes show a slightly stronger correlation coefficient. At first sight, these results do not seem to support the hypothesis that inappropriate processing is responsible for high AIME without learning effects. If the lack of effect of AIME on recall were due to inappropriate processing, then it is reasonable to expect a negative correlation between Transformation and AIME, which is not the case for the present study.

Close to the just described role of processing, another explanation comes from research on the role of purpose on processing and learning from text (Linderholm & van den Broeck, 2002; McCrudden, 2011; McCrudden & Schraw, 2007; Narvaez, van den Broek, & Ruiz, 1999; van den Broek, Lorch, Linderholm, & Gustafson, 2001). As the game used in this dissertation embedded all the content knowledge in terms of text, this line of research is highly relevant to explain the results of the present study. McCrudden and Schraw (2007) distinguished between two broad types of relevance based on the specificity of the experimental manipulation: Specific relevance instructions (“what” and “why” questions) and general relevance instructions (Perspective and Purpose). For the present study the Purpose is more relevant to describe given its similarity with the concept of perceived demand characteristics (PDC). Purpose refers to reading based on some clue or cue in the context, such as when ones read to study versus for entertainment. McCrudden and Schraw (2007) suggested that the assignment of a particular purpose (i.e., reading to study or for fun) influences the *inferential activities* used during learning. Among this activities, the researchers mention explanatory and predictive inferences, paraphrases, monitor, evaluation, associations and repetition (e.g., van den Broek et al., 2001). Research on the effects of purposes (i.e., reading for study or for fun) have found qualitative differences on participants inference generation, but no quantitative difference on recall (e.g., Narvaez et al., 1999), while others have found both types of differential patterns (i.e., inference generation and recall) in favor of the participants under the condition to read for study (van den Broek et al, 2001). Other studies have shown that individual high on working memory capacity (WMC) recalled more than the individuals low in WMC under the purpose to read for study, while low WMC individuals recall the same under both conditions (i.e., for study or for entertainment) (e.g., Linderholm & van den Broeck, 2002). These results make plausible the possibility of AIME affecting higher level thinking or "inferential activity" instead of the information processing strategies reported by the participants (i.e., selecting, connecting, and planning) under Corno and Mandinach's (1983) model.

6.1.3. General AIME, Self-efficacy and Cognitive Engagement: Participants' Initial Perceptions of Games

The third prediction of Salomon's model, that under an unspecific instruction (e.g., "play for fun") AIME, and therefore SCENG and Transformation processes, should be affected by individuals initial self-efficacy and their general AIME towards the medium (i.e., towards games). This hypothesis was only supported by the data in the case of General AIME and AIME Task, although some other relations were clearly different without reaching statistical differences (see Table 51 and Table 52). Results show that AIME with the tasks (AIME Task) had a positive Pearson's correlation with General AIME only in the condition for fun. On the other hand, in the condition to learn (i.e., who received an instruction designed to be specific and unambiguous) this relationship between their previous General AIME and their current AIME disappears. This means that the impact of individuals General AIME on the actual AIME reported was diminished by the instruction to learn. These results go in line with Salomon's (1984) general suggestion that when no clear instructions are given to the participants (i.e., "play for fun") there is more room left for participants to approach the task from their own schemata and/or scripts. Similarly, this finding is in agreement with Greene et al. (2004) and Gregoire et al. (2001) findings which showed positive correlations between individuals' perceptions, motivation and cognitive engagement. On the contrary, when the PDC is controlled by an external instruction (i.e., "play to learn physics…you will be tested"), AIME tends to be less sensitive to initial perceptions about games in general, and more sensitive to the specific stimulus demands under the externally controlled PDC.

Self-efficacy did not show a differential correlation with AIME between the two conditions. However, it showed a clear difference in terms of size and direction with AIME. In fact, the correlation coefficient under the condition to learn was negative, while under the condition for fun was positive. This suggests that participants' confidence on playing video games, which was not low for this specific sample (see Table 34), might have influenced positively the AIME actually used while solving the learning task. On the contrary, as the confidence of playing successfully a video game is less connected to trying to learn an unfamiliar and difficult subject such as physics, this relationship was negative on participants in the condition to learn physics. Participants instructed to play for fun might have approached the task from a more relaxed, everyday kind of mood, which did not involve self-efficacy judgments related to learning physics, so that they attempted the learning tasks in a more fearless and secured manner. In other words, they did not relinquish the investment of mental effort as the participant in the condition to learn might have. This lack of "negative" self-efficacy or more general conative judgment should have left more cognitive room for engaging with the tasks more appropriately. However, in the absence of differences in learning and AIME between the conditions it is hard to warrant such a claim. In summary, under the instruction to have fun, the level of AIME might have been influenced positively by participants' confidence in playing successfully a video game. This confidence might have produced positive feelings and less frequent negative judgments concerning the ability to learn physics. On the other hand, the decision to invest mental effort on participants instructed to learn could have been affected by self-efficacy judgments related to learning physics. This suggestion is in line with the general concept of self-efficacy (e.g., Meece et al., 2006). For the authors, self-efficacy represents judgments about one's ability to attain a specific level of performance within the context of a particular situation. Such judgments have been shown to be central mediator of achievement-related constructs such as persistence, effort and self-regulatory strategies (e.g., Bong & Skaalvik, 2003). In this sense, the situation and the task clearly should have been different in both groups. In the condition for fun the task might well have been perceived as consisting on solving "minigames" or puzzles that might have positive consequences on the game. In the condition to learn, the same task could have been perceived to be about learning the physic content behind the task.

Therefore, dissimilar tasks' perceptions may call different self-efficacy judgments, which in turn define the level of effort and persistence with the task. A similar argument can be found in Beentjes (1989). In his replication of Salomon's model in a sample of Dutch children, he suggested that the role of self-efficacy seemed to depend on the interaction between the topic and the medium, and not on the medium itself. Some topics are more easily learn than others and certainly more appropriate than others to a particular medium. In this case, having only a measure of self-efficacy with the medium (i.e., the game) and not with the topic (i.e., physics) limits the understanding of the role of self-efficacy on AIME in the context of the educational game used here (see section 6.2 for further elaboration of this point). It can be argued that the sample used in this study (i.e., college students) belong to the "digital natives" category and therefore have been exposed to different technologies, including games. These previous experiences have developed in these individuals some preconceptions about games that can be important when introducing these technologies for learning purposes. Salomon et al. (1989) in their study of metacognitive guidance using a computer tool (i.e., "Reading Partner") pointed out that despite the opportunity of such guidance to expend more mental effort, it was up to the participants to take this opportunity. It is reasonable to expect that the relatively low level of General AIME, that games require a low level of mental effort, played a role on individuals' decision to invest mental effort while playing *Genius Unternehmen Physik*.

6.1.4. Behavioral Engagement

As suggested by Salomon (1984), based on Kahneman's (1973) model of effortful attention, the attentional behavior of participant should also be affected by an effort of manipulating participants' perceived demands characteristics of the task (PDC). Results showed that Total Dwell Time was not affected, but fixation duration and reading depth did show statistically higher values in the condition to learn. Higher fixation durations during the reading of the content related to physic may suggest that participants' attention in the condition to learn was more focused and effortful on the learning aspects of the game. Similarly, reading depth as capturing how deeply a piece of text/image is read was higher in the condition to learn. Participants in this condition tend to read deeper into the texts and images selected in the AOIs than participants in the condition for fun. It should be remembered that the AOIs used for calculating these measures were selected so as to be relevant to both the learning tasks to solve and the questions in the content knowledge posttest. Therefore, these results suggest an effortful processing of the information relevant for learning. These findings go in line with previous studies conducted in educational games (e.g., Kickmeier-Rust et al., 2011). These authors found higher fixations duration for students that scored high on the learning tests. However, this study was conducted using only nine participants.

The three eye tracking measures (i.e., fixation durations, total dwell time, and reading depth) may represent different intensities of information processing. As described in Section 2.5.2 mean fixation duration is the average time a set of fixations hits a particular AOI. However, this measure does not say anything about the pattern of "hitting" that is, whether or not participants were watching the AOI and then to some other areas of the screen and then back again same AOI, and then to change the attention again and so on. In other words, two participants with the same mean fixation duration can have different patterns of fixations, so that one individuals jump among different AOIs and other participant may well stayed and inspect just one AOI. This could be interpreted as the former one being "distracted" and the second one being more focused. This level of "distraction versus focus" can be captured by the measure of total dwell time. That is why it represents a measure of more long-term processing. Following the same reasoning, reading depth should reflect even more focused processing given that is based on the measure of dwell time, but it is adjusted – divided – by the area in square centimeters of the AOI being looked. In other words, reading depth also reflect this "focus" cognitive processing, but considering the size of the AOI. To have higher values on the reading depth measures means to have a more focused attention. The measure of dwell time and therefore total dwell time inform only the extent at which participants tend to focus for longer times their attention in an AOI in more or less "continuous" way. However, an individual can look at an AOI for quite a while in a "day dreaming" type of state, and not being actually reading the information presented. By taking into account the area of the AOI, the reading depth measure can inform participants' general scan of the AOI or a deeper reading of the content. However, independently of the relationships between these three measures, the fact that fixation duration and reading depth show statistical difference while cognitive engagement measures and learning did not needs further exploration.

To provide with a more detailed perspective on the eye tracking measures used, correlations among these behavioral engagement measures (i.e., fixation duration, total dwell time and reading depth) and learning were differentiated in terms of whether the measures belonged to the Journal Mode (i.e., the part of the game in which all the content knowledge is embedded in the form of a “scientific journal”) or the Task Mode (i.e., the learning tasks to be actually solved). All the behavioral engagement measures related to the journal showed a significant positive correlation with the posttest of physics knowledge. This result reflects the fact that the actual information needed to answer the posttest was included in the journal pages and not in the tasks themselves. This implies that participants, who spent more time with the journal, in particular with the specific AOIs containing the information asked in the posttest, could obtain higher scores on the posttest. On the other hand, the behavioral engagement measures related to the learning tasks did not show the same pattern of relationships. This reflects the same situation already mentioned that the information needed to answer correctly the posttest was not on the tasks themselves but on the content of the journal. These results suggest that the higher the time spent in the tasks did not lead to higher posttest score if participant do *not* consulted carefully the content on the journal. Furthermore, the analysis of the correlations among the behavioral engagement measures with the journal and the tasks shows that the higher the participants’ fixation duration means when reading the journal, the higher the fixations duration means when reading the tasks. This suggests that the more effortful the processing of the information in the tasks or the more aware were participants of the need knowledge requirements, the more likely they were to open the journal and search for the information needed.

All in all, it is not possible to know with certainty whether the participants were making sense of the information, building conceptual schemas or simply trying to have a general sense of the information embedded in the game. In the absence of a consistent pattern of correlations with the cognitive engagement measures, and of statistical difference between groups on AIME, transformation processes and learning, what exactly the eye tracking measures mean in the context of this study remains an open question (e.g., Holmqvist et al., 2011; Law et al., 2010).

6.1.5. AIME Measures and Emotional Engagement: Competing goals or Intrinsic Design?

Finally, the anchored measure of AIME to different aspects (i.e., modes of play) of *Genius Unternehmen Physik*, that is, AIME with the learning tasks (AIME Task) and AIME with the simulation game (AIME Simulation) showed the expected positive correlations between them and with the measure of emotional engagement (see Table 50). These results, as expected, go in an opposite direction to the argument that educational games which do not show the “intrinsic” fantasy (Habgood & Ainsworth, 2011; Malone, 1981; Rieber, 1996) do not provide an integrated learning experience and are likely to break individuals flow experience and immersion (Egenfeldt-Nielsen, 2005; Habgood & Ainsworth, 2011; Kerres et al., 2009; Kiili, 2005). Egenfeldt-Nielsen (2005) proposed the name edutainment for these games, which according to the author offer “arbitrary rewards” (e.g., getting points to get to the next level instead of the feeling of mastery for completing the level). As suggested before (see Section 2.2.2) this position assumes that points and feeling of satisfaction cannot occur together. A more general argument in the same direction was advanced by Kerres et al. (2009). They contend that game and learning are highly likely to be experienced as two disruptive “modes” which undermine the flow and immersion of players. These theoretical considerations seem to be applicable to *Genius Unternehmen Physik* (Jantke, 2006). However, they all reflect to some extent the assumption that intrinsic fantasies are the way educational games should be designed in order to produce all the positive effects expected from them. The limitations of such assumption were extensively discussed on sections 2.2.2, 2.2.3 and 2.2.4. In this section the empirical evidence obtained regarding this issue is discussed. The fact that the two different measures of mental effort and emotional engagement showed positive and moderate correlations might be interpreted as an indicator that participants did not experience the game as “separated-disruptive”. Participants that reported to have tried to think and reflect as well as having invested mental effort and concentration while solving the learning tasks (AIME Task), also tended to have tried to think, reflect, concentrate and invest mental effort while engaged with the simulation game (AIME Simulation) designed in *Genius Unternehmen Physik*. This correlation can be interpreted as an “integrated” subjective experience of participants during the game. To make the argument more salient, what would have meant to have had a highly negative correlation between these two measures of AIME (i.e., with the tasks and with the simulation)? This would mean that participants would have chosen to cognitively engage only with either the learning tasks or the simulation game. On the other hand, the interview data (see Section 5.4.2) suggests that participants’ interest and curiosity could vary during the game session and when a learning task appeared they were actually more interested in playing the game. This situational interest in a particular activity within the game should not lead to conclude that the extrinsic features of the game led some participants to experience a disruptive experience. On the contrary, this suggests that it is more a matter of degree and not an either/or situation concerning the intrinsic/extrinsic integration or the flow/disruption experience. Of course, these fluctuations of the process of playing the game cannot be captured by questionnaires administered at the end of the game session. On the other hand, the positive correlation of the emotional engagement measure with the AIME measures provide further support for a more subjectively integrated experience of participants. Participants reported to have tried to think and reflect, concentrate and invest mental effort with the simulation game also reported to have felt enjoyment and totally involved in the game world. Again, to put the argument forward, what would mean a hypothetical negative relation between these variables? This would mean that the more participants try to think and reflect on the different aspects of the simulation game, the less enjoyment and involvement they feel. Similarly, this could also mean that participants highly involved in the game world, were so in a more mindlessly way – with low thinking, reflection and mental effort as measured by AIME simulation.

On the other hand, it could be possible that the positive correlation between both AIME measures reflect the diversity of goals that normally affect individuals' behavior and that are characteristics of games (Juul, 2005). Several goals may coexist in the game. For example, there is the overall goal of not "losing" in the sense of not going bankrupt. Then, there are sub-goals related to hiring more workers, expanding the business by building more fabrics, increasing the productivity of the business, reducing spending, cleaning up an area for expansion, helping the workers pull up weight through a pulley system (this fictional part is introduced in one of the learning tasks about pulley systems), getting more money, solving a task, etc. Similarly, there are also goals related to probing one's own abilities, skills or knowledge. The positive correlation among AIME Tasks, AIME Simulation and Emotional Engagement may well reflect how the game supported the coexistence of different goals. From this perspective negative correlations would have meant to have highly conflicting goals so that committing to one would mean to leave others behind. In summary, the data suggest that participants could invest mental effort with both the learning tasks and the simulation without hindering their perceived feelings of enjoyment and involvement in the game world. These findings do not seem to support the claims concerning the likelihood of these types of games to "break the flow" (Kiili, 2005) and to provide necessary with disrupting experiences during the game session

(Kerres et al., 2008). Although it is recognized that the emotional engagement measure used here and that taps enjoyment and involvement may represent only a proxy and imperfect measure of flow. Keeping this in mind, these results seem to support the notion that is more central to think in terms of design patterns to improve educational game design instead of in the intrinsic-extrinsic dichotomy which is mostly concerned with finding a place for content knowledge within a game. Such a proposal based on game design patterns is sketched in the implication section 6.3.2.

6.2. Limitations of the Study

6.2.1. *Genius Unternehmen Physik*

Strictly speaking, *Genius Unternehmen Physik* is a border line case of game and closer to what is usually called a business simulation game (Juul, 2005) with elements of city-building games. To a certain extent, *Genius Unternehmen Physik* resembles video games such as *Anno 1701, The Settlers series* or *Capitalism 2*[9]. In general terms, from the sessions conducted it can be said that the game story and its tasks were reasonably well embedded and on most participants had the initial effect of making meaningful the solving of the task. That is, the solving of the task had a clear connection with the game story and participants made reasonable assessment of the possible consequences of whether or not they were able to solve the task in terms of both changes in the game state (i.e., the amount of money left) or in the game story (e.g., help an apprentice or support the workers). However, a game stops to be a game if the player can no longer affect the game state in any significant way (Salen & Zimmerman, 2003). In technical terms the game has a weak “action<outcome choice molecule” (Section 2.1.4 & Table 4). When participants choose to perform an action they do not always know whether or not that action had any particular outcome in the game. Likewise, participants often do not know what to do next, which is a central issue in keeping the anatomy of a choice clean and clear. On the other hand, it was apparent from participants’ spontaneous verbalizations that although the game did not seem to support participants while attempting to solve the tasks, the tasks and the story of the game were clearly understood and accepted by the participants. Just a couple of them mentioned that they did not see any relationship between the tasks and the game. The rest of the participants shared the goal imposed by the game. This relation

[9] For *Anno 1701* see http://anno.de.ubi.com/history1701.php, for the *Settlers 7* see http://www.settlers7.settlersmaps.com/news.php, for *Capitalism 2* see http://www.enlight.com/capitalism2/

between tasks and the fictional world of the game suggests that even in apparently highly “extrinsic” game design like this one (Jantke, 2006), the integration also can occur in participants mind. In doing so, participants contribute to a more integrated and meaningful experience. Therefore, and following the lines of the game studies field (see Section 2.1), the central problem with *Genius Unternehmen Physic* is neither its exchange of correct task for money nor its extrinsic reward system, but the limited gameplay which lead to a very limited set of alternatives for the players. For example, in the game *Anno 1701* there are dozens of voluntary quests the players can choose from. Some of them seem very “extrinsic” to the extent that they are not related to the fictional world of the game. Of course they do have a connection to the whole game, but the strength of such connection is highly varied. In terms of game design patterns, *Genius Unternehmen Physic* presents a more simple structure of *Producers-Consumers* that do not evolve rich enough to provide with *Varied Gameplay* and with opportunities for *Risk/Reward* evaluations. These limitations hinder the emergence of the key patterns of *Tension*, *Emotional Immersion* and *Cognitive Immersion*. Section 6.3.2 provides with a few proposals for improving the design *of Genius Unternehmen Physik* from a game design patterns perspective.

There are several usability issues that might have hampered the smooth pace of the gameplay. The first one is the use of a bulb to announce a task. For some reason, when participants click on the button it would not always work. It seems that the bulb had a particular point on which to be clicked, any other point does not work. A related issue was the inclusion of the bulb only above one particular building (i.e., fabrics). As the game allows participants to move around and explore the territory, if they happen to be away from the fabric on which the bulb appeared, it was less likely that they realized the presence of a task waiting to be solved. These two issues (the bulb was hard to click effectively and the bulb could be easily gone unnoticed) are the two major usability factors explaining why some participants could not even open the tasks in the first place. Secondly, an important proportion of participants mentioned they had difficulties finding the right way to get more employees to work. Some spent literally half of the time trying to understand how to get more employees. Although the solution is fairly simply, it might be counterintuitive. The game design delivers an employee whenever two buildings are connected to one another by a street. Other games for entertainment have a more intuitive solution. They simply add a function to create employees (e.g., *Age of Empires*[10]). In the

[10] For Age of Empires see http://www.ageofempires.com

case of *Genius Unternehmen Physic* “magically” the employees appear once the player has created a lane connecting two buildings, so that the players’ action (build a lane) seem quite unrelated to the outcome (employees). This is another example the breaking the “action<outcome choice molecule”. Third, some the learning tasks also showed several difficulties. The main consequence here is that participants do not know exactly where to click to provide an answer, so that they have to click everywhere until they hit the right point on the screen. For example, in task number 1 (see Figure 19) participants have to select three images that represent the three states of matter (solid, liquid and gas). In particular the picture corresponding to the correct answer for the gas state – a weak steam coming out of a teapot – is hard to fine even though in most cases participants knew the answer. The usability issues presented here are simple to address: to announce the tasks in a location that is always visible (e.g., as part of the menu) and allow several access options to open the tasks (e.g., from the building and from the menu); the workers should be created with a particular function as in games for entertainment; the tasks should indicate which objects/elements are “clickable”. These indicators or signals should be consistent across tasks and across other features of the game. More about “signaling” as load-reducing method is provided below when discussing cognitive load issues in the game.

From a cognitive load and multimedia learning perspective (Kalyuga & Plass, 2009; Mayer & Moreno, 2003; Nelson & Erlandson, 2008), *Genius Unternehmen Physik* presents overload scenarios which may hinder individuals’ performance. The overall argument was already discussed above in Section 6.1. In *Genius Unternehmen Physik*, the learning material embedded in a scientific “Journal” and in the learning tasks can overload participants’ working memory capacity. For example, the “Journal” (i.e., a set of pages with content knowledge about physics) and the learning task (i.e., several problems/questions to be solved by the participants) represent *separated representations* (Kalyuga & Plass, 2009) requiring participants to search and match the information by accessing different parts of the game. For example, in order for a participant to solve a task, she must open the task. A pop-up window appears. Within this window the participant needs to browse the different pages to access the instruction of the task, and the main data that compose the problem. Then, she needs to either “go back” to the beginning or close the window, then click on the “office” button, then click on “Journal”, and start the search for the information that then matches the one she needs to answer the task. Once the information has been found, the participant should close the “Journal” and open again the task, and provide the answer. This may force participants to keep in mind the information they need to find and once they find it, they need to keep it in mind in order to provide the answer. This suggests a possible overload scenario in which resources employed to hold a representation in working memory could be better used for essential processing of the material (Mayer & Moreno, 2003).

