TAHA DHIAEDDINE AMDOUNI

# EVALUATION OF THE IMPACT OF MICRO ADAPTIVITY IN SERIOUS GAMES.

# EVALUIERUNG DER AUSWIRKUNG VON MIKRO-ADAPTIVITÄT IN LERNSPIELEN.

BY

TAHA DHIAEDDINE AMDOUNI

[ December 21, 2011 at 14:08 ]

# EVALUATION OF THE IMPACT OF MICRO ADAPTIVITY IN SERIOUS GAMES.

TAHA DHIAEDDINE AMDOUNI

Chair for Applied Software Engineering
Fakultät für Informatik
Technische Universität München
December 2011

[ December 21, 2011 at 14:08 ]

*I dedicate this work*

First to my parents who gave me the chance to study in Europe.

Second to the members of the Dance Research Group, colleagues and academic advisors from the Chair of Applied Software Engineering.

Third to the young generation gamers for whom the game was designed.

2011-2012

## ABSTRACT

Serious games do not only entertain.

Motivation is the most important factor in serious games!

Entertainment, enjoyment and fun keeps students motivated.

Motivation is the driving force behind our actions.

The difficulty of a task influences the learner's motivation.

Play is a powerful mediator for learning.

Serious games are successful learning tools.

Adaptivity helps to provide individual learning experience.

Using adaptivity for higher individuality.

Individual tutoring produces the best learning outcome.

"no one knows how best to make an interface than someone who uses it day after day after day. We play our games a lot, but we can never play them as much as our fans do." - Soren Johnson.

Games are highly interactive and the UI is important to the success and playability of a game.

As more and more computing becomes about ubiquitous personal mobile platforms like iPhone, iPad, Nintendo DS, Android, and many flavors of mobile phones & tablets, the opportunities to create unique serious games is more possible then ever before.

Imagine if your child learns maths with assistance through playing, so you can lean back and spare you a Lerner and forward effort.

Imagine if we could combine the gameplay with all the enjoyment & motivation in it and the learning with all the difficulties that children could encounter in their first steps; bringing out a serious game on the iPad.

With assistance we, "Dance Research Group", mean that: based on the theory of the zone of Proximal Development [Vygotsky´s Psychological and Pedagogical patterns], learns your child more with a little help when needed, what we call further dynamic adaptivity; this principal let the game adapt himself to the skills of your child giving him the favor to enjoy the game and learn through it effectively.

Contemporary educational researchers often suffer from being tagged as either positivist or interpretivist—two opposite epistemological paradigms in social studies—and there is no exception for game-based learning (GBL) research carried out in educational contexts. An alternative methodology which embraces pragmatism, called a 'Spiral Research Model' is proposed. It incorporates a mixed-methods approach and multiple case studies in a practical way, in which the research question along with its changes over time, determines the choice of research methods. The temporal focus shift enabled the combination of three types of inquiries— exploratory, confirmative

and explanatory in a single piece of research. The research began with a bottom-up analysis framework which provided the ground for the identification and classification of key issues and concepts. Two issues were identified and were used to direct the development of a questionnaire survey which was used to collect quantitative data. The questionnaire acted as a validating instrument for the qualitative findings drawn from the exploratory studies. The questionnaire findings were then deepened by follow-up semi-structured interviews; while the findings of the interviews were used to explain the rationale behind the views of the respondents in the questionnaire survey, with the support of document analysis. The explanation of the phenomena constructed the 'justified true belief', which is the perceived knowledge which supports the foundation of the conclusions of this research. The paper explains how the combination worked, how qualitative and quantitative data were collected, analysed and synthesized to form the research thesis, and how issues of reliability and validity were approached. [1]

## ZUSAMMENFASSUNG

Das aktive Kind in seinem Kontext als Untersuchungseinheit.

Der Weg vom Objekt zum Kind und vom Kind zum Objekt verlaeuft ueber eine andere Person. Das in einen Kontext eingebundene Kind ist die kleinste bedeutungshaltige Untersuchungseinheit fuer die Entwicklungspsychologie.

Waehrend andere Theorien Kind und Umwelt als zwei Entitaeten ansehen, die interagieren, haelt ein Kontexttheoretiker diese Trennung fuer kuenstlich und verzerrend.

"Der Mensch erschafft sich selbst von aussen durch psychologische Werkzeuge."

Wygotski definiert die Zone der proximalen Entwicklung als die Distanz zwischen dem aktuellen Entwicklungsniveau eines Kindes, bestimmt durch seine Faehigkeit, Probleme selbstaendig zu loesen , und der hoeheren Ebene als potentieller Entwicklung, die durch die Faehigkeit bestimmt wird, Probleme unter Anleitung Erwachsener oder faehigerer Kameraden zu loesen.

Der jeweilige Partner fuehrt das Kind also durch die Zone der proximalen Entwicklung, und zwar benutzt er dazu Stichworte, Hinweise, Modellbildung, Erklaerungen, Leitfragen, Diskussionen, Mitwirkung, Ermutigung, Aufmerksamkeitssteuerung, ... Der Helfer baut gewissermaßen Bruecken zwischen den vorhandenen Faehigkeiten des Kindes und neuen Fertigkeiten. Dieser Prozeß laeuft in verschiedenen Kontexten (oder Kulturen) unterschiedlich ab, da es in verschiedenen

1 A Spiral Research Model for Game-Based Learning Studies: A Pragmatic Educational Research Design in Practice Meyer [65]

Kulturen verschiedene Werkzeuge gibt (Sprache, Schrift, Abakus, Computer, ...). Außerdem entwickeln Kinder solche Faehigkeiten, die in ihrer Kultur hoch bewertet werden.

Wygotski ist der Ansicht, daß sich Entwicklung nur dann verstehen laesst, wenn man unmittelbar den Prozeß der Veraenderung betrachtet und nicht erst das Ergebnis dieser Veraenderung. Der Prozess ist wichtiger als sein Ergebnis. Motto: ... erst in der Bewegung gibt sich der Koerper zu erkennen.

## PUBLICATIONS OF THE D.R.G.

D.R.G. stands for **DANCE** Reserch Group, **D**eveloping **A**daptive **N**onlinear **C**ompeting **E**dutainment-Games

- 2008 - Barwin on the Beagle - Towards A Framework for Behavior Aware Intelligent Learning Systems Ismailović and Pagano [41]
- 2009 - On the Application of different Research Methods in Software Engineering Ismailović et al. [43]
- 2011 - WEMAKEWORDS – AN ADAPTIVE AND COLLABORATIVE SERIOUS GAME FOR LITERACY ACQUISITION Ismailović et al. [44]
- 2011 - ADAPTIVITY IN SERIOUS GAMES - Ph. D. Thesis Ismailović [40]
- [Ismailović 2011a] Sensing and adaptivity framework for learning in serious-games - Ph.D.
- [Ismailović 2011b] Developing adaptive nonlinear compelling edutainment technology/games
- Paper Damir & Barbara
- Paper Juan & Damir
- 2011 - Drehbuch Emil & Pauline in der Höhle Bartl [5]
- 2011 - Evaluation of a serious game Kolb [54]
- 2011 - An Adaptive Serious Game for Preschool Mathematics on Mobile End Devices Simeonova [88]
- Alex´s Thesis
- Juan´s Thesis
- 2011 - Bug Game: Learning Math in a Serious Game for iPad Farah George [27]
- 2011 - Domain-independent Framework for story driven game-based learning Barbara [3]
- Paper Taha & Damir (first discussion)

*"Watch a man*
*at play for an hour*
*and you can learn*
*more about him*
*than in talking to him*
*for a year."*
*(Plato, 427-348 v.Chr.)*

— Bopp [8]

## ACKNOWLEDGMENTS

I would like to thank Prof. Bernd Brügge for giving me the opportunity to research the emerging topic of serious games.

Foremost I offer my sincerest gratitude to my supervisor, Damir Ismailović, who has encouraged and supported me throughout my thesis with his knowledge and advice whilst allowing me the freedom to find my own path. I would like to express my sincere appreciation of my proofreaders for their time and effort.

I am also indebted to the stakeholders, children, kindergarten that participated in the evaluation and the experts and teachers who contributed to a better and easier analysis of the data gathered.

# CONTENTS


I INTRODUCTION 1

1 INTRODUCTION 3
- 1.1 Philosophy 3
- 1.2 Overview of the Scientific Research 4
- 1.3 Problem 5
- 1.4 Rationale, Significance and Importance of the Scientific Research 6
- 1.5 Purpose 6
- 1.6 Research Questions & Hypotheses 7
  - 1.6.1 Research Questions 7
  - 1.6.2 Research Hypotheses 10
- 1.7 Outline 10

II STATE OF THE ART: SERIOUS GAMES 13

2 SERIOUS GAMES 15
- 2.1 History and Evolution 15
- 2.2 Arts of Serious Games 17
  - 2.2.1 Video Games 18
- 2.3 Serious Games Background 19
  - 2.3.1 Pedagogical Background 21
  - 2.3.2 Psychological Background 23
  - 2.3.3 Socio-cultural Background 24
  - 2.3.4 Engagement 25
  - 2.3.5 Motivational Background 25
  - 2.3.6 Play/Gameplay in Child´s Development 25
  - 2.3.7 Cognitive Development Background 26
  - 2.3.8 Learning Theories & Process 26
  - 2.3.9 Effective Learning: Educational Attainment Level 28

3 DESIGN ESSENTIALS 31
- 3.1 Adaptivity 31
  - 3.1.1 Notion of Adaptivity 31
  - 3.1.2 Adaptive Gaming Environment 33
  - 3.1.3 Types of Adaptivity 34
  - 3.1.4 Tutor-Adaptivity (IntelligentTutoringSystem) 34
  - 3.1.5 Macro-Adaptivity 35
  - 3.1.6 Micro-Adaptivity 35
  - 3.1.7 Vygotsky´s Theory (Gameplay, Scaffolding & ZPD) 40
- 3.2 Embedded Assessment in Games 42
  - 3.2.1 Why to assess? 43
  - 3.2.2 What to assess? 45
  - 3.2.3 How to assess? 46


3.2.4 What Claims can be made Based on Performance in the Game? 48
3.2.5 What Observations of Learner Actions Would Provide Evidence for the Claims? 48
3.2.6 Research on Embedded Assessment in Educational Games 49
3.2.7 The challenges, Limitations and Disadvantages of Embedded Assessment 50
3.3 Assessment for Games 52
3.3.1 Why Engage in Assessment? 52
3.3.2 Which Measures Should I Select? 53
3.3.3 External Assessment of What is Learned in Games 55

III FOUNDATION: SCIENTIFIC RESEARCH IN SERIOUS GAMES 57

4 SCIENTIFIC RESEARCH 59
4.1 About Case Studies 59
4.2 About Evaluation 60
4.2.1 Definition of Evaluation 62
4.2.2 Evaluation: Aim & Objectives 63
4.2.3 Types of Evaluation 64
4.2.4 Qualitative vs. Quantitative 65
4.2.5 Evaluation Designs 67
4.2.6 Logic Model 69
4.2.7 Guidelines to Usability Engineering (Software Engineering Vs. Evaluation) 72
4.3 Guidelines and Limitations of Research in Serious Games 72
4.3.1 Guidelines 74
4.3.2 Methodologies 74
4.3.3 Challenge 75
4.3.4 Limitations 75
4.3.5 Agreement 76
4.3.6 Skepticism 77

5 TRENDS IN SCIENTIFIC STUDIES ON SERIOUS GAMES 79
5.1 Research on Serious Games 80
5.2 Evaluation Studies on Learning Enhancement and Motivation 82
5.3 Studies Evaluating Serious Games 84
5.4 Studies on Serious Games for Maths 89
5.5 Studies on Adaptive Serious Games 97
5.6 Studies on the impact of Adaptivity: ZPD 100
5.7 Recommendations for our Case Study 101

IV PILOT STUDY: BUG GAME ADAPTATION 103

6 INNOVATION 105
6.1 Context of the Study 105

6.1.1 Social 105
6.1.2 Technical 108
6.2 Game Development 108
6.2.1 Bug Game Scenarios & Requirements 109
6.2.2 Game Process 113
6.3 Adaptivity Model 114
6.3.1 Adaptive Approach 114
6.3.2 Model 116
7 EVALUATION STUDY: BUG GAME 119
7.1 Rigor of the Study 121
7.1.1 Restatement of the objectives and Research Questions of the Study 122
7.1.2 Researcher´s situated perspective 122
7.1.3 Aim 122
7.1.4 Evaluability Assessment 122
7.1.5 Logic Model 123
7.1.6 Evaluation Design 124
7.1.7 Site Selector 124
7.1.8 Participants Selection 124
7.1.9 Research Methodology 125
7.1.10 Research Design 125
7.1.11 Data Collection instruments 125
7.2 Qualitative Study: Experts´ interviewing 126
7.2.1 Interviews 127
7.2.2 Interviewees 128
7.2.3 Interview Structure 128
7.2.4 Interview Procedure 128
7.2.5 Interview Results 129
7.2.6 Validity 131
7.2.7 Reliability 131
7.3 Participant Observation Study 131
7.3.1 Video Material Permission 131
7.3.2 Participant Observation 132
7.3.3 Interviews Post-test 132
7.4 Research Product 133
V EVALUATION REPORT 135
8 DATA ANALYSIS & DISCUSSION 137
8.1 Summary of the methodology 137
8.2 Present evidence and ideas from sources 137
8.3 Discussion 139
8.3.1 Bug Game: Game Situation I 140

8.3.2 Bug Game: Game Situation II 142
8.3.3 Bug Game: Game Situation III 142
8.3.4 Bug Game: Game Situation IV 142
8.4 Findings in Chart form 142
8.5 Experience and Perception 142

9 CONCLUSION & RECOMMENDATIONS FOR FUTURE WORK 147
9.1 Restatement of the Research Question 147
9.2 Summary that systematically answer your Research Questions 147
9.3 Possibilities & Constraints 147
9.4 Recommendation for future Research 148
9.5 Contributions 148

VI APPENDIX 151

A APPENDIX TEST 153
A.1 Appendix Section Test 153
A.2 Another Appendix Section Test 153
A.3 Another Another Appendix Test 153

BIBLIOGRAPHY 155

# LIST OF FIGURES


Figure 1 Game Flow Corridor Masuch [63] 9
Figure 2 Games that entertain, interact and educate. 16
Figure 3 Arts of Serious Games & their operational areas Göbel [34] 17
Figure 4 Attributes of a serious game [92] 19
Figure 5 Serious in Game Development Masuch [63] 22
Figure 6 The flow channel of Csikszentmihalyi (Juul, 2009) Pomper et al. [73] 32
Figure 7 The Flow Theory Masuch [63] 33
Figure 8 Practitioner Perspective Vs. Researcher Perspective Among Evaluation 61
Figure 9 The Research Variables, Methods, and Results of Empirical Studies reviewed 81
Figure 10 Typical system architectures Sklar and Richards [89] 86
Figure 11 LeverageTM 98
Figure 12 Emotions & Facial Expressions 107
Figure 13 REDUCING UNPRODUCTIVE LEARNING ACTIVITIES IN SERIOUS GAMES P23 138
Figure 14 REDUCING UNPRODUCTIVE LEARNING ACTIVITIES IN SERIOUS GAMES P108 138
Figure 15 REDUCING UNPRODUCTIVE LEARNING ACTIVITIES IN SERIOUS GAMES P109 139
Figure 16 REDUCING UNPRODUCTIVE LEARNING ACTIVITIES IN SERIOUS GAMES P110 140
Figure 17 The Relations Among the Adaptive Variables and their effect on the motivation of children 142


# LIST OF TABLES


Table 1 Literature Review: Studies on the impact of Adaptivity 80
Table 2 Literature Review: Studies on the impact of Adaptivity 82
Table 3 Literature Review: Studies on the impact of Adaptivity 84
Table 4 Literature Review: Studies on the impact of Adaptivity 90

Table 5 Literature Review: Studies on the impact of Adaptivity 96
Table 6 Literature Review: Studies on the impact of Adaptivity 100
Table 7 Learning Content vs. Game Aspect Skills in Bug Game 113
Table 8 Events in Bug Game 114
Table 9 Adaptable variables & their properties in Bug Game 115
Table 10 Intervieews 128
Table 11 Adapt Events & their influence on the game process in Bug Game 130
Table 12 Game Situation 1 (3-5 years); Starter Level 141
Table 13 Game Situation 2 (4-6 years) 143
Table 14 Game Situation 3 (5-7 years) 144
Table 15 Game Situation 4 (6-8 years) 145

## LISTINGS

## ACRONYMS

# Part I

# INTRODUCTION

If we adapt, we should be aware of the Game Flow, we should adapt individually, with just the right amount and responding to the actual needs of the player. Making the Events and the Tasks more complex would motivate and encourage the child, otherwise, if too hard, that would interrupt the gamplay and annoy the child.

# 1 INTRODUCTION

## 1.1 PHILOSOPHY

The purpose of this work is to evaluate the impact of adaptivity which we would integrate in a serious game designed on the iPad for preschool children aged from 4 to 8.

Evaluation relies on objectivity, otherwise the outcomes wouldn´t be neither valid nor reliable. [Ref?]

Being objective consists in conducting an intensive literature Research on similar studies conducted earlier on serious games having the goal to proof and measure the effectiveness of serious games on the learning enhancement by preschool children.

Thus, i present my thesis this way:

- **an Exploratory Research** answering the question: What are Serious Games?
  Exploratory studies are essential whenever a researcher is breaking new ground, and they can almost always yield new in-sights into a topic for research.
  Exploratory studies are also a source of grounded theory.

  *Part 2; Chapter 2*

- **a Descriptive Research** answering the question: How are Serious Games designed and developed?

  *Part 2; Chapter 3*

- **an Explanatory Research** explaining the need of Research in Serious Games nowadays.

  *Part 3; Chapter 4*

- **a Comparative Research** making comparaisons across different studies made in this field.

  *Part 3; Chapter 5*

- **an Evaluation Framework** assessing the impact of Micro-Adaptivity on children´s motivation and enjoyment.

  *Part 4; Chapter 6, Chapter 7*

- and Finally delivering **an Evaluation Report**.

  *Part 5; Chapter 8, Chapter 9*

The game, we present here, consists of five mini-games, an intro game and an award game, have been designed in association with USM with their long-year experience in serious games, Serapion with their innovative Adaptivity Framework and TUM: Chair for Applied Software Engineering.

To the contribution of the design this game and the choose of the learning content have participated next to the Dance Research Group also Dr. Peter Pohl, Children Psychologist and Fr. Bartl, Children Pedagogue.

*Thoughts 10.09.11*

- Due to the limitations encountered conducting such an evaluation on a wide population of children (underage), i have chosen to conduct better a case study consisting in successive usan´ability Testing sessions which have been analyzed afterwards.
- Limitations consisting in: - Different interpretations of children´s emotions - High-Cost - Time - Population of underage children (50 to 100) with different cultural background
- Usability Testing could be analyzed thanks to the data gathered from observation sessions: - Enjoyment on the iPad / Killing the Bugs, interacting and moving to reach the whole of the screen - Skills acquired per Game - Skills growing through games - Tests to evaluate the acquired learning plus in maths, such as recognition of numbers (dots on the back of Bugs), arabic numbers in Labyrinth, writing numbers on the right way, we do all life, . . . ) Addition in Labyrinth, Rope and Flashlight Games
- I have chosen to conduct a test on each child (Case Study of 5 to 10 members) showing an increase in mastering both of recognition and addition & an increase in hand having the iPad as a learning instrument better than play consoles. Because of the fact that giving a child the possibility to try on early age to interact with new technologies such as tablets and smartphones enriches their Way and let them be interactive, play and enjoy the innovative gameplay & learn automatically. The learning occurs due to the tasks planned into the games.

## 1.2 OVERVIEW OF THE SCIENTIFIC RESEARCH

- First I have tried to research deeply the serious games field searching for both design and evaluation framework. But a general framework were never designed.
- Thus I contributed to make it more qualitatively by discussing how our design contribute to more enjoyed gameplay and an enhancement in learning skills.
- Towards a better cognitive development of preschool children we would give them through play the ability to develop their own learning strategy from young on, what helps them later on in their lives.

- Since our game would be adaptive, depending on who is playing and how much skills does she/he have, I have chosen to make the **quantitative part** as a Quasi-Experimental Case Study on one of the mini games, the Bug Game with the most of the adaptive variables.
- There for we would compile different Game Situations with different values of variables.
- Some usability testing sessions were first conducted to test the faisability such a pilot study.
- We have contacted psychologists to assist us on the evaluation days in understanding children's emotions.
- A group of helpers were mobilised to measure the impact of each variable on a group of children, another one to evaluate the emotions, a third one to measure the time taken for each level or until the first right answer or until the first disenjoyment/angryness per child.
- Here we would extract the effective time before activating the Adaptivity Framework and the best time to switch between levels predefined from the pedagogue Mme Bartl.

## 1.3 PROBLEM

*Evaluation of the Real-Time adaption mechanism in DEG's.*
*Developing playable games that adapt to the user.*
*Games learn to adapt themselves to satisfy and motivate the user.*

- Evaluation of the environment aspects, pedagogical aspects, usability, social presence.
- How to measure the children's emotions?
- How to measure the Learning effectiveness?
- What about gathering impressions from the teachers about the children Amelioration.
- „Evaluation ist die systematische und zielgerichtete Sammlung, Analyse und Bewertung von Daten zur Qualitätssicherung und Qualitätskontrolle. Sie gilt der Beurteilung von Planung, Entwicklung, Gestaltung und Einsatz von Bildungsangeboten bzw. einzelnen Maßnahmen dieser Angebote (Methoden, Medien, Programme, Programmteile) unter dem Aspekt von Aktualität, Funktionalität, Wirkungen, Effizienz und Nutzen." - Dr. Sigmar Tergan.
- Evaluation im Bildungsbereich: „das Erfassen und Bewerten von Prozessen und Ergebnissen zur Wirkungskontrolle, Steuerung und Reflexion" - Dr. Jost Reichmann.

- **The aim of this paper is threefold:**
  **First, to provide a framework for the design of good Serious Games.**
  **The second is to provide a framework for the evaluation of Serious Games.**
  **And the third is to conduct a case study evaluation on one mini game, willing to define the requirements of the adaptive mode.**

## 1.4 RATIONALE, SIGNIFICANCE AND IMPORTANCE OF THE SCIENTIFIC RESEARCH

- The primary research step is a **comprehensive literature study** that will cover many research disciplines, namely:
  - the development psychology by young children
  - the technology we develop on it
  - ...
- review relevant studies conducted in this field over the past twenty-five years and extract Guidelines and Limitations from them.
- Design guidelines will be extracted from prominent theories of Cognitive Development such as those of Piaget and Vygotsky.
- The validity of these guidelines will be tested empirically.
- We work from the assumption that the wealth of an intensive literature research would provide us with principal guidelines for the design of both of the game and the evaluation framework.
- Research Methodologie

## 1.5 PURPOSE

- Evaluation Goal
- In order to integrate adaptivity in our Game, we would evaluate the impact of this on:
  - Gameplay,
  - Challenges in Game,
  - Game Environment,
  - Children´s Motivation,
  - Children Skills´Acquisition

- The purpose of this study was to investigate the effects of the Micro-Adaptivity Framework integrated in a mathematics digital educational game on the behaviour and emotions of children playing with it. To fulfill this purpose the effects of learning achievement and motivation toward mathematics was investigated.

## 1.6 RESEARCH QUESTIONS & HYPOTHESES

### 1.6.1 *Research Questions*

All Research Questions from Damir´s dissertation would be integrated here in the introduction. The most of them will be referenced to Damir´s Ph.D. Thesis and the ones dealing with evaluation questions will be answered here after the chapter of Comparative Literature Research: Trends in Research in Serious Games.
[1]

- How would experts, psychologists, and pedagogues interpret the value of Micro-Adaptivity in our Game?
- How would we vary the adaptive variables, without annoying children or without affecting both Gameplay and Challenge?
- When do children loose the interest in the game?
- When should we activate the adaptive mode? (Developers´point of vue)
- When would experts/pedagogues/teachers suggest to activate/enable the adaptive mode? (Experts´point of vue)
- Which variable should we adjust and how much? (Developers´point of vue, Experts´point of vue)
- As "Bad games have little challenge, or too much challenge. Good games have just the right amount." Schell (Schell2008), how would we, by means of Adaptivity, adjust the challenge in the Game within the Game flow Corridor?
- How would we ensure an individual and continuous Gameplay through the Flow Channel?
- Which variable and with which amount is responsible for and how to maintain the challeges in Game continuous? Longer Gameplay & Enjoyment; Success in Tasks; Better Skill Acquisition?

1 Contribution from my comparative research: From [Reducing Unproductive Learning Activities in Serious Games Wu [101]] to [(2002) Why Scaffolding Should sometimes Make Tasks More Difficult for Learners Reiser [81]] are two worlds, we seek the middle, the right individual Adjustment.