Likewise, the Journal was designed using an intensive text format, presenting *excessive information at once* (Kalyuga & Plass, 2009). Participants are supposed to read through the whole text until they come across with the information they need. Although participant can controlled the pace with which they browse the journal content, they do not control how much information is presented at once on each page of the journal. Each page may present several elements of information that need to be processed at once. With a modest prior knowledge and familiarity with the topic of the sample used in this study (see Section 6.2.5 for issues related to the sample), it is reasonable to expect an overload situation for the participants given the overloading of one channel (i.e., visual) and the complexity of the topic at hand.

By applying the multimedia principles of Mayer and Moreno (2003) and the suggestions made by Kalyuga & Plass (2009), some solutions to the overload situations in *Genius Unternehmen Physik* could be the followings. The issues mentioned above could be reduced by using segmentation, signaling, redundancy and multimedia principles. These principles as applied to *Genius Unternehmen Physik* can guide participants, avoid diversions of cognitive resources, support the management of information and eliminate the temporal and spatial split of sources of information (Kalyuga & plass, 2009). Segmentation (i.e., learner's controlled of information) could be applied to the content embedded in the Journal as hints that participants can access while answering the task and not having to close the task window and then open the journal window and so on. The hints need to have the relevant information for dealing with the task. Signaling (i.e., provide learners with clues as to how to process information or to focus on relevant information) can be applied in some tasks and also in the text used in the journal. For the tasks, interactive signaling in the form of icons or glowing frames can indicate which objects are clickable and which are not. In the texts embedded in the journal, information signaling in the form of italics, headings and subheadings, together with indicators of relevance (i.e., therefore, in summary, the main points are…) can help participants to focus their attention on information relevant to the task at hand. By applying the redundancy principle, it is possible to avoid simultaneous image and text presentation. For example, the NPCs usually present the same information related to the game in the form of text and voice. It is easy to remove the text and leave just the narration for giving participants the information required at a given point. Finally, the *Genius Unternehmen Physik* needs to apply the central principle of multimedia learning (i.e., multimedia learning): people learn better from words and pictures together (Mayer & Moreno, 2003). This would support the rearrangement of the information embedded in the Journal. The information presentation should move from extensive plain text with static and unrelated pictures to text and (dynamic) pictures with a logical relationship among them. In this way the dual coding processing may increase the building of schemas in long-term memory without an overload of the processing information channels (i.e., visual and auditory).

6.2.2. Internal, External and Ecological Validity Issues

Internal validity. Three main threats to the internal validity of the present experiment might have been the *pretest sensitization,* the *effect of experimenter's hypothesis,* and the *demands characteristics* (Rosenthal & Rosnow, 2009). First, the pre-test could have affected individuals' expectations during the experimental session and therefore could have decreased the effect of the independent variable (i.e., perceived demands characteristics). On the one hand, as already discussed (see Section 6.1.1), it could have fostered individuals' uncertainty or lack of self-efficacy on their ability to learn physics. On the other hand, according to the verbalization of participants during the interview (see Section 5.4.2), they seem to have tried to compensate their initially felt "ignorance" so that the game session was an opportunity to check what were actually the answers to the test. Second, the experimenter's hypothesis refers to a differential expectation of the researchers concerning how the participant should respond or act as a function of his/her treatment condition. In the present study, it was expected that participants in the condition to learn would spend more time reading the content knowledge embedded in the game, should elicit during the interview more statements related to transformation processing strategies and, finally, participants should score higher in the posttest administered. From these three situations, it is more likely that the researcher's expectation could have had an impact during the interview and the posttest. In the interview, the researcher could have engaged in more extensive dialogs or could have also added more follow up questions following the hypothesis of more mental effort invested and therefore more statements of effortful processing. In other words, without clear consciousness, the researcher –who was not blind to the participants' experimental condition – could have cued or suggested some answers to the participants or could have searched more intensively for those answers during the interview. Likewise, during the posttest session the researcher, through verbal and/or nonverbal behavior, could have cued participants to take more or less seriously the answering of the posttest of physic knowledge. Finally, the demand characteristic refers to the tendency of participants to interpret the purpose of the experiment and act in consequence. In the present study, most of the sample of participants was current college students taking part of a course on games offered by the researcher. This might have led some of the participants to show a more positive attitude toward the study, the experimental session and the game itself as they would in another situation. For example, and given the low experience in games, most of the participants found the game to be interesting, good and maybe useful for learning. One participant at the end of the session revealed that she thought the researcher was the one who had designed the game she just played. As the researcher was teaching her in this course elements of game design, the likelihood that she would say more positive things or to adopt a more positive attitude seems very likely.

External validity. This validity refers to the extent that the present study's results can be generalized across populations and settings. In this study 67% of the sample came from the bachelor in education (Erziehungswissenschaft) and were attending a course on games offered by the researcher. The other 33% came from another bachelor (Angewandte Kognitions- und Medienwissenschaft –Applied cognition and media) and were collecting credits for participating in research. Likewise 85% were women. The interaction of this sample (low experience in games plus very low prior knowledge and interest in physics) with the treatment (i.e., play to learn physics) may have biased their general performance during the game session and the questionnaires and tests. Closely related to the issues discussed above about internal validity, the effect of the pretest could have changed the general behavior of the participants of the sample used in this study. Finally, the reactive effects of the experimental setting limits what can be predicted concerning the effect of the independent variable (i.e., play for fun or to learn) in participants when they are exposed to it in a non-experimental setting. Finally, the results obtained in this study could be limited to the game employed here (i.e., Genius Unternehmen Physik) and might not be generalizable to other educational games addressing the same or different subject matters.

Ecological validity. Concerning the decision to employ a lab experiment, a few commentaries follow. In the field of educational game research, lab experiments seem to be unpopular. The basic critique is that any result obtained in a lab experiment is hardly generalizable to any real-life situation. Although this may be true, the purpose of this dissertation was to explore a set of hypothesized psychological processes involved in learning from an educational game. These volitional processes of cognitive engagement (i.e., mental effort and information processing) are difficult to foster in real-life setting, in particular formal educational settings. Therefore, a field experiment or quasi-experiment, although ecologically more valid, does not guarantee that the processes of interest are going to happen. In other words, one may have ecological validity, but not psychological validity. As Kuhl and Beckmann (1985) put it: "If we want to develop a better understanding of the processes mediating between cognition and action in everyday behavior, we need to develop methods that invoke those processes in an experimental situation" (p. 271). A similar contention was made by Salomon and Globerson (1987). The authors considered experiments to be a valuable method because they inform researchers "about what *can* be made to happen, not about what actually happens under normal conditions" (Salomon & Globerson, 1987, p. 627. Emphasis in original). In other words, in the case of educational games, knowing what can be made to happen means creating the best conditions in terms of both the design and affordances of the game and the attitude of the players so that the ideal psychological processes behind it can be "invoked" and studied. This knowledge can help advance a theory of learning from games, but also in more practical terms, it means having a more cautious attitude towards the implementation of educational games in more "real life" settings. By knowing experimentally what can be made to happen and how, then it would be possible to design specific and appropriate supports for the real-life, normal conditions in which a game must be used. To some extent this dissertation has aimed at precisely that: proposing a set of processes that might play a role in learning from educational games. And given that even experimentally inducing such processes seems difficult, more difficult should be when leaving a group of individuals play an educational game in the context of a lab session in a school.

6.2.3. Experimental Manipulation

There seems to be certain evidence concerning the limited effect that the instruction to learn or for fun had on the participants' goals and behaviors. This result goes in line with other attempts of manipulating individuals' perception of the task (e.g., Heers, 2005). Two further issues are discussed below.

First, how individuals understand learning from a game or how learning is supposed to happen in a game may have affected how participants interpreted the instruction. Research on conception of learning has shown that the more sophisticated this conception is, the better the quality of the learning process (Säljö, 1979; Tsai, 2009). Tsai (2009) showed how college student had more sophisticated conceptions of learning from a web-based learning environment when compared to their general conceptions of learning. For example, more categories of learning as "seeing in another way" where found for learning from web-based environments. As the sample of participants in the present study had a very limited experience with games in general, it seems that some of them really struggle in trying to understand how one is supposed to learn something from a game. It might be the case that participants held or developed some conceptions of "learning from educational games" that preclude them to instantiate the instruction to learn as expected. However, no data is available to support this contention other than spontaneous comments during gameplay.

Second, at a more practical level the newness of the media to the participants and their need of more extensive practice and familiarity with the media could also be factors affecting the pace of gameplay and therefore hindering participants' intentions to stick to the instruction to learn. The game presented itself at the beginning as quite hectic with participants receiving at the same time text-based message below the screen, pictures of the NPCs with text and audio, and other small events. From a multimedia perspective, such situations known as the modality effect can be counterproductive if individuals do not compensate this situation with appropriate memory strategies (Seufert, Schütze, & Brünken, 2009). Finally, the usability issues discussed in Section 6.2.1 could have hindered and consumed an important part of participants' attention and effort (i.e., cognitive load). In summary, the usability and design issues together with participants own interpretations of the situation, could have made the verbal instruction to learn or for fun insufficient. In other contexts such as televiewing (Salomon & Leigh, 1984) or a game for entertainment (Elliot & Harackiewicz, 1994) in which the topic is not too remove from the participants' life (Salomon & Leigh's 1984 adventure story and Elliot & Harackiewicz' 1994 pinball game) and the stimulus are more or less simpler, this type of verbal instruction are more likely to have the expected effect.

6.2.4. Instruments and Measures

Although all the scales had acceptable levels of reliability, the coding frame of the interview showed the lowest inter-rater reliability of 69%. This result is in line with similar efforts attempted by Howard (1989) who obtained a reliability index of 72%. It is very difficult to achieve higher levels of reliability given the idiosyncrasy of participants' responses to the interview questions. Likewise, for some of the questions participants showed difficulties in recalling the events during the game, answering "do not know" or "I do not remember". This highlights the limitations of retrospective interviews for dynamic environments such as games. Likewise, it may be useful to add more specific questions concerning the details of each of the tasks. The cognitive task analysis approach (Clark, Feldon, van Merriënboer, Yates, & Early, 2007) should be a useful starting point to refine the content related questions. The present interview might have used questions at a level too general to better distinguish among the different information processing activities as proposed by Corno and Mandinach (1983), Howard (1989), and Rogers and Swan (2004).

The recall test developed from the content knowledge embedded in the game also present some important limitations. First, it can have the traditional problem of content validity, that is, it may not have covered enough content as to represent fairly the amount of content knowledge accessible for the participants. It might be the case that some participants just read and got interested for content that was not part neither of the Areas of Interest selected nor of the test questions. However, by having anchored most of the questions on the tasks and the information related to them, and given that these tasks are central in the design of Genius Unternehmen Physic, it is considered that the misrepresentation of content knowledge was lowered to the minimum possible. On the other hand, scores were low. This could be explained by the conscious attempt of the researcher to avoid "ceiling effects" given that *Genius Unternehmen Physik* was developed for a much younger population than the one used in the sample. The idea was to design a test not too easy for the college students who participated in the present study (cf. Cennamo et al., 1991). However, the final test could have turned out to be too ambitious. Another issue related to the recall test used here in the context of Salomon and colleagues research, is that the present study did not distinguished between recall and inference items (cf. Salomon, 1984; Cennamo et al., 1991). Previous research on AIME showed slightly higher correlation coefficients with inferences items rather than recall items (see Table 16), except for Cennamo et al. (1991) who found a higher correlation coefficient between AIME and recall. Given that the research on Salomon's model has frequently distinguished between recall (recalling information explicitly presented to the participants) and inferences (information deduced from a story not explicitly presented), it is fair to discuss why this was not the case in the present study. In the context of the stories used in the previous studies on AIME, developing open ended questions such as "what did the artist think when he was paid?" (Salomon, 1984, p. 652) is a relatively simple procedure. In this example, whatever the thinking of the artist was it would be considered an "inference" because the story as presented to the participants cannot "explicitly" show what the artist was thinking. Similarly, in the context of lecture-like procedures, questions such as "How are fission and fusion different?" (Cennamo et al., 1991, p. 10) are simple to ask and whether or not they represent inferences depends on whether or not the answers contain any explicitly presented information during the lecture. In other words, the material used in these examples lends itself for such distinctions and more or less straightforward operationalization. On the other hand, the material used in the present study, namely the learning tasks participants had to solve and the information embedded in the "Journal" were not that suitable for drawing such distinctions and finding an adequate operationalization. For example, one task asked "order the planets in the right position" and the information stated in the journal is "Mercur, Venus…Neptune (…) The inner planets are mercur and Venus…and closing follow Uranus and Neptune." (See Appendix H). If a researcher wants to distinguish between recall and inferences as defined above, it is easy to find the main obstacle concerning the simplicity of the task and the information. Recalling seems to be straightforward. However, what an inference from this task would look like? What an inference question could look like? Certainly, in this particular case it seems very difficult to find a solution. This example represented most of the material available. In summary, the type of material that made up the stimuli was insufficient to draw a meaningful distinction between recall and inference, therefore only recall items were developed.

The inclusion of the following complementary measures could have helped understand the experience of playing *Genius Unternehmen Physik*: 1) the study did not include a measure of academic self-efficacy or self-efficacy for learning physics. Although from the sample used it could be speculated a low level of self-efficacy for learning physic, a formal measure would allow the exploration of the role of self-efficacy on effort investment, as hypothesized by Salomon (see Section 2.4.1.1); 2) a lack of comprehensive and multidimensional measure of flow, beyond fun and involvement, precluded any comparison regarding the relative centrality of flow and cognitive engagement when learning from games; 3) the study gathered only mental effort measures related to the learning tasks, but not to other sources of instructional content embedded in the game; 4) the lack of a measure of the difficulty of the learning tasks. With such a measure could have been possible to explore its correlation with the mental effort measures and eventually compare either their strength or their slopes between the condition to learn and the condition for fun. In this manner it could have been possible to know whether or not the mental effort measures reflect individuals' voluntary intentions to invest mental effort, as hypothesized by Salomon (1984), or the level of difficulty of the task at hand, as hypothesized by Kahneman (1973).

6.2.5. The Sample

The qualitative and quantitative properties of the sample used in this study are discussed next. From a qualitative point of view, the sample was composed of college students with a very limited experience with computer games, although they have grown up surrounded by these technologies (Prensky, 2001). This could have had a negative impact on the results of the study and it is important to know to what extent previous experience with games are a necessary condition to learn from an educational game. Second, 85% of the sample was female, and only 15% corresponded to males. Research has shown a consistent difference between women and men in terms of their level of involvement with computer games (Hartman & Klimmt, 2006). These authors explored women' dislikes about mainstream computer games. They found that lack of meaningful social interaction, violent content, competitive elements, and sexual gender role stereotypes were the aspects women disliked about computer games. This may be relevant for the current study given that *Genius Unternehmen Physik*, although does not have elements of violence or sex-related stereotypes; it does lack the meaningful social interaction that women seem to enjoy more than men. Findings from neuroscience tend to support this claim. Spreckelmeyer et al. (2009) found gender-specific activation patterns for money and for social approval which may reflect differences in the motivational value of rewards. Men may be more activated by money and women by social rewards. The implications for *Genius Unternehmen Physik* are clear as it uses almost exclusively money as the reward system of the game. Similarly, Hoeft, Watson, Kesler, Bettinger, and Reiss (2008) found gender differences in the mesocorticolimbic reward system. More specifically, the study showed differences in the brain activation and the functional connectivity patterns between men and women while playing a computer game. The findings may explain why men are more intensively involved with computer games. These findings from neuroscience have direct implications for the present study to the extent that they explain factors that might have had an important effect on the performance of the females composing the sample.

From a quantitative perspective, the sample size was an issue concerning the statistical techniques that can be used and the type of analysis that can be implemented. A description of the issues related to the current sample size was already provided (see Section 4.5.2). This section provides a brief discussion of alternative statistical procedures that could not be conducted because of the relatively small sample size used in this study. On the one hand, several mediational analyses could have been helpful to better understand the results, but these analyses rely on the assumptions of regression analysis which could not be fulfilled by the present data. For example, mediational analysis between initial general AIME and self-efficacy, cognitive engagement and learning are promising. Another analysis could have entailed the measures of mental effort (AIME task, AIME simulation and SCENG) and learning. Finally, AIME task, transformation processes and learning could have been another possible mediational analysis. Concerning also the sample size, the implementation of a factorial analysis on the AIME task and SCENG items could have helped the development of the discussion concerning the possibility of both measuring different aspects of mental effort (see Section 6.1). To explore which items load on the same and/or different factors could initiate future studies of construct validity of these measures. Although the issue concerning the minimum sample size or the subjects-to-variables ratio it is open to discussion (cf. Bryant & Yarnold, 1995; McCallum, Widaman, Zhang, & Hong, 1999), there seems to be a general agreement that samples greater than 50 individuals should yield a more or less robust result. This unfortunately was not the case for the present study.

6.3. Implications

6.3.1. Implications for Theory Building

The construction of a conceptual framework from multiples perspectives on psychology and education has resulted in a number of implications for the theory of the respective models and for the theory of learning from educational games. First, having anchored the measures of cognitive engagement and behavioral engagement helped to understand how participant might have experienced the game and its different parts or "modes of play". When playing an educational game, does the cognitive and behavioral engagement measures across different modes of play reflect the design of the game itself or only provide valid information as to the game as experienced by the individuals? It can be argued that both. If the design of the game has parts that are not well orchestrated in the story, as in the case of the Statistics Mode of Play, this could result in lower mental effort during this mode of play or lower correlation with the mental effort invested during other modes of play.

First, this dissertation assumes that in order to advance the scientific knowledge about a phenomenon or "object of inquiry", it is necessary to have a clear – hopefully agreed upon – conceptual definition of this object of inquiry. Difference concepts lead to different operationalizations and research designs. In this context, any attempt to organize the scientific discourse and future research around educational games would contribute to the communication among researchers and designers, and to advance the current understanding of how, for whom and in what conditions a particular educational game could produce the expected benefits. In a modest way, this dissertation represents a movement in such direction. This dissertation finds its place at the intersection of technology and educational psychology. More concrete, in the intersection of educational games, engagement and learning. After an extensive review of the concept of game, educational game, and simulations, it was offered a conceptual definition of what is an educational game together with its concomitant definition of what learning is from a process perspective. Educational games refer to:

Digital applications consisting of a fictional world and a character enclosed by a rule-based system that provides a hierarchy of goals to instigate on individuals the voluntary investment of mental effort and deep processing of information in order to acquire knowledge and skills, and preparing individuals to apply those knowledge and skills in future real life situations.

When this definition is compared against the ones usually found in the field of educational games (see Section 2.2.1 and Table 5), several improvements seem to emerge in order to clarify this “object of inquiry”. First, this definition does not define educational games in terms of other constructs that are difficult to define, such as “virtual environment”, “immersive environments” and the like. Second, this definition incorporates the central aspects that define a game, that is, the idea of a rule-based system, a hierarchy of goals (i.e., “challenge” or “conflict”), players’ effort and commitment. Third, the present definition explicitly avoids the use of the word *immersive* and *chance* as encountered in some definitions, because their meaning and role in games is still a subject of debate.

This dissertation opens with a statement from John Dewey (1913) concerning the role of effort in education. Following Dewey, the main assumption of this dissertation is that learning relies on effort. Therefore, the concept of mental effort was extensively discussed from a psychological perspective. Following Dewey, this dissertation assumed that effort is valuable only to the extent that it is connected to deeper and broader thinking. As discussed in Section 2.4.1, for Dewey effort supposes a degree of mental stress experienced as a peculiar emotion coming from conflicting tendencies, that is, from the desire to reach understanding and from the aversion placed by the existence of obstacles to such desire. In this way, Dewey already established in its conceptualization of effort 1) the quantitative and qualitative aspects of mental effort, and 2) its connections with cognition, conation and affects. This dissertation emphasizes the first legacy of Dewey’s considerations about the role of effort in education. In order to understand this quantitative and qualitative dimension of effort two lines of research have been integrated in order to capture these two dimensions. One line of research corresponds to Salomon’s (1984) model of Amount of Invested Mental Effort (AIME) (Section 2.4.1.1) and the other line of research relates to Corno and Mandinach’s (1983) model of cognitive engagement (Section 2.4.2.1). The former asks for the *amount* and the latter discusses the *type* or *quality* of this effort as expressed by particular information processing strategies. These two notions together have formed the main dependent variable of this study: cognitive engagement.

As this dissertation has attempted to disentangle the notion of game and educational game, the same was done for the highly used notion of "engagement". After reviewing the literature on engagement from the perspective of educational psychology, and from the perspective of game theory, it is proposed that engagement is a multidimensional construct with three distinctive dimensions: cognitive, emotional and behavioral. Most importantly, this construct and its dimension is suggested as an improvement to the current and highly cited gaming model of Garris et al. (2002), whose limitations has been already discussed (see section 2.2.2 and 2.5). Furthermore, while other researchers consider engagement in terms of immersion, flow, and even motivation, this dissertation explicitly defined engagement as distinct from motivation and established it as a volitional construct related more with the implementation of learning goals during a post-decisional phase of human goal oriented behavior. In retrospective, one might ask, if the construct of "engagement" is still needed or if it could be replaced with the construct of volition (Section 2.5), since the construct of volition has been in the psychological research for some time now. Although from a rigorous and strictly scientific perspective, volition might suffice in order to explore, explain and predict any of the issues usually examined under the concept of engagement, it seems unlikely that the term "volition" would eventually replace the term "engagement". This might be due to the familiarity of the term engagement among the researchers on the field of educational games and its "face validity" in terms of its meaning and role on learning. However, as dissertation has attempted to show, when one asks what it is really meant by engagement, then engagement "face validity" is seriously threatened.

6.3.2. Implications for Practice

Educational games are concerned with the learning of content knowledge or the development of skills. However, research has described a general pattern of interaction of individuals with technology-enhanced environments. This pattern refers to the suboptimal use of resources such as instructional guidance (Clark, 2009) and scaffolding (Pea, 2004), limited help-seeking behavior (Aleven, Stahl, Schworm, Fischer, & Wallace, 2003), modest access to guidance and feedback in multiuser virtual environments (e.g., Hickey et al., 2009; Nelson, 2007) and/or the misuse of helping tools in cognitive tutors (Muldner et al, 2011). To the extent that these overt behaviors might reflect modest cognitive engagement, it seems that fostering such engagement could be a promising route to improve the effectiveness of educational games. Therefore, the question is: How can game design support individuals' cognitive engagement? In other words, how can game design help individuals to sustain a deep engagement with the content knowledge?

Genius Unternehmen Physik has been theoretically analyzed and criticized by Jankte (2006). The author basically pointed out that between the learning and the gaming part of this educational game there is an evident separation. Likewise, he suggested that the knowledge acquired in the game is not useful for gameplay. For example, Jantke mentioned that the learning tasks come as a punishment and where learning and game are two separated activities. As an example, he criticized the fact that when the "learning" part appears the game is paused. However, from comments of users in this study, this aspect was perceived positively. Secondly, Jantke suggested that in effective educational games the knowledge acquired should be useful for gameplay. In the case of *Genius Unternehmen Physik*, at least in the first half an hour of the game, the design encourage participants to engage with the learning tasks, even though whatever knowledge participants might have acquired is not clearly useful for engaging in other aspects of the game.

These critiques imply the centrality of the notions of endogenous fantasies and interactive narrative for the design of educational games (see sections 2.2.3 and 2.2.4). These notions reflect the belief that the way to design effective educational games is by "coupling" together the content material with the game, so that the players advance in the game by applying the appropriate knowledge or skills. For example, by attempting to destroy a balloon, players are supposed to employ their knowledge of fractional numbers (Dugdale 6 Kibbey, 1975). Similarly, in order to advance in the narrative of *Quest Atlantis* (QA), individuals are supposed to apply their content knowledge (Barab et al., 2010). In the context of Taiga (one of QA's virtual worlds), individuals are said to use academic content in order to make inform decisions that can change the environment and advance the narrative. For example, an individual must write a scientific report explaining why process (e.g., erosion) might be causing the fish to die. As argued elsewhere (Filsecker & Kerres, 2013), drafting such reports has little connections with other features of QA (e.g., cols and lumins – a sort of money within the game). An individual presumably reporting accurately the situation in Taiga could get a few cols and lumins, but what the affordances of those features are is not immediately transparent or obvious. These design approaches rely on the degree at which the "fantasy" is accepted by the individual. Likewise, as already discussed (Section 2.2.2, 2.2.3 and 2.2.4), these approaches with their focus on coupling content and game, have lost sight of *gameplay* (Fortugno & Zimmerman, 2005)and have narrowed the theoretically unlimited possibilities in which fiction and rules can interact to produce interesting gameplay (Juul, 2005). Finally, these current design approaches overlook a central fact in games: Fiction alone, without coupled to a particular design pattern, can be useful at the beginning of gameplay, but later could easily loose its power because the more players play a game, the more they focus on the *mechanics* and less on the game's fictional world (see section 2.1.2).

Therefore, this dissertation assumes another perspective on the issue of effective educational game design: an educational game should be first of all a *game*. In this context the role of content knowledge can take different functions in a game, other than getting points or "changing the narrative". Hopefully, the description that follows can illuminate what is meant by claiming that knowledge can have other functions in educational games as well. Below is a description of how *Genius Unternehmen Physik* could be improved through the implementation of specific game design patterns (Björk & Holopainen, 2005).

As an example of the type of analysis proposed here, the following discussion describes a few game design patterns already in use in games for entertainment and that could be implemented in *Genius Unternehmen Physik* if the goal is to make more likely that individuals 1) decide to access the resources (i.e., content knowledge) and 2) cognitively engage with them. A discussion concerning the management of cognitive load seems straightforward in the context of the discussion on sections 6.1 and 6.2.1 and will not be further addressed here.

First of all, it is important to remember that games are goal-oriented activities, which normally present players with a general goal or winning condition and several paths or sub-goals to achieve the main goal of the game. For example, the *Settlers 7* main goal is either to destroy the enemy or to accumulate a certain amount of *Victory Points*. These goals can be achieved by adopting one of three different strategies: Military, Science, and Trade. It is also possible to mix these strategies. *Settlers 7* offers different ways to get Victory Points, such as building a complete army, researching a specific technology, fulfilling a quest, gaining prestige, etc. In terms of prestige, players are offered several ways to get prestige points (e.g., acquiring prestige-related technology, building prestige objects, or establishing trading points). Through these specific activities player get different *Rewards:* “material rewards” and “prestige rewards”. The latter usually unlocks a new option. For example, the capacity to update a storehouse and the ability to hire a geologist. This goal structure gives meaning to players’ actions in the game. Even the fulfillment of small quest contributes, for example, in the later acquisition of Victory Points. In terms of game design patterns, the accumulation of Prestige in the *Settlers 7* gives players *New Abilities* and *Privileged Abilities*, within a broader context of a *Hierarchy of Goals*. On the other hand, the Victory Points represents the classic *Score* so characteristic of games. It is clear from this brief description of the *Settlers 7* that players are given a broad variety of interesting choices or *Freedom of Choice*. These terms in italics are the names of just a few design patterns employed in the *Settlers 7* that can be used to improve *Genius Unternehmen Physik*. What follows is a description of these patterns as could be instantiated in *Genius Unternehmen Physik.*

In terms of the *Hierarchy of goals (*players need to complete at least some lower-level sub-goals to continue)*,* it is necessary to give *Genius Unternehmen Physik* an encompassing goal that can be initially defined by the players and that can provide meaning for the achievements of sub-goals or *Supporting Goal*s (goals whose completion allow the achievement of more complex/difficult goals. In *Genius Unternehmen Physik* instead of mention the goal of “becoming a successful entrepreneur”, it could be stated that the main goal should be to dominate a particular industry or several industries. Then, player should have the possibility to choose at least financial or corporate goals and to set constrains such as the amount of time allowed to achieve those goals. Together it should be possible to add a clear winning and losing conditions (e.g., not being able to achieve the predefined goals within a predefined timeframe, going bankrupt or the players’ industry being taken over by another one). It should be possible to win when all the predefined goals are achieved. *Genius Unternehmen Physik* could also add other general goals such as the victory points used in the *Settlers 7* or “lobby points” so that it has a second via to win the game. For example, assuming other activities are implemented in the game, one important feature to add is the use of Prestige points that provide players with the pattern *New Abilities* or *Privileged Abilities*, (new abilities that allow performing actions not available to other players) which will function as *Rewards* and would unlock abilities critical to win the game. In this way players will more likely struggle to perform those actions that would provide them with the amount of prestige points needed. Consequently, players would unlock and perform those activities critical to win. Emulating the role of Science in the *Settlers 7*, and the role of Research & Development (R&D) on *Capitalism 2*, in *Genius Unternehmen Physik* could be established the possibility of initiating a R&D project for researching a technology not being researched yet or to purchase a technology. The assumption is that R&D improves the quality of the products to be sold, and they can also provide players with prestige points or furthermore with "lobby points" (depending on the fictional world suggested by the game). Among the new abilities could be the ability to get and train characters to act as a spy or specific scientists. A spy could emulate the role that clerics play in the *Settlers 7*, that is, they are in charge of researching old technology and they are also in charge of fulfilling different quest and they can move through neutral or enemy territory with no hindrance. The technologies that they research usually can improve the abilities of the soldiers or increase the players' prestige. In *Genius Unternehmen Physik*, clerics in the form of a spy could be used to fulfill quests, visit and obtain strategic information from other competing industries, or to have the ability to negotiate with politicians or other businessmen. Likewise, to have them around could mean the improvement of the general economy of the players business, the ability of the workers in their different functions (e.g., sellers, advertisers, creative team, etc.). On the other hand, clerics can act as specific scientists that can advance the technology needed to achieve the game's goal. For example, these scientists could be a bit expensive, but could be very useful in similar terms as the spy just described. From a game design pattern perspective, these features allow players to perform certain actions that affect the outcome of the game otherwise impossible (i.e., Gain Competence). At the same time, players feel they can influence the outcome with such actions (i.e., *Perceived Chance to Succeed*) and also being able to understand how these actions affect the game (i.e., *Predictable Consequences*). Through these patterns the players perceive which choices lead to the more meaningful and effective effects on the game state (i.e., *Freedom of Choice*). Following the example with *Genius Unternehmen Physik*, it may be possible to have these scientists work together with the player to solve a task that requires subject matter knowledge. It could also be the case that if these scientists find the player to be not that knowledgeable, they can decide to quit and go to another business, with all the consequences that can be deduced from the above description. All these patterns together may focus players' attention in solving a problem and applying abstract logical thinking (i.e., *Cognitive Immersion)*. If the learning tasks are allowed to have the role described above, it is more likely that the tasks are going to be addressed more "mindfully" or players will be more cognitively engaged with them. In Salen and Zimmerman's (2004) words, in this context the learning tasks may be more likely to be "discernible and integrated into the larger context of the game" (Salen & Zimmerman, 2004, p. 34).

In summary, this dissertation assumed that a promising approach seems to be to explore more carefully the concept of *gameplay* and *game design patterns* in order to make more meaningful the solving of learning tasks, the use of resources (i.e., content knowledge) and feedback. The central goal here is to study the different design patterns and how are they implemented and enacted by players. In time, with the adequate research methods such as the ones used in this dissertation, it could be possible to identify design patterns that work for educational purposes. In a nutshell, the goal for future research should be the building of an empirically based *repository of educational game design patterns* that can be shared and improved by the educational community, similar to the Design Principles Database.[11]How this could be achieved is further described below.

6.3.3. Implications for Future Research

As discussed in section 2.2.5, modest empirical evidence supporting the effectiveness of educational games has been found. In order to understand this limited evidence in favor of educational games, it is necessary to move beyond the "black box" approach to educational game research (Honey & Hilton, 2011) and attempt to propose and measure the mechanisms underlying the process of learning from educational games. In other words, there is a need to develop a theory of learning from educational games. Such a theory should offer a description of empirically testable mechanisms by which an educational game can produce a particular outcome of interest. The knowledge of these mechanisms should lead to guidelines for the design of effective educational games and their inclusion in formal and informal learning contexts. Therefore, a line of research in educational games should focus on the development of an *empirically testable theory of learning from educational games* which can better inform designers to produce more responsive solutions to the users' psychological states, needs and actions. One approach is to employ the conceptual framework proposed in Section 2.5 to the extent that suggests

[11] See http://www.edu-design-principles.org

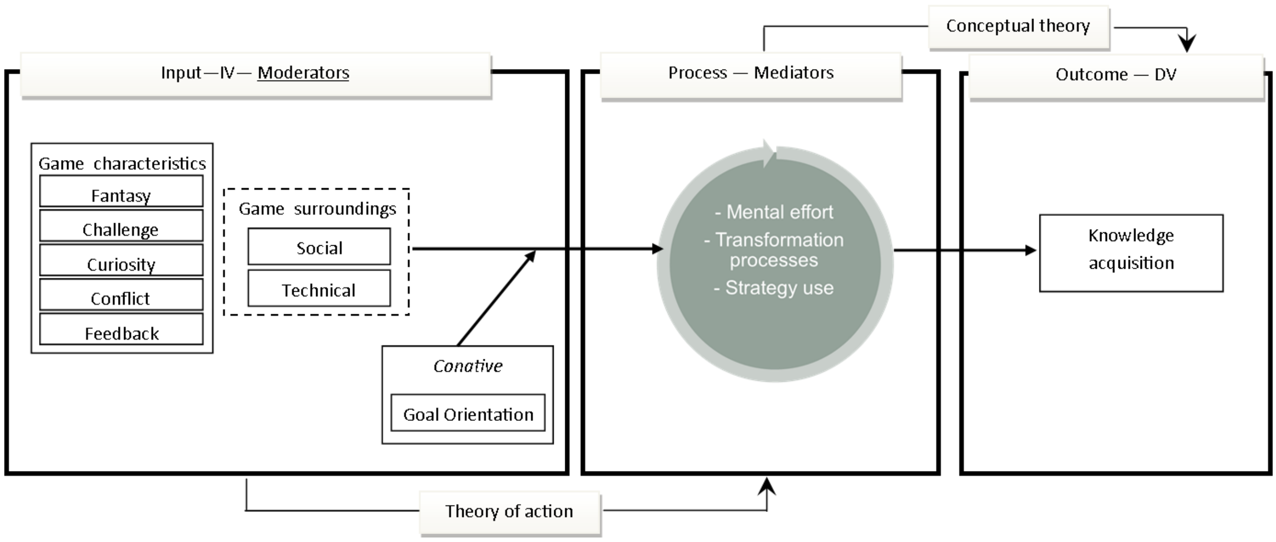
central mediators (e.g., engagement) and possible moderators (e.g., prior knowledge or task perceptions).

The present study represents an attempt at employing such conceptual framework of studying engagement and some of its possible determinants. Consequently, the present study has attempted to assess what individuals actually did while playing *Genius Unternehmen Physik*, especially by the eye tracking measures employed. These "doings" – such as the depth with which individuals read the information in the game – seem to have contributed to the outcome (i.e., learning). It also seems that individuals' current actions and cognitive engagement are a function of individual differences in terms of the perceived investment of effort necessary to learn from a game. From the perspective of the conceptual framework proposed, it is possible to suggest that this perception of the effort required to learn from a game corresponds to a moderator variable (i.e., Task perception) that warrants further examination. The same could be – with caution – suggested for the perceived self-efficacy on computer gaming. Likewise, reading depth could be hypothesized to be a possible mediator, despite the lack of differences in learning outcomes (assuming that mediational processes occur also in absence of an effect of the independent variable on the dependent variable – Mackinnon, 2008). Certainly, there are other possible moderators and mediators to examine in future studies. For example, several proximal and distal factors affecting cognitive engagement have been proposed (Salomon & Globerson, 1987). Among the distal factors important to understand is how individuals need for cognition affect the investment of mental effort. In a similar vein, the impact of action-control/volitional styles on individuals' decisions to invest mental effort in a task warrants further research. Finally, motivational variables such as self-efficacy and goal orientation seem worth exploring in the context of educational games together with surveys and interviews to examine individuals' perceptions of educational games and their conceptions of learning from educational games.

Figure 54 illustrates how the conceptual framework can be used (see Figure 15, page 136) to examine the effects of the social game surroundings on individuals learning. Here the question can be whether collaboratively playing an educational game has a higher impact on learning as opposed to solo playing (Input). It can also be of interest whether collaboratively playing a game impacts learning by increasing individuals' cognitive engagement (Mediator). Finally, it may be useful to explore whether or not the hypothesized effect of collaborative playing is moderated by individuals goal orientation

(Moderator). The full rationale of this hypothetical study is that collaborative activities invite individuals to set learning goals and to implement those goal by mutually supporting each other (i.e., asking/answering questions, providing encouragement to each other, having opportunities to self-explain, etc.) increasing their level of cognitive engagement, but only for those who hold learning goal orientations. As can be seen, the degree of complexity of any single study depends on the research question examined. However, when comparing the original conceptual framework with one of its instantiations as depicted in Figure 54, it is hopefully clear to see its usefulness to accumulate, organize and compare multiple studies in time so as to provide cumulative evidence to sketch a theory of game-based learning.

Figure 54: Illustration of the Conceptual Framework for Future Research



Apart from theoretical and empirical considerations, the study of engagement presents also diverse methodological challenges to overcome. For example, capturing individuals' cognitive processes is a very complex endeavor. In this case, the suggestion is to combine think aloud and retrospective interviews in successive stages within and between studies. In this manner, a richer data to assess individuals' cognitive processes may be available. In particular, future studies should consider longer periods of time for conducting these types of interviews from a more explorative approach complementing the focus oriented interviews conducted in the present study. Similarly, future studies should consider information processing strategies such as the one examined in this dissertation together with other cognitive activities such as inferential thinking and reasoning. As suggested in the interpretation of the results (Section 6.1) it is important to explore the extent at which individuals engage in thinking and information processing unrelated to the task and more related to negative thoughts, ruminations and the like. As already pointed out, these processes have also a role to play on individuals' performance and learning.

Likewise, another important research issue is the differentiation of mental effort as cognitive load and mental effort as AIME or mindfulness. A possible future study to tackle this issue could be based on an experimental study would consist of the presentation of a game in two modes (i.e., playing mode and vicarious mode) (Mattheiss et al., 2010) with two instructions (i.e., to play or watch for fun or to learn/remember the content). This latter should induce on individuals the perception of a more demanding task (i.e., the Perceived Demands Characteristics or PDC) and therefore will cognitively engage in a deeper way (Glaser et al., 2012; Salomon, 1984). This represents a 2x2 ANOVA factorial design with mode (Playing versus Observing) and PDC (High versus Low) as between subjects' factors. According to Mattheiss et al. (2010) individuals watching the game should learn if they "cognitively engage" with the content of the game as suggested by the idea of vicarious learning. Therefore, a higher cognitive engagement should be expected in the vicarious condition (H1). On the other hand, the "observers" do not need to deal with the decision making involved in games so that their cognitive load should be lower and therefore more resources should be left to process the content of the game (Plass et al., 2010). This implies that individuals in the playing condition should show higher levels of cognitive load (H2). Likewise, in the playing condition the level of flow experienced should be higher than in the vicarious condition (H3). Eye tracking measures as the one used in this dissertation should show positive correlations with cognitive engagement (H4) and learning (H5). As for the interactive effects expected, the central ones are related to flow, cognitive engagement and learning. The highest values for flow should occur under the condition to play with Low PDC (H6), while cognitive engagement should be high under the vicarious condition with high PDC (H7). Finally, learning also should be the highest under the vicarious condition with the high PDC (H8). This design is depicted from the perspective of the

Further research should be done to validate the eye tracking measure of reading depth as an indicator of effortful information processing. Likewise, future investigations might use a different research design to eliminate the effect of the pretest on individuals' behavior and performance. Such as design could be the Solomon design (see sections 5.4.2 and 6.2.2). Similarly, replicating the present study but using other strategies for manipulating individuals' perceived demands characteristics of the task (PDC). Here research on reading text and its use of pre-reading relevant questions (see McCrudden, 2011) could be a strategy to explore by simply adding those questions just prior to the presentation of the learning content in the game. Finally, quasi-experimental designs can help understand the role of task value on cognitive engagement and learning, issue that in the context of this lab experiments was not possible to address. In this context, research questions that could be asked include the relationship between the perceived value of the task of playing a game and individuals' cognitive engagement and learning.

As already suggested in the introduction, these lines of research may provide useful information to modern educational data mining techniques that are concerned with detecting engagement/disengagement from the analysis of log-files. The importance of this type of research lies in the assumption that any model that can be developed and tested through data mining techniques will be able to produce as an output the input with which the model was conceived. Therefore, the more relevant the inputs for such models, the richer the models and therefore, higher will be their predictive power. For example, being able to detect from individuals' behavior a state of mindlessness and generate an appropriate response seems to be a useful application of this line of research. This brings together the central issue of formative assessment and feedback. Echoing the Bloom's (1984) research on the "two-sigma problem" (i.e., individuals under a one-to-one tutoring situation perform two standard deviations better than individuals in other instructional situations), current research on "what works best" (Hattie, 2009) have identified feedback and formative assessment as important aspects of successful teaching (Köller, 2012). Therefore, understanding what type of cognitive feedback is timely needed is a promising area of research that can inform the design and the development of responsive models of individuals' behavior on a moment-to-moment basis. Likewise, there is a need to also explore conative and affective feedback in situations of relatively high complexity and where individuals engage in high stakes decision making processes. For example, Economides (2009) found that students who received conative feedback showed higher scores during a computer-based assessment environment. The article distinguishes between 1) positive conation feedback and 2) control of negative conation feedback. The former attempts to develop and increase self-awareness, interest, self-efficacy and volition. This line of research should be a priority for scholars interested in developing adaptive systems in the context of educational games.

7. References

Aarseth, E. (2004). Genre trouble: Narrativism and the art of simulation. In N. Warpdrip-Fruin & P. Harrigan (Eds.), *First person. New media as story, performance, and gam*e, (pp. 45-55). Cambridge, MA: MIT Press.