- 
  - What effects do the games have on the students' academic mathematics achievement, as measured by (a) the school district-wide benchmark exam, and (b) the treatment game performance test, provided in Appendix A?
  - What effects do the games have on students' motivation as measured by a motivation survey, Course Motivation, developed based on Keller's ARCS Model (1987a) and provided in Appendix A?
  - How do individual differences of prior knowledge, computer experience, and language background affect students when using the game? In this study, prior knowledge is referred to preexisting mathematics knowledge and language background is referred to English fluency. These two factors were determined based on the participants' school record. Computer skill was determined by the Demographic survey provided in Appendix A.

*RQs: From General to Specific*

Therefore, the current study investigated the application of different Game Situations of the Bug Game on a single-group of children, by focusing on these research questions:

- When do children loose the interest in the game?
- When should we activate the adaptive mode? (Developers´point of vue)
- When would experts/pedagogues/teachers suggest to activate/enable the adaptive mode? (Experts´point of vue)
- Which variable should we adjust and how much? (Developers´point of vue, Experts´point of vue)
- As "Bad games have little challenge, or too much challenge. Good games have just the right amount." Schell [86], how would we, by means of Adaptivity, adjust the challenge in the Game within the Game flow Corridor?
- 
- Which variable and with which amount is responsible for and how to maintain the challeges in Game continuous?
  - longer Gameplay & Enjoyment
  - Success in Tasks ()
  - Better Skill Acquisition
- How would we ensure an individual and continuous Gameplay through the Flow Channel (ZPD Concept)? 6 on page 32

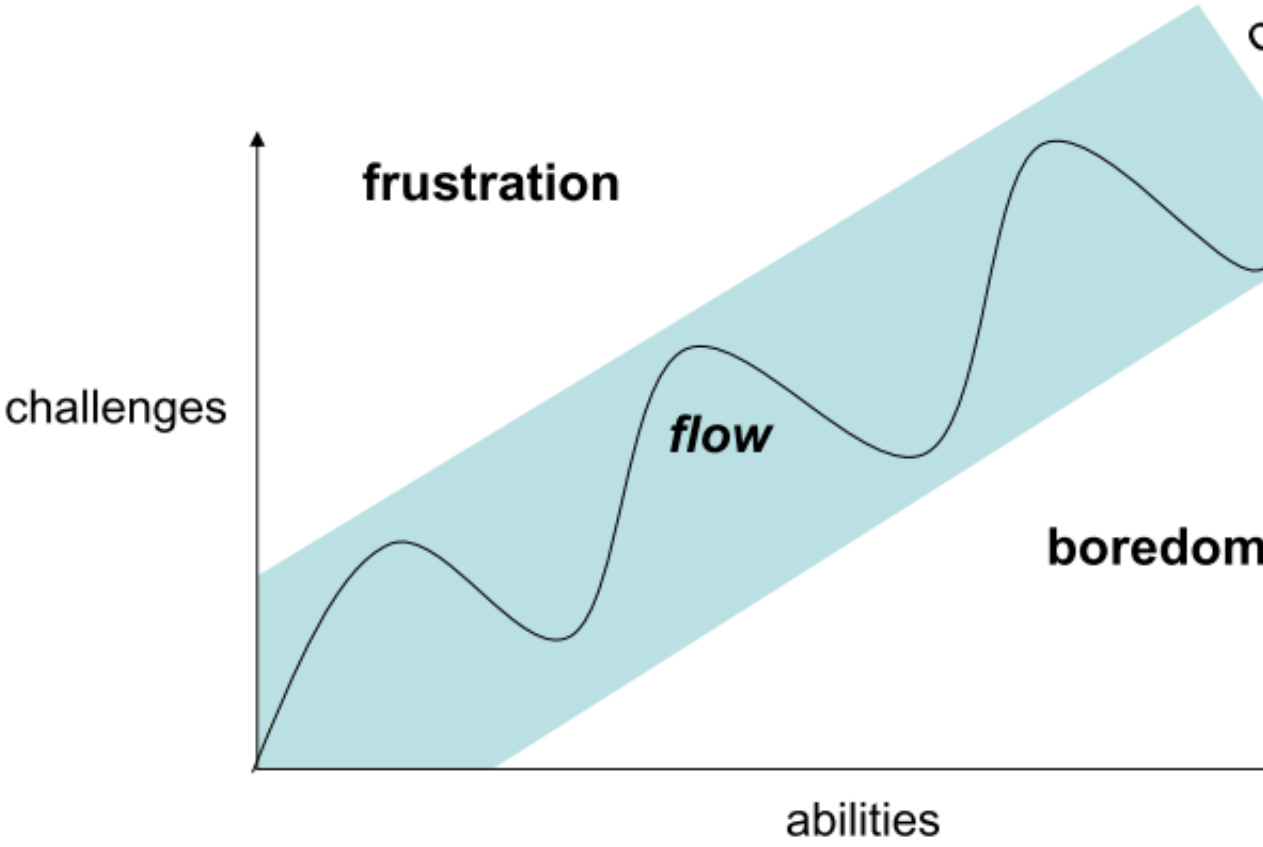



Figure 1: Game Flow Corridor Masuch [63]

### 1.6.2 *Research Hypotheses*

- Since every Mini Game have different adaptive variables, we would evaluate them separately
- Damir & i have chosen the Bug Game with the most number of adaptive variables as the favorite to make a case study on it, so we can compile different Game Situations or levels, that would be more suitable to children.
- 
  - Three hypotheses were posed to help answer the research questions.
    - There is no significant difference between learners' achievement of the experimental group, who receive the pre-Algebra and/or Algebra I instructional games, versus the control group, who do not receive the games.
    - There is no significant difference between learners' motivations of the experimental group, who receive the pre-Algebra and/or Algebra I instructional games, versus the control group, who do not receive the games.
    - There is no significant difference between effects of the games on students with differences in (a) prior knowledge, (b) computer experience, and (c) language background.
- The hypotheses we wish to evaluate are causal claims and thus an intervention study is warranted.
- Test Variables: self-confidence, value, enjoyment, motivation, challenge, skills, tasks, time, score, ...

## 1.7 OUTLINE

1. Introduction
2. State of the Art Serious Games
   a) Serious Games: Definition, Art, Background theories (Game-Flow-Theory & Motivation), ...
   b) Adaptivity: Definition, Art, Theory & Background, Notions of Adaptivity (ZPD & Scaffolding), Tutor, Macro, Micro, ...
3. Scientific Research in Serious Games
   a) Literature Review about SR
   b) Trends in SR in SG

c) Conclusion, Contributions, My methodology & Why?

4. Evaluation **{Bug Game}**

a) Game Development, Game Process, Model Adaptivity (Tasks, Events, Skills, Assessment in Games, ...) **{Bug Game}**

b) Study I: **Qualitative**: Interviews with experts, Analysis, Results (Level Simulation / Game Situations), Discussion, ... **{Bug Game}**

c) Study II: **Qualitative**: Quasi-Tests: Children Observation while playing Game Situations, Record (No DATA / Only Video), Experts Analysis (Videos interpretation / Theories about correlaion between variables & their impact on children), Results (from hypotheses to theories about adaptive variables / No Charts), Discussion, ... **{Bug Game}**

5. Report

a) Results (AgeDerivedLevels, theories about variables correlation, theories about their impact on children), Reuse of Results in planned Future Work

b) Contributions about Assessment within Games

Part II

# STATE OF THE ART: SERIOUS GAMES

Guidelines to Serious Games Development. Foundation à propos Evaluation Framework in this field. Exploratory (WHAT) & Descriptive (WHAT & HOW) Research.

# 2

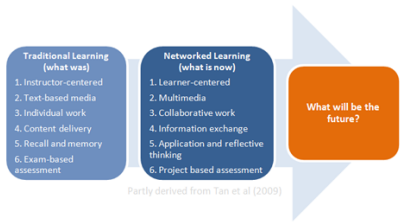
# SERIOUS GAMES

By citing Clark Abt's definition from 1970: Serious games are games that ... "have an explicit and carefully thought-out educational purpose and are not intended to be played primarily for amusement. This does not mean that serious games are not, or should not be, entertaining." (Abt 1970, p9).

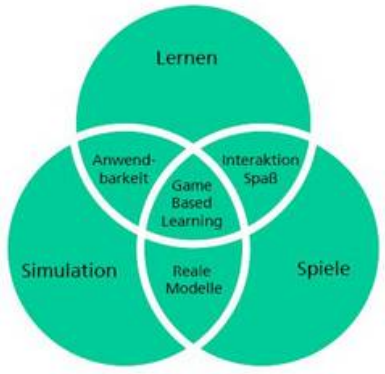
## 2.1 HISTORY AND EVOLUTION

- Serious games, skill games, immersive digital educational games, ...
- Gameplay & Motivation
- in the last two decades there has been concern that play is being replaced by other activities such as computer games, ranging from commercial arcade games to different kinds of educational software. [96]
- Educational software developers have struggled to present a substantial amount of content and context without sacrificing the degree of control game players expect.

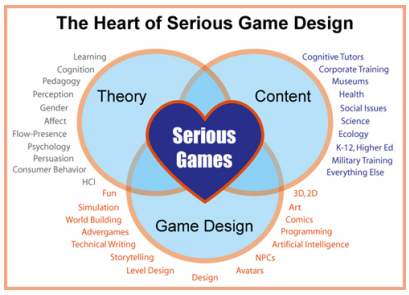
- So Serious Games are not about putting learning courses on a game; you should not think about Educational Games as delivery of courses. Games Based Learning GBL are about augmenting our learning and our performance. This includes a role in formal learning and, occasionally can be the delivery mechanism for a full learning solution, but the real opportunity is augmenting learning and performance, not learning delivery. To be fair, I also argue that Games are not about learning, but rather that they are about fun. We do not learn for intellectual self-gratification, we learn because we want or need to do something that we cannot do now.
  Serious Games really are about learning, when you use a broad definition of learning. Problem solving, creativity, information access, collaboration, innovation, experimen- tation, and more are all part of learning in my definition.- mLearning
- Parallels between good gaming and good learning experiences [Gee, J. 2003] suggest the use of computer games in the education sector. In this paper A Framework for the Development, Design



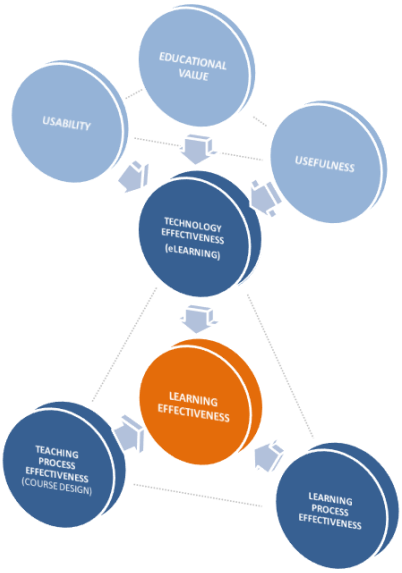

(a) Traditional vs. Futuristic




(b) Simulation, Games & GBL




(c) Serious Games




(d) Learning Effectiveness

Figure 2: Games that entertain, interact and educate.

and Deployment of Customisable Mobile and Hand Held Device Based Serious Games by Hanno Hildmann, Cherif Branki, C Pardavila and Daniel Livingstone published in [19].

- "Nowadays, we are surrounded by technological and multimedia products in every aspect of our everyday lives. One of these products is digital games, which have seen a gradual increase in use over the last twenty years for multiple purposes (Jeong, Park, Ryu & Lee, 2008). Digital games occupy an important part of most children's leisure time and an important part of our culture as a whole (Kirriemuir & McFarlane, 2004)." [70]

- Educational games are being used by many different people (Rieber, 1996): parents buy games for their children, teachers find games online that address the content they are focusing


Figure 3: Arts of Serious Games & their operational areas Göbel [34]

on, and many students find these environments engaging and worthwhile. Underwood et al. [93]

- Serious games (with rare exceptions) are developed outside the mainstream game studio and game publishing industries with the intent to teach, persuade, or promote using established videogame dynamics. Iuppa and Borst [45]

## 2.2 ARTS OF SERIOUS GAMES

- "... mLearning is not about courses and delivery of formal learning, mLearning really is about learning, when you use a broad definition of learning. Problem solving, creativity, information access, collaboration, innovation, experimen- tation, and more are all part of learning in my definition." [76]

- "Throughout the literature, several classifications of digital games can be found that do not meet a general consensus among researchers.
  Taking into account the studies of Kirriemuir & McFarlane (2004),

Schiffler (2006), Felicia (2009), Kickmeier-Rust (2009), and Whitton (2010), we can classify digital games in the following **seven categories**:

- Action Games (interactive gameplay requiring fast reflexes and hand-eye coordination). Included in this category are reaction-based, platform, maze, shooting, fighting, racing and sport games.
- Role-Playing Games (interactive fantasy gameplay requiring planning). Research on this field led to the development of MMORPGs (Massive, Multiple, Online Role-Playing Games).
- Strategy Games (requiring analytical thinking, reasoning and planning). These games formed the basis for the development of RTS (Real Time Strategy) type.
- Adventure Games (interactively experiencing narratives with cognitive/reasoning aspects).
- Simulation Games (replay real or fictitious situations).
- Puzzle Games (matching or constructive puzzles).
- Educational Games (games developed for educational purposes); Games that teach through lively activities combining education with entertainment (Yanhong, Liming & Lifang, 2010). This combination transforms the game into an effective tool for educational purposes, with elements such as immediate feedback, interactivity, and challenge (Annetta et al., 2006; Amr, 2007; Yusoff, 2010).

### 2.2.1 *Video Games*

*[Wikipedia]*

Video games Educational video games are considered a type of serious game, as these games have a strong purpose other than pure entertainment.[Video games 'stimulate learning', BBC News, March 18, 2002] Some people call these types of games edutainment because they combine education and entertainment. An educational computer game can be defined as an electronic medium with all the characteristics of a gaming environment that have intended educational outcomes targeted at specific groups of learners. Video games can aid the development of proficiency by allowing users to interact with objects and manipulate variables. They are said to be particularly effective when designed to address a specific problem or teach a certain skill in curriculum subjects, where specific objectives can be stated and when deployed selectively within a context relevant to the learning activity and goal. Simple types of games can be designed to address specific learning outcomes such as recall of factual content. For instance, the Nobel Prize Foundation website uses on-line games to aid children in understanding the discoveries made by its laureates by embedding

| Area | Pedagogy | Game elements | Reality |
|---|---|---|---|
| Attributes | Attributes | Harmony | Learning objectives |
| | Experience | Uncertainty | Target group |
| | Low resource demanding | Interactivity | Challenge |
| | Exploration | Engaging | Clients |
| | Incremental | Flow | Organisation |

Figure 4: Attributes of a serious game [92]

the scientific knowledge as part of the game environment. To aid in educating students and adults about the finer details of different political systems, numerous companies have developed simulations that immerse the player into different political systems by forcing them to make realistic political decisions. These games vary from running an actual election campaign to games that allow the player to make the day-to-day decisions of running a country, as seen in Democracy. These types of games are targeted at students, educators and adults alike.

## 2.3 SERIOUS GAMES BACKGROUND

- Harteveld et al. 2007;
  Their goal was not only to develop the game but to guide future developers in making entertaining and educational serious games.
  They theorise that a player should understand that a game has specific learning goals in order to appreciate the results. However, without this explicit framing, which may detract from the play element (games are usually played voluntarily with player control), there could be less learning as the player focuses on the goals and rules of the game66. Thus there is a fine balance when designing or using games with an educational focus.
  Their approach was to divide serious games into three areas for consideration.

- Gameplay, Fun and Motivation Prensky [75]
  - Why is motivation such a big problem? Because all learning requires effort, and, like crime, people rarely do it without a motive.
  - The TRUE 21st century learning revolution is that learning is finally throwing off the shackles of pain and suffering that have accompanied it for so long.
  - Fun : The Great Motivator
  - The Oxford English Dictionary 4 defines fun as:

    - A cheat or trick; a hoax, a practical joke
    - Diversion amusement, sport; also boisterous jocularity or gaiety, drollery. Also, a source or cause of amusement or pleasure.
    - to make fun of, poke fun at (a person, etc): to ridicule. For or in fun: as a joke, sportativly, not seriously. (he, it is) good, great fun: a source of much amusement. Like fun: energetically, very quickly, vigorously. What fun!: how very amusing 1 for the fun of the thing: for amusement; to have fun with: to enjoy (a process); spec. to have sexual intercourse. c. Exciting goings on. Also fun and games, freq. Used ironically; spec. amatory play. Colloq.
  - Microsoft's Encarta World English Dictionary defines fun as:
    - amusement: a time or feeling of enjoyment or amusement. Just for fun, we wore silly hats.
    - something amusing: something such as an activity that provides enjoyment or amusement. Skiing is fun for the whole family.
    - mockery: playful joking, often at the expense of another. What's said in fun can still hurt.
  - dare I ask – what is "learning?"
  - Fun and Learning
    Without being able to define either "fun" or "learning" specifically, can we say anything about the relationship between them? Does having fun hurt learning, or help it? Some researchers have looked at this:
    - "In simple terms a brain enjoying itself is functioning more efficiently."... "When we enjoy learning, we learn better" (Rose and Nicholl, 1998)
    - "Enjoyment and fun as part of the learning process are important when learning new tools since the learner is relaxed and motivated and therefore more willing to learn." ... "The role that fun plays with regard to intrinsic motivation in education is twofold. First, intrinsic motivation promotes the desire for recurrence of the experience...Secondly, fun can motivate learners to engage themselves in activities with which they have little or no previous experience." (Bisson and Luckner, 1996)
    - Bisson and Luckner also cite Middleton, Littlefield & Lehrer, 1992; Datillo & Kleiber, 1993; and Hastie, 1994; as concluding that fun increases learners' motivation.

  - So fun in the learning process creates relaxation and motivation. Relaxation enables learners to take things in more easily; motivation enables them to put forth effort without resentment.
  - Gameplay
    Which leads me to a perhaps unfamiliar word: "Gameplay." Remember, game designers focus primarily on motivation; educators don't. The most important thing that educators can learn from game designers is how they keep the player engaged. Gameplay is all the activities and strategies game designers employ to get and keep the player engaged and motivated to complete each level and an entire game.
  - Gameplay and Education
    How can we bring more Gameplay – **continuous**, **appropriate challenge** – into the educational process? If we make it our goal, we can easily insert Gameplay into almost everything we do.
  - In the jargon of today's students, "Gameplay Rules!"
    Why? Because Gameplay Motivates!

- Cognitive Development through interaction.
- Vygotsky´s Theory.
- Paiget´s Theory.
- Bruner´s Theory.
- Try to find Similarities and what is intersting for our Design

*Damir gefällt 5*

### 2.3.1 *Pedagogical Background*

- Games are the language of children.
- The pedagogical aspects focused on scaffolding, usability, social presence and embedding games within the curriculum.
- The motivation to play and therefore to learn.
- The need for pedagogical support during learning - and therefore during gaming.
- "Flintoff (2002) argues that computer games can be useful in enhancing memory capacity, in concentration of attention and in the problem solving strategies of young children, which can indirectly affect their academic achievements." [96]

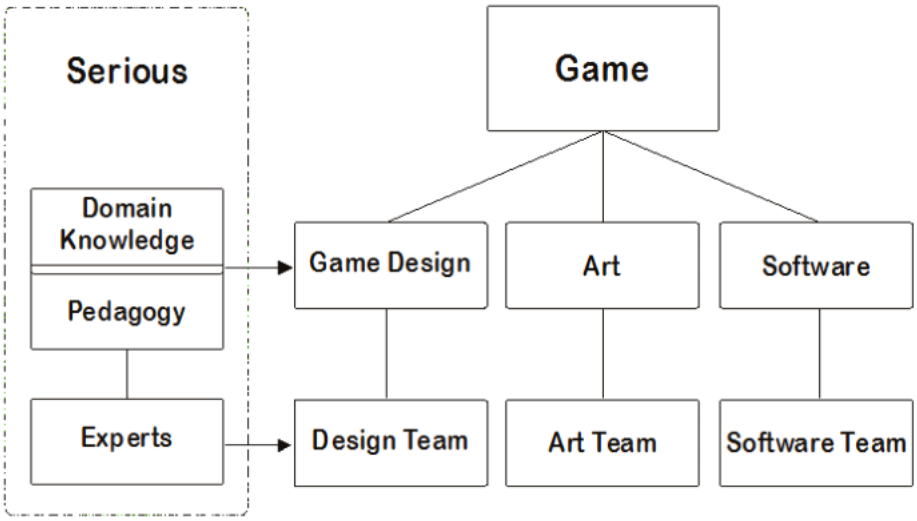



Figure 5: Serious in Game Development Masuch [63]

*limitations*

- Serious games are pedagogical multimedia products made to help learners develop specific competencies. Their use has proven to be promising in many domains, but is at present restricted by the time consuming and costly nature of the developing process. When developing Serious Games (SGs) for academic purposes, not only is there a budgetary challenge, but there is also the challenge of integrating enough educational value without sacrificing the fun characteristics. [1]

1. Potential Learning Tool

   - "Recently computer games have been proposed as a **potential learning tool** by both educational researchers (e.g., Barab, Thomas, Dodge, Carteaux, & Tuzun, 2005; Betz, 1995; Gee, 2003; Kafai, 1995; Malone, 1981; Rieber, 1996; Squire, 2003) and game developers (e.g., Aldrich, 2004; Prensky, 2001)." [48]
   - "Frequently-cited arguments held by these researchers for using computer game in education are:
     - (a) computer games can invoke intense engagement in learners (Malone, 1981; Rieber, 1996)
     - (b) computer games can encourage active learning or learning by doing (Garris, Ahlers, & Driskell, 2002)
     - (c) empirical evidence exists that games can be effective tools for enhancing learning and understanding of

1 Tools and Methods for Efficiently Designing Serious Games Meyer [65]

complex subject matter (Ricci, Salas, & Cannon-Bowers, 1996)

– (d) computer games can foster collaboration among learners (Kaptelin & Cole, 2002)." [48]

2. Mutimodality

   - "Multimodality in a game increases likelihood of knowledge acquisition, knowledge application, and enhances understanding of complex and abstract phenomena (Klimmt, 2009)." [38]

3. Interactivity

   - "Interactivity has a similar effect as Multimdality since it caters to the individual learner's speed and capacity (Klimmt, 2009)." [38]

- Enhancing learning management systems to better support computer science education Röß ling et al. [83]

### 2.3.2 *Psychological Background*

Developmental psychology provides us with an extensive knowledge base on children's development. Although some designers do refer to the work of developmental psychologists like Piaget and Vygotsky, they could make more extensive use of the results of a decade of research on children's development.

1. Brain Health

   - On top of that, certain studies have proved that regular play of brain challenging games can contribute to brain health. Just like a muscle can keep fit when it's used regularly, brain can be also kept health by performing simple exercises, just like our game does, also including the additional benefit of knowledge increase and versatility and interesting contents. in this paper Geography Educational Gaming System for Mobile Devices by Juan Manuel de Blas Quintana, Salvador Otón, Roberto Barchino, and José Ramón Hilera published in [19]
   - "... Our goal is letting us do what our brains do well and providing support for what our brains do not do well. Our brains are very good at pattern matching and reasonably good at executive monitoring, but not so good at performing rote operations." [76]

*Damir gefällt: Play is a Pattern*

2. Play

- "Early childhood educators know how important play is in children's lives. Play is not only an enjoyable and spontaneous activity of young children but it also 99 contributes significantly to children's psychological development." [96]
- "The toys and gestures with which children play are seen to be significant artefacts from their social and cultural settings—so, in play, children are acquiring the tools and meanings of their culture." [96]
- "Play prepares children for adulthood (**Groos** 1898, 1901). Based on the assumption that play is unique to childhood, Groos argued that play develops children's physical and mental capacities that will serve them as adults." [96]
- Surplus Energy: "Schiller, a German philosopher, defined play as 'the aimless expenditure of exuberant energy' (Schiller 1875:112, in Dockett and Fleer 1999:24)" [96]
- "Some Play theoreticians (Lazarus 1883, in Dockett and Fleer 1999, Patrick 1916) have argued that play is used to restore energy—this theory of play is known as Recreation or Relaxation play theory." [96]
- "Vygotsky (1977, 1978) viewed play as highly significant to development. 'Play contains in a concentrated form, as in the focus of a magnifying glass, all developmental tendencies.' (Vygotsky 1978:74). Vygotskians view play as the most significant "leading" activity of the early childhood years (Vygotsky 1977, Bodrova and Leong 1996). This means that the most significant psychological achievements of the early childhood age occur while children engage in play." [96]
- "Play serves to either raise or lower levels of stimulation, depending on whether a child is under- or over-stimulated. Play provides novelty, uncertainty and complexity at optimal levels for children; these qualities in optimal amounts are seen to be most conducive to individual functioning. The balance between the new and the familiar is often seen applied to education." [96]
- "Piaget (1962) shifted the focus of study from social and emotional aspects of play to children's cognition. He placed play within his stage-based theory of cognitive development and assigned it a significant role in the growing of children's minds." [96]

### 2.3.3 *Socio-cultural Background*

- GBL are an effective strategy for teaching socio-emotional skills to children and young people.