Admiraal, W., Huizenga, J., & Akkerman, S. (2011). The concept of flow in collaborative game-based learning. *Computers in Human Behavior, 27*(3), 1185–1194. doi: 10.1016/j.chb.2010.12.013

American Psychological Association (2009). *Publication manual of the American Psychological Association (6th ed.).* Washington, D.C.: Author.

Aleven, V., Stahl, E., Schworm, S., Fischer, F., & Wallace, R. M. (2003). Help seeking and help design in interactive learning environments. *Review of Educational Research, 73*(2), 277–320. doi:10.3102/00346543073003277

Alkan, S., & Cagiltay, K. (2007). Studying computer game learning experience through eye tracking. *British Journal of Educational Technology, 38*(3), 538–542. doi:10.1111/j.1467-8535.2007.00721.x

Anderson, J.R., Reder, L.M., & Simon, H.A. (1996). Situated learning and education. *Educational Researcher, 25*(4), 5-11. doi: 10.3102/0013189X025004005

Apperley, T. H. (2006). Genre and game studies: Toward a critical approach to video game genres. *Simulation & Gaming, 37*(1), 6–23. doi: 10.1177/1046878105282278

Appleton, J., Christenson, S., Kim, D., & Reschly, A. (2006). Measuring cognitive and psychological engagement: Validation of the student engagement instrument. *Journal of School Psychology, 44*(5), 427-445. Doi:10.1016/j.jsp.2006.04.002

Azevedo, F. S. (2006). Personal excursions: investigating the dynamics of student engagement. *International Journal of Computers for Mathematical Learning*, *11*, 57-98. doi: 10.1007/s10758-006-0007-6

Bandura, A. (1982). Self-efficacy mechanism in human agency. *American Psychologist, 37*, 122-147.

Barab, S, A., & Dede, C. (2007). Games and immersive participatory simulations for science education: An emerging type of curricula. *Journal of Science Education and Technology, 16*(1), 1-3. doi: 10.1007/s10956-007-9043-9

Barab, S, Dodge, T., Ingram-Goble, A., Pettyjohn, P., Peppler, K., Volk, C., & Solomou, M. (2010). Pedagogical dramas and transformational play: Narratively rich games for learning. *Mind, Culture, and Activity, 17*, 235–264. doi: 10.1080/10749030903437228

Barab, S. A., Sadler, T. D., Heiselt, C., Hickey, D., & Zuiker, S. (2007). Relating narrative, inquiry, and inscriptions: Supporting consequential play. *Journal of Science Education and Technology, 16*(1), 59–82. doi: 10.1007/s10956-010-9220-0

Barab, S., Thomas, M., Dodge, T., Carteaux, R., & Tuzun, H. (2005). Making learning fun: Quest atlantis, a game without guns. *Educational Technology Research and Development, 53*(1), 86–107. doi: 10.1007/BF02504859

Bauerlein, M. (2008). *The dumbest generation: How the digital age stupefies young americans and jeopardizes our future (Or, don't trust anyone under 30).* New York, NY: Jeremy P. Tarcher/Penguin.

Becker, K. (2010). Distinctions between games and learning: A review of the literature on games in education. In R. V. Eck (Ed.), *Gaming & cognition: Theories and perspectives from the learning sciences* (pp.22-54). Hershey, PA: IGI Global Publishing.

Beentjes, J.W. (1989). Learning from television and books: A dutch replication study based on Salomon's model. *Educational Technology Research and Development, 37*(2), 47-58. doi: 10.1007/BF02298289

Bender, R., & Lange, S. (2001). Adjusting for multiple testing- when and how? *Journal of Clinical Epidemiology, 54*(4), 343-349. doi: 10.1016/S0895-4356(00)00314-0

Bennett, S., Maton, K., & Kervin, L. (2008). The ‘‘Digital Natives'' debate: A critical review of evidence. *British Journal of Educational Technology, 39*, 775–86. doi: 10.1111/j.1467-8535.2007.00793.x

Bereiter, C., & Scardamalia, M. (1989). Intentional learning as a goal of instruction. In L. B. Resnick (Ed.), *Knowing, learning, and instruction: Essays in honor of Robert Glaser* (pp. 361–392). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Berlyne, D. E. (1960). *Conflict, arousal and curiosity.* New York: McGraw-Hill.

Berthold, K., & Renkl, A. (2010). How to foster active processing of explanations in instructional communication. *Educational Psychology Review, 22*, 25-40. doi:10.1007/s10648-010-9124-9

Bitzer, D.L., & Johnson, R.L. (1971). *Plato. A computer-based system used in the engineering of education*. Proceedings of the IEEE, 59, 6, 960–968.

Björk, S. & Holopainen, J. (2005). *Patterns in game design*. Boston, MA: Charles River Media.

Bloom, B. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one–to–one tutoring. *Educational Researcher, 13*(6), 4–16. doi: 10.3102/0013189X013006004

Blumenfeld, P., Kempler, T., & Krajcik, J. (2006). Motivation and cognitive engagement in learning environments. In R. K. Sawyer (Ed.), *The cambridge handbook of the learning sciences* (pp. 475–488). Cambridge: Cambridge University Press.

Blumenfeld, P., Modell, J., Bartko, W. T., Secada, W. G., Fredricks, J. A., Friedel, J., & Paris, A. (2005). School engagement of inner–city students during middle childhood. In C. R. Cooper, C. T. Garcia Coll, W. T. Bartko, H. Davis, & C. Chatman (Eds.), *Developmental pathways through middle childhood: Rethinking contexts and diversity as resources* (145–170). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Blumenfeld, P. C., Soloway, E., Marx, R. W., Krajcik, J. S., Guzdial, M., & Palincsar, A. (1991). Motivating project-based learning: Sustaining the doing, supporting the learning. *Educational Psychologist, 26*(3), 369–398. doi: 10.1080/00461520.1991.9653139

Bohnert, A., Fredricks, J., & Randall, E. (2010). Capturing unique dimensions of youth organized activity involvement: Theoretical and methodological considerations. *Review of Educational Research, 80*, 576–610. doi: 10.3102/0034654310364533

Bong, M., & Skaalvik, E.M. (2003). Academic self-concept and self-efficacy: How different are they really? *Educational Psychology Review, 15*(4), 1–40. doi: 10.1023/A:1021302408382

Bordeaux, B. R., & Lange, G. (1991). Children's reported investment of mental effort when viewing television. *Communication Research, 18*(5), 617–635. doi: 10.1177/009365091018005003

Boucheix, J.M., & Lowe, R. K. (2010). An eye-tracking comparison of external pointing cues and internal continuous cues in learning with complex animations. *Learning and Instruction, 20*(2), 123–135. doi: 10.1016/j.learninstruc.2009.02.015

Bourg, D. M., & Seemann G. (2004). *AI for game developers—Creating intelligent behavior in games*. Sebastopol, CA: O'Reilly & Associates.

Bowers, V.A., & Snyder, H.L. (1990). Concurrent versus retrospective verbal protocols for comparing window usability. *Proceedings of the Human Factors and Ergonomics Society, USA, 34*, 1270–1274. doi: 10.1177/154193129003401720

Bransford, J. D., & Schwartz, D. L. (1999). Rethinking transfer: A simple proposal with multiple implications. *Review of Research in Education, 24*, 61–100. doi: 10.3102/0091732X024001061

Bransford, J. D., Brown, A. L., & Cocking, R. R. (1999). *How people learn: Brain, mind, experience and school*. Washington, DC: National Academy Press.

Brophy, J.E. (2010). *Motivating students to learn*. New York, NY: Routledge.

Brookhart, S. M., & Durkin, D. T. (2003). Classroom assessment, student motivation, and achievement in high school social studies classes. *Applied Measurement in Education, 16*(1), 27–54. doi: 10.1207/S15324818AME1601_2

Brookhart, S. M., Walsh, J. M., & Zientarski, W. A. (2006). The dynamics of motivation and effort for classroom assessments in middle school science and social studies. *Applied Measurement in Education, 19*, 151–184. doi: 10.1207/s15324818ame1902_5

Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher, 18*(1), 32–42. doi: 10.3102/0013189X018001032

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

Bruner, J. S. (1996). *The culture of education*. Cambridge, MA: Harvard University Press.

Bryant, F. B., & Yarnold, P. R. (1995). Principal components analysis and exploratory and confirmatory factor analysis. In L. G. Grimm & R. R. Yarnold (Eds.*), Reading and understanding multivariate statistics* (pp. 99–136). Washington, DC: American Psychological Association.

Buchanan, K. (2006). *Beyond attention-getters: Designing for deep engagement* (Unpublished doctoral dissertation). Michigan State University, Michigan.

Buckingham, D. (2007). *Beyond technology*. Cambridge: Polity Press.

Byrne, B. (2012). *Structural equation modeling with Mplus: Basic concepts, applications and programming*. New York: Routledge.

Cacioppo, J. T., Petty, R. E., & Morris, K. J. (1983). Effects of need for cognition on message evaluation, recall, and persuasion. *Journal of Personality and Social Psychology, 45*, 805–818. doi: 10.1037/0022-3514.45.4.805

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

Callejas, G. (2007). *Digital games as designed experience: Reframing the concept of immersion* (Unpublished doctoral dissertation).Victoria University of Wellington, New Zealand.

Cennamo, K. S. (1993). Learning from video: Factors influencing learners' preconceptions and invested mental effort. *Educational Technology Research and Development*, *41*(3), 33–45. doi: 10.1007/BF02297356

Cennamo, K. S., Savenye, W. C., & Smith, P. L. (1991). Mental effort and video-based learning: The relationship of preconceptions and the effects of interactive and covert practice. *Educational Technology Research and Development, 39*(1), 5–16. doi: 10.1007/BF02298103

Charles, D., Charles, T., McNeill, M., Bustard, D., & Black, M. (2010). Game-based feedback for educational multi-user virtual environments. *British Journal of Educational Technology, 42*(4), 638–654. doi: 10.1111/j.1467-8535.2010.01068.x

Chein, J. & Schneider, W. (2012). The brain's learning and control architecture. *Current Directions in Psychological Science, 21*(2), 78–84. doi: 10.1177/0963721411434977

Chen, M., Kolko, B. E., Cuddihy, E. and Medina, E. (2011). *Modeling but NOT measuring engagement in computer games*. Paper presented at the seventh International Conference on Games+Learning+Society. Madison, USA. Retrieved from http://dl.acm.org/citation.cfm?id=2206383

Chi, M. T. H. (1997). Quantifying qualitative analyses of verbal data: a practical guide. *Journal of the Learning Sciences, 6*(3), 271–315. doi: 10.1207/s15327809jls0603_1

Clark, R. (1994). Media will never influence learning. *Educational Technology Research and Development*, 42(2), 21–29. doi:10.1007/BF02299088

Clark, R.E. (Ed.). (2001). *Learning from media: Arguments, analysis, and evidence*. Greenwich, CT: Information Age Publishing.

Clark, R.E. (2007). Learning from serious games? Arguments, evidence, and research suggestions. *Educational Technology*, (May-June), 56–59.

Clark, R. E. (2009). How much and what type of guidance is optimal for learning from instruction? In S. Tobias, & T. M. Duffy (Eds.), *Constructivist Theory Applied to Instruction: Success or Failure?* (pp. 158–183). New York: Routledge.

Clark, R. E., Feldon, D. F., Van Merriënboer, J. J. G., Yates, K. A., & Early, S. (2007). Cognitive task analysis. In J. M. Spector, M. D. Merrill, J. J. G. Van Merrienboer & M. P. Driscoll (Eds.), *Handbook of research on educational communications and technology* (pp. 577–594). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Cocea, M. & Weibelzahl, S. (2009). Log file analysis for disengagement detection in e-Learning environments. *User Modeling and User-Adapted Interaction, 19*(4), 341–385. doi: 10.1007/s11257-009-9065-5

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*, 155–159. Retrieved from http://www2.psych.ubc.ca/~schaller/349and449Readings/Cohen1992.pdf

Connolly, T., Boyle, E., MacArthur, E., Hainey, T., & Boyle, J. (2012). A systematic literature review of empirical evidence on computer games and serious games. *Computers & Education, 59*, 661–686. doi: 10.1016/j.compedu.2012.03.004

Cordova, D. I. & Lepper, M. R. (1996). Intrinsic motivation and the process of learning: beneficial effects of contextualization, personalization, and choice. *Journal of Educational Psychology, 88*(4), 715–730. doi: 10.1037/0022-0663.88.4.715

Corno L. (1993). The best-laid plans: modern conceptions of volition and educational research. *Educational Researcher, 22*, 14–22. doi: 10.3102/0013189X022002014

Corno, L., & Mandinach, E. (1983). The role of cognitive engagement in classroom learning and motivation. *Educational Psychologist, 18,* 88–108. doi: 10.1080/00461528309529266

Craik, F. I., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning & Verbal Behavior, 11*, 671–684. doi: 10.1016/S0022-5371(72)80001-X

Crawford, C. (1984). *The art of computer game design.* Berkeley, California: Osborne/McGraw-Hill.

Crookall, D., Oxford, R. L., & Saunders, D. (1987). Towards a reconceptualization of simulation: From representation to reality. *Simulation/Games for Learning, 17*, 147–171. Retrieved from http://www.eric.ed.gov/ERICWebPortal/search/detailmini.jsp?_nfpb=true&_&ERI

CExtSearch_SearchValue_0=EJ363822&ERICExtSearch_SearchType_0=no&accno=EJ363822

Csíkszentmihályi, M. (1990). *Flow: The psychology of optimal experience*. New York: Harper and Row.

de Freitas, S., (2006). *Learning in immersive worlds. A review of game-based learning*. Bristol. Joint information systems committee. Retrieved from http://www.jisc.ac.uk/media/documents/programmes/elearninginnovation/gamingreport_v3.pdf

DeLeeuw, K. E., & Mayer, R. E. (2008). A comparison of three measures of cognitive load: Evidence for separable measures of intrinsic, extraneous, and germane load. *Journal of Educational Psychology, 100*, 223–234. doi: 10.1037/0022-0663.100.1.223

de Jong, T. (1991). Learning and instruction with computer simulations. *Education & Computing, 6*, 217–229. doi: 10.1016/0167-9287(91)80002-F

de Jong, T. & Joolingen, W.R. van (1998). Scientific discovery learning with computer simulations of conceptual domains. *Review of Educational Research, 68*, 179-202. doi: 10.3102/00346543068002179

Dempsey, J. V., Rasmussen, K. & Lucassen, B. (1994). *Instructional gaming: Implications for instructional technology*. Paper presented at the annual meeting of the Association for Educational Communications and Technology, Nashville, TN.

De Rivecourt, M., Kuperus, M., Post, W., & Mulder, L. (2008). Cardiovascular and eye activity measures as indices for momentary changes in mental effort during simulated flight. *Ergonomics*, *51*(9), 1295–1319. doi: 10.1080/00140130802120267

Dewey, J. (1913). *Interest and effort in education*. Carbondale: Southern Illinois University Press.

Dewey, J. (1938). *Experience and education*. New York: Collier Books.

diSessa, A. A. (2000). *Changing minds: Computers, learning, and literacy*. Cambridge, MA: The MIT Press.

Dickey, M.D. (2005). Engaging by design: how engagement strategies in popular computer and video games can inform instructional design. *Education Training Research and Development, 53*(2), 67–83. doi: 10.1007/BF02504866

Dow, S., Mehta, M., Harmon, E., MacIntyre, B., & Mateas, M. (2007). Presence and engagement in an interactive drama. *Proceedings of the SIGCHI Conference on Human factors in Computing Systems (CHI), USA*, 1475-1484. Doi:10.1145/1240624.1240847

Duchowski, A. (2007). *Eye tracking methodology: theory and practice*. London: Springer-Verlag.

Dugdale, S., & Kibbey, D. (1975). *Fractions curriculum of the PLATO elementary school mathematics project*. Computer-based Education Research Laboratory Technical Report. University of Illinois, Urbana, Ill.

Economides, A. A. (2009). Conative feedback in computer-based assessment. *Computers in the Schools, 26*(3), 207–223. doi: 10.1080/07380560903095188

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

Egenfeldt-Nielsen, S. (2007). Third generation education use of computer games. *Journal of Educational Multimedia and Hypermedia, 16*(3), 263–281. Retrieved from http://www.editlib.org/p/24375?nl

Eisenhart, M. (1991). Conceptual frameworks for research circa 1991: Ideas from a cultural anthropologist; Implications for mathematics education researchers. In R. Underhill (Ed.), *Proceedings of the Thirteenth Annual Meeting of the International Group for the Psychology of Mathematics Education* (pp.202–219), Blacksburg, VA.

Elliot, A. J., & Harackiewicz, J. M. (1994). Goal setting, achievement orientation, and intrinsic motivation: A mediational analysis. *Journal of Personality and Social Psychology*, *66*(5), 968–980. doi: 10.1037/0022-3514.66.5.968

Engeser, S., & Rheinberg, F. (2008). Flow, performance and moderators of challenge-skill balance. *Motivation and Emotion, 32*(3), 158–172. doi:10.1007/s11031-008-9102-4

Engle, R., & Conant, F. (2002). Guiding principles for fostering productive disciplinary engagement: Explaining an emergent argument in a community of learners' classroom. *Cognition and Instruction, 20*(4), 399–483. doi: 10.1207/S1532690XCI2004_1

Ennemoser, M. (2009). Evaluating the potential of serious games. What can we learn from previous research on media effects and educational intervention? In U. Ritterfeld, M. Cody, & P. Vorderer (Eds.), *Serious games: Mechanisms and effects* (pp. 344–373). Mahwah, New Jersey: Routledge/LEA.

Federation of American Scientists (2006). *Summit on Educational Games: Harnessing the power of video games for learning.* Washington, DC: Author. Retrieved from http://www.fas.org/programs/ltp/policy_and_publications/summit/Summit%20on%20Educational%20Games.pdf

Filsecker, M., Bormann, M., & Kerres, M. (2011, April). *Avoiding learning in game-based learning environments: Gaze patterns at tasks' vs. play elements*. Paper at the annual meeting of the American Educational Research Association, New Orleans, LA.

Filsecker, M. & Hickey, DT. (submitted). Digital game-based learning in elementary science education: The effects of incentives on students' engagement and ecology learning. *Manuscript submitted for publication*.

Filsecker, M. & Kerres, M. (2013). Designing and studying educational games: Limitations of current design and research approaches in game-based learning. In B. Bigl and S. Stoppe (Eds.), *Playing with virtuality, Theories and methods of computer game studies* (pp. 349–368). Frankfurt: Peter Lang.

Foerde, K., Knowlton B.J., Poldrack R.A. (2006). Modulation of competing memory systems by distraction. *Proceedings of the National Academy of Sciences, USA, 103,* 11778–11783. doi: 10.1073/pnas.0602659103

Fortugno, N., & Zimmerman, E. (2005). Learning to play to learn – lessons in educational game design. *Gamasutra*. Retrieved from http://www.gamasutra.com/view/feature/2273/soapbox_learning_to_play_to_learn_.php

Frasca, G. (2001). Simulation 101: Simulation versus representation [Website]. Retrieved from http://www.ludology.org/articles/sim1/simulation101.html

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research, 74*(1), 59–109. doi: 10.3102/00346543074001059

Garris, R., Ahlers, R., & Driskell, J. E. (2002). Games, motivation, and learning: A research and practice model. *Simulation & Gaming, 33*(4), 441–467. doi: 10.1177/1046878102238607

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

Gee, J. P. (2005). Learning by design: Good video games as learning machines. *E-Learning and Digital Media, 2*(1), 5–16. Retrieved from http://www.wwwords.co.uk/elea/content/pdfs/2/issue2_1.asp#2

Gee, J.P. (2011). Reflections on empirical evidence on games and learning. In S. Tobias & J.D. Fletcher (Eds.), *Computer games and instruction* (pp.223–231). Charlotte, NC.: Information Age Publishing.

Girard, C., Ecalle, J., & Magnan, A. (2012). Serious games as new educational tools: how effective are they? A meta-analysis of recent studies. *Journal of Computer Assisted Learning, 29*(3), 207–219. doi: 10.1111/j.1365-2729.2012.00489.x

Glaser, M., Garsoffky, B., & Schwan, S. (2012). What do we learn from docutainment? Processing hybrid television documentaries. *Learning and Instruction, 22*, 37–46. doi:10.1016/j.learninstruc.2011.05.006

Gredler, M. E. (1996). Educational games and simulations: A technology in search of a (research) paradigm. In D. H. Jonassen (Ed.), *Handbook of research for educational communications and technology: A project of the association for educational communications and technology* (pp. 521–540). New York: Simon & Schuster Macmillan.

Gredler, M. E. (2003). Games and simulations and their relationships to learning. In D. H. Jonassen (Ed.), *Handbook of research for educational communications and technology* (pp. 571–581). Mahwah, New Jersey: Lawrence Erlbaum, Inc.

Greenblat, C. S. (1981). Teaching with simulation games: A review of claims and evidence. In C. S. Greenblat & R. D. Duke (Eds.), *Principles and practices of gaming-simulation* (pp. 139–153). Beverly Hills, CA: Sage Publications.

Greenfield, P. M. (1984). *Mind and media: The effects of television, video games, and computers*. Cambridge, MA: Harvard University Press.

Greene, B.A., Miller, R.B., Crowson, H.M., Duke, B.L., & Kristine, L.A. (2004). Predicting high school students' cognitive engagement and achievement:

Contributions of classroom perceptions and motivation. *Contemporary Educational Psychology, 29*, 462–482. doi:10.1016/j.cedpsych.2004.01.006

Greeno, J.G. (1997). On claims that answer the wrong questions. *Educational Researcher, 26*(1), 5–17. doi: 10.3102/0013189X026001005

Greeno, J. G. (1998). The situativity of knowing, learning, and research. *American Psychologist, 53*(1), 5–26. Retrieved from http://psycnet.apa.org/index.cfm?fa=buy.optionToBuy&id=1997-42695-001

Greeno, J.G. (2006). Learning in activity. In R.K. Sawyer (Ed.), *The cambridge handbook of the learning sciences* (pp. 79–96). New York: Cambridge University Press.

Gregoire, M., Ashton, P., & Algina, J. (2001, April). *The role of prior and perceived ability in influencing the relationship of goal orientation to cognitive engagement and academic achievement.* Paper presented at the Annual Meeting of the American Educational Research Association, Seattle, WA.

Gresalfi, M., Barab, S., Siyahhan, S., & Christensen, T. (2009). Virtual worlds, conceptual understanding, and me: designing for consequential engagement. *On the Horizon, 17*(1), 21–34. doi:10.1108/10748120910936126

Guan, Z., Lee, S., Cuddihy, E., & Ramey, J. (2006). The validity of the stimulated retrospective think-aloud method as measured by eye tracking. *Proceedings of the SIGCHI conference on human factors in computing systems, USA*, 1253–1262. doi:10.1145/1124772.1124961

Gunter, G. A., Kenny, R. F., & Vick, E. H. (2008, October). 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. doi:10.1007/s11423-007-9073-2

Habgood, M.P. (2007). *The effective integration of digital games and learning content* (Unpublished doctoral dissertation). University of Nottingham, Nottingham.

Habgood, M. P. & Ainsworth, S.E. (2011). Motivating children to learn effectively: Exploring the value of intrinsic integration in educational games. *The Journal of the Learning Sciences 20*(2), 169–206. doi: 10.1080/10508406.2010.508029

Habgood, M. P., Ainsworth, S.E., & Benford, S. (2005). Endogenous Fantasy and Learning in Digital Games. *Simulation & Gaming 36*(4), 483–498. doi: 10.1177/1046878105282276

Hancock, G. R., & Mueller, R. O. (2001). Rethinking construct reliability within latent systems. In R. Cudeck, S. D. Toit, & D. Sörrbom (Eds.), *Structural equation modeling: Present and future—a festschrift in honor of Karl Jöreskog* (pp. 195–216). Lincolnwood, IL: Scientific Software International.

Hannafin, M.J. (1989). Interaction strategies and emerging instructional technologies: Psychological perspectives. *Canadian Journal of Education Communication, 18*(3), 167–179. Retrieved from http://cjlt.csj.ualberta.ca/index.php/cjlt/article/view/274/208

Hartman, I., & Klimmt, C. (2006). Gener and computer games: Exploring females' dislikes. *Journal of Computer-Mediated Communication, 11*, 910–931. doi: 10.1111/j.1083-6101.2006.00301.x

Hattie, J. (2009). *Visible learning: A synthesis of over 800 Meta-Analyses relating to achievement*. London: Routledge.

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

Heers, R. (2005). *Being There: Untersuchungen zum Wissenserwerb in virtuellen Umgebungen* (Unpublished doctoral dissertation). Eberhard-Karls-University, Tübingen.

Hickey, D. T., & Kindfield, A. (1999). *Assessment-oriented scaffolding of student and teacher performance in a technology-supported genetics environment*. Paper presented at the annual meeting of the American Educational Research Association, Montreal, Quebec, Canada.

Hickey, D.T., Ingram-Goble, A., & Jameson, E. (2009). Designing Assessments and Assessing Designs in Virtual Educational Environments. *Journal of Science and Educational Technology, 18*, 187–208. doi: 10.1007/s10956-008-9143-1

Hmelo-Silver, C. E. (2004). Problem-based learning: What and how do students learn? *Educational Psychology Review, 16*(3), 235–266. doi: 10.1023/B:EDPR.0000034022.16470.f3

Hoeft, F., Watson, C.L., Kesler, S.R., Bettinger, K.E., & Reiss, A.L. (2008). Gender differences in the mesocorticolimbic system during computer game-play. *Journal of Psychiatry Research 42,* 253–258.doi: 10.1016/j.jpsychires.2007.11.010

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van de Weijer, J. (2011). *Eye Tracking - A Comprehensive Guide to Methods and Measures*. New York: Oxford University Press.

Holsanova, J., Rahm, H., & Holmqvist, K. (2006). Entry points and reading paths on the newspaper spread: Comparing semiotic analysis with eye-tracking measurements. *Visual Communication, 5*, 65–93. doi: 10.1002/acp.1525

Honey, M. A., & Hilton, M. (eds.). (2011). *Learning science through computer games and simulations*. Washington, DC. National Academies Press.

Howard, D.C. (1989). *Variations of cognitive engagement as evidence of self-regulated learning* (Unpublished doctoral dissertation). Simon Fraser University, Vancouver.

Howard-Jones, P.A., Demetriou, S., Bogacz, R., Yoo, J., & Leonards, U. (2011).Toward a science of learning games. *Mind, Brain, and Education*, *5*, 33 – 41. doi:10.1111/j.1751-228X.2011.01108.x

Huang, W. H. (2010). Evaluating learners' motivational and cognitive processing in an online game-based learning environment. *Computers in Human Behavior, 27(2)*, 694–704. doi:10.1016/j.chb.2010.07.021

Huizinga, J. (1949). *Homo ludens: A study of the play-element in culture*. New York, NY: Routledge & Kegan Paul Ltd.

Jantke, K.P. (2006). Digital games that teach: A critical analysis. TU Ilmenau, IfMK, Diskussionsbeiträge. Retrieved from http://www.db-thueringen.de/servlets/DerivateServlet/Derivate-14245/Diskussionsbeitrag_22_Digital%20Games%20that%20each_Jantke.pdf

Järvelä, S. & Salovaara, H. (2004). The interplay of motivational goals and cognitive strategies in a new pedagogical culture – a context oriented and qualitative approach. *European Psychologist, 9*, 4, 232–244.doi: 10.1027/1016-9040.9.4.232

Järvelä, S., Veermans, M. & Leinonen, P. (2008). Investigating students' engagement in a computer-supported inquiry – a process-oriented analysis. *Social Psychology in Education, 11*, 299–322. doi: 10.1007/s11218-007-9047-6

Jennett, C., Cox, A. L., Cairns, P., Dhoparee, S., Epps, A., Tijs, T., et al. (2008). Measuring and defining the experience of immersion in games. *International Journal of Human- Computer Studies, 66*, 641–661. doi:10.1016/j.ijhcs.2008.04.004

Jimerson, S. R., Campos, E., & Greif, J. L. (2003). Toward an understanding of definitions and measures of school engagement and related terms. *California School Psychologist, 8*, 7–27. Retrieved from http://casel.org/

Jones, B., Valdez, G., Norakowski, J., & Rasmussen, C. (1994). *Designing learning and technology for educational reform*. Oak Brook, IL: North Central Regional Educational Laboratory.

Judd, T. & Kennedy, G. (2011). Measurement and evidence of computer-based task switching and multitasking by 'Net Generation' students. *Computers & Education, 56,* 625–631. doi:10.1016/j.compedu.2010.10.004

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixation to comprehension. *Psychological Review, 87*(4), 329–354. doi: 10.1037/0033-295X.87.4.329

Juul, J. (1998, November). *A clash between game and narrative*. Paper presented at the Digital Arts and Culture Conference, Bergen, Norway.

Juul, J. (2001). Games telling stories?—A brief note on games and narratives. *Game Studies: The International Journal of Computer Game Research, 1(1).* Retrieved from http://www.gamestudies.org/0101/juul-gts/

Juul, J. (2005). *Half-real: Video games between real rules and fictional worlds*. London, England: the MIT Press.

Kahneman, D. (1973). *Attention and effort*. New York: Prentice Hall.

Kallinen, K., Salminen, M., Ravaja, N., Kedzior, R., Sääksjärvi, M. (2007). Presence and emotion in computer game players during 1st person vs. 3rd person playing view: Evidence from self-report, eye-tracking, and facial muscle activity data. *Proceedings of the 10th International Workshop on Presence*, Barcelona, Spain.

Kalyuga, S., & Plass, J. (2009). Evaluating and managing cognitive load in games. In R. E. Ferdig (Ed.), *Handbook of research on effective electronic gaming in education* (pp. 719–737). Hershey, PA: Information Science Reference.

Kebritchi, M. (2008). *Effects of a computer game on mathematics achievement and class motivation: An experimental study* (Unpublished doctoral dissertation). University of Central Florida, Florida.

Kebritchi, M., Hirumi, A., & Bai, H. (2010). The effects of modern mathematics computer games on mathematics achievement and class motivation. *Computers & Education*, *55*(2), 427–443. doi:10.1016/j.compedu.2010.02.007

Kerres, M., Bormann, M., & Vervenne, M. (2009). Didaktische Konzeption von Serious Games: Zur Verknüpfung von Spiel-und Lernangeboten. *MedienPädagogik. Zeitschrift für Theorie und Praxis der Medienbildung*. Retrieved from http://www.medienpaed.com/2009/kerres0908.pdf

Kerres, M., Ojstersek, N., & Stratmann, J. (2009). Didaktische Konzeption von Angeboten des Online-Lernens. In L. Issing & P. Klimsa (Eds.), *Online_Lernen: Handbuch für wissenschaft und praxis* (pp. 263–271). München:Oldenbourg Verlag.

Ketelhut, D. J. (2004). *Assessing science self-efficacy in a virtual environment: A measurement pilot* (Unpublished qualifying paper). Harvard University, Cambridge, MA.

Kickmeier-Rust, M. D., Hillemann, E., & Albert, D. (2011). Tracking the UFO's paths: Using eye-tracking for the evaluation of serious games. In R. Shumaker (Ed.), *Virtual and mixed reality - New trends* (pp. 315–324). Lecture Notes in Computer Science, 6773. Berlin: Springer.

Kiili, K. (2005). Digital game-based learning: Towards an experiential gaming model. *The Internet and Higher Education, 8*(1), 13–24. doi:10.1016/j.iheduc.2004.12.001

Kirriemuir, J. & McFarlane, A. (2004). *Literature review in games and learning*. (Report No. 8). Bristol: Nesta Futurelab.

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist, 41*, 75–86. doi: 10.1207/s15326985ep4102_1

Klopfer, E., Osterweil, S. & Salen, K. (2009). *Moving learning games forward: Obstacles, opportunities, & openness.* The Education Arcade: MIT. Retrieved from http://education.mit.edu/papers/MovingLearningGamesForward_EdArcade.pdf

Köller, O. (2012). What works best in school? Hatties Befunde zu Effekten von Schul- und Unterrichtsvariablen auf Schulleistungen. *Psychologie in Erziehung und Unterricht, 59*, 72 – 78.

Kolb, D. A. (1984). *Experiential learning: experience as the source of learning and development*. Englewood Cliffs, N.J: Prentice-Hall.

Koole, S. L., Kuhl, J., Jostmann, N., & Vohs, K. D. (2005). On the hidden benefits of state orientation: Can people prosper without efficient affect regulation skills? In A.

Tesser, J. Wood, & D. A. Stapel (Eds.), *On building, defending, and regulating the self: A psychological perspective* (pp. 217–243). London: Taylor & Francis.

Kuhl., J. (1987). Action control: the maintenance of motivational states. In F. Halisch and J. Kuhl (Eds.), *Motivation, intention, and volition* (279–307). Berlin: Springer-Verlag.

Kuhl, J. & Beckmann, J. (Eds.). (1985). *Action control: From cognition to behavior.* Berlin: Springer.

Kunkel, D., & Kovaric, P. (1983). *Mental effort and learning from TV: comparing expectations for PBS and commercial network programming*. Paper presented at the annual conference of the International Communication Association, Dallas, Texas.

Kreimeier, B. (2002, March). The case for game design patterns. *Gamasutra*. Retrieved from http://www.gamasutra.com/features/20020313/kreimeier_01.htm

Lankoski, P. (2011). Player character engagement in computer games. *Games and Culture*, *6*, 291–311. doi: 10.1177/1555412010391088

Larson-Hall, J. (2010). *A guide to doing statistics in second language research using SPSS*. New York and London: Routledge.

Läuter, J. (1978). Sample size requirements for the T^2 test of MANOVA (Tables for one-way classification). *Biometrical Journal, 20*(4), 389-406. doi: 10.1002/bimj.4710200410

Law, E. L.-C, Mattheiss, E., Kickmeier-Rust, M. D., & Albert, D. (2010). Vicarious learning with a digital educational game: Eye-tracking and survey-based evaluation approaches. In G. Leitner, M. Hitz & A. Holzinger (Eds.), *Lecture Notes in Computer Science - HCI in Work and Learning, Life and Leisure* (Vol. 6389, pp. 471–488). Berlin: Springer.

Lee, J. (1999). Effectiveness of computer-based instructional simulation: A meta analysis. *International Journal of Instructional Media, 26*(1), 71–85. Retrieved from http://www.questia.com/library/p533/international-journal-of-instructional-media/i2627014/vol-26-no-1-winter

Leemkuil, H., & de Jong, T. (2011). Instructional support in games. In S. Tobias & J.D. Fletcher (Eds.), *Computer games and instruction* (pp. 353-370). Charlotte, NC.: Information Age Publishing.

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

MacKinnon, D. P. (2008). *Introduction to statistical mediation analysis*. New York, NY: Erlbaum

MacCallum, R.C., Widaman, K.F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods, 4*(1), 84–99. doi: 10.1037/1082-989X.4.1.84

Maietta, R. (2008). *Media Review: MAXQDA 2007*. Marburg, Germany: Verbi Software. http://www.maxqda.com. *Journal of Mixed Methods Research*, *2*, 193–198. doi: 10.1177/1558689807314014

Malone, T. W. (1981). Toward a theory of intrinsically motivating instruction. *Cognitive Science*, *5*(4), 333–369. doi:10.1016/S0364-0213(81)80017-1

Malone, T. W., & Lepper, M. R. (1987). Making learning fun: A taxonomy of intrinsic motivations for learning. In R. E. Snow & M. J. Farr (Eds.), *Aptitude, learning and instruction: Vol. 3. Cognitive and affective process analysis* (pp. 223–253). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Marshall, S. P. (2005). Assessing cognitive engagement and cognitive states from eye metrics. In D. Schmorrow (Ed.), *Foundations of augmented cognition* (312–320). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Marton, F. M., & Säljö, R. (2005). Approaches to learning. In F. M. Marton, D. Hounsell, & N.J. Entwistle (Eds.), *The experience of learning: Implications for teaching and studying in higher education* (pp. 106–125). Edinburgh: Centre for Teaching, Learning and Assessment, University of Edinburgh.

Mattheiss, E., Kickmeier-Rust, M., Steiner, C., & Albert, D. (2009). Motivation in game-based learning: It’s more than „flow“. In A. Schwill, & N. Apostolopoulos (Eds.), *Lernen im Digitalen Zeitalter - Workshop-Band Dokumentation der Pre-Conference zur DeLFI2009 Die 7. E-Learning Fachtagung Informatik der Gesellschaft für Informatik e.V*(pp. 77–84).

Mayer, R. E. (2010). Unique contributions of eye-tracking research to the study of learning with graphics. *Learning and Instruction, 20*(2), 167–171. doi: 10.1016/j.learninstruc.2009.02.012

Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist, 38*(1), 43–52. doi: 10.1207/S15326985EP3801_6

McCrudden, M. T. (2011). Do specific relevance instructions promote transfer appropriate processing? *Instructional Science, 39*, 865–879. doi: 10.1007/s11251-010-9158-x

McCrudden, M. T., & Schraw, G. (2007). Relevance and goal-focusing in text processing. *Educational Psychology Review, 19*, 113–139. doi: 10.1007/s10648-006-9010-7

Meece, J.L., Glienke, B. B., & Burg, S. (2006). Gender and motivation. *Journal of School Psychology, 44*, 351–373. doi: 10.1016/j.jsp.2006.04.004

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning & Verbal Behavior, 16,* 519–533. doi: 10.1016/S0022-5371(77)80016-9

Muir, M. & Conati, C. (2012). An analysis of attention to student-adaptive hints in an educational game. In S.A. Cerri, W.J. Clancey, G. Papadourakis, K. Panourgia (Eds.), *Lecture notes in computer science* (pp. 112–122). Berlin: Springer-Verlag.

Mulder, G. (1986). The concept and measurement of mental effort. In G. R. J. Hockey, A. W. K. Gaillard, M. G. H. Coles (Eds.), *Energetical issues in research on human information processing* (pp. 175–198). Dordrecht, The Netherlands: Martinus Nijhoff.

Muldner, K., Burleson, W., van De Sande, B., & Vanlehn, K. (2011). An analysis of students' gaming behaviors in an intelligent tutoring system: Predictors and impacts. *User Modeling and User-Adapted Interaction, 21*(1–2), 99–135. doi:10.1007/s11257-010-9086-0

Murphy, K.R. & Myors, B. (2004) *Statistical power analysis. A simple and general model for traditional and modern hypothesis tests*. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Nacke, L., & Lindley, C. (2008). *Boredom, immersion, flow - A pilot study investigating player experience*. IADIS Gaming 2008: Design for Engaging Experience and Social Interaction, Amsterdam, The Netherlands.

Nacke, L. E., Drachen, A. & Goebel, S. (2010). Methods for evaluating gameplay experience in a serious gaming context. *International Journal of Computer Science, 9*(2), 1–12.

Nacke, L., Stellmach, S., Sasse, D., Niesenhaus, J., Dachselt, R. (2011). LAIF: A logging and interaction framework for gaze-based interfaces in virtual entertainment environments. *Entertainment Computing, 2*, 265–273. doi:10.1016/j.entcom.2010.09.004

Narvaez, D., van den Broek, P., & Ruiz, A. B. (1999). The influence of reading purpose on inference generation and comprehension in reading. *Journal of Educational Psychology, 91*(3), 488–496. doi: 10.1037/0022-0663.91.3.488

Nelson, B. (2007). Exploring the use of individualized, reflective guidance in an educational multi-user virtual environment. *Journal of Science Education, 16*(1), 83–97. doi: 10.1007/s10956-006-9039-x

Nelson, B. & Erlandson, B. (2008). Managing cognitive load in educational multi-user virtual environments: reflection on design practice. *Educational Technology Research and Development, 56*, 619–641. doi: 10.1007/s11423-007-9082-1

Newmann, F., Wehlage, G. G., & Lamborn, S. D. (1992). The significance and sources of student engagement. In F. Newmann (Ed.), *Student engagement and achievement in American secondary schools* (pp. 11–39). New York: Teachers College Press.

Olsson, P. (2007). *Real-time and offline filters for eye tracking* (Unpublished master thesis). KTH Royal Institute of Technology, Sweden.

O'Neil, H. F., Wainess, R., & Baker, E. L. (2005). Classification of learning outcomes: Evidence from the computer games literature. *The Curriculum Journal, 16*(4), 455–474. doi: 10.1080/09585170500384529

Paas, F., Tuovinen, J. E., Tabbers, H., &Van Gerven, P. W. M. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist, 38*, 63–71. doi: 10.1207/S15326985EP3801_8

Papert, S. (1998). Does easy do it? *Games Developer*. Retrieved from http://www.papert.org/articles/Doeseasydoit.html

Pavlas, D. (2010). *A model of flow and play in game-based learning: The impact of game characteristics, player traits, and player states* (Unpublished doctoral dissertation). Florida University, Florida.

Pavlas, D., Bedwell, W., Wooten, S., Heyne, K., & Salas, E. (2009). Investigating the attributes in serious games that contribute to learning. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *USA, 53,* 1999-2003. 10.1177/154193120905302705

Pea, R. (2004). The social and technological dimensions of scaffolding and related theoretical concepts for learning, education and human activity. *Journal of the Learning Sciences, 13*(3), 423–451.doi: 10.1207/s15327809jls1303_6

Perkins, D. N. (1985). The fingertip effect: How information-processing technology shapes thinking. *Educational Researcher, 14*(7), 11 –17. doi: doi: 10.3102/0013189X014007011

Perneger, T.V. (1998). What's wrong with bonferroni adjustments. *British Medical Journal*, 316, 1236–1238. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1112991/

Petko, D. (2009). Unterrichten mit Computerspielen. Didaktische Potenziale und Ansätze für den gezielten Einsatz in Schule und Ausbildung. *MedienPädagogik. Zeitschrift für Theorie und Praxis der Medienbildung*. Retrieved from http://www.medienpaed.com/15/petko0811.pdf

Pintrich, P. R., & Garcia, T. (1991). Student goal orientation and self-regulation in the college. In M. Maehr & P. R. Pintrich (Eds.), *Advances in motivation and achievement: Goals and self-regulatory processes* (pp. 271–402). Greenwich, CT: JAI Press.