- "Sociocultural theorists discuss the overarching role of play in child development and view it as the most significant "leading" activity of the early childhood years (**Vygotsky** 1977, Bodrova and Leong 1996)." [96]

### 2.3.4 *Engagement*

- Gee (2003) admits three realizations he had for why games possess engaging qualities: [1. The computer game] requires the user to learn and think in ways in which I am not adept ... [2.] learning is or should be both frustrating and life enhancing ... [3.] game designers keep making the games longer and more challenging (and introduce new things in new ones), and still man- age to get them learned. (p. 6) Albers and Still [1]

### 2.3.5 *Motivational Background*

- "Digital games can motivate players by providing them with appropriate levels of challenge, curiosity, control, and fantasy (Malone & Lepper, 1987; Balasubramanian & Wilson, 2005). Habgood (2007) argues that a digital game is "an interactive challenge on a digital platform, which is undertaken for entertainment"." [70]
- Gaming is evidently associated with fun and fun is associated with motivation. Michael & Chen explain the relevance of fun in serious gaming. Fun is a result of the positive feedback somebody gets from a game. Pomper et al. [73]

### 2.3.6 *Play/Gameplay in Child´s Development*

- When I watch children playing video games at home or in the arcades, I am impressed with the energy and enthusiasm they devote to the task. Why can't we get the same devotion to school lessons as people naturally apply to the things that interest them? – Donald Norman, CEO, uNEXT. Prensky [75]
- "(Garvey 1977). Play has been characterised as a spontaneous, self-initiated and self-regulated activity of young children, which is relatively risk free and not necessarily goal-oriented. Play is intrinsically motivated: normally children have an internal desire and interest to engage in play, they are actively involved in creating their play and are in control of it." [96]
- „Der Mensch ist nur da ganz Mensch, wo er spielt" Friedrich Schiller

- „It is a paradoxical, that many educators and parents still differentiate between a time for learning and a time for play. Without seeing the vital connection between them.“ Leo Buscaglia
- "High Score Education: Games, not schools, are teaching kids to think." Wired Magazine Title Story, 2003

### 2.3.7 *Cognitive Development Background*

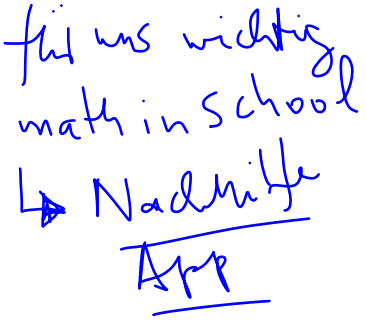
- ...
- "Modern and classical theories of play have identified many ways in which play may advance the cognitive, social and emotional development of children." [96]

### 2.3.8 *Learning Theories & Process*

- "The **traditional view of learning** is represented by the **behaviorism model** and **cognitivism model**.
  The **behaviorism model**, which is derived from the **stimulus** and **response** theory, suggests that learning is a change in the behavioral disposition of an organism that can be shaped through reinforcement (Jonassen, 1993).
  From the perspective of **cognitivism**, learning refers to the processing and transfer of new knowledge into long-term memories. The mind is perceived as an information processor with short-term and long-term memories. Knowledge is a storehouse of representations (patterns) which can be called upon for use in reasoning and translated into language (Hung, 2001). Stimulus can be provided by instructors to improve recall and retention of knowledge." [87]
- The **convergence of learning theories** and advances in technology created a space for the growth of innovative educational design and intervention, which is called **Emergent Design**. "It is an approach used for educational intervention; the claim is a more general one, however, in that the strategy is appropriate in settings for technologically enabled paradigmatic change (...).
  Through **Emergent Design** experiences it is possible to create a balance between digital technology and the approach to management of organization and of organizational change that has come in the wake of the technology.
  A distinction must be made because the temptation to use either of them alone has led to failure. It is the combination that offers an optimistic vision for the future of learning—the combination of these two products of the digital age along with a theoretical framework based on the work of pre-digital-age thinkers who

knew what to do but did not have the means to do it. Among these the most central is **Paulo Freire**, but also represented are **John Dewey** and **Jean Piaget**, although he did not focus on education per se [Cavallo, D.; Emergent Designing and learning environments: Building on indigenous knowledge; IBM Systems Journal, Vol. 39, NOS 3&4, 2000]. [28]

- Learning Principles: Albers and Still [1]
  - By learning and identity, Gee means learners are "willing to commit themselves fully to the learning in terms of time, effort, and active engagement.... [They must be willing to] learn, use, and value the new semiotic domain" (p. 59). The concept requires an engaging game that demands users learn the game and learn the content.
  - Situated meaning and learning suggest that "humans learn, think, and solve problems by reflecting [and connecting] their previous embodied experiences in the world" (p. 74); situated meaning is how students bridge game system goals and educational goals to make meaning in their own learning situation.
- Learning Principles by Gee 2003
  - Active, critical learning principle: All aspects of the learning environment (including the ways in which the semiotic domain is designed and presented) are set up to encourage active and critical, not passive learning.
  - Psychological moratorium principle: Learners can take risks in a space where real-world consequences are lowered.
  - Committed learning principle: Learners participate in an extended engagement (lots of effort and practice) as extensions of their real-world identities in relation to a virtual identity to which they feel some commitment and a virtual world that they find compelling
  - Identity principle: Learning involves taking on and playing with identities in such a way that the learner has real choices (in developing the virtual identity) and ample opportunity to mediate on the relationship between new identities and old ones. There is a tripartite play of identities as learners relate, and reflect on, their multiple real-world identities, a virtual identity, and a projective identity.
  - Self-knowledge principle: The virtual world is constructed in such a way that learners learn not only about the domain but also about themselves and their current potential capacities.



  - The 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; reprobing the world to test this hypothesis, and then accepting or rethinking the hypothesis.

*[Wikipedia]*

- Components of Interactive Learning: The concept of serious games involves immersing students in virtual worlds by means of role-playing and community interactive games. For learning, this means that the cooperative, critical-thinking, and problem-solving practices encouraged in digital games make serious games a key form of pedagogy. Adapting gaming to a form of experiential learning brings real-world issues into education within the structure of a planned curriculum. Along with their intrinsically engaging properties, games have been touted for their ability to teach ill-defined problem-solving skills, elicit creativity, and develop leadership, collaboration, and other valuable interpersonal skills.Gee, J (2003). What Video Games Have to Teach Us About Learning and Literacy. New York: Palgrave Macmillan.

### 2.3.9 *Effective Learning: Educational Attainment Level*

- "A recent review of game-based learning research indicated that **most gaming studies focus on learning conceptually** – concepts like general reasoning, creativity, system understanding and decision making, which does not demand special knowledge of subject areas (Bateson, 1972)." [48]

- "An extensive body of research demonstrates the positive effects of digital games on child and adolescent players' learning (Blumberg & Ismailer, 2009). This was even empirically supported by Robertson & Miller (2009) and Owston et al. (2009), especially for less able children." [38]

- "The nature of games promotes several vital skills for deep learning, such as metacognition, selective attention, problem solving, perspective taking, a chance to practice, thinking of alternative solutions, multiple modularities, multiprocessing, information literacy and are motivating (Blumberg & Ismailer, 2009; Charsky, 2010; Mason & Rennie, 2008; Davidson, 2008). James Paul Gee (2003), too, finds serious games to be an effective teaching tool that involve 36 learning principles, but not just by virtue of being a game." [38]

- "The use of information visualization systems, interactive techniques and multimedia tools in teaching and learning experiences tends to be effective. This statement is supported by the

fact that these tools can be effective in improving students' learning and data retention for 90% of learners when they verbalize something and then implement ..." [28]

# 3

# DESIGN ESSENTIALS

- This Chapter describes an intensive literature research aiming to formulate a set of guidelines for designing Serious Games for children aged four to eight.
- "Software designers aim to present educational content in a playful way to make it more attractive and accessible to its young audience." Verenikina et al. [96]
- Developing digital educational games (DEGs) that can cost-effectively foster learning with fun and pleasure is a vision for researchers and practitioners in the field of HCI and technology-enhanced learning (TEL). DEGs offer exciting and dynamic environments which engage players in meaningful and motivating learning activities. Law and Rust-kickmeier [55]

## 3.1 ADAPTIVITY

### 3.1.1 *Notion of Adaptivity*

- By generative, I mean that the system creates teaching materials and events as it progresses, it is not merely storing material prepared by someone.
- By adaptive, I mean that the teaching interaction is adjusted, both in concept and form to the current needs and abilities of each pupil individually.

*Damir´s Definition of Adaptivity Ismailović [40]*

- **Adaptivity** is a non-invasive approach that enables a serious game to learn from learner's behavior by intelligently monitoring and interpreting learner's actions in the game's environment and adjusts automatically learning and game elements according to the student's individual ZPD as necessary.
- **Adaptive serious game**: is a serious game that implements the *Adaptivity* approach.

*HCI*

- Given the growth of Child Computer Interaction research and of the rapid adoption of interactive technologies as teaching tools, next generation HCI technologies play an important role in the future of education. Educators rely on technology to improve/adapt learning to the pedagogical needs of learners, thus

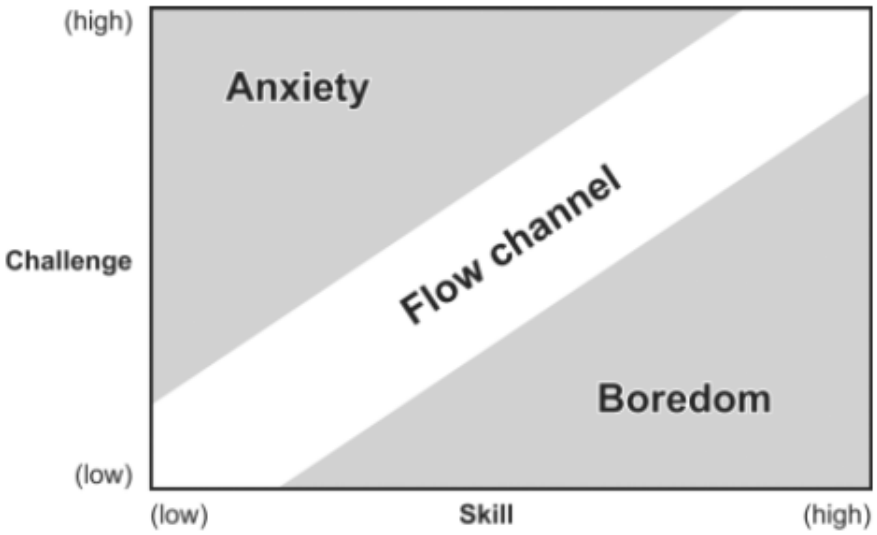



Figure 6: The flow channel of Csikszentmihalyi (Juul, 2009) Pomper et al. [73]

the HCI community needs to examine how these concepts can be matched to contemporary paradigms in Educational pedagogy. The classroom is a challenging environment for evaluation, thus new interaction techniques need to be established to prove the value of new HCI interactions in the educational space. Tse et al. [91]

*personal, unique, individual*

- Constructivist educators stress that learning is **personal**, **unique**, and **contextualized** for each learner (Reeves & Okey, 1996). Toward that end, we argue that games, along with technology that is finally available today, provide environments that can be adapted to the individual learner according to constructivist maxims. Underwood et al. [93]

*Flow Theory*

- A third important element of a serious game is the **immersion** of the player. A serious game stimulates that the attention and actions of the players/trainees remain focused on the **didactic goals of the game**. Game design theory refers to the **'Flow Theory'** of psychologist Mihály Csíkszentmihályi (2002) when describing player immersion. Csíkszentmihályi describes flow as the ultimate state of immersion. Pomper et al. [73]

*Damir gefällt: Challenge vs. Skill im Flow Channel*

*Flow Channel*

- According to this theory a player will stay immersed (in the **'Flow channel'**) when he constantly experiences the right balance between the challenges of the game and his own skills (Juul,

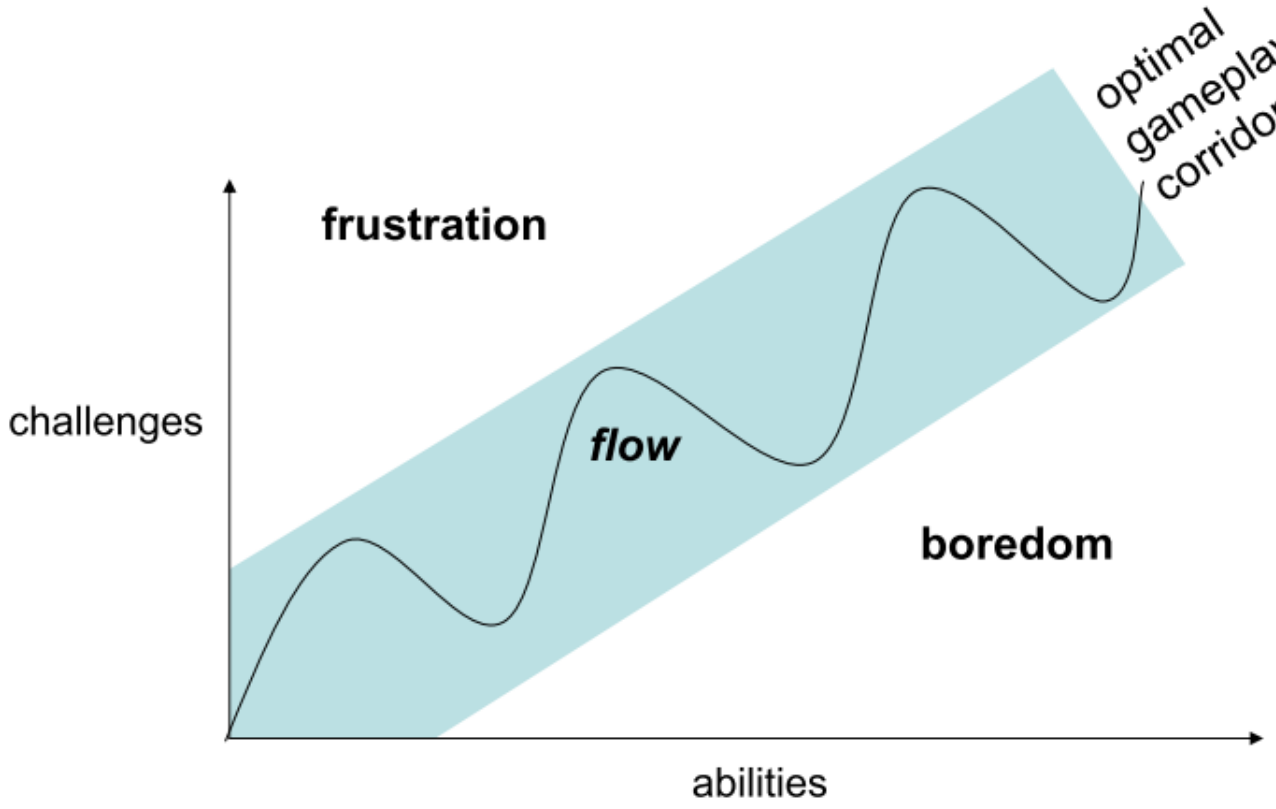



Figure 7: The Flow Theory Masuch [63]

2009).
For this reason a serious game has to be designed in such a way that it **constantly adapts the challenge to the skills of the player**.
When the challenge gets relatively too high, anxiety is the result, whereas boredom results from the players skill level being much higher than needed to overcome the challenges of the game. Pomper et al. [73]

- Gee (2003) is amazed at the effort users put into learning a game despite how increasingly difficult a game becomes. Albers and Still [1]
-

### 3.1.2 *Adaptive Gaming Environment*

- In a DEG, adaptive and interactive digital storytelling serves two essential purposes:

1. it strongly supports a **personalized learning experience** by adapting the game's story to individual preferences and by providing the possibility of explorative learning processes.
2. Second, it serves the **re-usability of learning material** by enabling the realization of different stories and entirely different games, also for a variety of different learning domains based on more or less the same pool of **story units**, **patterns** and **structures** as well as **learning** and **gaming concepts** and elements/objects.

### 3.1.3 *Types of Adaptivity*

Three different adaptivity modes are possible:

- Micro adaptivity; This is the adaptivity of tasks
- Macro adaptivity;In Macro adaptivity
- Tutor adaptivity Basically this adaptivity mode comes from intelligtent tutoring systems.
- To **help** learners across games, **suggestions** can be made for games learners seem to be ready for, either for remediation or challenge. Guiding these selections, learning is optimized when **the difficulty of the activity lies in the learner's zone of proximal development** (Vygotsky, 1978), where it is a challenge for the learner, but can still be accomplished with scaffolding via **hints** or **peer discussions**. Underwood et al. [93]
- The term Adaptivity is also similar to Performance.
  **In-game performance** can be used to **adapt** the game environment **to suit the learner**.
  Adaptations can take the form of dynamically:
  - swapping media (e.g., auditory learner gets audio cues, visual learner gets visual cues),
  - making content simpler or harder (e.g., using an aggressive boss in a simulation of "learning to be agreeable" if the player is successful with an approachable boss)
  - "re-skinning" graphical elements in a game (e.g., changing genders of non-player characters, altering billboards and street scenes). Underwood et al. [93]

### 3.1.4 *Tutor-Adaptivity (IntelligentTutoringSystem)*

- Many educators and researchers have long sought to use assessment in tutoring environments (e.g., Snow & Mandinach, 1991), leading to the development of several systems, including:

- **Cognitive Tutors** (Anderson, et al., 1995; Koedinger et al., 1997), which keep track of students' knowledge through the use of a cognitive model to offer immediate feedback and guidance;
- **ASSISTments** (Razzaq et al., 2005), which represent a derivation of the cognitive tutoring approach by merging instructional assistance with robust assessment; and
- **Assessment and Learning in Knowledge Spaces** (ALEKS;Falmagne, et al., 2004), which dynamically generates interactive courses adapted to a student's goals, preferences, capabilities, and knowledge. Underwood et al. [93]

  - While these tutoring systems have shown significant positive learning effects, they focus on procedures, constrain students to particular problem-solving paths, and eventually provide answers to problems instead of letting students attempt solutions and learn from their mistakes. Many lessons can be learned from these efforts to provide more open-ended environments for students to explore subject matter knowledge, procedures, and 21st century skills. Underwood et al. [93]

### 3.1.5 *Macro-Adaptivity*

*very short, should be extended*

- Macro-adaptivity refers to traditional techniques of adaptation such as adaptive presentation and adaptive navigation on the level of learning objects (LOs) (or learning situations (LeSs) in a DEG). Generally, macro-adaptive interventions are based on a fixed learner model (e.g., traits) or adaptation model (e.g., pedagogical implications) and on typical (knowledge) assessments (via test items). Law and Rust-kickmeier [55]

### 3.1.6 *Micro-Adaptivity*

- One big european project called ELEKTRA was focusing on tutor adaptivity, where a non playing character "Galileo" is interacting adaptively with the player/learner.
- In ELEKTRA, a project funded by the European Commission we developed a formal cognitive framework for the non-invasive assessment and interventions within complex learning situations, that is, **micro adaptivity**. Kickmeier-Rust et al. [51]
- The results indicate that (micro) adaptive interventions (i.e., appropriate and meaningful interventions/feedback for an individual learner, his/her knowledge and learning progress) are superior to neutral (i.e., non-individualized but semantically

correct interventions) and inappropriate interventions (i.e., non-individualized, unsuited interventions) in terms of learning and gaming measures. In addition, we analysed the relationships between learning progress and socio-emotional variables. The results indicate that adaptive feedback not only facilitates learning but also attitude and immersion. Kickmeier-Rust et al. [51]

- While adaptivity in Technology Enhanced Learning (TEL) has been an important topic in research and development for over a decade now, game-based learning is a rather recent advance in the field. The underlying idea is to make use of young people's motivation in computer games for learning purposes. If this is to be successful, the learning nature of the game should be as unapparent as possible. On the other side, adaptation to the learners' individual knowledge is very important in learning. Its presence ensures that the learning game, based on the learner's actual state of knowledge, provides personalized challenges and support. Albert et al. [2]

- The original idea behind developing the concept of microadaptivity was to present adaptive hints to the learner depending on her/his progress in the learning game situation. This idea has been generalized to the concept of adaptive interventions some of which depend on the learner's skill state while others do not. Albert et al. [2]

  - A skill activation adaptive intervention may be applied if a user gets «stuck» in some area of the problem space and some skills are not used although the user model assumes that the user masters these skills.
  - A skill acquisition adaptive intervention may be applied in a similar situation where, however, the user model assumes that the user does not master the unused skill.
  - Basically independent of the model is the application of motivational adaptive interventions. These might be applied, e. g., if the user does not act at all for a certain, unexpectedly long time.
  - Assessment clarification adaptive interventions may be applied, e. g., if the user's actions give contradicting support for and against the assumption of him/her mastering a certain skill.

- The architecture consists of four modules or engines.
  The learner is connected to the ELEKTRA system through the **game engine (GE)**. It provides the non-adaptive parts of the game, and as such it is also the user interface to the system. The GE provides information on the learner's action in the game to

the **skill assessment engine (SAE)**. The SAE updates the learner model (i.e. the skill state likelihoods) and the information it has in the skill ontology.
The resulting information about the learner's skill state and its changes are then forwarded to the **Educational Reasoner (ER)**, the pedagogical part of microadaptivity. Based on pedagogical rules and learning objectives, the **ER** gives recommendations on adaptive interventions to the **adaptation realization (AR)** module which maps the abstractly formulated educational recommendations onto more concrete game recommendations. Albert et al. [2]

- 80Days (http://www.eightydays.eu/): The project 80Days is built upon the results of its predecessor ELEKTRA which has made significant contributions to advancing the state-of-the-art of competitive DEGs in terms of educational game design, integration of pedagogical models and taxonomies, and personalization by adaptive technology. Pedagogically, 80Days is grounded in the framework of self-regulated personalized learning (SRPL) [5] which propagates the importance of self-regulation through mindful and meaningful choice and exploration, reflection, and self-personalization in the learning process. It contributes towards the creation and sustainability of intrinsic motivation, a key factor of effective game-based learning.Law and Rustkickmeier [55]

- The ELEKTRA project (http://www.elektra-project.org/) Albert et al. [2]

- With the **shared view** that immersive DEG can make learning engaging, inspiring and presumably effective, enthusiasms and efforts over game-based learning have soared in the recent years. The major strengths of DEGs include a high level of intrinsic motivation to play and to proceed in the game; clear goals and rules; a meaningful yet rich appealinglearning context; an engaging Storyline with random elements of surprise; immediate feedback; a high level of interactivity, challenge and competition. These characteristics are compatible with Merrill's [11] odel for successful learning.
**Nevertheless**, DEGs have some drawbacks such as difficulties in providing an appropriate balance between gaming and learning activities or between challenge and ability, in aligning the game with national curricula, and in affording extensive costs of developing high quality games. Besides, the lack of sound instructional models is seen a common weakness of most educational games, leading to separation of learning from gaming.
**Nevertheless**, commercial computer games are tremendously successful and game industry constantly increases sales to sev-

eral billions of euros. A large number of youngsters spend many hours a week playing games. These observations corroborate the presumption that utilizing gaming activities for educational purposes and exploiting the educational potential of computer games is a highly promising approach to facilitate learning by making it enjoyable and pleasant. Law and Rust-kickmeier [55]

- Individualization of learning experiences and adaptation to personal aims, needs, abilities and prerequisites entail in-depth understandings of individual learners and their behaviour with a DEG. It is critical not to destroy the immersion and gaming experience by intervening knowledge assessments, which are commonly used in traditional approaches to adaptive educational technologies that focus on knowledge and learning progress. In contrast, DEGs take into account a broader scope of issues such as individual preferences (e.g. visual styles or gaming genre). Prevailing cognitive models for adaptive educational technologies (which are primarily competence-based) should thus be merged with theories of achievement motivation and models of interactive and adaptive storytelling to establish a comprehensive theoretical framework for combining learning and gaming.

- Techniques of adaptation and individualization are essentially
  - **adaptive presentation**,
  - **adaptive navigation support** and
  - **adaptive problem solving**. Law and Rust-kickmeier [55]

- Micro-adaptive interventions are non-invasive (meaning that an overall narrative is not compromised) and affect the presentation of a specific LO or LeS. Law and Rust-kickmeier [55]

- Due to the nature of immersive DEGs the adaptation within such games needs to be **continuous** and **less periodic**; it needs to occur **more frequently than on a task by task level**. This issue can be resolved by integrating micro-adaptivity into the environment, where adaptation occurs within the various learning situations as opposed to around them. Micro-adaptivity creates challenges of its own due to the nature of the experience of game play, and the impact that game world changes can have on a player's experience. Law and Rust-kickmeier [55]

- The concept of micro-adaptivity, an approach that enables an educational game to intelligently monitor and interpret the learner's behaviour in the game's virtual world in a non-invasive manner. Kickmeier-rust [50]

- The basic idea of the micro-adaptivity concept is to provide the learner with appropriate educational support without corrupting immersion and the flow of the gaming experience. Both the assessment procedures and the system's responses to the learner must be educationally meaningful and suitable. As a consequence, the game can be equipped with a set of potential interventions (including feedback) that can be triggered in an intelligent manner. The conditions under which a certain adaptive intervention is given are developed on the basis of pedagogical and didactic rules while considering a strong integration in the game play context. The nature of this set of rules may vary depending on the nature of the learning situation, ranging from if–then clauses to alterations of interventions (e.g. varying the intensity of an intervention). Different learning situations and different types of rules may also induce differ- ent types of interventions.
  In the context of DEGs, we can classify the following intervention types:
  - **Competence activation interventions** may be applied if a learner becomes stuck in some area of the problem space and some competencies are not used even though the system assumes that the learner possesses them.
  - **Competence acquisition interventions** may be applied in situations when the system concludes that the learner lacks certain competencies.
  - **Motivational interventions** may be applied, for example, if the learner unexpectedly fails to act for a certain long period of time.
  - **Feedback** may be utilized to provide the learner with information about the learning progress or the game.
  - **Assessment clarification interventions** may be applied, for example, in the form of a query, if the learner's actions provide contradicting support for the assumption of a certain competence state. Kickmeier-rust [50]
- Pedagogical interventions and feedback guide the learning process and inform the learner about the learning progress and the possible deviations from a planned learning path. These interventions also aim to provide the learner with appropriate information and direct the learner's view on important information. In the context of computer-based learning, studies have demonstrated the effectiveness of interventions and feedback (cf. Azevedo & Bernard 1995). Moreno (2004), for example, used feedback to decrease the cognitive load of learners in discovery-based learning environments. In a recent study, Tan et al. (2006) showed

that interventions could improve immediate goal achievement. Kickmeier-rust [50]

- **Updating competence state probabilities**
  As a next step, we must transfer our assumptions of available and lacking competencies to entire competence states and the likelihood of those states. To identify a learner's current competence state at the beginning of a gaming/learning episode, we assume an initial probability distribution over the competence structure. There exist an arbitrary number of possibilities for this initial probability distribution. Kickmeier-rust [50]

### 3.1.7 *Vygotsky´s Theory (Gameplay, Scaffolding & ZPD)*

*Zusammenfassen, Teil in Chapter 2 (SG Background), Teil in Chapter 3 (Adaptivity)*

- "To apply educational models of the Zone of Proximal Development (ZPD) and approaches to assisted performance to existing knowledge about surgical skill acquisition, and thus provide a framework for potentially enhancing surgical education.
  The stages of the ZPD appear to be consistent with existing psychomotor theories of skill acquisition. The fourth phase of the ZPD, where the performance of skills can deteriorate, and which can occur for a number of reasons such as ill health or absence from routine duties, is not well described in the surgical literature. Approaches to assisted performance may provide additional insights into the complexities of how to teach surgical skills, how to approach the process of hand-over of responsibility from teacher to learner, and the importance of enhancing teaching skills of surgical educators.
  The ZPD and approaches to assisted performance would appear to provide a valuable framework for planning and assessing surgical education programs.
  Assisted performance and the four stages of the zone of proximal development
  Assisted performance defines what a learner can do with help, with the support of the environment, of others and of the self. The contrast between assisted performance and unassisted performance identified the fundamental nexus of development and learning that Vygotsky describes as the zone of proximal development (ZPD) (Vygotsky, 1978). For any domain of skill, a ZPD can be created. In the ZPD, assistance is provided by the teacher, the expert, and the more capable peer. Distinguishing the proximal zone from the developmental level by contrasting assisted versus unassisted performance is of major importance in understanding approaches to education. It is in the proximal zone that teaching may be defined in terms of learner development. Vygotsky (see Vygotsky, 1956) stated that teaching is good only when it "awakens and rouses to life those functions which

are in a stage of maturing, which lie in the zone of proximal development". Consequently, teaching consists in assisting performance through the ZPD. Teaching can be said to occur when assistance is offered at points in the ZPD at which performance requires assistance." C and Stacey [11]

*Understanding Scaffolding and the ZPD in Educational Research - Verenikina [95]*

Research with Scaffolding & ZPD Concepts.

- Socio-Cultural Notion von ZPD
  Zone of proximal development – **learning potential**
  The notion of proximal development is probably one of the most famous concepts introduced by Vygotsky. Initially he used the concept as an index for intellectual potential, arguing that developmental potential cannot be assessed in absolute terms (for example, by a score on an intelligence test): it should be based on what a child can learn under optimal conditions (for example, with assistance). The zone developed into an educational concept. The zone is a shared activity, collaboratively produced in the interaction between a child and more knowledgeable others – imitation, though not to be confused with copying.

- GamePlay
  Vygotsky was a strong advocate of play as a means of promoting **spontaneous learning**.

*INSTRUCTIONAL GUIDANCE*

- "Vygotsky espoused the notion of the Zone of Proximal Development defined as the difference between a child's actual and potential levels of development (known, for example, by what a child can do alone and with the assistance of an expert other). According to Vygotsky (1978), play creates a broad zone of proximal development, both in cognitive and socio-emotional development. In make-believe play children perform above their own cognitive abilities logical thinking, memory and attention. Their ability for deliberate behaviour and self-regulation in make-believe play is also beyond their everyday norm." [96]

- "Congruent with the perspective of Vygotsky's (1978) ZPD model, it can be argued that an instructor or facilitator, rather than a peer expert or a **technological learning tool**, has acted as the major scaffold for game-playing participants' cognitive development. This observation confirms the claim by Garris et al. (2002) that the value of the instructor in scaffolding learners is a critical (and somewhat overlooked) component in the use of educational games, as are learner support strategies such as online help, cues/prompts, informative feedback, and other activities." [48]

- **ZPD** is the difference between what a learner can do without help and what he or she can do with help:

ZPD is the distance between the actual developmental level as determined by independent problem solving and the level of potential development as determined through problem solving under adult guidance, or in collaboration with more capable peers. [REF Vygotski]

- On the Concept of Cultural Age in Vygotskys Cultural Historical Theory For the purpose of a clearer understanding of the concept ZPD Kazuo [47]

## 3.2 EMBEDDED ASSESSMENT IN GAMES

*What is embedded assessment? First a definition with references*

- Assessment is not new. Tests are used successfully as gauges and gateways. Assessment is also done successfully in the classroom, both formally and informally – teachers give quizzes and tests, they look into faces to gauge students' grasp of the material, and they ask questions. Underwood et al. [93]
- Need of Assessment in Serious Games.
- The educational research community has been experimenting with educational games with a focus on **pedagogy** and **curriculum**, but little effort has been made to assess what students are actually learning in these environments.
  Designing **embedded assessments into games** is one of the critical gateways to creating learning tools that are maximally engaging for the learner, using **sound pedagogical methodology** as a foundation.
  It´s a technology that facilitates near real-time data collection through embedded assessments, visual data mining, inference mechanisms, and dynamic individualization.
  It describes a methodology for creating valid embedded assessments and identify types of data that can be collected from gaming environments along with approaches for analysis, all toward the goal of individualized adaptation. Underwood et al. [93]

*Damir gefällt das: My Contribution*

- People are starting to realize that **integrating assessments into gaming environments** is important and that external multiple-choice questionnaires and human-scored artifacts are not sufficient to assess all that people may be learning in games, since the latter are time-wise inefficient and do not scale easily.
  Both the Digital Media Collaboratory (out of U. Texas at Austin) and the EduMetrics Institute advocate addressing assessment issues as **an initial part of serious game design** (Chen & Michael, 2005).

Shute et al. (2008) propose an approach for embedding assessments into games to reveal what is being learned during the gaming experience, where the focus is not only on content knowledge, but also **higher-order skills**.
Some efforts are interested in analyzing a sequence of events to infer abilities such as problem-solving strategies (e.g., Ketelhut, Nelson, & Schifter, 2009).
Others have begun to implement embedded assessments to **gauge content learning** (e.g., Zapata-Rivera, 2009).
PIXELearning creates sophisticated training simulations that give the player **appropriate feedback** so that players come to understand the connection between their in-game actions and the outcomes (Chen & Michael, 2005). Underwood et al. [93]

- Shute et al. (2008) and Shute (2009) propose an approach for embedding assessments into immersive games to reveal what is being learned during the gaming experience, with a focus on creative problem-solving and systems thinking.
  Their work takes a similar approach to that described here, such as using evidence-based assessment design and doing statistical analysis to make inferences based on in-game behavior. However, our paths quickly diverge. In this theoretical work (i.e., not yet tested), they overlay an **expert student model** onto the learner's experience to determine whether the student is doing the right thing at the right time (in the same way intelligent tutoring systems work), and to identify misconceptions. Where they need to build a new expert student model for each problem to be solved, we are interested **in building a general model of mathematical knowledge** that will **apply to many learning situations**. And where they focus on correcting misconceptions, we are interested in determining **learners' prior knowledge** so the system can help them build on that directly. Our efforts, described in later sections, build on these experiences by providing data analysis and mining capabilities, and adapting the environment to the individual learner. Underwood et al. [93]

### 3.2.1 *Why to assess?*

- To make the most of digital learning environments, it would be useful to know what happened inside the game, when it happened, and how it happened. How can the learner get more of what they want or need, and less of what they do not want or need? To answer these questions, more **in-game data** needs to be collected, and sense made of the data. No small task, to be sure. Underwood et al. [93]

- Why assess learning in games?
  While they are recommended as engaging approaches to learning, the adoption in education of games is slow because they have **not yet been shown to be effective** (Federation of American Scientists, 2006). One way to show they are effective is to measure the learning that occurs during play. Underwood et al. [93] **How to measure the learning that occurs during play?**

- What do We want to do With the Assessment Results?
  How can assessments be designed to be embedded into games?
  Valid assessments are best designed in a **top-down fashion**, starting with the **goal of determining what to do with the assessment results**.
  Experience shows that if this part is postponed until the end, the data to be collected would likely need to be redefined.
  It is simply more efficient to define these, for example, as **pedagogical goals** such as feedback or adaptivity, at the beginning, and to **periodically** revisit them to ensure that they use the types of data that can be collected. It is an iterative process.
  Much can be learned from modern entertainment games which already employ embedded assessment and feedback, though for different purposes. Their goal is to foster **high engagement** in order to **extend the playability** of a title.
  For example, **tutorials** in games slowly introduce features of the game to the player and offer players a chance to practice certain skills. They provide **scores as feedback** to demonstrate what is important in game play, **leaderboards** for comparison to other players, and **persistent scoring** that allows players to monitor their own progress.
  These features essentially show the results of embedded assessments and if used appropriately in educational games may, in fact, greatly **enhance levels of learner interest and engagement**. Underwood et al. [93]

- Custom **feedback** during games includes suggestions to help learners overcome weaknesses – perhaps via non-player agents, providing practice areas, or other creative mechanisms.

- Similarly, **persistent scores** can be displayed as an ongoing gauge of **proficiency of necessary knowledge**, **process**, and **skills**.
  Underwood (2008) presents some guidelines for pedagogically sound feedback.

- For learners who have a competitive streak, it is helpful to provide **leaderboards** or other mechanisms after the game by which learners can compare their performance with others.

- High scores may not be the only thing learners are interested in, however, so it would also be useful to provide such measures as accuracy (e.g., knowledge displayed), efficiency (e.g., in procedures), behavior, creativity, or other things different types of learners will value.
- **Four general pedagogical goals** that can leverage the results of embedded assessments in games were identified:
  - Provide custom feedback to learners during games
  - Provide custom feedback to learners after games
  - Tailor which games learners experience
  - Tailor how learners experience the games Underwood et al. [93]

### 3.2.2 *What to assess?*

- Educational games, in order to be effective, should be designed as constructivist learning environments where learners can explore, experiment, and actively solve problems (Reeves & Okey, 1996), with a focus on both learning outcomes and the actions and processes taken toward a goal.
  In this section we will briefly list the types of things that can be measured, then describe an approach to designing assessments and embedded assessments, including challenges, technology to address these challenges, measurement, and validation. Underwood et al. [93]
  - **Process** is a bit more difficult to measure. **Process** includes such things as procedures for solving algebraic equations, and paths taken toward a goal like racing a car or constructing a bridge. It is generally easy to discern whether the outcome is correct, but it is more difficult to discern whether the path taken was good or efficient, as in the case where mistakes in process lead to correct answers.
  - In an interactive environment, **fine-grained actions** can be collected to help measure **process**.
  - Even more challenging to measure are **skills**, including **higher-order skills** (e.g., critical think- ing, decision-making, and problem solving) and **behavioral skills** (e.g., teamwork, leadership). It is possible to directly test these skills in an interactive environment, **but a theory of how these skills can be demonstrated is still needed**.
- For games that lend themselves to learning by drill, content knowledge (e.g., math facts, following foreign language directions, other right/ wrong content) can be readily assessed. It is

simple only in the sense that assessment design research has a long history of focusing on content knowledge. There are still design principles to follow to make sure the assessments are valid, that is, that they measure what is intended to assess. Underwood et al. [93]

- Embedded assessment is the process of measuring **knowledge** and **ability** as part of a **learning activity** rather than after the fact, when it is only an approximation of learner behavior. Student actions can be evaluated within context while carrying out tasks, or otherwise interacting in a gaming environment. Underwood et al. [93]

- These actions can be collected, viewed, and analyzed either immediately or after the session. Sometimes "so-called" embedded assessments are implemented as pop-up quizzes that a learner cannot bypass. We identify these as possible "tools" an educator might employ within a larger framework, but not as fully-realized embedded assessment.
  Embedded assessments in games can focus on such things as content **knowledge**, **process** and **procedures**, and higher-order 21st century **skills** such as collaboration and strategic thinking (hereafter called **"higher-order skills"**; Partnership for 21st Century Skills, 2009). Underwood et al. [93]

### 3.2.3 *How to assess?*

- Considering the importance of not destroying immersion with the game, the assessment of the learning progress and psychopedagogical feedback must occur in a **non-invasive way**. This requires an **intelligent system** that is capable of assessing individual competencies and learning progress by **observing** and **interpreting** behaviour **within** the game. [51]

- Adaptivity concept should be aware of both **engagement** & **learning**. the challenge must be neither too easy (promoting boredom) nor too hard (discouraging continued play).

- The focus is on how to use game environments to provide one-on-one learning experiences. Even with advances in technology, there is still a need for humans to oversee individual learning, but digital games can scale, in ways that teachers cannot, to provide one-on-one experiences in the form of adaptive environments from the results of embedded assessments. This approach is key to fostering both **engagement** and **learning**, where the challenge must be neither too easy (promoting boredom) nor too hard (discouraging continued play). Underwood et al. [93]

- Embedded assessments provide **information** to **adapt** the game environment to the learner. Games can be, by their very nature and design, outstanding forums for providing formative feedback to the learner in a familiar, low-stress fashion, creating a space in which learners are free to explore and are receptive to help. Underwood et al. [93]

- **How** can games, including puzzles, simulations, and other immersive spaces, be designed so that embedded assessments are seamless, expedient, and valid?
  **How** can **higher-order skills** be assessed?
  **How** can we assess process and infer behavior, in addition to content knowledge?
  **How** can embedded assessments be technologically implemented?
  **How** can games be adapted to individuals?

- Assessment Design Methodology is an approach to assessment design based on **Educational Testing Service's evidence-centered design** (ECD; Mislevy, Steinberg, & Almond, 2003), which provides an **evidentiary argument for a test's validity**.
  The questions to be answered, in order, are:

  - What do we want to do with the results of the assessment?
  - What claims do we want to make about the learner based on performance in the game?
  - What observations of learner actions would provide evidence for the claims? Underwood et al. [93]

- Here we will describe some mechanisms to measure content knowledge, process, and higher-order skills.

  - Content knowledge is relatively easy to measure. Right/wrong knowledge can be compared to correct answers, just as in traditional tests; a structured formula can give weight to more complex problems than easier ones;
  - Content analysis can determine that certain incorrect solutions showed misconceptions or careless errors, and weight them accordingly. Psychometricians often use item response theory to guide them.
  - As complicated as these measurement approaches can get, procedures and skills are even more complex to assess, and other approaches to measurement are required. Underwood et al. [93]

- Through Evidence-centered design, a game can be designed to facilitate the measurement of process for students taking known paths. However, in a well-designed game, learners will often come up with approaches not considered by the game designer.

Visual IQ and Inflection can help address the "long tail" of learners, that is, the many individual learners who do not fall into statistical groups. Research in this area is just beginning. We are investigating ways to find patterns in solution paths as they are related to correct answers. Underwood et al. [93]

- Even more challenging than assessing process are higher-order skills (e.g., leadership, strategic thinking). A mix of structured and probabilistic formulae may be the best way to measure these skills. Traditional strategic thinking tests are usually in the form of **asking what someone would do in a particular situation**, sometimes called "**situational judgment**" tests. These tests can provide some information about how people think they would act in a situation, but they can have validity issues (Ployhart & Ehrhart, 2003). Underwood et al. [93]

- Game environments create the opportunity to more directly test these skills. Similar to measuring process, measurement can begin by defining rules for known or expected skill performance as defined in the assessment design, supplemented by visual data mining tools and more open-ended pattern matchers. Underwood et al. [93]

### 3.2.4 *What Claims can be made Based on Performance in the Game?*

- Claims are statements we want to be able to make about a learner's knowledge, higher-order skills, or other attributes based on their performances.

- Claims are a way to communicate performance and may be very general (e.g., learner can read at the second grade level) or specific (e.g., learner can decode initial consonants).

- It is useful to break that down into strengths and weaknesses of sub-skills, including communication, delegation, and team-support, in order to reach measurable actions.

- We are both identifying the skills we want to measure and how we want to report them. Underwood et al. [93]

### 3.2.5 *What Observations of Learner Actions Would Provide Evidence for the Claims?*

- We identify **ideal types of evidence** that would support the claims about learner performance. This is immediately balanced with the practicality of measuring the skill in the game, since there are certainly limitations.
  Embedded into the game, the tests can appear as problems to

be solved in order to do such things as build or to break a code. Also **needed to be identified is what behavior shows lack of the skill**.
Given a list of **observables** and **actions** that will provide **data** to support **claims**, we must also determine what is feasible to measure in a gaming environment and how to design it unobtrusively using normal game mechanics. Underwood et al. [93]

### 3.2.6 *Research on Embedded Assessment in Educational Games*

- Video games that were designed for entertainment actually provide an environment in which players learn (Gee, 2003).
  **Tutorials** are meant to help players learn to play the game, typically presenting a few new things at a time so the player can steadily assimilate them.
  Games also assess the player through **levels**, **scoring**, and other **in-game resources**.
  The challenge of educational games is to use these methods to help people learn **particular content**, **skills**, and **behaviors**. Many existing efforts use mechanisms such as engaging online quizzes to assess student learning, and that is the extent of the "game."
  Following is a review of the research and development efforts of both the educational research and serious games communities, who are both addressing very similar issues related to the use of games and of embedded assessment for learning. Underwood et al. [93]

- Research efforts are beginning to address how to embed assessment validly, reliably, and without interrupting play.
  These efforts mostly use games as assessment, and may provide engaging virtual environments and feedback (e.g., Clarke & Dukas, 2009; Hickey, 2009).
  These are generally performance assessments that take place after a course of learning, another example of external measures.
  These efforts, while successfully assessing such competencies as critical thinking, are addressing many good questions for how to have engaging assessments in a virtual environment, **but they are not addressing the issue of how to integrate them into a learning environment**. Underwood et al. [93]

- **Validation**
  The main challenges in designing valid assessments are captured by the following questions:

  – How do you know that students are learning what you claim they are learning?

  - How do you know you are measuring what you claim you are measuring? Underwood et al. [93]

- Assessment validity is an area that psychometricians have been researching for many years. The main purpose of assessment design is to have a test (e.g., college entrance exam) where student performance can **confidently predict** the relationship to a goal (e.g., performance in the first year of college).
  For these high-stakes cases, **specialized statistical analysis** is used and **continual adjustments** of the test are made so that it maintains correlation with the ultimate goal.
  Traditional psychometricians seem to over-emphasize measurement of decontextualized knowledge, assuming that the less contextualized their instruments, the more likely they are to assess generalizable knowledge and skills. Proponents of authentic assessment seek to estimate learning within specific contexts that approximate the ill-defined, uncontrollable aspects of the real world, a world in which the generalizability of standardized tests may have little relevance (Reeves & Okey, 1996).
  The goals of educational games are not as critical (i.e., high-stakes) as college entrance exams, but the assessments still need to be valid.
  Evidence-centered design helps ensure validity, leaving an evidentiary trail of design decisions. However, it is advisable to do explicit validation.
  **For such things as math knowledge, running pilot studies with pre- and post-tests can provide the proper level of validation**.
  **For higher-order skills, expert opinions on what constitutes the skill will be sufficient**; these experts can also assign ratings to game behavior. Learners and teammates can rate themselves and each other, and these can be compared to computed scores. Learners can also be asked to react to the scores. As always, the game and embedded assessments should continually be adjusted to get closer to more accurate measurement. Underwood et al. [93]

### 3.2.7 *The challenges, Limitations and Disadvantages of Embedded Assessment*

1. **Challeges:**

   - We just described an evidence-centered methodology for assessment design.
     The points that focus on creating an immersive and motivating educational game are:

     - Adapt the environment to the learner.

  - Assess process and skills, not just content. Some of these are easy to measure, and some more challenging.
  - Embed assessments into the fabric of the game. If a test feels like school, learner motivation can be lessened. Embedding assessments can help overcome this problem. Underwood et al. [93]
- These are the challenges in creating embedded assessments in gaming environments. They include how to:
  - Maintain flow while collecting in-game data. This is both a design and technological problem. The game has to be designed so that actions in the game provide evidence of a pedagogical goal to be measured. The technology has to collect data in real-time so as not to interrupt gameplay.
  - Analyze the data. Appropriate techniques should be used to measure performance, including **statistical**, **visual**, and other **data mining approaches**. Analyzing content knowledge is relatively simple; process and skills are harder. Underwood et al. [93]

2. **Limitations:**

- Digital gaming environments make it easy to collect every keystroke learners make, but it is not clear that this is the best approach.
  Focusing on the **wrong data** could lead to **incorrect conclusions about learner needs**.
  Theories for what to look at still need to be developed. Underwood et al. [93]
- Another limitation is that, even with all the data now available, we still do not know other things about the learners, like their feelings at the moment, what else is on their screen, as well as other distractions and pressures, opening up other areas of research. Underwood et al. [93]
- How reliable are embedded assessments? This is an open question, since each is as reliable as the assessment itself. Underwood et al. [93]
- There have been many studies on **how effective educational games are for learning**, but they generally lump all games into one category (e.g., Virvou, Katsio- nis, & Manos, 2005).
  Furthermore, they do **not evaluate the learning that occurs during gameplay**.
  **In a game that is well-designed, immersive, and engaging, embedded assessments should be valid and reliable and**

**not distract the learner, but these are also areas ripe for research.** Underwood et al. [93]

3. **Disadvantages:**
   - If embedded assessments are not truly part of the game environment, they start to seem like "school" by popping out or otherwise taking away the learners' focus from the game, where learners will lose the flow of the game, and therefore, the flow of learning. Underwood et al. [93]
   - Another disadvantage is that it could be costly to design embedded assessments properly since they need to be designed by people knowledgeable about game design in order to be done well. Underwood et al. [93]

## 3.3 ASSESSMENT FOR GAMES

- As with any other educational intervention, games for learning will need to be evaluated to ensure that they help learning and that they do so in a way that as far as possible does not favour one group over another.
- In serious games, assessments are used to measure whether or not the product does the job, teaches the required skills, and changes behavior. More important than that, assessment testing should measure whether or not those skills and behaviors transfer to the real world and solve the performance problems that were identified in the analysis phase of the project. Assessment testing will tell the client whether or not the serious game was worth the cost of investment. We should note that instructional designers typically use the term summative evaluations for these assessments. But we'll be going beyond the strict definition of a research-oriented summative evaluation, so as to incorporate the kinds of assessments we'd like to do for promotional and persuasive games as well.

### 3.3.1 *Why Engage in Assessment?*

- The primary reasons for assessment identified in the literature are to screen for disabilities; to assess kindergarten readiness; to assist in developing curriculum and daily activities; to evaluate the effectiveness of a project or a program; and to provide feedback to parents. Bowers [9]
- Review of research indicates assessment techniques, both standardized and informal have been, and will remain, an important tool for early childhood professionals.