Pintrich, P., & De Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology, 82*(1), 33–40. doi: 10.1037/0022-0663.82.1.33

Pintrich, P. R., & Schrauben, B. (1992). Students motivational beliefs and their cognitive engagement in classroom academic tasks. In D. H. Schunk & J. L. Meece (Eds.), *Student perception in the classroom* (pp. 149–184). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Pintrich, P. R., & Schunk, D. H. (2002). *Motivation in education: Theory, research, and applications*. Upper Saddle River, NJ: Pearson Education.

Plass, J.L., O'Keefe, P.A., Homer, B.D., Hayward, E.O., Stein, M., & Perlin, K. (2011, April). *Motivational and Educational Outcomes Associated with Solo, Competitive, and Collaborative Game Play.* Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA.

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

Prensky, M. (2001b). Digital natives, digital immigrants. *On the Horizon, 9*(5), 1–6. doi: 10.1108/10748120110424816

Prensky, M. (2011). Comments on research comparing games to other instructional methods. In S. Tobias & J.D. Fletcher (Eds.), *Computer games and instruction* (pp.251–278). Charlotte, NC.: Information Age Publishing.

Randel, J. M., Morris, B. A., Wetzel, C. D., & Whitehill, B. V. (1992). The effectiveness of games for educational purposes: A review of recent research. *Simulation & Gaming, 23*(3), 261–276. doi: 10.1177/1046878192233001

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin 124*(3), 372–422. doi: 10.1037/0033-2909.124.3.372

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology, 62*, 1457–1506. doi: 10.1080/17470210902816461

Renkl, A. (1997, April). *Intrinsic motivation, self-explanations, and transfer*. Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, IL.

Richardson, J., & Newby, T.J. (2006). The role of students' cognitive engagement in online learning. *American Journal of Distance Education, 20*(1), 23–37. doi: 10.1207/s15389286ajde2001_3

Rieber, L. P. (1996). Seriously considering play: Designing interactive learning environments based on the blending of microworlds, simulations, and games. *Educational Technology Research and Development, 44*, 43–58. doi:10.1007/BF02300540

Rieh, S. Y., Kim, Y. M., & Markey, K. (2012). Amount of invested mental effort (AIME) in online searching. *Information Processing and Management, 48*(6), 1136–1150. doi: 10.1016/j.ipm.2012.05.001

Ritterfeld, U., Cody, M., & Vorderer, P. (Eds.), (2009). *Serious games: Mechanisms and effects*. Mahwah, NJ: Routledge, Taylor and Francis.

Rogers, D., & Swan, K. (2004). Self-regulated learning and internet searching. *Teachers College Record, 106*(9), 1804–1824. Retrieved from http://www.tcrecord.org/library/abstract.asp?contentid=11671

Rose, N.S. & Craik F.I.M. (2012). A processing approach to the working memory/long-term memory distinction: Evidence from the levels-of-processing span task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*(4), 1019–1029. doi: 10.1037/a0026976

Rosenthal, R., Rosnow, R. L. (2009). *Artifacts in behavioral research*. New York: Oxford University Press.

Rotgans, J. I., & Schmidt, H. G. (2011). Cognitive engagement in the problem-based classroom. *Advances in Health Sciences Education, 16*(4), 465–479. doi:10.1007/s10459-011-9299-y

Salen, K., & Zimmerman, E. (2003). *Rules of play. Game design fundamentals.* Cambridge, MA: MIT Press.

Säljö, R. (1979). *Learning in the learner's perspective 1. Some commonsense conceptions*. Gothenburg, Sweden: Institute of Education, University of Gothenburg.

Salomon, G. (1981). Introducing AIME: The assessment of children's mental involvement with television. In H. Gardner & H. Kelly (Eds.), *New directions in child development. Vol. 13. Viewing children through television* (pp. 89–102). San Francisco, CA: Jossey Bass.

Salomon, G. (1983). The differential investment of mental effort in learning from different sources. *Educational Psychologist, 18*(1), 42–50. doi: 10.1080/00461528309529260

Salomon, G. (1984). Television is "Easy" and print is "Tough": The differential investment of mental effort in learning as a function of perceptions and attributions. *Journal of Educational Psychology, 76*(4), 647–658. doi: 10.1037/0022-0663.76.4.647

Salomon, G., & Almog, T. (1998). Educational psychology and technology: A matter of reciprocal relations. *The Teachers College Record, 100*(2), 222–241. Retrieved from http://www.tcrecord.org/content.asp?contentid=10310

Salomon, G., & Globerson, T. (1987). Skill is not enough: The role of mindfulness in learning and transfer. *International Journal of Educational Research, 11*, 623–637. doi: 10.1016/0883-0355(87)90006-1

Salomon, G., & Leigh, T. (1984). Predispositions about learning from print and television. *Journal of Communication, 34*(1), 119–135. doi: 10.1111/j.1460-2466.1984.tb02164.x

Salomon, G., Globerson, T., & Guterman, E. (1989). The computer as a zone of proximal development: Internalizing reading-related metacognitions from a reading partner. *Journal of Educational Psychology, 81* (4), 620–627. doi: 10.1037/0022-0663.81.4.620

Salomon, G., & Perkins, D. N. (1989). Rocky roads to transfer: Rethinking mechanism of a neglected phenomenon. *Educational Psychologist, 24*(2), 113–142. doi: 10.1207/s15326985ep2402_1

Salomon, G., Perkins, D. N., & Globerson, T. (1991). Partners in cognition: Extending human intelligence with intelligent technologies. *Educational researcher, 20*(3), 2–9. doi: 10.3102/0013189X020003002

Sankoh, A.J., Huque, M.F., & Dubey, S.D. (1997). Some comments on frequently used multiple endpoint adjustment methods in clinical trials. *Statistics in Medicine, 16*(22), 2529–2542. doi: 10.1002/(SICI)1097-0258(19971130)16:22<2529::AID-SIM692>3.0.CO;2-J

Sarter, M., Gehring, W. J., & Kozak, R. (2006). More attention must be paid: The neurobiology of attentional effort. *Brain Research Reviews, 51*, 145–160. doi:10.1016/j.brainresrev.2005.11.002

Savery, J., & Duffy, T. (1996). Problem based learning: An instructional model and its constructivist framework. In B. Wilson (Ed.), *Constructivist learning environments: Case studies in instructional design* (pp. 135–148). Englewood Cliffs, NJ: Educational Technology Publications.

Scardamalia, M., & Bereiter, C. (1993). Technologies for knowledge-building discourse. *Commun. ACM, 36*(5), 37–41. doi:10.1145/155049.155056

Schank, R.C. and Abelson, R.P. (1977) *Scripts, plans, goals and understanding: An inquiry into human knowledge structures*. Hilllsdale, New Jersey: Lawrence Erblaum Associates.

Schneider, W. and Shiffrin, R.M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review, 84*(1), 1–66. doi: 10.1037/0033-295X.84.1.1

Schlechty, P. C. (1997). *Inventing better schools: An action plan for educational reform*. San Francisco, CA: Jossey-Bass.

Schoenau-Fog, H. (2011, September). *The player engagement process - an exploration of continuation desire in digital games*. Paper presented at the Think Design Play: Digital Games Research Conference, Hilversum, Holland.

Schrader, C., & Bastiaens, T. J. (2012). The influence of virtual presence: Effects on experienced cognitive load and learning outcomes in educational computer games. *Journal of Computers in Human Behavior, 28*, 648–658. doi: 10.1016/j.chb.2011.11.011

Schuurink, E.L. & A. Toet (2010). Effects of third person perspective on affective appraisal and engagement: Findings from second life. *Simulation and Gaming, 41*(5), 724–742. doi: 10.1177/1046878110365515

Schweinle, A., Turner, J. C., & Meyer, D. K. (2006). Striking the right balance: Students' motivation and affect in upper elementary mathematics classes. *Journal of Educational Research, 99*(5), 271–293. Doi: 10.3200/JOER.99.5.271-294

Seufert, T., Schütze, M., & Brünken, R. (2009). Memory characteristics and modality in multimedia learning: An aptitude–treatment–interaction study. *Learning and Instruction, 19*(1), 28–42. doi: 10.1016/j.learninstruc.2008.01.002

Serrano, E. L., & Anderson, J. E. (2004). The evaluation of food pyramid games, a bilingual computer nutrition education program for Latino youth. *Journal of Family and Consumer Sciences Education, 22*(1), 1–16. Retrieved from http://www.natefacs.org/JFCSE/v22no1/v22no1Serrano.pdf

Shaffer, D. W. (2007). *How computer games help children learn*. New York: Palgrave Macmillan.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin.

Shernoff, D. J., Csikszentmihalyi, M., Schneider, B., & Shernoff, E. S. (2003). Student engagement in high school classrooms from the perspective of flow theory. *School Psychology Quarterly, 18*(2), 158–176. doi: 10.1521/scpq.18.2.158.21860

Sheskin, D.J.(2007). *Handbook of parametric and nonparametric statistical procedures*. Boca Raton, FL: Chapman & Hall/CRC.

Skinner, B. F., & Belmont, M. J. (1993). Motivation in the classroom: reciprocal effects of teacher behavior and student engagement across the school year. *Journal of Educational Psychology, 85*(4), 571–581. doi: 10.1037/0022-0663.85.4.571

Spreckelmeyer, K.N., Krach, S., Kohls, G., Rademacher, L., Irmak, A. et al. (2009). Anticipation of monetary and social reward differently activates mesolimbic brain structures in men and women. *Social Cognitive and Affective Neuroscience, 4*, 158–165. doi: 10.1093/scan/nsn051

Squire, K. D. (2008). Video game-based learning: An emerging paradigm for instruction. *Performance Improvement Quarterly*, *21*(2), 7–36. doi: 10.1002/piq.20020

Steinkuehler, C. A. (2006). Massively multiplayer online video gaming as participation in a discourse. *Mind, Culture, and Activity, 13*(1), 38–52. doi: 10.1207/s15327884mca1301_4

Suits, B. (2006). Construction of a definition. In K. Salen & E. Zimmerman (Eds.), *The game design reader. A rules of play anthology* (pp. 172–191). London, England: The MIT Press.

Supinski, S. B. (1996). *A comparison of simultaneous versus sequential use of interactive video instruction and cooperative learning: Effects on achievement, amount of invested mental effort, and attitudes* (Unpublished doctoral dissertation). The Florida State University, Tallahassee.

Tabachnick, B., & Fidell, L. (1996). *Using multivariate statistics*. New York, NY: HarperCollins College Publishers.

Tobias, S., & Duffy, T. M. (Eds.). (2009). *Constructivist theory applied to instruction: Success or failure?* New York, NY: Taylor & Francis.

Tobias, S., & Fletcher, J.D. (Eds.). (2011). *Computer games and instruction*. Charlotte, NC.: Information Age Publishing.

Tsai, C. (2009). Conceptions of learning versus conceptions of web-based learning: The differences revealed by college students. *Computers & Education, 53*, 1092–1103. doi: 10.1016/j.compedu.2009.05.019

Tüzün, H., Yılmaz-Soylu, M., Karakus, T., Inal, Y., & Kızılkaya, G. (2009). The effects of computer games on primary school students' achievement and motivation in geography learning. *Computers & Education, 52*, 68–77. doi: 10.1016/j.compedu.2008.06.008

Uriccio, W. (2005). Simulation, history, and computer games. In J. Raessens & J. Goldstein (Eds.), *Handbook of computer game studies,* (pp. 327–334). Cambridge, MA: MIT Press.

van den Broek, P., Lorch, R. F. Jr., Linderholm, T., & Gustafson, M. (2001). The effects of readers' goals on inference generation and memory for texts. *Memory & Cognition, 29*(8), 1081–1087. doi:10.3758/BF03206376

van den Haak, M.J., & de Jong, M.D.T. (2003). Exploring two methods of usability testing: concurrent versus retrospective think-aloud protocols. *IEEE Professional Communication Conference.* Proceedings. IEEE International.

van Gog, T., Paas, F., van Merrienboer, J. J. G., & Witte, P. (2005). Uncovering the problem-solving process: Cued retrospective reporting versus concurrent and retrospective reporting. *Journal of Experimental Psychology, 11,* 237–244. doi: 10.1037/1076-898X.11.4.237

van Gog, T., & Scheiter, K. (2010). Eye tracking as a tool to study and enhance multimedia learning. *Learning and Instruction, 20*, 95–99. doi: 10.1016/j.learninstruc.2009.02.009

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Vogel, J. J., Vogel, D. S., Cannon-Bowers, J., Bowers, C. A., Muse, K., & Wright, M. (2006). Computer gaming and interactive simulations for learning: A meta-analysis. *Journal of Educational Computing Research, 34*(3), 229–243. doi: 10.2190/FLHV-K4WA-WPVQ-H0YM

Wainess, R. (2010, May). *A framework for research on motivation and game-based learning.* Paper presented at the Games for Learning Conference, New York University, New York.

Wang, H., Shen, C. & Ritterfeld, U. (2009). Enjoyment in games. In U. Ritterfeld, M. Cody, & P. Vorderer (Eds.), *Serious games: Mechanisms and effects* (pp. 25–47). Mahwah, New Jersey: Routledge/LEA.

Wertheimer, M. (1945). *Productive thinking*. New York: Harper & Row.

Whitton, N. (2011). Game engagement theory and adult learning. *Simulation & Gaming, 42*(5), 596–609. doi: 10.1177/1046878110378587

Whitson, C., & Consoli, J. (2009). Flow theory and student engagement. *Journal of Cross-Disciplinary Perspectives in Education, 2*(1), 40–49. Retrieved from http://jcpe.wmwikis.net/file/view/whitsonconsoli.pdf

Wilcox, R.R. (2012). *Modern statistics for the social and behavioral sciences: A practical introduction*. Boca Raton, FL: CRC Press.

Wilson, K. A., Bedwell, W. L., Lazzara, E.H., Salas, E., Burke, C.S., Estock, J.L., Orvis, K.L et al. (2009). Relationships between game attributes and learning outcomes. *Simulation & Gaming, 40*(2), 217–266. doi: 10.1177/1046878108321866

Young, M.F., Slota, .S, Cutter, A.B., Jalette, G., Mullin, G., Lai, B., & Yukhymenko, M. (2012). Our princess is in another castle: a review of trends in serious gaming for education. *Review of Educational Research, 82*(1), 61–89. doi: 10.3102/0034654312436980

Zyda, M. (2005). From visual simulation to virtual reality to games. *IEEE Computer, 38*(9), 25–31.

8. Appendices

8.1. Appendix A. Game Features by Author

Comparison of main game features by author

Salen & Zimmerman (2003)	Authors														Juul (2005)
Features	*Parlett*	*Abt*	*Huizinga*		*Caillois*		*Suits*		*Crawford*		*Avedon/Sutton-Smith*		*Costikyan*		Features
Creates special social groups			*Yes*	Yes											Social grouping
Proceeds according to rules that limit players	*Yes*	*Yes*	*Yes*	Yes	*Yes*	Yes	*Yes*	Yes	*Yes*	Yes	*Yes*	Yes		Yes	Rules
System of parts/Resources and token									*Yes*				*Yes*		
Goal-oriented/outcome oriented	*Yes*	*Yes*				Yes	*Yes*				*Yes*	Yes		Yes	Outcomes
Uncertain					*Yes*										
Inefficient							*Yes*	Yes							Less efficient means
Involves decision-making		*Yes*							*Yes*	Yes			*Yes*		Interaction
Never associated with material gain			*Yes*	Yes	*Yes*	Yes									"Separate"
Artificial/safe/Outside of ordinary life															
Not serious and absorbing			*Yes*			Yes						Yes		Yes	"Not work"
Voluntary					*Yes*		*Yes*				*Yes*				
Conflict or contest	*Yes*							Yes	*Yes*	Yes	*Yes*	Yes		Yes	"Goals"
Make-believe/Representational					*Yes*				*Yes*						Rules? Or Separate?
A form of art													*Yes*		
Activity, process, event		*Yes*							*Yes*		*Yes*				
			Huizinga		Caillois		Suits		Crawford		Avedon/Sutton-Smith			Kelley	

Note. The first left column represent the features of games as identified by Salen and Zimmerman (2003). The last column on the right represents the features identified by Juul (2005). Both set of features have been organized and grouped (shaded versus non-shaded rows). The "Yes" in italics shows the authors' features as described in Salen and Zimmerman (2003). The non-italic "Yes" shows the authors' features as described in Juul (2005).

8.2. Appendix B. Patterns Template

Pattern template comparison:

Kreimeier (2002)	Björk & Holopainen (2005)
Name: Predictable consequences **Problem:** The player has to perceive failure as a consequence of her mistakes, not as a random. **Solutions:** The player cannot take a meaningful decision to act (or not to act) if the result of a possible action can not be anticipated. **Consequences:** It is no longer possible to implement surprises that are impossible to anticipate. Ultimately, the designer might arrange for actions that the informed player will never perform **Examples:** examples include crates and barrels that are marked as explosive actually explode, fireballs that hurt a player when they hit, etc. **References:** Doug Church introduced the concept of perceived/perceivable consequence, which he uses also do describe predictable outcomes.	**Name:** Predictable consequences **Core definition:** Players can predict how the game state will change if they perform actions. **General description:** When players can understand how actions and events affect the game state of a game, those actions and events have Predictable Consequences. **Examples:** The actions in first-person shooters often contain no elements of chance and thereby they have totally predictable consequences. **Using the pattern:** The most Predictable Consequences (although maybe only in the short term) are the players' own actions when they have Perfect Information of the game state and the evaluation function is static. **Consequences:** Predictable Consequences let players predict future game states and thus have Anticipation and notice Hovering Closures in games. **Relations:** Instantiates: Perceived Chance to Succeed, Anticipation, Investments, Experimenting, Cognitive Immersion, among others.

8.3. Appendix C. Gee's 36 principles

Gee's 36 principles organized by key authors

Principles, their main idea & core Authors
Main ideas: How games foster critical thinking and learning (Bereiter & Scardamalia, 1989).Games as preparing for future learning (Bransford & Schwartz, 1999):
1) Active, Critical Learning Principle: All aspects of the learning environment are set up to encourage active and critical, not passive, learning
2) Design Principle: Learning about and coming to appreciate design and design principles is core to the leaning experience
3) Semiotic Principle: Learning about and coming to appreciate interrelations within and across multiple sign systems
4) Semiotic Domains Principle: Leaning involves mastering semiotic domains
5) Meta-level thinking about Semiotic Domain Principle: Learning involves active and critical thinking about the relationships of the semiotic domain being learned to other semiotic domains
Main idea: the role of identity in individuals engagement in terms of Committed learning and working within their "Regime of Competence" (diSessa, 2000):
6) "Psychosocial Moratorium" Principle: Learners can take risks in a space where real-world consequences are lowered
7) Committed Learning Principle: Learners participate in an extended engagement (lots of effort and practice) as an extension of their real-world identities
8) Identity Principle: Learning involves taking on and playing with identities in such a a way that the learner has real choices
9) Self-Knowledge Principle: The virtual world is constructed in such a way that learners learn not only about the domain but also about themselves
10) Amplification of Input Principle: For a little input, learners get a lot of output
11) Achievement Principle: For learners of all levels of skill there are intrinsic rewards from the beginning
12) Practice Principle: Learners get lots and lots of practice in a context where the practice is not boring
13. Ongoing Learning Principle: Vague distinction between the learner and the master

14) "Regime of Competence" Principle: The learner gets ample opportunity to operate within, but at the outer edge of, his or her resources

Main idea: Meaning is situated in individuals Experience – diSessa (2000), Schön (1987), Lakoff, 1987):

15) Probing Principle: Learning is a cycle of probing the world (doing something); reflecting in and on this action and, on this basis, forming a hypothesis

16) Multiple Routes Principle: There are multiple ways to make progress or move ahead.

17) Situated Meaning Principle: The meanings of signs are situated in embodied experience. Meanings are not general or decontextualized.

18) Text Principle: Texts are not understood purely verbally but are understood in terms of embodied experience.

19) Intertextual Principle: The learner understands texts as a family ("genre") of related texts and understands any one text in relation to others in the family

20) Multimodal Principle: Meaning and knowledge are built up through various modalities (images, texts, symbols,, etc.), not just words

21) "Material Intelligence" Principle: Thinking, problem-solving and knowledge are "stored" in material objects and the environment.