Assessment methods can be used to screen for disabilities, to assess kindergarten readiness, to assist in developing curriculum and daily activities, to evaluate the effectiveness of a project or a program, and/or to provide feedback to parents. Traditional questions surrounding assessment centered on three key questions. In the future, these questions will continue to be asked, along with many new questions, which accompany our increasingly complex knowledge of child development. Bowers [9]

### 3.3.2 *Which Measures Should I Select?*

- Perhaps one of the most confusing aspects of assessment is choosing which method(s) to use.
  Assessment measures can be grouped into two categories.
  The first category includes **standardized tests** and **inventories** that are available for purchase and used primarily to compare children against developmental norms or to other children.
  The second category consists of **informal methods**.
  **Observations** may be obtained from informal methods in a methodical way, but the results are not usually compared to standardized norms or averages.
  Today, hundreds of standardized tests exist, with the majority being designed for use by persons specifically trained in their use and interpretation.
  Studies suggest early childhood professionals tend to choose standardized tests that are already being used by others (Johnson & Beauchamp, 1987). Many individuals in early childhood rely on a combination of assessment methods, including some of the common informal methods, as well, as a source of collecting data about young children. Bowers [9]

- **Standardized Tests and Inventories**
  One of the major advantages of standardized tests is that the results can be used to compare a child to developmental norms or to children in similar circumstances.
  A norm is an average or series of averages obtained on the sample of children used in developing the test.
  A second advantage sometimes cited is the predictive validity of such tests.
  That is, children who perform well on standardized tests in the preschool years tend to also perform well on tests in kindergarten and in the early elementary years (Vacc, Vacc, & Fogleman, 1987). Bowers [9]

  – Common tests for preschoolers include the Battelle (Newborg, Stock, & Wnek, 1984), the Child Behavior Checklist (Achenbach, 1992), the DIAL-R (Mardell-Czudnowski &

Goldenberg, 1990), the Miller Assessment for Preschoolers (Miller, 1982), the Peabody Developmental Motor Scales (Folio & Fewell, 1984) and the Peabody Picture Vocabulary Test (Dunn & Dunn, 1981). Tests often used with infants include the Bayley Scales of Infant Development (Bayley, 1993), and the Denver Developmental Screening Test (Frankenburg et al., 1975). Bowers [9]

- One of the major disadvantages of standardized tests is how to interpret the data obtained. Results from administration of such tests must be considered in comparison to similar children in similar circumstances. Often, this comparison is relatively difficult to achieve. Bowers [9]
- The **predictive validity of standardized tests**, that is their ability to forecast achievement in kindergarten, may not be the most important variable we want to know (Pilkington, 1988; Quay & Steele, 1998; Rudner, 1996). More and more, **alternatives to formal tests are being explored**, including measures which are thought to be more holistic and developmentally appropriate in their focus and approach. Bowers [9]
- Standardized tests are accompanied by information regarding their **validity** and **reliability**.
  **Validity information**, indicating whether the test really "measures what it is supposed to measure," (Witt, Elliott, Kramer & Gresham, 1994, p. 103) offers teachers the opportunity to evaluate whether the test is appropriate for its intended use.
  **Reliability information**, indicating the test's ability to produce "the same result when repeatedly measuring the same thing," (Witt et al., 1994, p. 95) helps determine a teacher's degree of confidence in the information that will be obtained. Information on both validity and reliability can be found in the manual that accompanies the test when purchased. Bowers [9]

- **Selecting Informal Methods**
  The criteria used for selecting any method should be based on the purpose of the assessment itself.
  Questions asked prior to engaging in assessment should include:
  - Why do I need this information? What is the purpose of my efforts?
  - Based on what I need to know, what kind of information will be helpful? Test scores? Written records? Works found in a portfolio?

  - How often and when do I need to collect such information, and how can I best assure the information is accurate and valid? In addition, any method used should be selected for its appropriateness for the children on whom it will be used.
    Two important criteria are developmental appropriateness, e.g., "Is it designed for the age of child I'm testing?" and cultural appropriateness, e.g., "Is it relevant to the background and daily circumstances of the child?" Bowers [9]

- **Using Informal Methods**
  To some extent, any method is only as good as the person performing it. Thus, assessments are most appropriate when every effort has been made to insure **objectivity** (Salvia & Ysseldyke, 1995).
  Objectivity refers to the process of examining a child or event without pre-formed ideas about outcome. When we are objective we report only the facts, and, in turn, interpret those facts based only on what was observed, rather than feelings or attitudes about the child. Bowers [9]

- One example of an **innovative method** is play-based assessment, which evaluates children's developmental skills and their social interactions, learning styles and behaviors through play (Lowenthal, 1997, p. 1). Bowers [9]

### 3.3.3 *External Assessment of What is Learned in Games*

- There has been a fair amount of research on external assessment of learning in games.
  - Some efforts require students to submit solutions, responses, or summaries of game interactions to an instructor or other real person, who then grades them (e.g., Barab, Arici, & Jackson, 2005; Chen & Michael, 2005).
  - Other efforts do conversation analysis of discussions that take place in-game (e.g., Shaffer, 2007) and out-of-game (e.g., Steinkuehler, 2008), but neither looks at game activity directly.
  - Another approach is to carry out a pre- and posttest experimental design to determine what the student has learned (e.g., Zapata-Rivera, VanWinkle, Shute, Underwood, & Bauer, 2007). Underwood et al. [93]

- By administering external tests, teachers may lose much of the value of the games since learners are already inherently demonstrating their knowledge and skills by interactions during gameplay.

Assessments embedded into games will, in addition, reduce the often artificial separation between performance and assessment in school (Reeves & Okey, 1996). Each of these approaches has shown that students are learning in the gaming environment. Underwood et al. [93]

- As **Jim Gee** indicated at the American Educational Research Conference (Gee, 2009), **"If you are testing outside the game, you had better have good reason for doing it. The very act of completing a game should serve as an assessment of whatever the intervention was designed to teach or measure."** Underwood et al. [93]

Part III

# FOUNDATION: SCIENTIFIC RESEARCH IN SERIOUS GAMES

The Scientific Research world contains a rich tapestry of variation that almost makes you want to pull your hair out! "Scientific research is defined by the strategy of understanding how everything works through logical thinking." "Research into questions posed by scientific theories and hypotheses." Beginning with introspection, finishing with statistics. Explanatory (WHAT, HOW & WHY) & Comparative Research.

# 4

# SCIENTIFIC RESEARCH

## 4.1 ABOUT CASE STUDIES

- "Case studies of professional software development practices describe how real (or realistic) projects are planned and executed. Cases provide engaging models of the activities and materials of software development to students and other novice practitioners. They vividly remind learners of the possibilities for meaningfully applying knowledge and skills in the world beyond the classroom. During the past six years, we have developed and used a collection of usability engineering case studies for teaching human-computer interaction, primarily to upper-level undergraduates in computer science and in information sciences and technology.
  Case studies are descriptions of a specific activity, event, or problem, drawn from the real world of professional practice.
  Cases incorporate vivid background information and personal perspectives to elicit empathy and commitment, and present contingencies, complexities, and often dilemmas intended to evoke integrative analysis and critical thinking." [13]

- "Case studies are widely used in professional education — in business, medicine, law, and engineering [Williams 1992]." [13]

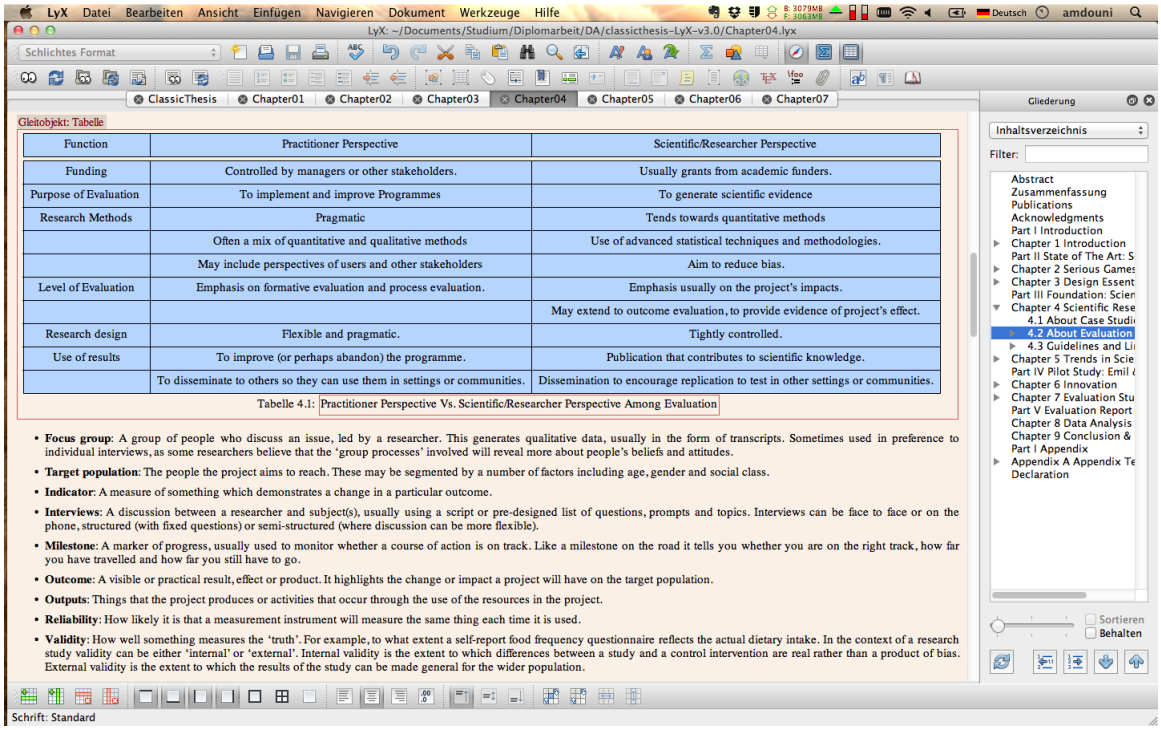
- Prior investigations of case-based learning in the computer and information science and engineering (CISE) disciplines have been quite encouraging, primarily in the arena of professionalism and computer ethics.
  Case-based approaches have also been developed and explored in more technical CISE topic areas such as computer graphics [Shabo et al. 1996]; design [Guzdial et al. 1996]; **ubiquitous computing** [McCrickard and Chewar 2004]; and usability engineering [Rosson et al. 2004a; 2004b].
  This research has also provided evidence that case-based learning promotes key **meta-cognitive skills**, including **cognitive elaboration**, **error management**, **reflection**, **self-regulation**, and **transfer of knowledge** [Carroll and Rosson 2005; Kolodner et al. 2004]. [13]

- Some of the core concepts in our usability-engineering course are scenarios, design trade-offs (couched as claims analysis), user interface metaphors, mental models, collaborative coordination and awareness, ubiquitous computing, graphical design, direct

manipulation, Fitts' law, information architectures, and information visualization. Students learn techniques associated with the various phases in the usability-engineering process, for example, requirements-interviewing and ethnographic observation, task analysis and user-modeling, paper-based and scenario-machine prototyping, evolutionary development, and usability evaluation methods like experimental analysis of user performance, collection and interpretation of thinking-aloud protocols and field study data, and survey construction and analysis. [13]

## 4.2 ABOUT EVALUATION

- **Evaluation**: Establishing whether a project has achieved what it set out to do.
- The topic of evaluation can sometimes seem quite daunting, full of jargon and with a baffling variety of different methodologies. But evaluation is actually quite simple: it is **judging the value of something and looking at the extent to which a programme or project has achieved its objectives**.
- **Monitoring** checks the extent to which a project is proceeding according to plan.
  A collection of routine data that helps you assess whether projects are proceeding to plan. A sub-set of evaluation.
- DO
  - Set a budget for evaluation.
  - Build evaluation into the start of a project.
  - Bring all stakeholders together and agree aims and objectives.
  - Set out how the project will achieve its objectives.
  - Find out what stakeholders think a successful project will look like.
  - Agree what will be measured and how it will be measured.
  - Use data collection methods which are appropriate to the available resources and will help answer the key evaluation questions.
  - Consider the value of both quantitative and qualitative information.
  - Scope out timing and logistical issues, and consider the impact they will have on the evaluation.
  - Think creatively: there is no single way to evaluate a project.
  - Keep evaluations simple and useful.



| Function | Practitioner Perspective | Scientific/Researcher Perspective |
|---|---|---|
| Funding | Controlled by managers or other stakeholders. | Usually grants from academic funders. |
| Purpose of Evaluation | To implement and improve Programmes | To generate scientific evidence |
| Research Methods | Pragmatic | Tends towards quantitative methods |
| | Often a mix of quantitative and qualitative methods | Use of advanced statistical techniques and methodologies. |
| | May include perspectives of users and other stakeholders | Aim to reduce bias. |
| Level of Evaluation | Emphasis on formative evaluation and process evaluation. | Emphasis usually on the project's impacts. |
| | | May extend to outcome evaluation, to provide evidence of project's effect. |
| Research design | Flexible and pragmatic. | Tightly controlled. |
| Use of results | To improve (or perhaps abandon) the programme. | Publication that contributes to scientific knowledge. |
| | To disseminate to others so they can use them in settings or communities. | Dissemination to encourage replication to test in other settings or communities. |

Tabelle 4.1: Practitioner Perspective Vs. Scientific/Researcher Perspective Among Evaluation

- **Focus group**: A group of people who discuss an issue, led by a researcher. This generates qualitative data, usually in the form of transcripts. Sometimes used in preference to individual interviews, as some researchers believe that the 'group processes' involved will reveal more about people's beliefs and attitudes.
- **Target population**: The people the project aims to reach. These may be segmented by a number of factors including age, gender and social class.
- **Indicator**: A measure of something which demonstrates a change in a particular outcome.
- **Interviews**: A discussion between a researcher and subject(s), usually using a script or pre-designed list of questions, prompts and topics. Interviews can be face to face or on the phone, structured (with fixed questions) or semi-structured (where discussion can be more flexible).
- **Milestone**: A marker of progress, usually used to monitor whether a course of action is on track. Like a milestone on the road it tells you whether you are on the right track, how far you have travelled and how far you still have to go.
- **Outcome**: A visible or practical result, effect or product. It highlights the change or impact a project will have on the target population.
- **Outputs**: Things that the project produces or activities that occur through the use of the resources in the project.
- **Reliability**: How likely it is that a measurement instrument will measure the same thing each time it is used.
- **Validity**: How well something measures the 'truth'. For example, to what extent a self-report food frequency questionnaire reflects the actual dietary intake. In the context of a research study validity can be either 'internal' or 'external'. Internal validity is the extent to which differences between a study and a control intervention are real rather than a product of bias. External validity is the extent to which the results of the study can be made general for the wider population.

Figure 8: Practitioner Perspective Vs. Researcher Perspective Among Evaluation

  - Share your findings as widely as possible.

- DON'T
  - Start the project without collecting baseline data.
  - Try to measure everything.
  - Only have one person responsible for the evaluation.
  - Spend so long designing a questionnaire that you do not have the time to use it.
  - Collect data that will not be used.
  - Construct a comprehensive evaluation plan then forget it as soon as you get the funding.
  - Make claims from the evaluation that cannot be substantiated.

- When measuring the **broader effects**, an evaluation can reveal **unintended consequences**. These are outcomes, both positive and negative, that might arise but were not originally intended.

- **Focus group**: A group of people who discuss an issue, led by a researcher. This generates qualitative data, usually in the form of transcripts. Sometimes used in preference to individual interviews, as some researchers believe that the 'group processes' involved will reveal more about people's beliefs and attitudes.

- **Target population**: The people the project aims to reach. These may be segmented by a number of factors including age, gender and social class.
- **Indicator**: A measure of something which demonstrates a change in a particular outcome.
- **Interviews**: A discussion between a researcher and subject(s), usually using a script or pre-designed list of questions, prompts and topics. Interviews can be face to face or on the phone, structured (with fixed questions) or semi-structured (where discussion can be more flexible).
- **Milestone**: A marker of progress, usually used to monitor whether a course of action is on track. Like a milestone on the road it tells you whether you are on the right track, how far you have travelled and how far you still have to go.
- **Outcome**: A visible or practical result, effect or product. It highlights the change or impact a project will have on the target population.
  Changes in clients or communities associated with program activities and outputs.
- **Outputs**: Things that the project produces or activities that occur through the use of the resources in the project.
  Products and services delivered to a program's clients.
- **Reliability**: How likely it is that a measurement instrument will measure the same thing each time it is used.
- **Validity**: How well something measures the 'truth'. For example, to what extent a self-report food frequency questionnaire reflects the actual dietary intake. In the context of a research study validity can be either 'internal' or 'external'. Internal validity is the extent to which differences between a study and a control intervention are real rather than a product of bias. External validity is the extent to which the results of the study can be made general for the wider population.
- **Case study**. A rich description and analysis of a program in its context, typically using multiple modes of qualitative data collection.

### 4.2.1 *Definition of Evaluation*

- There are a lot of different definitions of evaluation. Here is one of the best, because it touches on the most important aspects of evaluation:

"Evaluation is a collection of methods, skills and sensitivities necessary to determine whether a human service is needed and likely to be used, whether it is conducted as planned, and whether the human service actually does help people" (Posavac and Carey, 1980, p.6).
This definition encompasses the two main types of evaluation: process and summative.

### 4.2.2 *Evaluation: Aim & Objectives*

- Projects need clear objectives that describe what they aim to do and how they will do it. In general, the clearer the objectives, the easier they will be to measure. If it is not clear what the project is trying to achieve, it will not be possible to measure whether or not it has been successful.

*Where is our **Project´s Logic Model**? The term is **usual** when conducting **Evaluation**.*

- This is why evaluation needs to be considered right at the start, and built into a **project's logic model**.
  A simple way to set objectives is to use **SMART objectives**:
  - **S**pecific: objectives should specify what you want to achieve;
  - **M**easureable: measure whether or not objectives are being met;
  - **A**chievable: are the objectives achievable and attainable?
  - **R**ealistic: can the objectives be realistically achieved with the available resources?
  - **T**ime-bound: when should the objectives be met?
- Aim
  A broad statement of intent setting out the purpose of the project.
- A Goal-centered Evaluation is the process of assessing the extent to which the goals and objectives of a program were met. (Tyler) Canada [12]
- User-centered Evaluation
- An Evaluation design should begin with an **evaluability assessment** (also called pre-evaluation assessment), the front-end planning process that helps determine how best to evaluate the program in order to generate an Evaluation Framework.
  A pre-evaluation assessment thus generates a thorough description of the program's structure — its objectives, logic, activities, and indicators of successful performance. Canada [12]
- In clarifying program structure and intent, it can determine the plausibility of the program achieving its intended goals,

identify opportunities to improve program performance, and serve to ensure credible and useful evaluations (Wholey, 1987). It considers such factors as the program's characteristics, available research methodology, cost, and constraints on the use of the desired research methods in determining the **best evaluation design**. Canada [12]

### 4.2.3 *Types of Evaluation*

- Formative, process and impact/outcome evaluations seek to answer different types of question about the project.
  None is superior to the others. Rather, they complement each other and, ideally, should all be conducted at the appropriate stages in a project's cycle.
  Often an evaluation will seek to answer both process and impact/outcome questions.
  - **Formative evaluation** starts during a project's development stages and uses theory to develop and plan the project's components, development and pilot testing. It informs the direction a project will take. Pre-testing is a type of formative evaluation and involves trying out some of a project's parts before it is launched in full. It assesses a project's relevance to identified health problems, and the practicality of different intervention methods.
    It might identify features from a literature search that support a particular approach. For example, motivational interviewing, combined with reinforcement and follow-up.
  - **Process evaluation**, also known as implementation evaluation, begins at the start of a project. It assesses implementation and delivery, and identifies factors and conditions relating to how a project is being implemented. It aims to see why a project does or does not meet its aims and objectives, and can show whether the project deviated from the original plan.
    An evaluation which focuses on the process used throughout a project: it aims to see why the project meets or does not meet its aims and objectives; what went right and what went wrong; what can be learnt for future projects.
    Process evaluation. An assessment that compares actual with intended inputs, activities, and outputs.
  - **Impact/outcome evaluation** focuses on whether a project met its aims and objectives. This might be in terms of health outcomes, such as obesity, or impacts on health behaviours such as increased physical activity or better nutrition.

- Although the literature includes over a hundred different kinds of evaluation (see Patton, 1982), the vast majority boil down to two types: those that aim to determine if the program has been implemented as planned, and those that measure its success in achieving its objectives (i.e., its impact). The label most often associated with the first type is "process evaluation," although it is sometimes called formative evaluation. The latter type is known as "summative evaluation," also known as impact, outcome, or effectiveness evaluation.

  - Process Evaluation: The main objective is to provide feedback to managers on whether the program is being carried out as planned and in an efficient manner. Guidance should be provided for modifying the program to help ensure it meets its objectives. With this information, the program can be modified so it is carried out as planned, or the plan itself can be modified if it is found lacking.
    How is the program operating and how can it be made better?
    Process evaluations are directed at three key questions:

    1. the extent to which a program is reaching the appropriate target population;
    2. whether or not its service delivery is consistent with program design; and
    3. what resources are being expended (Rossi and Freeman, 1993).

  - Summative Evaluation: Does the program achieve its objectives? The purpose of summative evaluations is to assess the impact of the program; that is, ascertain the extent to which the program meets its objectives, and the needs of its target group. As well, it should provide advice for modifying the program so that it will better serve the needs of its clients and become more cost-effective (Stufflebeam and Shinkfield, 1985).

### 4.2.4 *Qualitative vs. Quantitative*

- Refer to Damir´s paper "On the Application of different Research Methods in Software Engineering" for Scientific research. Ismailović and Pagano [42]

*If you are not measuring, why bother? Ellen Wagner*

- If you are developing mobile solutions, how do you know what is happening with them?
  Without data, how do you know whether your mobile initiative is going well, needs tweaking, or should be put out of everyone's

misery?
Data can be of many types, and different initiatives can have different evaluation schemes, but you really should be looking to see what the outcomes are.
**Qualitative data**, such as user reports, is one form of valid data. Representative samples of the audience can be surveyed for feedback.
**Quantitative data**, such as percentages of the potential population of children finishing a level under the estimated time by the pedagogists/psychologists, is another form of data. Game logs can be collected and analysed. [76]

- When it comes to evaluating the impact of a learning initiative, I strongly believe that Donald Kirkpatrick's four levels of evaluation (Evaluating Training Programs, 1994) make sense, if used properly.
  That is, figure out what the organizational change needs to be (for example, "If we could just impact X, we could save/earn the organization $Y million a year," where X could be better decisions).

- That, ultimately, is the goal you want to achieve, and is your level 4 objective. Then you work back through observable behaviors (level 3, seeing whether the performer is persistently demonstrating the changed behavior), and an assessment of learner performance after the learning intervention (level 2, via a summative assessment), to a subjective evaluation of the learning experience by the learner (level 1, mostly useful for improving the learner "experience"). (I like one extra level I heard a client mention, level 0: essentially "Are they even showing up for the learning experience?") Level 2 is not necessarily useful when you move to a performance support use instead of a formal learning experience. [76]

- **Qualitative Research Design** is Research Design that uses systematic observation and focuses on the meanings people give to their social actions.
  If I measure something an inch wide and a mile deep; that's generally qualitative.
  It maximizes the possibility of obtaining valid answers to research questions or hypotheses.
  The primary focus of the reader is on the validity of the conclusion of the experimental Treatment.
  A method for controlling factors that could interfere with the accuracy of the Findings.
  The plan used to obtain valid and reliable answers to research questions according to the canons of Science.
  A set of instructions that tells the researcher how data should

be collected and analyzed in order to answer a specific research question.
Must be defined for each study.

- **Quantitative Research Design** is a Research Design that emphasizes the use of numbers and statistics to analyze and explain social events and human behavior.
  If I measure something or investigate something a mile wide and an inch deep; that's generally quantitative.
  - Experimental is a Research Design that attempts to discover a cause&effect relationship (Causation) between variables.
  - Quasi-Experimental
  - Non-Experimental
  - True-Experimental
- **Qualitative evaluations** use narrative or descriptive data rather than numbers.
  An information that is reported in narrative form or which is based on descriptive information, such as diaries, open ended responses to questions and field notes.
- **Quantitative evaluations** give numerical results.
  An information that is reported in numerical form, such as number of people attending and drop out rates.

### 4.2.5 *Evaluation Designs*

- There is no single correct evaluation design for impact evaluations. The goal is to come up with the best design possible under the circumstances. Almost all designs represent a compromise dictated by many practical considerations such as how much money and time are available, what the client considers compelling, how much a design might interfere with the normal operation of the program, and so on. In deciding on the design, the credo is to maximize the credibility and usefulness of the findings. The evaluator must anticipate the kind of arguments that will be used to dismiss the findings.
- Evaluation design is a plan for conducting an evaluation that specifies:
  1. a set of evaluation questions,
  2. the data that will be collected and analyses that will be undertaken to answer the evaluation questions,
  3. the estimated costs and time schedule for the evaluation study, and

4. how the evaluation information will be used.

**Research** is not the same as **evaluation**. The two are very different activities.
In research projects, an intervention tends to be designed and controlled by researchers who aim to ensure that it is delivered in a standardised way.
When evaluating a project in real-life, the challenge is usually to investigate a project that is being delivered by other people in a real-life settings. This tends to lead to more varied results and requires a more pragmatic approach. An aim of Game Developer Research is to identify those adaptive interventions that are 'effective'. Strictly speaking, effectiveness can only be demonstrated through controlled research designs which are expensive and time-consuming.

In general, a stronger evaluation design increases the confidence with which conclusions can be drawn from findings. In particular, a strong evaluation can indicate that a project's outcomes were caused by the project itself rather than by chance.

*Should we agree or decide on making the design more Experimental (**randomised controlled trials**) or Quasi-Experimental?*

- **Experimental designs**
  It is generally acknowledged that the **strongest scientific evidence comes from experimental designs**, and especially **randomised controlled trials** (RCT).
  Participants or groups of people are randomly allocated to receive an intervention (intervention group) or not (control group). Changes in the intervention group are compared against measures in the control group. This reduces the possibility that the changes occurred by chance, and increases confidence that they were caused by the intervention itself rather than an external factor.
  Experimental design. An assessment design that tests the existence of causal relationships by comparing outcomes for those randomly assigned to program services with outcomes for those randomly assigned to alternative services or no services.

- **Quasi-experimental designs**
  These types of designs tend to be more **feasible** in evaluations of projects in **real-life settings**. Quasi- experimental designs may use a control group but, unlike the RCT, do not randomly allocate participants to either the intervention or control groups.
  Quasi-experimental design. An assessment design that tests the existence of a causal relationship where random assignment is not possible. Typical quasi-experimental designs include pre-post designs, comparison group designs, and interrupted time-series designs.

- **The quasi-experimental evaluation will explore whether the dynamic switching (Adaptivity) between levels in the Bug**

**Game will improve the gameplay and the learning expeience by evaluating which adaptive variables and with which amount are responsable for the challenge made by children while playing the adaptive version of the game.**

- Quasi-Experimental Design is Practical, feasible, generalizable, used in "real world" practice settings.
- Limitations: Unable to make clear cause-effect statements, but can increase knowledge.
- Threats to validity need to be specified.
- **Pre-experimental designs**
  These types of studies provide the **weakest evidence** and should only be used when all other possibilities have been explored. Pre-experimental designs include a **pre-post study** (where data are collected before and after an intervention).
- **Pre-post design** is an assessment design that compares outcomes before and after the program.
- **Randomized experiment**. An experiment, such as a randomized controlled trial, that randomly assigns units to different conditions.

### 4.2.6 *Logic Model*

*Logic Model & Theory-Driven Evaluation*

- **Theory-Driven Evaluation**
  Assumptions about resources and activities and how these are expected to lead to intended outcomes are often referred to as program theory. A **logic model** is a useful tool for describing program theory. The hypothesis, often implicit, is that if the right resources are transformed into the right activities for the right people, then these are expected to lead to the results the program was designed to achieve. Some evaluators believe that making explicit the underlying assumptions about how a program is supposed to work increases the potential for evaluation utility. Although developing the program theory prior to the evaluation is considered most beneficial for predicting relationships, developing program theory at the end of the evaluation helps explain observed causal relationships.
  Leeuw (2003) provides an excellent review of three approaches to restructuring program theories after the program has been implemented:
  - The policy-scientific approach is more empirical than the other approaches and consists of generating a series of propositions, or assumptions, about how the program is

supposed to work. The evaluator then tests these propositions through a review of relevant scientific research, interviews with key staff, and document reviews.

  - The strategic assessment approach is driven through conversations or dialogues with program staff and participants. The focus is to draw out the underlying assumptions about how the program works and then subject these to open debate among stakeholders and staff.
  - The elicitation approach aims at recovering the mental models or cognitive maps that program staff hold about their program. The various maps are then com- pared, contrasted, and assessed for their validity through open dialogue and reviews of existing related research. Wholey et al. [98]

- The central theme in all three approaches is discovering the underlying assumptions held about how the program is believed to be working to achieve its outcomes and then testing these assumptions once they have been made public. Wholey et al. [98]

- A logic model is useful for focusing on an intervention's likely impact. A logic model describes the relationships between each element in a project or intervention, and the likely direction of change.

  - Ressources / Input
  - Activities
  - Initial Outcomes
  - Short-term Outcomes
  - Intermediate Outcomes
  - Long-term Outcomes

- A program logic model is a flowchart that summarizes key elements of a program: resources and other inputs, activities, outputs (products and services delivered), and intermediate outcomes and end outcomes (short-term and longer-term results) that the program hopes to achieve. Logic models should also identify key factors that are outside the control of program staff but are likely to affect the achievement of desired outcomes. A logic model shows assumed cause-and-effect linkages among model elements, showing which activities are expected to lead to which outcomes, and it may also show assumed cause-and-effect linkages between external factors and program outcomes. Wholey et al. [98]

- In laying the foundation for any evaluation, an evaluator must learn the answers to three critical questions:

1. Exactly what is being evaluated?
   This question implies that the evaluation designer must have or develop a good understanding of the program to be evaluated. This is simple common sense. Without a knowledge of the issues and problems impelling a program, the objectives of the program devised to address the problems, the activities and services developed to reach the objectives, the target group of the program and so on, it would be impossible to determine which evaluation questions are important to answer. To proceed with a vague understanding of the program guarantees the failure of the evaluation. To understand why this is so, ponder an analogy posed by Rutman (1980). Do drugs work? "Such a question is obviously ludicrous since it is necessary to identify the drug, its dosage, how it is administered, for which problem, and to what type of people" (pp. 45-46). A clear description of the program, its components, services, goals and target group forms the basis for developing evaluation issues and methodology. This information is also used to produce the **logic model**.
2. Why is it being evaluated?
   It is also crucial to find out why the program is being evaluated. Only with this information can the evaluator design an evaluation that is of use to program sponsors or administrators. Often, there is no single reason for wanting a program evaluated; there are almost always different audiences for an evaluation and each may have its own reasons for wanting (or not wanting) an evaluation. Policy makers may want to know about the adequacy of design; sponsors may wish to know how the program compares to less expensive alternatives; program managers may want to know how to iron out specific problems; staff may want roles and responsibilities clarified; all are likely interested in suggestions for improving program operations; and all would like to know if the objectives are being met.
3. Who will use the Information and how?
   This is a critical question to ask of all key stakeholders. If there is no satisfactory response to this question, chances are the information will never be used and there is little reason to spend the time and money to gather it. Never overlook the possibility of multiple purposes and hidden or conflicting agendas of key stakeholders concerning the evaluation (which may be uncovered through interviews with the stakeholders). Evaluators cannot meet the relevant needs of their clients unless they know who will use the information and how.

### 4.2.7 *Guidelines to Usability Engineering (Software Engineering Vs. Evaluation)*

- "Case studies of professional software development practices describe how real (or realistic) projects are planned and executed. Cases provide engaging models of the activities and materials of software development to students and other novice practitioners. They vividly remind learners of the possibilities for meaningfully applying knowledge and skills in the world beyond the classroom. During the past six years, we have developed and used a collection of usability engineering case studies for teaching human-computer interaction, primarily to upper-level undergraduates in computer science and in information sciences and technology."
- Evaluation of systems usability Savioja et al. [85]
- A Case Library for Teaching Usability Engineering: Design Rationale, Development, and Classroom Experience Carroll and Rosson [13]
- Verification, Validation and Testing in Software Engineering Dasso and Funes [21]
- Praise for Usability Testing Essentials Barnum [4]
- Usability of Complex Information Systems Evaluation of User Interaction Albers and Still [1]
- HANDBOOK OF PRACTICAL PROGRAM EVALUATION Wholey et al. [98]
- Software Testing Testing Across the Entire Software Development Life Cycle Everett and McLeod [25]

## 4.3 GUIDELINES AND LIMITATIONS OF RESEARCH IN SERIOUS GAMES

Taken from KEBRITCHI [49] About prooving learning effectiveness through serious games.

A cursory literature review indicated that out of 40 reviewed studies related to Serious Games, only 14 empirical studies and 4 literature reviews focused on the use of Serious Games for facilitating learning in a formal school setting.

Methodological flaws in empirical studies are another factor that hinders reaching solid conclusions about the effects of instructional games. One of frequent problems is lack of control groups in the studies. Examining the effect of a treatment without comparison with a control group is problematic (Mitchell & Savill-Smith, 2004; Vogel et al., 2006). Out of the 14 empirical studies, only five studies used

experimental research design incorporating control and experimental groups.

Furthermore, the results of the literature review indicated mixed results related to the effectiveness of games as learning tools. Although the majority of the reviewed empirical studies, 9 out of 11, indicated that using serious games improves learning and learning environments, the literature reviews indicated mixed results.

The findings of empirical studies revealed that instructional games promoted learners' attention (Yip & Kwan, 2006), state of flow (Kiili, 2005b), motivation (Rosas et al., 2003), delayed retention (Cameron & Dwyer, 2005), mathematics performance (Ke & Grabowski, 2007; Lopez-Moreto & Lopez, 2007; Shaffer, 1997), knowledge transfer (Shaffer, 2006), decision making (Corsi et al., 2006), expert behavior development (VanDeventer & White, 2002), and spatial skills and brain oscillation (Natale, 2002). In addition, using games created dynamic (Rosas et al., 2003) and collaborative (Squire, Giovanetto, Devane, & Durga, 2005) learning environments which positively affected learning.

However, the literature reviews indicated the results were not always positive. Randel's et al. (1992) study that reviewed 67 empirical studies, from 1984 to 1991, compared the instructional effectiveness of the games with conventional classroom instruction and indicated that out of 67 studies, 38 showed no differences between the game and conventional studies, 27 favored games, but again 5 were questionable in terms of their method and 3 favored conventional instruction. The two reviews conducted by Emes (1997) and Harris (2001) found no clear causal relationship between academic performance and the use of computer games. A slight improvement in learning and attitude of learners toward the subject matter was found as a result of using the games instead of traditional teaching methods (VanSickle, 1986). Finally, a recent literature review which analyzed 32 studies concluded that interactive simulation and games were more effective than traditional classroom instruction on learners' cognitive gains (Vogel et al., 2006). The contradicting views of the literature review, the existence of relatively few empirical studies in the reviews, and the cited methodological flaws in the empirical studies necessitate further rigorous empirical study to help educators and instructional designers reach better conclusions about the effects of instructional games so that they may better understand, implement, and facilitate the games in classroom setting. As suggested by Van Eck (2006), instructional games would likely experience widespread development and use if persuasive examples of empirical studies could show the enhancement of learning by using instructional games. This study was conducted in response to such needs to provide solid results by implementing experimental method.

### 4.3.1 *Guidelines*

- "According to Ross and Morrison (2004), and Savenye and Robinson (2004), researchers in instructional technology should employ mixed, parallel methods to produce the most convincing body of evidence.
  Therefore, although the study adopted **qualitative case study** as the dominant paradigm in order to investigate a contemporary phenomenon – game-based learning – within its real-life context (Yin, 1984, p. 23), where possible, quantitative procedures were employed in the method of this study to corroborate and extend the primarily qualitative approach. **Qualitative data** were collected in multiple forms – **in-field observation**, **document analysis**, and **think-aloud verbal protocol** – to achieve triangulation of data. The researcher subsequently conducted an analysis of themes in order to capture the essence of game-based learning through the voices of those who have participated directly in its implementation. **Quantitatively**, a **within-group pretest–posttest comparison** was made to investigate if students' **math** test performance, learning attitudes, and metacognitive awareness level improved through a game-based summer math program." [48]

### 4.3.2 *Methodologies*

1. An experimental design
   - Pre-test and Post-test method with a control group in this paper An Experimental Study of Game-Based Music Education of Primary School Children by Sibel Çoban and İmge Tuncer published in [19]with a total of 52 students 26 of whom were in the "study group" and 26 in the "control group"
2. A quasi-experimental design
   - A quasi-experimental design is followed in which data have been gathered on the process of gaming and learning of 458 pupils from twenty classes of five schools. Qualitative and quantitative data have been gathered. In this paper Cognitive and Affective Effects of Learning History by Playing a Mobile Game by Jantina Huizenga, Wilfried Admiraal, Sanne Akkerman and Geert ten Dam published in [19].

### 4.3.3 *Challenge*

- Three major challenges for developing and evaluating immersive digital educational games (DEGs) are identified, including
  - improving adaptive technologies, especially digital storytelling, to shape learning experience;
  - providing technological approaches to reduce developments costs for DEGs, and
  - developing robust evaluation methodologies for DEGs. Law and Rust-kickmeier [55]

### 4.3.4 *Limitations*

- The integration of external resources (e.g., learning media) with a game engine into a coherent and immersive game environment is difficult.
  Since the development and application of immersive DEGs is still at an early stage, to date no appropriate methodologies exist which enable an effective integration of existing learning resources and their (re)use in DEGs.
  Hence, it is deemed critical to analyze the technological and didactic demands and mutual dependencies between learning resources, learning activities, pedagogical models, and narrative/game engines. An approach of resource harmonization, resource symbolization, and ontological resource description, should also be established. Law and Rust-kickmeier [55]

- Exploiting up-to-date 3D computer games for educational purposes is a dawning technology but still in its fledgling stage. It is one of the challenges to be addressed in 80Days, which also aims at augmenting and integrating the related theoretical frameworks in cognitive psychology. Exploiting the desirable features of digital games to be effective learning tools entails well-orchestrated efforts of a highly interdisciplinary team to tackle the three major challenges addressed above. Concrete empirical results will be documented in the near future. Law and Rust-kickmeier [55]

1. Cost & Time
   - A study to proove the effectiveness of Serious Games in learning enhancement is a long-term research experience.
     In the study presented in this paper "The Impact of Mind Game Playing on Children's Reasoning Abilities: Reflections from an Experience by Rosa Maria Bottino, Michela Ott and Mauro Tavella" published in [19]; For more than 3

years, a study was conducted with standardized tests carried out at national level.

- The development of competitive DEGs is cost-intensive, and the markets are narrow because DEGs may relate to limited age groups or specific curricula. Thus, the integration of existing learning resources is a crucial aspect of efficient and cost- effective learning design and game development. Law and Rust-kickmeier [55]

2. Intervening Variables

   - With respect to the relation to the variable of gender and the variable the financial status of the family and the variable of the parents' educational background and the variable of the parents' occupation point out to slight differences in the achievement rate of the students. In this paper An Experimental Study of Game-Based Music Education of Primary School Children by Sibel Çoban and İmge Tuncer published in [19].
   - "As Squire (2003) discovered, bringing a computer game into classrooms may raise as many issues as it solves.
     First, playing games does not appeal to every student.
     Second, students may be distracted by game-playing, and thus, not achieving the learning goals (Miller, Lehman, & Koedinger, 1999).
     Further, students may fail to extract intended knowledge from a complicated gaming environment (Squire, 2003).
     Finally, game design researchers (Smith & Mann, 2002) are worried that making games where the objective is to facilitate students' learning will risk **sacrificing the game part along the way**." [48]

3. Lack of general frameworks providing guidelines for structured GBL evaluation.

   - In this paper Development of a General Framework for Evaluating Games- Based Learning by Thomas Connolly, Mark Stansfield and Thomas Hainey published in [19].

#### 4.3.5 *Agreement*

- "Many researchers and educators have noted the positive effects of educational games in thinking and learning, and as a result, many researchers have developed games for educational purposes in many fields (Oblinger, 2004; Virvou, Katsionis & Manos, 2005; Amr, 2007; Pivec & Kearney, 2007). Contemporary research indicates that educational games have the potential to improve learning (Srinivasan, Butler-Purry & Pedersen, 2008)

and to increase the likelihood that the desired learning outcomes will be achieved (Pivec & Kearney, 2007)." [70]

- "More specifically, educational games can improve learning because they:
  - increase motivation (Garris, Ahlers & Driskell, 2002; Lee & Peng, 2006; Graesser, Chipman, Leeming & Biedenbach, 2009),
  - create cognitive conflict (Chen, Lien, Annetta & Lu, 2010),
  - can improve and enhance visual-spatial perception ability (Greenfield, Brannon & Lohr, 1994; Lee & Peng, 2006),
  - can generate various competencies (such as motor, cognitive, emotional, social, personal, etc.) (Kretschmann, 2010),
  - can stimulate intensive mental engagement (Prensky, 2001), and
  - create conditions for better memory retention (Oblinger, 2004; Lee & Peng, 2006)." [70]
- "Considering children's natural love for playing games, the growing availability of electronic media, and the fact that children spend a great deal of time interacting with electronic media (Fisch, Lesh, & Motoki, 2009), the educational value of digital games is very clear." [70]

### 4.3.6 *Skepticism*

- "Skeptics toward game-based learning contend that **the effectiveness of computer games on learning is still a mystery**.
  Several major reviews on educational games (Dempsey, Rasmussen, & Lucassen, 1996; Randel, Morris, Wetzel, & Whitehall, 1992; Vogel et al., 2006) indicated **no clear causal relationship between academic performance and the use of computer games**.
  A common skepticism on using computer games for learning purposes lies in the lack of an empirically-grounded framework for integrating computer game into classrooms." [48]
- "According to Hodgson, Man & Leung (2010), the wisdom of using digital games for learning is debatable, because games are commonly perceived as tools for having fun. Although games can be effective learning environments, not all games are effective, nor are all games educational and beneficial for all learners or for all learning outcomes (Oblinger, 2006)." [70]
- "Nevertheless, there exist quite a **few empirical studies** relating to questions of exactly how and why children act one way or

another within the evolving dynamics of an educational gaming environment (Gee, 2003; Ke, 2008; Conati & Zhao, 2008; Fisch, Lesh, & Motoki, 2009). In addition, many researchers contend that **the effectiveness of computer games on learning is still unknown** (Ke, 2006, 2008)." [70]

# 5

# TRENDS IN SCIENTIFIC STUDIES ON SERIOUS GAMES

- 
- The literature study is divided into three parts: part one reviews relevant theories and empirical studies from developmental psychology; part two discusses the literature on young children and technology and investigates existing guidelines; and part three surveys the literature on doing research and usability testing with children.
- This Chapter presents a **comparative Research** on Evaluation Frameworks designed for Serious Games.
  I started from the analysis of ... studies and developed a preliminary theory of adaptive factors affecting user satisfaction.
- Maybe we should call it comparative Studies!?

*Literature Review; Number of reviewed Studies; Aim; methodologies & Results*

- This Chapter will present the results of an extensive literature search to identify measurements that have been taken in relevant studies.
  The literature has a wealth of articles suggesting ways that GBL can be evaluated in terms of particular areas with particular measurements, experimental designs and analytical techniques.
- Little has been published about ...
- "The most successful and the most common form of educational games are the **mini games** targeting at **preschool age** and **primary education** level children. These games attempt to help young children to obtain **basic skills such as numbers**, letters, **simple maths**, and reading, offering entertainment and instruction for their target audience (Kickmeier-Rust, 2009)." [70]

  The purpose of this section is to provide an overview about research conducted on effects of instructional games. This overview indicated major concerns and needs for further research in the field of instructional games, thus, guided this study and formed major variables for this study. The findings revealed that research investigated learning effect of the games from various perspectives by using different research methods. The results are presented based on research methods in two sections of non-empirical or literature review and empirical research.

– *non-empirical or literature review vs. empirical research*

| ... | STIMULUS PRESENTED | STIMULUS WITHDRAWN |
|---|---|---|
| Pleasant stimulus | Positive reinforcement | Negative punishment (Omission) |
| Unpleasant stimulus | Positive punishment | Negative reinforcement |

Table 1: Literature Review: Studies on the impact of Adaptivity

- – *literature review*

As a result of the literature review, a number of literature reviews or non-empirical studies were found that examined the effect of instructional games by using the results of reported research to analyze, compare and integrate the findings and reach a conclusion. The significance of the literature reviews is that they provide an overview about the effect of the games based on reviewing of a relatively large number of articles. Table 3 summarizes the literature reviews that were found along with the number of studies reviewed by them and their conclusions on the effect of instructional games on learning.

- The majority of the literature reviews, five of eight, reported mixed results (...), or no positive results (...) on effectiveness of serious games.
- The Research Variables, Methods, and Results of Empirical Studies reviewed as shown in 9

## 5.1 RESEARCH ON SERIOUS GAMES

- Evaluating Serious Games Confusing Terminology Youngblood [102]
- Design and Evaluation of a Serious Game for Immersive Cultural Training Froschauer et al. [29]
- Challenges in the Development and Evaluation of Immersive Digital Educational Games Law et al. [56]
- Design and evaluation challenges of serious games Raybourn and Bos [78]
- Methods for Evaluating Gameplay Experience in a Serious Gaming Context Nacke et al. [68]

The Research Variables, Methods, and Results of 14 Empirical Studies

| Study | Research Variables | | | Research method, Game subject, Educational level | Results of facilitating learning with game |
|---|---|---|---|---|---|
| | Dependent variables | Individual differences | Other variables | | |
| 1. Rosas et al. (2003) | Achievement Motivation | | Learning environment | Mixed method, Experimental Mathematics & Reading K-12 | Positive results |
| 2. Sedighian and Sedighian (1996) | Motivation | | | Qualitative Mathematics K-12 | positive |
| 3. Klawe (1998) | Achievement Motivation | Gender | Game design | Qualitative Mathematics K-12 | Positive |
| 4. McDonald and Hanaffin (2003) | Achievement Motivation | | | Qualitative Quasi-Experimental Social studies K-12 | Mixed results |
| 5. Lopez-Moreto and Lopez (2007) | Motivation | Computer experience | | Quantitative Mathematics K-12 | Positive |
| 6. Ke and Grabowski (2007) | Achievement | Prior knowledge computer experience language background | | Quantitative Experimental Mathematics K-12 | Positive results |
| 7. Moreno (2002) | Achievement | Prior knowledge computer experience Socio-status language background | | Quantitative Experimental Mathematics K-12 | Positive results |
| 8. Din and Calco (2000) | Achievement | | | Quantitative Experimental Mathematics K-12 | Mixed Results No improve in Mathematics |
| 9. Laffey, Espinosa, Moore, and Lodree (2003) | Achievement | At risk behavioral problem | | Quantitative Experimental Mathematics K-12 | Mixed results due to method flaw |

Figure 9: The Research Variables, Methods, and Results of Empirical Studies reviewed

| ... | STIMULUS PRESENTED | STIMULUS WITHDRAWN |
|---|---|---|
| Pleasant stimulus | Positive reinforcement | Negative punishment (Omission) |
| Unpleasant stimulus | Positive punishment | Negative reinforcement |

Table 2: Literature Review: Studies on the impact of Adaptivity

## 5.2 EVALUATION STUDIES ON LEARNING ENHANCEMENT AND MOTIVATION

Research on the effects of computer use on children's cognitive, social and physical development is ambiguous – it has advantages and disadvantages. Based on existing research (such as [1] and [2]) we assume that technology can potentially support the development of young children. We also assume that children are attracted to technology and, in general, want to 'play' with it. Many of the available products are weak in addressing specific developmental needs and skills of their intended users [3, 4]. Our primary concern is to establish how interaction designers can maximise the developmental spin-offs and make it fun to use. Gelderblom [33]

- Learning from Computer Games Nählinder and Oskarsson [69]
- Video Games as an Education Tool Robertson and Hart [82]
- "The 'holy grail' for training professionals is to harness the motivational properties of computer games to enhance learning and accomplish instructional objectives." (Garris, Ahlers, & Driskell, 2002, p. 442)
- Learning Language through Video Games: A Theoretical Framework, an Evaluation of Game Genres and Questions for Future Research Jonathan deHaan Games [30]
- Gaming for reading Clarke and Treagust [17]

1 Liang, P and Johnson, J. Using technology to enhance early literacy through play. Computers in the Schools. 15,1 (1999), 55-64.
2 Papert, S. Mindstorms: Children, computers and powerful ideas. New York, Basic Books, 1980.
3 Scaife, M. and Rogers, Y. Informing the design of a virtual environment to support learning in children. International Journal of Human-Computer Studies. 55 (2001), 115-143.
4 Wartella, E.A. and Jennings, N. Children and Computer Technology: New Technology - Old Concerns. The Future of Children. 10,2 (2000), 31-42.