22) Intuitive Knowledge Principle: Intuitive or tacit knowledge built up in repeated practice and experience, often in association with an affinity group

Main idea: Overt information, guidance and transfer (Bransford & Schwartz, 1999):

23) Subset Principle: Learning even at its start takes place in a (simplified) subset of the real domain

24) Incremental Principle: Learning situations are ordered in the early stages so that earlier cases lead to generalizations that are fruitful for later cases

25) Concentrated Sample Principle: The learner sees, especially early on, many more instances of the fundamental signs and actions than should be otherwise

26) Bottom-up Basic Skills Principle: Basic skills are not learned in isolation or out of context; rather, what counts as a basic skill is discovered bottom up

27) Explicit Information On-Demand and Just-in-Time Principle: The learner is given explicit information both on-demand and just-in-time

28) Discovery Principle: Overt telling is kept to a well-thought-out minimum, allowing ample opportunities for the learner to experiment and make discoveries

29) Transfer Principle: Learners are given ample opportunity to practice, and support for, transferring what they have learned earlier to later problems

Main idea: Linking every day and scientific understanding (diSessa, 2000):

30) Cultural Models about the World Principle: Learning is set up in such a way that learners come to think consciously and reflectively about some of their cultural models

31) Cultural Models about Learning Principle: Learning is set up in such a way that learners come to reflect about their cultural models about learning

32) Cultural Models about Semiotic Domains Principle: Learning is set up in such a way that learners come to reflect about their cultural models about a particular semiotic domain

Main idea: Thinking and reasoning (and therefore knowledge) are social and distributed (Brown, Collins and Duguid, 1989):

33) Distributed Principle: Meaning/knowledge is distributed across the learner, objects, tools, symbols, technologies, and the environment

34) Dispersed Principle: Meaning/knowledge is dispersed in the sense that the learner shares it with others outside the domain/game

35) Affinity Group Principle: Learners constitute an "affinity group," that is, a group that is bonded primarily through shared endevours

36) Insider Principle: The learner is an "insider," "teacher," and "producer" (not just a consumer) able to customize the learning experience

8.4. Appendix D. Games Reviewed for the Present Study

List of the games revised for the study

N	Game	Institution	Topic
1	Waker	Singapore-MIT GAMBIT Game Lab	Velocity
2	Vanished	MIT Education Arcade	Science
3	Osy Osmosis	University of Georgia	Osmosis process
4	Dreambox	Dreambox learning	
5	Selene	http://cygames.cet.edu/	Geological process of the moon

6	SURGE	Vandervilt University	Intuitive understanding of scientific processes
7	Lemonocity Descends (for Apps).	Twist Education	Motion
8	SAVE Science	Temple University & Arizona State University	Scientific concepts
	Timez Attack (4th graders)	Bigbrainz	Multiplication & Division
	2weistein	Cornelsen & Braingame	Math
9	Genius Unternemen Physik	Cornelsen	Physik
10	Genius Task Force Biolgie	Cornelsen	Ecology
11	Physikus	Braingame	Physics
12	Biolab	Braingame	Biology
13	Chemicus	Braingame	Chemistry
14	History of Biology	Spongelab	History of scientific discoveries
15	InnerCell	US Navy	Immune System
16	Immune Attack	National Science Foundation	Immune System
17	Zombie Division	U. of Nottingham	Algebra
18	Quest Atlantis	Indiana University	inquiry skills and science standards-based content
19	River city	Harvard	scientific inquiry and 21st century skills.
20	Supercharged!		Electromagnetism
21	Music games	http://www.joytunes.com	Play instruments
22	Kabongo games	https://www.kabongo.com	Multiple skills
23	Pink Panther's Passport to Peril	Wanderlust® Interactive, Inc	Geography
24	Logical Journey Of The Zoombinis	The Learning Company	Math and Logic

25	Algebra vs. the Cockroaches	http://catchupmath.com	y=ax+b
26	Spent	Urban Ministries of Durban	Basic economy
27	Powerup	http://www.powerupthegame.org	
28	DimensionM	www.dimensionu.com	Algebra
29	**DOG***eometry*	http://dogeometry.autoteles.org/	Geometry
30	CyberCIEGE	US Navy	computer and network security concepts

8.5. Appendix E. AIME Questionnaire

Amount of Invested Mental Effort Questionnaire (AIME)

GEISTIGER/INTELLECTUELLER ANSTRENGUNG/EINSATZ FRAGEBOGEN

Beantworten Sie bitte die folgenden Fragen bezüglich des Computerspiels, das Sie schon gespielt haben. Geben Sie Ihre beste Schätzung auf der 5-Punkte Skala an, wo 1 stellt ein minimale Rating und 5 eine maximale Rating dar. Die Fragen wird in 3 entsprechenden Gruppen geteilt, die die verschiedene Aspekte des Spiel darstellen: die Wirtschaftssimulation, die Wissensaufgaben, und die Wirtschaftsstatistiken

Answers‘ scale:

NIEDRIG 1 2 3 4 5 HOCH

I. Die Wirtschaftssimulation (es geht darum, z.B., die bestimmten Gebäude zu bauen, sich um die Gebäudesstatus und die Aufgabenbenachrichtung zu kümmern, die Kontostand, Bevölkerung und Arbeitsplätze im Auge zu behalten)

a. Wie sehr haben Sie versucht, die Wirtschaftssimulation zu verstehen?

b. Wie schwierig war es, die Wirtschaftssimulation zu verstehen?

c. Wie sehr haben Sie sich beim Spielen konzentriert?

d. Wie viel haben Sie sich über die Wirtschaftssimulation nachgedacht?

e. Wie viel glauben Sie, dass Sie sich an den Wirtschaftssimulation erinnern können?

f. Wie viel geistige/intellektuelle Anstrengung/Einsatz haben Sie investiert, um die Wirtschaftssimulation zu verstehen?

II. Die Wissensaufgaben (es geht darum, die verschiedenen wissenschaftlichen Probleme zu lösen).

a. Wie sehr haben Sie versucht, die physikalischen Wissensaufgaben zu verstehen?
b. Wie schwierig war es, die physikalischen Wissensaufgaben zu verstehen?
c. Wie sehr haben Sie sich konzentriert, während des Lesens der physikalischen Wissensaufgaben?
d. Wie viel haben Sie sich über die physikalischen Wissensaufgaben nachgedacht?
e. Wie viel glauben Sie, dass Sie sich an die physikalischen Wissensaufgaben erinnern können?
f. Wie viel geistige/intellektuelle Anstrengung/Einsatz haben Sie investiert, um die physikalischen Wissensaufgaben zu verstehen?

III. Die Wirtschaftsstatistiken (es geht darum, z.B., die allgemein Bilanz des Unternehmens, die Werte über die Investitionen, Profit, etc. zu beaufsichtigen und zu interpretieren)

a. Wie sehr haben Sie versucht, die Wirtschaftsstatistiken zu verstehen?
b. Wie schwierig war es, die Wirtschaftsstatistiken zu verstehen?
c. Wie sehr haben Sie sich konzentriert, während des Lesens der Wirtschaftsstatistiken?
d. Wie viel haben Sie sich über die Wirtschaftsstatistiken nachgedacht?
e. Wie viel glauben Sie, dass Sie sich an die Wirtschaftsstatistiken erinnern könnten?
f. Wie viel geistige/intellektuelle Anstrengung/Einsatz haben Sie investiert, um die Wirtschaftsstatistiken zu verstehen?

8.6. Appendix F. Situational Cognitive Engagement Questionnaire

Situational Cognitive Engagement Questionnaire

Bitte kreuzen Sie an, inwieweit Sie die folgenden Aussagen hinsichtlich Ihrer Erfahrungen mit der Wissensaufgabe zutreffend finden oder nicht.

Answers‘ scale:

gar nicht zutreffend 1 2 3 4 5 sehr zutreffend

a. Ich habe mich sehr stark in dieser Wissensaufgabe eingearbeitet

b. Ich habe mich sehr bemüht diese Wissensaufgabe zu lösen

c. Ich würde mich gerne noch weiter mit diesen Wissensaufgabe beschäftigen

d. Ich war so in der Wissensaufgabe vertieft, dass ich alles um mich herum vergessen habe.

8.7. Appendix G. Focus Interview

Focus Interview

Über das Spiel

1) Was war dein Ziel, allgemein?

2) Welche Strategie hast du umgesetzt/benutzt, um dein Ziel zu erreichen?

3) Was war besonders schwierig in dem Spiel und wie hast du das überwunden?

4) Welche Informationen oder Ereignissen hast du hauptsächlich nachverfolgt?

Für jeden Lernaufgabe

1) Was ist dein Ziel hier? Was mochtest du hier erreichen?

2) Was war besonders schwierig mit dieser Aufgabe und wie hast du das überwunden?

3) Worauf hast du hauptsächlich geachtet (auf welche Information) (9) und inwiefern hast du diese Info gestalten, um besser zu verstehen das Problem oder die Aufgabe? (4) T

4) Könntest du beschreiben welche Information war für dich wichtig und welche war nicht? („welche Info benötige ich…“) (9) T

5) Inwiefern hast du dein Vorwissen benutzt, um diese zu beantworten? (3, 7) T

6) Könntest du beschreiben, der Plan und Strategie, die du erstellt/eingesetzt hast, um diese Aufgabe zu lösen. (5) T

7) Welche Informationen oder Ereignissen hast du hauptsächlich nachverfolgt? (6, 9) A

8) Inwiefern hast du gedacht, ob du wirklich das Problem oder die Aufgabe verstanden hattest? (10, 14). A

9) Warum hast du diese Aufgabe versuchen zu antworten, statt einfach schließen die Aufgabe/Fenster?

Vielen Dank!

8.8. Appendix H. Recall Test Items

Recall test items and its corresponding content in the game

Liebe(r) TeilnehmerIn:

Mit ohne Hilfe, versuchen Sie, so schnell als möglich, die Fragen zu beantworten.

Beantworten Sie bitte kürzlich:

1. Welche materialen können das Klohäuschen am besten isolieren?

____

2. Ordnen Sie die Planeten in der richtigen Reihenfolge von der Sonne ausgehend auf ihre Bahnen zu.

____

3. Berechnen Sie die maximale Last und Trage das Ergebnis ein. Zugkraft F2=2 Newton tragende Seilstücke n=4. Maximale Last F1= ?

____

4. Wer formulierte die Hebelgesetze?

____

5. Könnten Sie die Hebelgesetze Gleichung aufschreiben?

____

6. Wovon wird die physikalische Größe der Dichte definiert?

____

7. Bezeichnen Sie die Aggregatzustände des Stoffes.____

8. Wer hat das alte geozentrische Weltbild zu Fall gebracht?

____

9. Ergänzen Sie: Je ___________ der Druck nämlich ist, desto ___________ ist die Siedetemperatur

10. Beschreiben Sie die drei Planetengesetze von Johannes Kepler.

11. Beschreiben Sie wie kann man einen Falschmünzer identifizieren.

12. Es gibt zwei Sorten von Flaschenzügen, festen Rollen und losen Rollen. Erklären Sie (1) was wird sich bei festen Rollen geändert und was wird sich bei losen Rollen geändert, und (2) Warrum mit einem Flaschenzuge von 4 Seilstucken kann man mehr Gewicht heben**.**

Example of test question (number 5), possible responses and the related information as embedded in *Genius Unternehmen Physik*

Test question	Possible Responses	Type of Response
5. Könnten Sie die Hebelgesetze Gleichung aufschreiben?	- Kraft F1 x Kraftarm = Last F2 x Lastarm L2	Recall

Passages from the game that correspond to the question:	
Tasks	Journal
<u>Flaschenzug:</u> Page 2. Image with data: Zugkraft F2=200 Newton tragende Seilstücke n=6 maximale Last F1=____N Page 2. Feedback: Wenn ihr mich fragt, dann solltet ihr maximal 120 kg Holz in den Korb laden. Ihr könnt ja mit maximal 200 Newton ziehen und habt 6 tragende Seile, das macht 1200 Newton also rund 12 0 kg.	Page 2.…die berühmten Hebelgesetzte. (Kraft F1 x Kraftarm L1 = Last F2 x Lastarm L2). Page 3: Ein Hebel ist nämlich genaue dann im Gleichgewicht, wenn das Produkt aus Kraft und Kraftarm des Hebelarms der einen Seite gleich dem Produkt aus Last und Lastarm auf der Gegenseite ist (Kraft F1 x Kraftarm = Last F2 x Lastarm L2). Page 4. Beispiel: Man hat einen Flaschenzug mit 4 tragenden Seilen und will eine 240 Kilogramm schwere Kiste 2 Meter hoch heben. Ihre Gewichtskraft beträgt 2400 Newton. Nach F1= n x F2 bedeutet das, dass man 8 Meter Seil „herausziehen“ und eine Gewichtskraft von 600 Newton einsetzen muss.

8.9. Appendix I. Translation of Questionnaires

Original and translated items of the control and dependent variables

Self-efficacy on Computer Gaming

Source: Ketelhut, 2011

I can learn how to play any computer game if I don’t give up.	Ich kann jedes Computerspiel spielen lernen, wenn ich nicht aufgebe.
I am very good at building things in simulation games.	In Simulationspielen kann ich erfolgreich etwas aufbauen
I can figure out most computer games.	Ich kann die meisten Computerspiele verstehen.
No matter how hard I try, I do not do well when playing computer games.	So sehr ich mich auch anstrenge, in Computerspielen habe ich keinen Erfolg.
I can keep winning at computer games for a long time	Ich kann bei Computerspielen ständig für eine lange Zeit spielen.

Mental effort Questionnaire

Source: Cennamo, dissertation.

How much of the lesson do you think you can remember?	Was denken Sie, an wie viel Inhalt aus Computerspielen kann man sich erinnern?
	Wie sehr, denken Sie, kann Inhalt von Videospielen nützlich sein, um eine tiefere Lernerfahrung zu haben?
How hard did you concentrate while watching the lesson?	Wie stark konzentrieren Sie sich, wenn Sie Computerspiele spielen?
How much did the lesson make you think?	Wie sehr machen Sie Computerspiele nachdenklich?
How much mental effort did you use in comprehending the lesson?	Wie viel geistige/intellektuelle Anstrengung investieren Sie, wenn Sie Computerspiele spielen?
How hard did your try to understand the lesson?	Wie sehr versuchen Sie gewöhnlich, in einem Computerspiel behandelte Inhalte nachzuvollziehen? (z.B. über Natur, Geschichte, Städte, etc.).

Preconceptions of Difficulty

Source: Cennamo, Savenye, & Smith, 1991

How difficult would it be for you to learn to solve a math problem from [medium]?	Wie schwierig denken Sie wäre es, in einem Computerspiel eine Mathematische/ wissenschaftliche Aufgabe lösen lernen?
How difficult would it be for you to learn about the life of a famous person from [medium]?	Wie schwierig denken Sie wäre es, etwas über eine geschichtliche Epoche in einem Computerspiel zu lernen?
How difficult would it be for you to learn to build a model from [medium]?	Wie schwierig wäre es für Sie, eine Vorstellung oder Model der Inhalte eines Computerspiels zu entwickeln?

8.10. Appendix J. AOI in % and Squared Centimeters

Below the AOIs common to all modes of play in *Genius Unternehmen Physik*:

AOI	%	squared cm
Worker_houses	0.89	7.96
Account	0.33	2.95
Resources	0.89	7.96
Menu_BuildingA	3.69	33.02
Menu_BuildingB	1.84	16.47
Back and Next	0.62	5.55
Office	0.87	7.79
Stats	0.96	8.59
Mail	0.42	3.76
News	10.37	92.80
All	39.16	350.46

Below the AOIs used to calculate the mean fixation durations, the mean dwell time and the mean reading depth for each page (e.g., J1, J2, etc.) and for the whole Journal Mode of Play, which entails the twelve pages of the Journal.

Page	AOI Name (% area, squared cms)	Page	AOI Name (% area, squared cms)
J1	Titel (5.45, 48.77), Subtitel1 (3.17, 28.37), Subtitle2 (2.47, 22.10), Subtitle3 (4.1, 36.69)	J8	Q8 (4.46, 39.91)
J2	Q7 (4.85, 43.40), Q8 (5.49, 49.13), Q4_Q5 (3.37, 30.16)	J9	Image1 (4.85, 43.40), Image2 (3.93, 35.17), Image3 (4.37, 39.11), Q10a (3.43, 30.70), Q10b (5.58, 49.94), Q10c (3.39, 30.34)
J3	Q5 (6.81, 60.94), Q12 (3.99, 35.71)	J10	Q2_Plna (2.46, 22.02), Q2_Plnb (3.03, 27.12), Q2_Plnc (8.17, 73.12)
J4	Q3_FZI (4.61, 41.26), Image (8.49, 75.98), FZII (5.39, 48.24)	J11	Pmb (35.07, 313.85)
J5	Q11a (3.28, 29.35), Q6_Dh (3.84, 34.37), Q11b (5.44, 48.68), Q11_Dha (6.49, 58.08), Q11_ Dhb (10.20, 91.28)	J12	Pmb (6.86, 61.39), Q1_Ins (2, 17.90)
J7	Q9_Tpf (3.24, 29)		

Below the AOIs used to calculate the mean fixation durations, the mean dwell time and the mean reading depth for each Task (e.g., Task1, Task2, etc.) and for the whole Task Mode of Play.

Page	AOI Name (% area, squared cms)	Page	AOI Name (% area, squared cms)
Task0	Alternatives A through E (0.48, 4.30), A_fig (2.30, 20.58), B_fig (2.93, 26.22), C_fig (1.94, 17.36), D_fig (4.98, 44.57), E_fig (2.84, 25.42), Instruction (1.14, 10.20)	Task 6	Instruction (1.36, 12.17), Data (1.68, 15.03), Answer (1.02, 9.13), Image (1.32, 11.81)
Task1	Instruction (0.69, 6.18), Solid (1.48, 13.24), Liquid (1.18, 10.56), Gas (0.73, 6.53), Obj1 (0.79, 7.07), Obj2 (0.67, 6), Obj3 (1.31, 11.72), Obj4 (1.11, 9.93), Obj5 (1.46, 13.07)	Task7	Instruction (1.43, 12.80), AltA (1.43, 12.80), AltB (0.76, 6.80), AltC (0.99, 8.86), AltD (0.74, 6.62)

Task2	DchAluminium (0.52, 4.65), DchKupfer (0.40, 3.58), DchStahl (0.39, 3.49), DchBlei (0.31, 2.77), DchGold (0.37, 3.31), Instruction (1.11 ,9.93), DchFormula (0.96, 8.59)
Task3	Instruction (1.12, 10.02), AltA (1.79, 16.02), AltB (1.87, 16.74), AltC (1.94, 17.36), AltD (1.82, 16.29), AltE (1.58, 14.14)
Task4	Instruction (2.57, 23), Planetlist (5.35, 47.88), SolarSystem (8.48, 74.89)
Task5	Instruction (1.02, 9.13), Data (0.86, 7.70), Answer (0.61, 5.46), Image (1.44, 12.89)

8.11. Appendix K. Gaming Questionnaire

Bitte kreuzen Sie an, inwieweit Sie die folgenden Aussagen hinsichtlich Ihrer Erfahrungen mit dem Computerspiel zutreffend finden oder nicht. Bitte lassen Sie keine Fragen aus und ändern Sie nach Möglichkeit einmal getroffene Aussagen nicht mehr – wir sind interessiert an Ihrer spontanen Einschätzung.

Answers‘ scale:

Sehr Zutreffend			Etwas zutreffend			Gar nicht zutreffend
1	2	3	4	5	6	7

1. Ich konnte stets antizipieren, was, in Reaktion auf die von mir gestartete Aktionen als nächstes passieren würde.

2. Ich habe innerhalb des Spiels eine kurze Verzögerung zwischen meinen Aktionen und den erwarteten Ergebnisse gemerkt.

3. Ich hatte das Gefühl, gut darin zu sein, das Spiel zu steuern.

4. Die Interfacegestaltung (Graphik und Toneffekte) des Spiels fand ich sehr gut.

5. Ich fühlte mich total in das Spielerlebnis hineingezogen.

6. Ich fühlte mich so stark in das Spielerlebnis, dass ich die Zeit vergessen habe.

7. Ich fühlte mich, als ob ich innerhalb der Spielwelt war.

8. Inhaltlich und thematisch hat mir das Spiel viel Spaß gemacht.

9. Die Figuren bereichern die Geschichte/den Plot („Storyline“) erheblich.

10. Die Wissensaufgaben bereichern die Geschichte/den Plot („Storyline“) erheblich.

11. Das Spiel hat mich voll und ganz eingenommen.

8.12. Appendix L. Scoring Rubric Recall Test

Scoring rubric for Recall Test

Question	Points						
	0,2	0,5	1	1,5	2	3	4
1		Korkplatten, kork, korwand	doppelte korkplatten		doppelte korkplatten mit luftpolster		
2	For every correct sequence. An amount of 0,2 points get discounted if “Pluto” is mentioned						
3			For the correct answer = 8		The correct answer plus the formula employed F1= n x F2		
4			If the answer is Archimedes				
5					The answer is F1*F2 = L1*L2 or F1=n*F2	Kraft F1 x Kraftarm l1 = Last F2 x Lastarm L2 or F1*L1 = F2*L2	
6		The answer mention just one to the components	The answer mention the components but without explicit relation or a false relation, “Masse und Volumen” or V/M		The answer is the “Masse durch Volume” or m/v or g/cm3		
7		The answer has only one of the three: “aggregatszustände: fest, flüssig, gasförmig	The answer has only two of the three: “aggregatszustände: fest, flüssig, gasförmig	The answer has the three “aggregatszustände: fest, flüssig, gasförmig			
8		Galilei		Kopernikus			