- Designing games for learning: insights from conversations with designers Isbister et al. [39]
- Can Educational Research Be Both Rigorous and Relevant? Reeves [80]
- Societal impact of a serious game on raising public awareness: the case of FloodSim Rebolledo-Mendez et al. [79]

*FloodSim Serious Game*

  - This paper presents an evaluation of the societal impact of a **simulation-based Serious Game**. FloodSim was developed with the aim of raising awareness of issues surrounding flooding policy and citizen engagement in the UK. The game was played by a large number of users (N=25,701) in a period of 4 weeks. Quantitative and qualitative analyses (on a reduced data set) were carried out in order to explore the impact of FloodSim play in raising the general public awareness around flooding in the UK. The results suggest FloodSim was hugely successful in generating general public interest and there was evidence that (a) FloodSim increased awareness at a basic level and (b) that despite the simplicity of the simulation, players perceived FloodSim to be an accurate source of information about flood risk and prevention.
  - Methodology: The evaluation was based on three sources of data: a) the demographic information left on the site, b) the feedback left on the site, and c) telephone semi-structured interviews.
  - The impact of a serious game on flooding in raising the general public's awareness about flooding in the UK was studied.
- Enhancing the educational value of video games Bellotti et al. [6]
- Reducing Unproductive Learning Activities in Serious Games for Second Language Acquisition Wu [101]
- Understanding game design for affective learning Dormann and Biddle [23]
- A framework for evaluating the quality of multimedia learning resources Leacock and Nesbit [57]
- Profiling academic research on digital games using text mining tools Bragge and Storgå rds [10]
- ASSESSING WHAT PLAYERS LEARNED IN SERIOUS GAMES: IN SITU DATA COLLECTION, INFORMATION TRAILS, AND QUANTITATIVE ANALYSIS Loh et al. [59]

| ... | STIMULUS PRESENTED | STIMULUS WITHDRAWN |
|---|---|---|
| Pleasant stimulus | Positive reinforcement | Negative punishment (Omission) |
| Unpleasant stimulus | Positive punishment | Negative reinforcement |

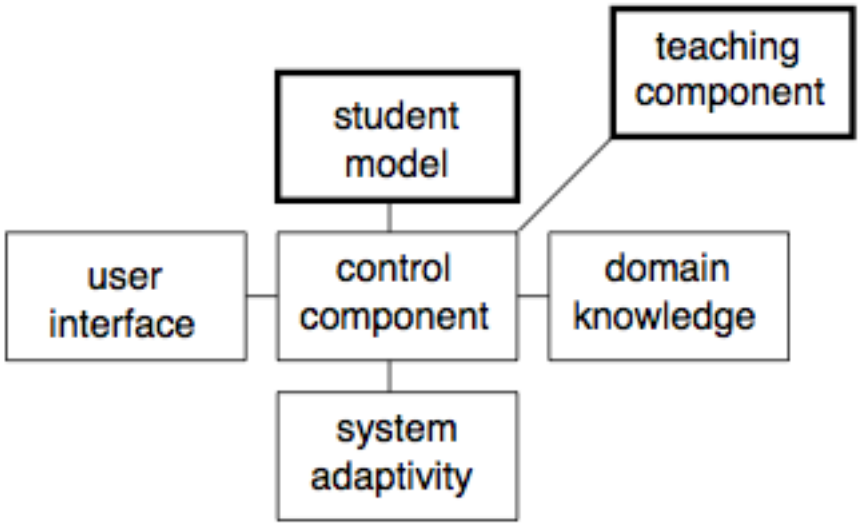
Table 3: Literature Review: Studies on the impact of Adaptivity

- Digital games in education: The design of games based learning environments Gros [35]
- A conceptual model for game based intelligent tutoring systems Mills and Dalgarno [66]
- Serious video game effectiveness Wong et al. [100]
- Digital Game Based Learning: It is Not Just the Digital Natives Who Are Restless Van Eck [94]
- Proof of Learning: Assessment in Serious Games Chen and Michael [14]
- Serious games: Serious opportunities Stapleton [90]
- The motivation of gameplay: The real twenty first century learning revolution Prensky [75]
- Games, Motivation, and Learning: A Research and Practice Model Garris et al. [31]
- Immersive Didaktik: Verdeckte Lernhilfen und Framingprozesse in Computerspielen Bopp [7]

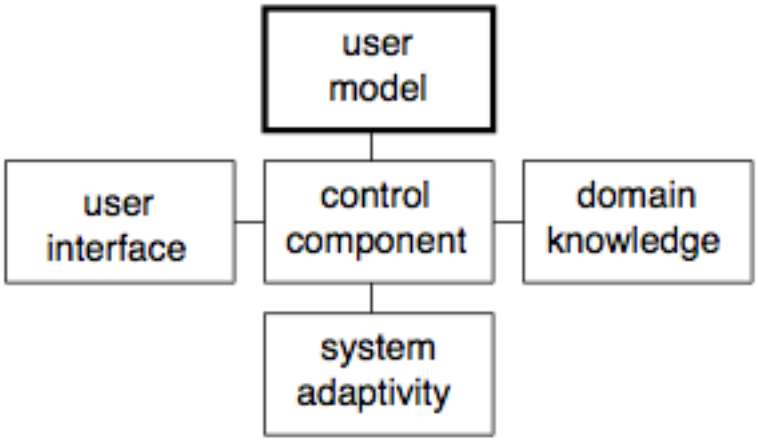
## 5.3 STUDIES EVALUATING SERIOUS GAMES

- Triadic Game Evaluation: A Framework for Assessing Games with a Serious Purpose Harteveld [36]
- Just in Time Learning: Implementing Principles of Multimodal Processing and Learning for Education Massaro [62]
- GameBased Learning for Knowledge Sharing and Transfer: the eVITA project Pappa [71]
- Advanced Frontiers of eAccessibility Serious Games and Interactive DevicesEvett [26]

- Using Games as a Method of Evaluation of Usability and User Experience in Human Computer Interaction Design Collins et al. [18]
- A QuasiExperimental Evaluation of an Online Formative Assessment and Tutoring System Koedinger et al. [53]
- The use of a digital dance mat for training kindergarten children in a magnitude comparison task Cress et al. [20]
- Family Story Play: Reading with Young Children and Elmo Over a Distance Raffle et al. [77]
- Increasing Childrens Social Competence Through Games, an Exploratory Study Hendrix et al. [37]
- The effects of individualized feedback in digital educational games Kickmeier-Rust et al. [51]
- Experiences from the use of a shared multimedia space for eLearning in Brazil primary schools de Oliveira et al. [22]
- Language Teaching in a Mixed Reality Games Environment Powers et al. [74]
- Investigating the Educational Effectiveness of Multiplayer Online Games for Children Garzotto [32]
- The Use of Agents in Human Learning Systems Sklar and Richards [89]
  - This paper examines agent-based systems designed for a variety of human learning tasks. These are typically split into two areas: "training", which generally refers to adult learning of job-related skills, frequently but not exclusively in military settings; and "education", which generally refers to child and adult learning in academic settings, including primary and secondary schools, colleges and universities. While the terms may indicate diverse areas within the field of human learning, from the standpoint of agent-based systems development, many of the more prominent issues are held in common; as well, these issues can be generalized to most interactive agent-based environments. Here, we categorize three major trends in development of agents to assist human learners: pedagogical agents, peer learning agents and demonstrating agents. We highlight recent work within each of these categories, bringing to light common themes and issues.
  - Pedagogical Agents
    Much like a narrator in a movie who provides voice-over




(a) interactive learning system




(b) interactive system

Figure 10: Typical system architectures Sklar and Richards [89]

to explain scenes but never actually appears in the film, a pedagogical agent will "pop-up" when the learner indicates that she needs help. There are three methods used to determine when this help is needed:

  * directly, upon request by the user, for example by clicking a "help" button;
  * indirectly, by the system monitoring the learner's performance and au- tomatically detecting when she seems to need assistance;
  * mixed initiative, which relies on a combination of directly and indirectly.

- The work of Lucas et al. [5] involves the use of pedagogical agents that assist learners when the agent perceives that the user needs help, based on a model of the learner and the learning goals of the system.
- The primary goal in a human learning environment is for the learner to advance.
- One of the more prominent issues that separates human learning systems from other agent-based and interactive system development is the aspect of testing and evaluation. It is expected that not only the software learning environment be fully debugged and tested, but also, particularly amongst education researchers, the system must be evaluated with respect to its effectiveness as a learning environment.

*no fixed standard for evaluating the effectiveness of interactive learning systems*

- While there is no fixed standard for evaluating the effectiveness of interactive learning systems, there are two generally accepted categories of assessment [6]:
  * formative assessment tests the design and behavior of a system in-progress, generally performed by computer scientists, system designers and builders; and
  * summative assessment evaluates the effectiveness of a completed system, generally performed by educators and/or psychologists.
- Researchers begin by identifying what is being evaluated. Design and performance aspects need to be examined differently.

5 [J. P. Lucas, B. Wilges, and R. A. Silveira. Inserting animated pedagogical agents inside distributed learning environments by means of fipa specifications. In Agent-based Systems for Human Learning, AAMAS Workshop, 2005.]

6 [D. Littman and E. Soloway. Evaluating itss: the cognitive science perspective. In M. C. Polson and J. J. Richardson, editors, Foundations of Intelligent Tutoring Systems. Lawrence Erlbaum Associates, Hillsdale, NJ, 1988.] & [M. A. Mark and J. E. Greer. Evaluation methodologies for intelligent tutoring systems. Journal of Artificial Intelligence and Education, 4, 1993.]

- System adaptivity can be compared using interactions at different skill levels.
  The control component can be evaluated using various system performance measures, such as speed.
  Finally, and probably the most important, improvement in student knowledge (i.e., learning) can be measured using the same criteria in a computer-based environment that are employed within standard educational and/or psychological testing.
  These include:
  * validity: does the test show evidence that it measures what it says it measures?
  * reliability: are multiple results for the same subject consistent?
  * objectivity: is the test administered and scored the same way for every participant?
  * standardisation: can results be translated into a meaningful representation of student performance?
- The techniques for performing assessments vary depending on which component is being evaluated, where in the system development cycle the evaluation is being performed and who is performing the evaluation.
- Evaluation of learning systems typically follows an iterative cycle. Beginning with system development and extending through to experimental research, steps may be revisited at any time during the formative phases of system development. Once summative assessment begins, in the experimental research phase, the system cannot change; otherwise, the summative results will be invalid. Pilot testing often occurs late during formative assessment, bridging the gap to summative assessment.
  There are three methods of pilot testing:
  * one-to-one, which is performed early in the development cycle, with one student, instructor, trainer or researcher providing feedback;
  * small-group, which is performed later in the development cycle, with a small group of students, instructors or trainers providing feedback; and
  * field, which is performed near the end of development, emulating experimental conditions with teachers or trainers and students in a "live" (i.e., class-room) setting.
- While developing relevant criteria is not an easy task, this method may prove useful in formative assessment and in

comparing different systems. With **expert knowledge** and **behavior assessment**, system performance is compared to that of a human expert performing the same task. Software systems may be subjected to a standard certification process, through careful examination by qualified human experts. In sensitivity analysis, the **responsiveness** of a system is tested on a variety of different user behaviors.

- After system development and pilot testing are complete, experimental research can begin.
- Two mechanisms for collecting evaluation data are common:
  - Quantitative, in which numerical data is analysed, frequently by comparing scores on pre- and post-tests and surveys, to measure changes in student performance and attitudes;
  - Qualitative, in which interviews and surveys are conducted and observations are made.
- Mixed methods research [7] combines the two, but typically, at least in the education arena, researchers tend to adhere to the methods of one category or the other.
  - Quantitative methods rely on standards type testing, with multiple choice style questioning and Likert scale surveys. System logs are also examined.
  - Qualitative, or "open", methods encompass data taken in both written and oral forms, as part of interviews, questionaires and open surveys containing short-answer questions (rather than multiple choice). Transcripts are "coded" and analysed based on measures such as frequency of broad term usage, often borrowing techniques from natural language processing in order to compute semantic similarity between answers.

- The Crucial Role of Animated Childrens Educational Games McCue [64]
- Edutainment: A Case Study of Interactive CDROM Playsets Egloff [24]

## 5.4 STUDIES ON SERIOUS GAMES FOR MATHS

1. **Study I: A case study of computer gaming for math: Engaged learning from gameplay? by Fengfeng Ke** [48]

7 [R. B. Johnson and A. J. Onwuegbuzie. Mixed methods research: a research paradigm whose time has come. Educational Researcher, 33(7), October 2004.]

| ... | STIMULUS PRESENTED | STIMULUS WITHDRAWN |
|---|---|---|
| Pleasant stimulus | Positive reinforcement | Negative punishment (Omission) |
| Unpleasant stimulus | Positive punishment | Negative reinforcement |

Table 4: Literature Review: Studies on the impact of Adaptivity

- "Employing **mixed-method approach**, this **case study** examined the in situ use of educational computer games in a summer **math program** to facilitate 4th and 5th graders' cognitive math achievement, metacognitive awareness, and positive attitudes toward math learning.
  The results indicated that students developed **more positive attitudes** toward math learning through **five-week computer math gaming**, but there was no significant effect of computer gaming on students' cognitive test performance or metacognitive awareness development.
  The **in-field observation** and **students' think-aloud protocol** informed that not every computer math drill game would engage children in committed learning.
  **The study findings have highlighted the value of situating learning activities within the game story, making games pleasantly challenging, scaffolding reflections, and designing suitable off-computer activities**." [48]
- "... the gaming researchers' proposition that we should integrate learning content into the game world by stealth so that younger learners, subconsciously, are **engaged in repetitive play** and hence repetitive learning." [48]
- "the very argument for using games for learning, that they are engaging, vanishes along with the game part.
  Therefore, the key question remains misty: **Do computer games really foster an engaging, effective learning experience in classrooms?**
  Limited studies were conducted to explore the above question." [48]
- "Certain researchers, such as Barab et al. (2005) and Squire (2003), did start to examine what happens with students and their learning processes in game-based curricula of mathematics, science, and history.
  Barab et al. (2005) built an educa- tional adventure game (Quest Atlantis) from scratch while Squire (2003) customized

a commercial off-the-shelf role-playing game (Civilization) for classroom application.
Games used in both studies could be classified into **simulation** genre."
As a complement to their works, the present study examined the use of puzzle games for **drill and practice purpose**.
Two reasons underlie the selection of drill and practice games:
(a) computer games have been used in education primarily as tools for supporting drill and practice, yet limited research has been done on the effectiveness of these games;
(b) in comparison with simulation games, drill and practice games are easier to be introduced in a classroom and integrated into a traditional curriculum (Squire, 2003)." [48]

- "To fully understand whether the computer math games **promote learning**, an evaluation of comprehensive learning outcomes of game-playing is necessary.
Researchers (e.g. Barkatsas & Hunting, 1996; Mayer, 1998) have claimed that **math learning depends on the development** of all three components: skill, metaskill, and will, or in other terms, the integration of cognitive, metacognitive, and affective affects (Mayer, 1998, p. 51). Although the games used this study concentrate on learners' practicing skills until they are nearly automatic (thus without conscious cognition or metacognition), this automatic level of mastery, as gaming researchers claimed, is achieved by iterative game cycles of cognitive assimilation or accommodation (Gee, 2003; Van Eck, 2006). That is, game play at each level exposes the players to new skills and allows them to practice this new skill set to an automatic level of mastery; they then move on to a new level and see the old skills be challenged and have to think again and learn anew (Gee, 2003)." [48]
- "Malone (1981), when exploring the motivational aspects of digital games, concluded that an engaging game should first present an **obvious and compelling goal**." [48]
- "It should be emphasized that this is a case study of one specific set of learning games with a small sample of students who, even though diverse in gender, socio-economic status, and prior math abilities, were from the same school that has consistently achieved higher levels of student proficiency score than demographically similar schools. Cautions should be exercised when generalizing the study findings to interpret the interaction between other types of games (e.g.

multiplayer role-playing games) and student population of different characteristics." [48]

2. **Study II: Applying a Conceptual Mini Game for Supporting Simple Mathematical Calculation Skills: Students' Perceptions and Considerations by Chris T. Panagiotakopoulos** [70]

   - "In this study, an educational game addressing mathematics was designed, developed and evaluated by a sample of 33 students of the fifth grade of a primary school. Each student played the educational game "Playing with Numbers" (PwN), performing additions with integers, additions with decimals and multiplications with integers for a total of one hour, divided into four sessions." [70]
   - One of the first fields where educational games were applied was **mathematics education**. [70]
   - "In the field of **mathematics**, many researchers propose that digital games are the perfect tools, with great motivational appeal for improving mathematical learning (Conati & Zhao, 2008; Lee & Cheng, 2009; Eliens & Ruttkay, 2009a). Ke (2006) also notes that it seems, from a few empirical studies, that digital games can be **effective tools** for understanding **arithmetical concepts** or for **problem solving**." [70]
   - "Kebritchi (2008) and Kebritchi, Hirumi and Bai (2010) indicated a significant improvement in the mathematical skills of participants playing a specific mathematical digital game, compared to those who did not play the computer game. This finding is supported by a number of reasons, which,
     - according to teachers, are:
       * the experiential nature of the learning,
       * the alternative method of teaching and learning,
       * the students dealing with mathematics in order to progress in the game (score),
       * the transformation of mathematics phobia to a projection of the relationship between mathematics and real life, and
       * the students feeling that they were not working with mathematics concepts.
     - According to the students, the reasons for the improvement in their mathematical skills were:
       * the combination of learning and entertainment,
       * the adventurous and exploratory context in which they played with mathematics, and

    * the challenge that existed in the game's environment." [70]

3. **Study III: Getting Serious about Math: Serious Game Design Framework & an Example of a Math Educational Game by Layla Husain** [38]

    - "In both instruction and gaming, it is important for the player/learner to have access to **assistance** if they are stuck (Wang et al., 2010; Charsky, 2010).
      **Assistance** can come from peers, instructors, virtual characters, **AI calibration of difficulty level**, **hints from the system** and many other sources.
      It is very common for gamers to begin playing without ever reverting to the instruction manual. Through **trial-and-error**, and consulting others on gaming forums and YouTube, they are able to get the help they need to proceed (Meegen & Limpens, 2010).
      **Gamers often only seek assistance when they need it**.
      One way for the system to **help the learner** is that for players that need more time to carry out a task, they could be allowed to stay at a certain level until they are prepared to advance.
      Instruction should be dispersed throughout the game and come '**on demand**' and '**just in time**' (Gee, 2003; 2004), or what gamification calls '**cascading information theory**'. The instructions in many games are often cleverly disguised as part of the game, so as not to bore players with a tutorial (Thomas & Young, 2010)." [38]
    - "Feedback is essential in SG, especially because the role of the learner and player are fused (Bente & Breuer, 2009)." [38]
    - "When the system responds to the player's actions and input, the player will better understand how the system works and feel consequential in the virtual world. Backlund et al. (2008) noted that in their virtual driving game, feeling 'less in control' of the car led to a decrease in self-efficacy. Therefore, the design of an **appropriate feedback system is vital for learning and having players continue playing**. **Statistics** can also inform the designer how to further development the game by **tracking** where the gamers start to **lose interest**, or **which segments are too difficult**. [38]
    - One of the key advantages of using computers is that they can keep track of students activities and provide immediate feedback about their progress, their understanding of the rules and for encouragement (Scarlatos, 2009)." [38]

- "In order to maintain the motivation to collect points or tokens, they can be associated with bonus material. There are many ways to implement this. The best way to decide is to conduct some **field studies** with the end users." [38]
- "In order to sustain motivation, **feedback** needs to be aligned with **reward** mechanisms in games. Rewards are fundamental to game design as well (gamification.org). Moreover, 'having the right reward is key to making sure players feel there is value to their actions' (gamification.org)." [38]
- "As important and helpful as feedback can be, we should be careful not to overwhelm the user with information, as it can be annoying and condescending, especially for advanced learners (Andre et al., 2009; Scarlatos, 2005). Magnus Haake remarked that when testing their game with children, they would sometimes get annoyed if there were too many comments made by the game that interrupted the game flow." [38]
- "Providing challenge in a game is another feature coveted by gamers that is also important in learning." [38]
- ""Moderate levels of complexity create intermediate levels of cortical arousal, which is both optimally pleasing to most interpreters and maximally efficient for learning in most instances" (Bryant & Fondren, 2009, p109)." [38]
- "Keeping the learner adequately challenged by staying within the proximal zone of development (ZPD) is essential for learning, according to the Vygotsky approach. Making a lesson too difficult frustrates learners, but making it too easy bores them. The ideal balance is when students are given a task at one level above their ability. Finding the right level is difficult in pre-scripted games, which are most common. Ideally, the program would model the user's level and adapt the game scenario and narrative, using artificial intelligence." [38]
- "'The biggest issue limiting SEGs is the lack of good artificial intelligence to generate good and believable conversations and interactions' (Annetta, 2010, p105)." [38]
- "Incorporating adaptive narrative and level adjustment is not only pedagogically beneficial (Lieberman, 2009; Wang et al., 2010), but also contributes to flow and intrinsic motivation (Pierce et al., 2008). Adaptive instruction keeps the material challenging, but not too difficult (Lieberman, 2009)." [38]

- "According to Squire (2008), there are no educational games that use realtime data, or that provide such adaptive content." [38]
- "Pierce et al. (2008) describe a way of non-invasively adapting a game using the ALIGN (Adaptive Learning in Games through Noninvasion) system." [38]
- "Part of the challenge of adapting game play is to find indicators in player behaviour that determine the level of the player and problem areas that need more practice." [38]
- "A well-designed game fosters the learning process by providing a safe environment where new skills can be tried out and practiced and where the virtual world's parameters can be tinkered with, as in Gee's 'sandbox' (Gee 2003; Lieberman, 2009)" [38]
- "There have been several studies that propose using SG's as **assessment tools** (Shute et al., 2009; Ramani & Sirigiri, 2008; Annetta, 2010)." [38]
- "**Stealth Assessment**, or **embedding assessment into game play** (Shute et al., 2009), is important to maintain **flow** because players are tested for educational content without even realising it." [38]
- "Serious games, done well, incorporate **challenge**, **fantasy**, **curiosity**, and **control**, which Malone & Lepper (1987) believe are common to all intrinsically motivating environments. Together, these characteristics induce **flow**, where the player is so engrossed in the activity that there is no sense of time or any distractions, only focus on the task at hand (Shute et al., 2009)." [38]
- ""Since video games are "action-and-goal-directed preparations for, and simulations of, embodied experience" they situate the target content" (Gee, 2008b, p18).
  Gee (2008b) adds that players are exercising their learning muscles, though often without knowing it and without having to pay overt attention to the matter.
  Charsky (2010) notes that while edutainment teaches lower order thinking skills, facts, concepts, and procedures, serious games facilitate higher order thinking skills, such as how to apply their knowledge, analyze their understanding, and evaluate their learning.
  Annetta (2010) describes how video games exploit both verbal and visual information, each processed by different cognitive subregions. "Words are processed only in the verbal region, whereas images are processed in both regions.., the other region is the visual region allowing for greater

| ... | STIMULUS PRESENTED | STIMULUS WITHDRAWN |
|---|---|---|
| Pleasant stimulus | Positive reinforcement | Negative punishment (Omission) |
| Unpleasant stimulus | Positive punishment | Negative reinforcement |

Table 5: Literature Review: Studies on the impact of Adaptivity

depth of processing and increased availability of multiple retrieval cues" (Annetta, 2010, p111)." [38]

- "Backlund et al. (2008) recommend **adapting the challenge**, or **difficulty levels** and providing appropriate semantic feedback, comprising of both instructional and performance feedback screens." [38]
- "Motivation can lead students to make greater effort, seek greater challenges, and attain higher achievement (Schunk et al., 2007)." [38]
- "One way to increase motivation comes from having goals. There are two types of goals: mastery goals and performance goals, described by Baron & Harackiewicz (2000). Those pursuing performance goals, however, "engaged in more superficial or strategic learning strategies, chose easier tasks, and withdrew effort when difficulty was encountered" (Baron & Harackiewicz, 2000, p232)." [38]
- "An important condition for learning is the **affective state of the learner**. There are many who appreciate the role of **emotional attachment** to learning and memory (Rosenfield, 1988; Caine & Caine, 1991).
  Gee (2009), borrowing from Damasio, shows that there is deeper learning when there is an emotional attachment to their learning and problem solving, and when something is at stake for the learner personally (Gee, 2009a; 2009b)." [38]
- "Self-efficacy can greatly impact how students perceive and enjoy **math**. Stodolosky et al. (1991) surveyed 60 students to find that the main negative feeling associated with math is the **fear of failure** because it was **difficult**, leading to **frustration** and **anxiety**." [38]

## 5.5 STUDIES ON ADAPTIVE SERIOUS GAMES

- WEMAKEWORDS AN ADAPTIVE AND COLLABORATIVE SERIOUS GAME FOR LITERACY Ismailović et al. [44]

*Damir, Dennis & Brügge*

*Leverage TM; www.pr-sol.com 2011*

- Study I: Moving to the Next Level: Designing Embedded Assessments into Educational Games by Jody S. Underwood. Underwood et al. [93]
  - Pragmatic Solutions (aka Pragmatic; www.pr-sol.com) has developed a suite of tools (LeverageTM) that facilitate the implementation of embedded assessment, analysis and reporting, and adaptive gaming environments (Underwood, Kruse, & Jakl, 2009).
  - Leverage has been developed to support America's Army. With each new release, Leverage scaled to support more and more in-game actions and more players. With the release of America's Army 3 (AA3), Leverage now fosters learning by filtering a player's in-game actions through a complex set of rules related to the Army's "core values" (loyalty, duty, respect, selfless service, honor, integrity, personal courage), with specific actions serving as a measure of a player's understanding and demonstration of core values, which in turn impacts future actions in the game, providing **implicit feedback as authentically as possible**.
  - Pragmatic is partnering with educational institutions (e.g., NYU) to integrate Leverage into educational games, continuing to discover new insights into collaborative learning and how people make decisions and acquire knowledge.
  - I would describe the technological capabilities of Leverage and how it supports our assessment methodology and addresses the challenges of embedded assessment.
  - Done in real-time through Pragmatic's infrastructure, in-game data is collected to update learner profiles without impeding the flow of the environment (Csikszentmihalyi, 1990), a requirement that is critical to fostering both engagement and the potential for learning (Shute, et al., 2008).
  - Leverage supports the dynamic delivery and modi- fication of any and all elements in the game, with the capability of providing individualized content to a large user base in near real-time. Elements dynamically delivered into the environment may include static or scrolling textual content, images, audio, video, or other in-game objects, as defined by the pedagogical framework of the game.