9	When the answer says either “höher” or “größer” in one of the blanks	When the answer says either “höher” or “größer” in both blanks				
10		The answer has only one of the following options: 1. Gesetz. Alle Planeten bewegen sich auf Ellipsen. In einem der beiden Brennpunkte der Ellipse steht die Sonne 2. Gesetz. Die Bahngeschwindigkeit der Planeten ist nicht konstant. Sie wird größer, wenn der Planet der Sonne näher kommt und kleiner, wenn er sich von der Sonne entfernt. Dabei gilt: Der vor der Sonne zum Planeten gezogene Strahl überstreicht in gleichen Zeiten gleiche Flächen (Flächensatz). 3. Gesetz. Die Quadrate der Umlaufzeiten zweier Planeten verhalten sich wie die dritten Potenzen ihrer großen Halbachsen		The answer has only two of the following options: 1. Gesetz. Alle Planeten bewegen sich auf Ellipsen. In einem der beiden Brennpunkte der Ellipse steht die Sonne 2. Gesetz. Die Bahngeschwindigkeit der Planeten ist nicht konstant. Sie wird größer, wenn der Planet der Sonne näher kommt und kleiner, wenn er sich von der Sonne entfernt. Dabei gilt: Der vor der Sonne zum Planeten gezogene Strahl überstreicht in gleichen Zeiten gleiche Flächen (Flächensatz). 3. Gesetz. Die Quadrate der Umlaufzeiten zweier Planeten verhalten sich wie die dritten Potenzen ihrer großen Halbachsen	The answer has all the options: 1. . Gesetz. Alle Planeten bewegen sich auf Ellipsen. In einem der beiden Brennpunkte der Ellipse steht die Sonne 2. Gesetz. Die Bahngeschwindigkeit der Planeten ist nicht konstant. Sie wird größer, wenn der Planet der Sonne näher kommt und kleiner, wenn er sich von der Sonne entfernt. Dabei gilt: Der vor der Sonne zum Planeten gezogene Strahl überstreicht in gleichen Zeiten gleiche Flächen (Flächensatz). 3. Gesetz. Die Quadrate der Umlaufzeiten zweier Planeten verhalten sich wie die dritten Potenzen ihrer großen Halbachsen	
11	The answer only states “an der Dichte der münze” or similar,i.e., “volume berechnen”	The answer refers either to the use of Volume or Masse or mentions the use of a weight/table		The answer refers to the use of Volume or Masse and mentions the provides the right formula	The answer refers to the use of Volume or Masse and mentions the use of a weight/table and provides the formula	
12		The answer either has one of the following ideas:		The answer has both of the following ideas:		The answer has both of the following ideas:

	1.that for the “feste rollen verändert sich die Richtung” 2. that for „lose Rollen verringet sich die Kraft“	1.that for the “feste rollen verändert sich die Richtung” 2. that for „lose Rollen verringet sich die Kraft“ OR The answer has the following idea: In some way the “kraft/Last wird geteilt”	1.that for the “feste rollen verändert sich die Richtung” 2. that for „lose Rollen verringet sich die Kraft“ AND The answer has the following idea: In some way the “kraft/Last wird geteilt”

8.13. Appendix M. Code Frame for the Interview Analysis

Code frame for exploring Corno and Mandinach's (1983) model of cognitive engagement

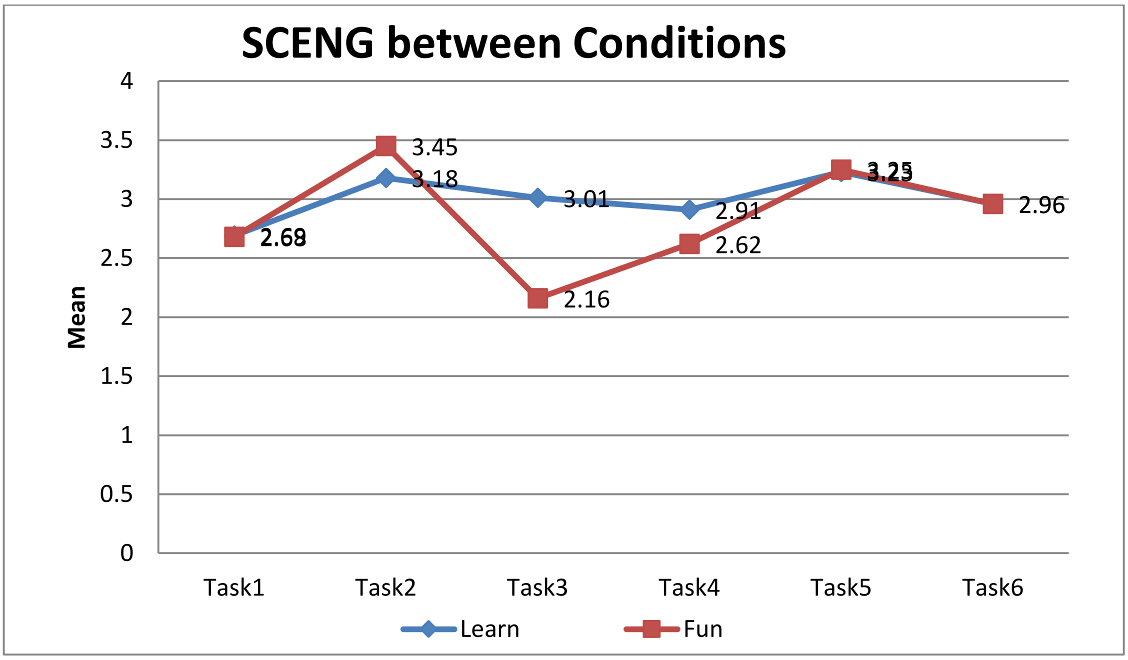
Processes	Examples
1. Acquisition	
Attention (+-)	
Receiving stimuli	Nee, also ich hab' das ja öfter nachgelesen
Attention (-)	Also ich hab' das überflogen.
Monitoring	
General	Also ich wusste das jetzt nicht ganz genau.
Specific	Also ich hab' das extra nochmal nachgelesen, damit ich weiß wie ich darauf komme, ob das jetzt Gold ist oder nicht Gold ist
2. Transformation	
Selectivity	Ähm, weil mich ja nicht alle Informationen interessiert haben
Connecting	
General	Also ich ich hatte das schon mal gehört, aber ich wusste nicht mehr wie das jetzt war.
Specific	Und wie sich das zusammensetzt. Also mit dem Volumen und der Maße,
Planning (+-)	
General	Ähm, ja das hab' ich nochmal benutzt um zu gucken welche mir noch, also was mir noch fehlt
Specific	
Planning (-)	ich habe greaten
3. Conative processes	
Motivation	Interesse. Also ich möchte das dann halt gerne vollständig lösen und nicht nur so halb.
4. Emotional processes	das hat mich geärgert

Results of the final coding of the interviews:

Overcode	Code	All codings	All codings %
Summary			
	Task related	1379	36%
	Reflections	375	9.79%
	Others (Goals, Instruction)	394	10.29%
	Uncodable	1683	43.93%
	Total	3831	100%

8.14. Appendix N. SCENG values by Task

	Condition to Learn		Condition for Fun					
Dependent variables	M	SD	M	SD	*df*	*t*	*p*	Cohen's *d*
Task 1	2.69	.71	2.68	.78	20	.027	.979	.01
Task 2	3.18	.83	3.45	.88	17	-.668	.512	.32
Task 3	3.01	.88	2.16	.74	29	2.87	.008	1.06
Task 4	2.91	1.16	2.62	.20	13	.602	.561	.33
Task 5	3.23	.73	3.25	.75	27	-.054	.958	.02
Task 6	2.96	.51	2.96	.89	15	-.863	.401	.44
Task 7	2.57	.55	2.57	.99	12	-.356	.726	.20

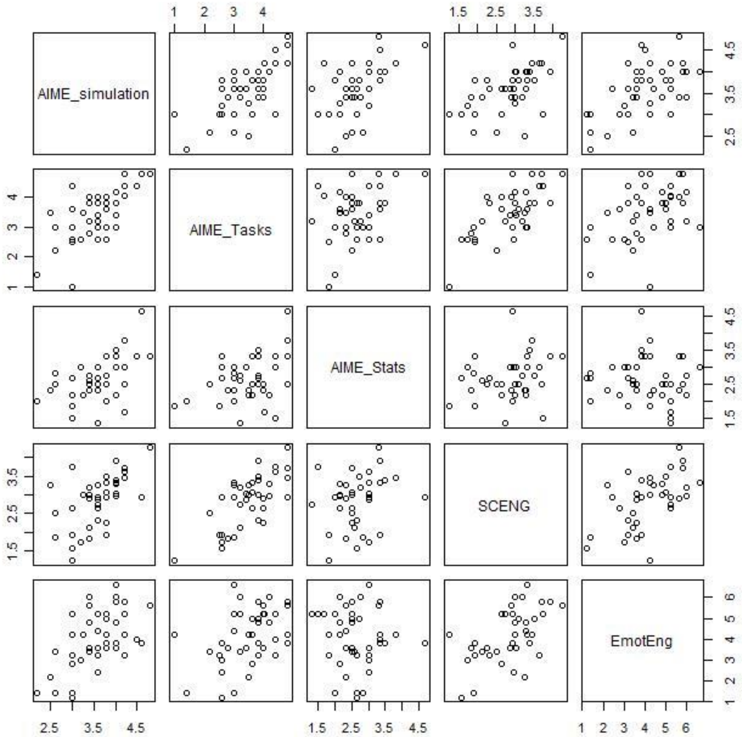


8.15. Appendix O. Pearson Correlation Coefficients and Scatterplots among the Dependent Variables for the Total Sample

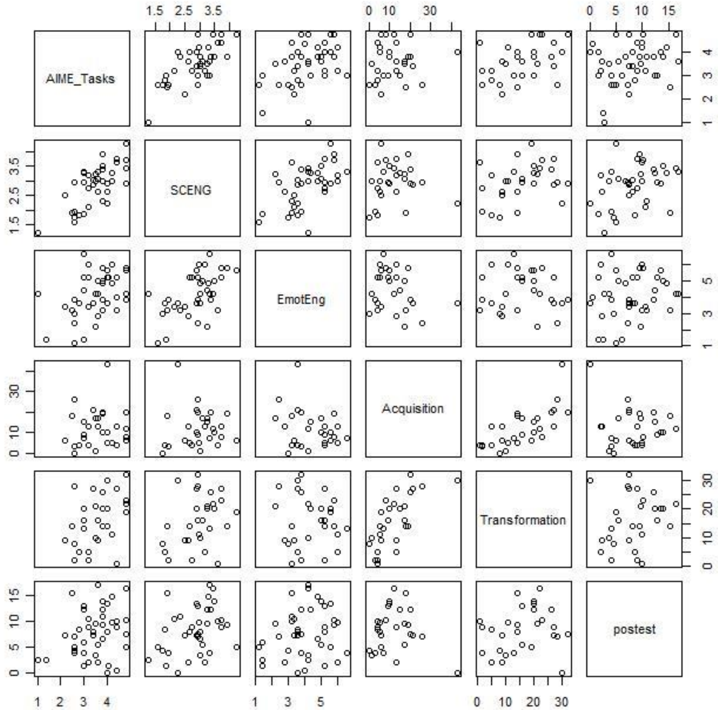
	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1.AIME_sim	.692**	.499**	.559**	.570**	-.163	.038	.066	.152	-.186	.202	-.020	-.225	-.238	-.230	.170
2.AIME_Tasks		.331*	.715**	.438**	.038	.305	-.051	.139	-.189	.214	.047	.032	.043	-.010	.257
3.AIME_Stats			.219	-.083	-.022	.061	.138	.368*	.202	.318*	.167	.206	-.119	.075	.057
4.SCENG				.562**	.083	.401*	.152	.128	-.053	.212	.236	.212	.299*	.193	.353*
5.EmotEng					-*.352*	-.051	.056	.081	-.080	.013	-.020	-.239	-.034	-.088	.197
6. Acquisition						.595**	-.023	-*.338*	-.196	-.207	.316	.187	.394*	.123	-.029
7. Transformation							.253	-.090	-.048	.302	.252	.235	.351	.163	.158
8.FD_Tasks								.645**	.102	.733**	.561**	.252	.275	.381*	-.097
9.DT_Tasks									.270	.759**	.368*	.150	-.021	.241	-.134
10.TDT_Tasks										.126	.004	.264	-.048	.260	-.087
11.Depth_Tasks											.331*	.120	.098	.163	-.180
12.FD_Journal												.602**	.617**	.679**	.355*
13.DT_Journal													.706**	.882**	.303*
14.TDTJournal														.707**	.325*
15.DepthJourn															.361*

Note. *p<.05;*p<.01.1.AIME_sim= Amount of Invested Mental Effort on the Simulation; 2.AIME_Tasks= Amount of Invested Mental Effort on the Tasks; 3.AIME_Stats= Amount of Invested Mental Effort on the Statistics; 4.SCENG=Situational Cognitive Engagement; 5.EmotEng=Emotional Engagement; 6.Acquisition=acquisition information processes; 7. Transformation =Transformation information processes; 8.FD_Tasks=Fixation Durations on Tasks; 9.DT_Tasks=Dwell Time on Tasks; 10.TDT_Tasks=Total Dwell Time on Tasks; 11.Depth_Tasks= Reading depth on Tasks; 12.FD_Journal=Fixation Durations on Journal; 13.DT_Journal=Dwell Time on Journal; 14.TDTJournal=Total Dwell Time on Journal; 15.DepthJourn=Reading depth on Journal. 16. TotalPosttest=Scores on the posttest.

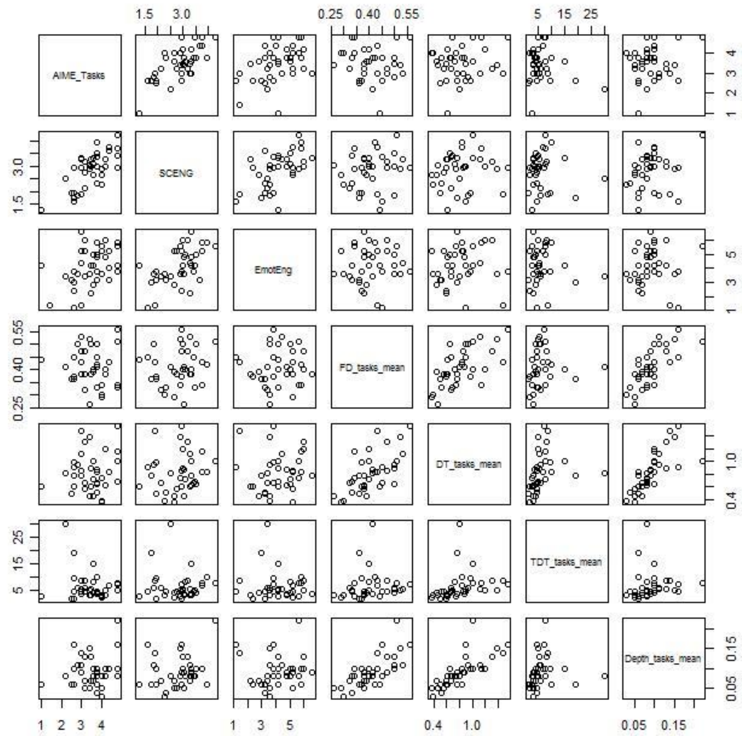
Scatterplot for the Dependent Variables for the Total Sample



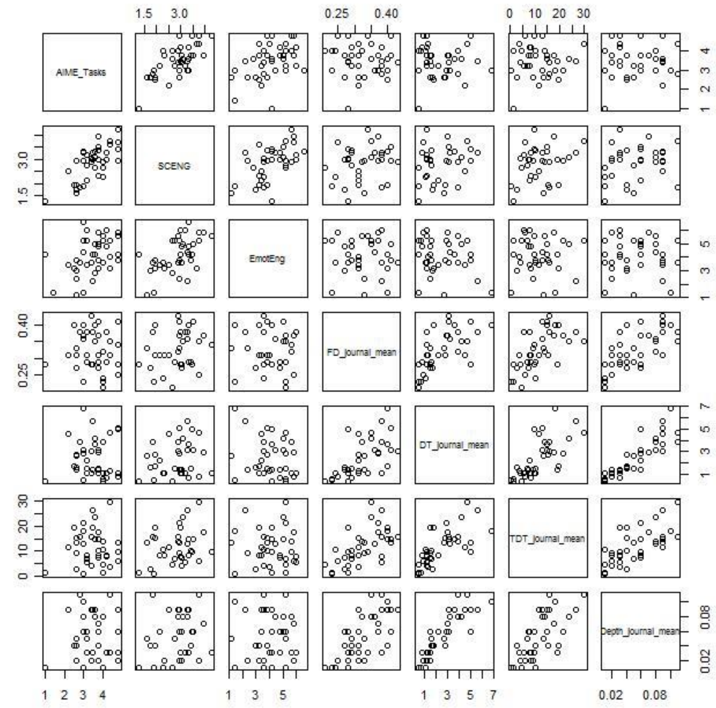
Note. AIME_simulation= Amount of Invested Mental Effort on the Simulation; 2.AIME_Tasks= Amount of Invested Mental Effort on the Tasks; 3.AIME_Stats= Amount of Invested Mental Effort on the Statistics; 4.SCENG=Situational Cognitive Engagement; 5.EmotEng=Emotional Engagement.



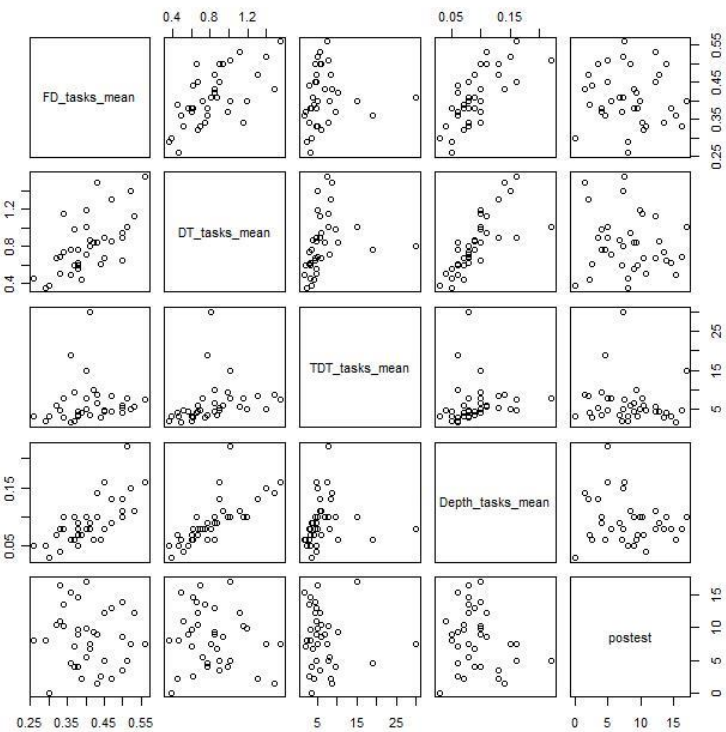
Note.AIME_Tasks= Amount of Invested Mental Effort on the Tasks; SCENG=Situational Cognitive Engagement; EmotEng=Emotional Engagement; Acquisition=acquisition information processes; Transformation =Transformation information processes; Posttest=Scores on the posttest.



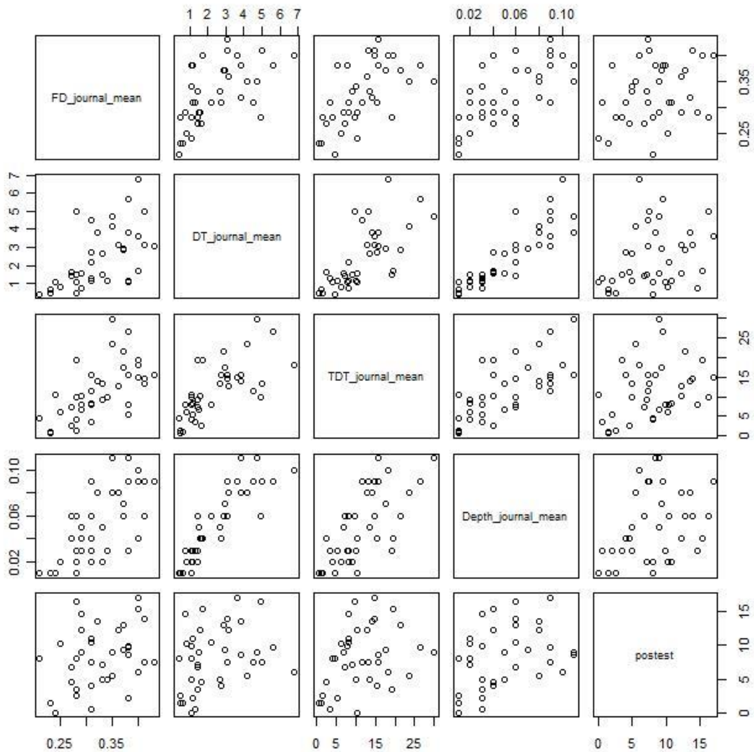
Note. AIME_Tasks= Amount of Invested Mental Effort on the Tasks;.SCENG=Situational Cognitive Engagement; EmotEng=Emotional Engagement; FD_Tasks=Fixation Durations on Tasks; DT_Tasks=Dwell Time on Tasks; TDT_Tasks=Total Dwell Time on Tasks; Depth_Tasks= Reading depth on Tasks



Note. AIME_Tasks= Amount of Invested Mental Effort on the Tasks; SCENG=Situational Cognitive Engagement; EmotEng=Emotional Engagement; FD_Journal=Fixation Durations on Journal; DT_Journal=Dwell Time on Journal; TDTJournal=Total Dwell Time on Journal; DepthJourn=Reading depth on Journal.



Note. FD_Tasks=Fixation Durations on Tasks; DT_Tasks=Dwell Time on Tasks; TDT_Tasks=Total Dwell Time on Tasks; Depth_Tasks= Reading depth on Tasks; Posttest=Scores on the posttest.



Note. FD_Journal=Fixation Durations on Journal; DT_Journal=Dwell Time on Journal; TDTJournal=Total Dwell Time on Journal; DepthJourn=Reading depth on Journal. Posttest=Scores on the posttest