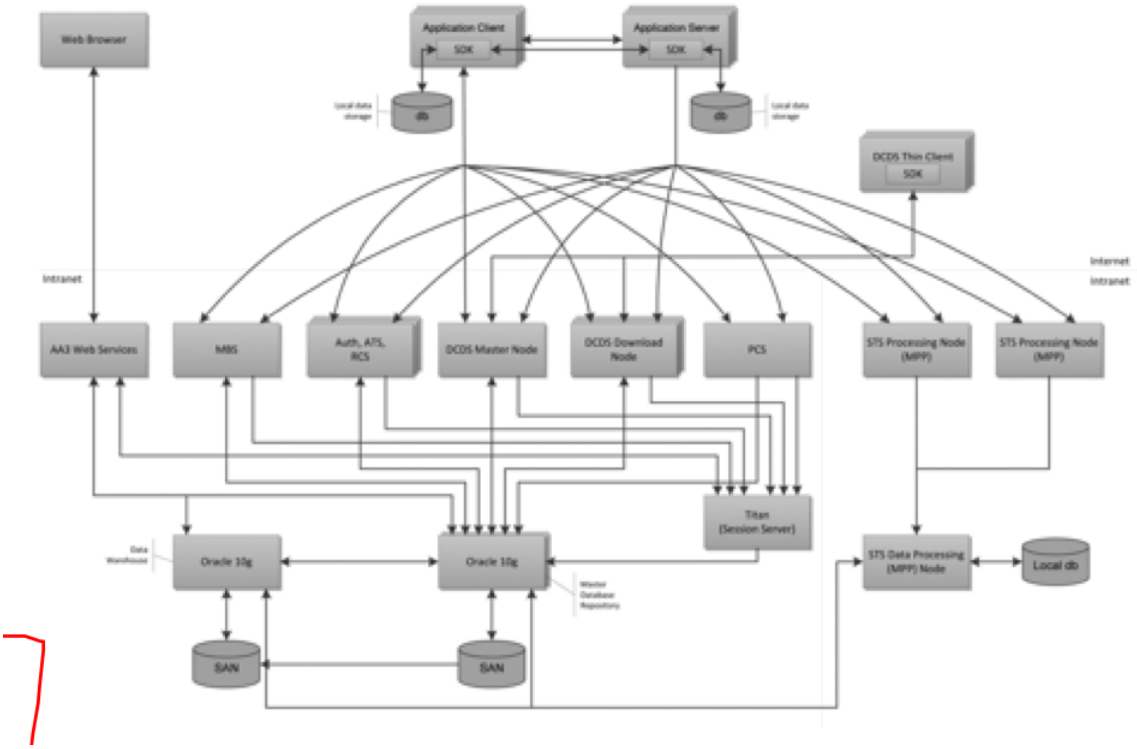



Figure 11: LeverageTM

– Leverage includes an **attribute tracking system (ATS)** to collect and organize data dynamically, a **rule-based statistical tracking system (STS)** to summarize acquired data, a **dynamic content delivery system (DCDS)**, a behavioral inference engine (Inflection), and an **interactive data visualization application (Visual IQTM)**.

  * **ATS**, the heart of this database-driven application, collects data of any grain-size in real time from embedded assessments and manages the results to adapt the game to an individual learner's needs.
  * **STS** tracks in-game events and is highly scalable using massive parallel processing methodology.
  * **STS** summaries, which include such things as correlation, regression, and factor analysis, are automatically fed into Visual IQ.
  * **DCDS** delivers individualized content to any game.
  * **Inflection** uses genetic algorithms and other data mining techniques to construct inferences and make predictions about learner behavior and performance.
  * **Visual IQ** provides data visualization and real-time analytic support tools, allowing **stakeholders** to analyze

data of a learner population, drill-down into individual learner profiles, and define custom groups.
**Visual IQ** provides various interactive charting and graphing tools, custom reporting options, and input of external events (e.g., completion of a course unit, a holiday) against which to track data.

– Leverage collects information about what learners do in the game, processes that information, and returns a best course of action back to the learner. As more data is collected, the system continues to be refined to improve inference and adaptivity.

*ASSISTment 2010*

- A Quasi-Experimental Evaluation of an On-line Formative Assessment and Tutoring System by Koedinger, K., McLaughlin, E. & Heffernan, N. Koedinger et al. [53]

  – ASSISTment is a **web-based math tutor** designed to address the need for timely student assessment while simultaneously providing instruction, thereby avoiding lost instruction time that typically occurs during assessment.
  This paper presents a **quasi-experiment** that evaluates whether ASSISTment use has an effect on improving middle school students' year-end test scores.
  The data was collected from 1240 seventh graders in three treatment schools and one comparison school. Posttest (7th grade year-end test) results indicate, after adjusting for the pretest (6th grade year-end test), that students in the treatment schools significantly outperformed students in the comparison school and the difference was especially present for special education students.
  A usage analysis reveals that greater student use of ASSISTments is associated with greater learning consistent with the hypothesis that it is useful as a tutoring system. We also found evidence consistent with the hypothesis that teachers adapt their whole class instruction based on overall student performance in ASSISTments.
  Namely, increased teacher use (i.e., having more students use the system more often) is associated with greater learning among students with little or no use suggesting that those students may have benefited from teachers adapting their whole-class instruction based on what they learned from ASSISTment use reports.
  These results indicate potential for using technology to provide students instruction during assessment and to give teachers fast and continuous feedback on student progress.

| ... | STIMULUS PRESENTED | STIMULUS WITHDRAWN |
|---|---|---|
| Pleasant stimulus | Positive reinforcement | Negative punishment (Omission) |
| Unpleasant stimulus | Positive punishment | Negative reinforcement |

Table 6: Literature Review: Studies on the impact of Adaptivity

## 5.6 STUDIES ON THE IMPACT OF ADAPTIVITY: ZPD

The difficulty of integrating gaming and learning in an engaging and educationally meaningful way has been recognized by several research and commercial projects, such as ELEKTRA (Kickmeier-Rust et al. 2006), 80 Days and the commercial "Global Conflicts" computer game series (http://www.globalconflicts.eu).

- (2002) Why Scaffolding Should sometimes Make Tasks More Difficult for Learners Reiser [81]
- (2002) Toward Measuring and Maintaining the Zone of Proximal Development in Adaptive Instructional Systems Murray and Arroyo [67]
- Using Software Scaffolding to increase Metacognitive skills amongst young learners Luckin et al. [60]
- (2008) Toward a theory of how young children learn to read in the ZPD: Implications for research and practice Wiles [99]
- (2010) Promoting Links and Developing Students Criteria for Visualizations by Prompting Judgments of Fidelity Chiu et al. [15]
- (2010) SURGE: Integrating Vygotskys Spontaneous and Instructed Concepts in a Digital Game? Research Goals and Theoretical Framework Clark et al. [16]
- (1997) Individual and Collective Activities in Educational Computer Game Playing Kaptelinin and Cole [46]
- Literary Creativity in School Age Children Vygotsky [97]
- (2004) Collaborative Scaffolding in Synchronous Environment: Congruity and Antagonism of TutorStudent Facilitation Acts Pata et al. [72]

- (2003) Assisted Performance and the Zone of Proximal Development ZPD; a Potential Framework for Providing Surgical Education C and Stacey [11]
- (2009) Bringing Tabletop Technologies to Kindergarten Children Marco et al. [61]

## 5.7 RECOMMENDATIONS FOR OUR CASE STUDY

- Refer to Christina´s Thesis "Evaluation of a serious game" for the first Evaluation Framework already designed. Kolb [54]
- Designing a performance augment has to start with the cognitive skill that is being executed and how it needs support. [76]
- "Focus more on making the interaction fun or useful—not on the learning aspects. In our community games, people are thinking, learning, experimenting, but we don't point to learning as an outcome at all. Learning takes place informally in all of the games." [76]
- Provide a balance between teacher-assigned and student-selected tasks.
- Each child is unique.
- As a future implication of my literature research, perhaps, the criteria of evaluating Educational Games can be utilised as a guideline for designing developmentally appropriate software for young children.

Part IV

# PILOT STUDY: BUG GAME ADAPTATION

Presenting the intaeractive game Emil & Pauline. A quite innovative initially non-adaptive game on the iPad. For one of the mini games, Bug Game, an Evaluation Framework have been designed to measure and discuss the dynamic amout of Adaptivity needed to make the game playable and so the learning experience more attractive.
Should it be a Quasi-Experimental or only Qualitative?
An interaction between Experts (Pedagogues, Psychologues and Gaming Industry Experts), Teachers, Gamers and Serious Games Developers.

# 6 INNOVATION

## 6.1 CONTEXT OF THE STUDY

Relying on these two statements *"Software can be made age appropriate even for children as young as three or four (NAEYC 1996)." [96]* & *"Digital devices are the perfect augment. Digital devices can be programmed to "say" or do anything you want and can execute reliably and repeatedly. They will remember arbitrary facts, and respond appropriately to the input you give them every time. However, they are not particularly good at making decisions or detecting the nuances of a conversation." [76]*, we are going to develop a game aimed at small culturally specific user populations respecting "the state of the art guidelines" to designing serious games.

### 6.1.1 *Social*

*Preschool children (4 to 8 years old)*

When designing software for young children, we should focus on a specific age group and try to establish a user profile of typical preschool gamer including their sociocognitive development, development of play and the cultural aspects of development, because children of different ages have vastly different preferences, development factors and levels of skills.

Although the most research has been done on games for older children and no guidelines exist, that distinguish between different age groups, what makes our task as developers and reserchers more complex, we focus on the 4 to 8 age group for the following reasons:

- Each child is unique and has an individual growth and development pattern, personality, temperament, learning style and background. The child's age can therefore be regarded as only a rough index of developmental status.
- At this age, children start their schooling experience and appropriate products can enhance school readiness and support the acquisition of cognitive skills such as reading, writing and story construction.
- Fine motor skills and hand-eye coordination are, at this age, adequately developed to use digital devices.

*Fostering emotional literacy in young children is the next complex task*

When observing children, we may encounter difficulty, trying to understand children´s emotions.

There are many factors that impact our ability to understand children´s emotions like the body language, the tone of voice, the facial expressions and the psychological responses (laughing, ...).

The ability of labeling emotions consists in identifying (Which emotions do the following faces express?), understanding (How do you know?) and responding to them in a healthy manner and this is hard to manage because:

- all emotions are valid; it is how we interpret them that counts,
- emotions may change through the experiment,
- children can have more than one emotion about the game state, and
- we can feel differently than someone else about the same emotion.

We could identify 6 principle states of emotions in young age going from:

- happy
- sad
- embarrassed
- scared
- nervous
- mad

and for each state we could identify the corresponding facial expressions and the clues that experience that:

- Happy - smiling
- Sad - frowning
- Embarrassed - cheeks, lips, eyes rolled
- Scared - mouth opened, eyes wide
- Nervous - sweating, teeth, eyebrows
- Mad - tight mouth, eyebrows

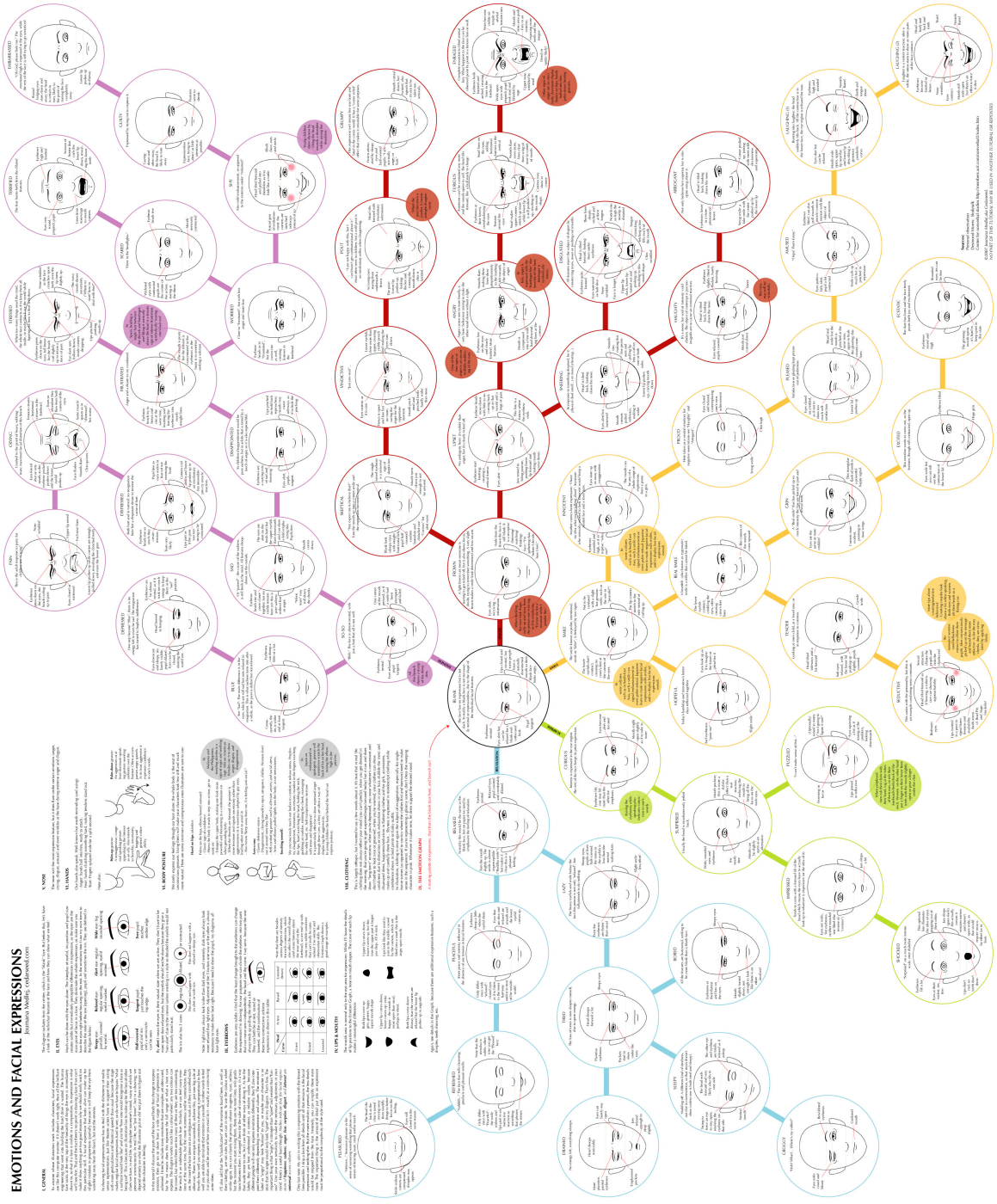



Figure 12: Emotions & Facial Expressions

### 6.1.2 *Technical*

*Hardware & Software used; Tablet Touch iPad & mobile learning with adaptivity*

In the last few years digital educational games have gained attention as a tool for facilitating learning in different sectors of society including but not limited to military, health, and education.

A number of factors have made Digital Educational Games´ attractive learning tools. The advancement of technology has made it possible to play games on simple platforms such as mobile devices, provided children with opportunities for engaging in activities associated with computers and since the use of new technological and learning methods has prooved an improvement of the learning process, a wide range of software has been designed with young children in mind.

With an iPad application, we're reaching a specific user demographic, but if the demographic makes sense, this is an attractive platform because of its mobility, crisp graphics, gestural interface, and other unique interface elements. The iPad has quickly become a prime platform for advergames and persuasive games, and more robust serious games and simulations are being developed for it. The gestural interface and accelerometer open the door for innovative new applications. Because of its handheld nature, certain serious games and simulations can seem even more immersive.

## 6.2 GAME DEVELOPMENT

In DANCE Research Group, we're responsible for designing serious educational iPad games for children. Some of these games are categorized to join the "Emil & Pauline" series of games. Emil and Pauline are two actors present in these games and they're responsible for children guidance and motivation.

The games were designed to reinforce academic standards for mathematics and target a variety of math skills, such as recognition of numbers, the right way to write them, making addition/substraction until 10, 15 or 20, geometry and so on.

This game consists of 5 mini games, an intro game () and an award game (Ufos). These 5 mini games are the "Rope Game", the "Bug Game", the "Math Labyrinth Game", the "Torch-Light Game" and the "Magnet Game".

There are also two main parts that are designed in parallel to these mini games. The first part is a game framework that is responsible for putting all the five mini games together with the intro and the bonus games to represent the whole big game. The second part, and

the most important, is the adaptivity model for the big game. This model is responsible for detecting the player's performance in each mini game and accordingly increases or decreases the game level.

Our game would mediate communication through the main characters in the game, support activity since the child should interact with the iPad and encourage productive learning through mathematical tasks.

For the further of the Study, we concentrate us on the Bug Game with the **most manipulable variables**.

First, the non adaptive game had Gamelog and score keeping and employed progressive difficulty levels – harder problems were presented at the higher levels of the game.

### 6.2.1 *Bug Game Scenarios & Requirements*

[27]The "Bug Game" is a serious game used for learning. It should teach the children counting numbers from 1 to 6. The game has two kinds of bugs, good and bad bugs. Each bug is carrying a number of dots on its back. The two actors, Emil and Pauline should be holding one or two numbers which represent the number of dots carried by the good bugs. The main goal of the game is that the child has to recognize the numbers held by the actors and count the number of dots carried by each bug to know whether it is a good or a bad bug. When the child differentiates between good and bad bugs, he should be able to save as many good bugs as he can by sending them through an exit cave before being eaten by the bad bugs. The child wins the game if he succeeds to save more than half of the good bugs.

The "Bug Game" has to be developed with a lot of features and behaviors. Examples of these features are the bugs movement and animations as well as their behavior when they meet each other. Moreover, the game should have many adaptable elements that can be changed easily. Consequently, by integrating it with the adaptivity model, the game should change its behavior and difficulty according to the child's performance.

The setting is a cave. Fireflies enter the cave through a hole in one corner. Each firefly has a certain number (between one and six) of glowing-dots on its back. In the other corner is an exit. At the bottom left and right corner the two main characters, Emil and Pauline, are positioned each holding up a sign showing a different digit between 1 and 6. Only fireflies which have the same number of dots on their

back as the numbers held by Emil and Pauline are allowed to leave the cave. The child's task is to help the 'good' flies to leave the cave by 'throwing' them towards the exit and prevent the 'bad' flies from leaving by squishing them. If 'bad' flies hit 'good' flies they eat them; if they hit each other they multiply. Flies that hit the border of the screen bounce off. The flying speed of the individual fly depends on the number on its back – flies with a higher number are faster. If a 'bad' fly manages to leave the cave, the game ends. A score about the number of 'good' flies that left the cave is kept. 'Good' flies that leave the cave by chance are counted as well. Thirty 'good' flies enter the cave and when they all have left the cave or been eaten the game ends.

1. Scenarios:

    a) Scenario A: Saving a Good Bug
    A child plays the bug game. He observes Emil and Pauline on both sides of the screen. Emil holds a card with the number 4 on it which means that each bug with the number 4 is a good bug. The child sees a bug with the number 4, he drags it or sends it to the cave and saves it.

    b) Scenario B: Interacting with a Bad Bug
    A bug with a number different than the ones held by Emil and Pauline means that its a bad bug. The child hits the bug, the bug is splashed and killed. He drags a bad bug to the cave and saves it or he fails to drag it far from the cave, the game is over and the child loses.

    c) Scenario C: Behavior of Bad Bugs
    When a bad bug comes near a good one and the child fails to move them away from each other, they collide, the bad bug opens its mouth and eats the good bug. If two bad bugs come near each other and the child fails to move them away from each other, they collide, the 2 bugs duplicate and become 4 bugs.

    d) Scenario D: Winning and Losing the Game
    When the child saves all the good bugs or more than half of them, he wins the game. When he saves only one bad bug or fails to save more than half the good bugs, the game is over and he loses.

2. Functional Requirements:

    a) Good Bugs and Bad Bugs
    This game consists of two kinds of bugs, good and bad bugs. A bug should hold a certain number of dots which are responsible for the teaching process. Bugs in this game should be moving through all parts of the screen, rotating according to the direction of motion and should be animated in different game situations. A specific number of

bugs should enter the screen with a pre-determined rate. A player should be able to hit a bug or drag any bug through any place on the screen. He should be able to save good or bad bugs by letting them exit the cave and he can kill bad bugs by hitting them by his finger. When a good bug and a bad bug meet each other, the bad bug should eat the good bug. When two bad bugs meet each other, they should duplicate to become 4 bugs.

b) Multitouch
The game should be a multitouch game so that more than one player can play together at the same time or a player can play with both his hands.

c) Adaptability
This game should be adaptable (not adaptive). The game should provide a sort of different easy and advanced behaviors to what is called the adaptive model (not a part of this thesis). This model is to decide which behavior should be changed, and when according to the player's performance during the game.

d) Emil & Pauline
Two main actors sitting on both sides of the screen, Emil and Pauline. One of them or both should be holding the number of dots that are carried by the good bugs. They should be animated in different game situations (animations are enabled but still the needed images are not provided). They should be talking to the player, guiding and motivating him during the game.

e) Winning and Losing
A score of the game, which is equal to the number of saved good bugs, should be displayed for the player on the screen. The game should tell the player one of 4 situations that end the game. First, when one bad bug exits the cave, the game is over and the player loses. Second, when all good bugs are saved, the player wins the game with full score. The third situation is that when no more good bugs are there and the score is less than half the full score, the player loses the game. Finally, when no more good bugs are there and the score is more than or equal half the full score, the player wins the game.

3. Nonfunctional Requirements:

a) Usability
Usability is the ease with which a user can interact and operate with the system. Since this game is for children, it should be more usable to allow the child to play understand

its purpose and play it easily. Moreover, usability should be measured by the amount of time the player should wait to interact with the system and observe the result of this interaction. For example, killing a bug shouldn't take more than 1 second.

b) Reliability
Reliability is an important factor of software quality. This game has to be reliable meaning that it should run continuously for 1 hour or until the game is over without crashing.

c) Performance
A major factor in determining the overall productivity of a system, performance is primarily tied to availability, throughput and response time. This game should have a good performance. During the game, there are many good and bad bugs on the screen and many others entering the screen. The game should handle the motion and the behavior of all of them. Also, the game should guarantee the player a quick response for his interaction with it. He should only wait for 2-3 seconds maximum in some cases depending on the animations running before the actual result. This is a similar idea to usability mentioned in subsection 3.4.1 but from the system perspective rather than the player's perspective. Morever, the performance can be measured with the frame-rate. So the game should have a minimal frame-rate about 30fps in some cases, and average frame-rate should be between 40 and 60 frames per second (fps).

d) Supportability & Extendability
Supportability is all the actions related to the inherent quality of the system. To deal with fast changing development requirements, the system has to be extensible. This extensibility could be considered as a part of supportability. The game should be able to easily integrate current and future technologies without major modifications of the existing system. The game has a lot of elements to be adapted and can be used by any other system. An external adaptivity model can be easily connected and integrated with the game such that the game behavior is changed depending on the player's performance.

e) Implementation Requirements
Implementation requirements are constraints on the implementation of our target system which specify tools and hardware platforms to be used. This game is designed only for iPad and cannot be run on any other iOS device. Also, it is restricted that the development of this game is done us-

| SKILLS |
| --- |
| Learning Content Skills<br>• Recognizing {(dots in order), (dots without order)}<br>• Counting {(dots in order), (dots without order)} |
| Game Aspect Skills<br>• Squashing bad bugs {(to limitate collisions & duplications), (to avoid bad bugs exit)}<br>• Saving good bugs {(by moving toward exit)} |

Table 7: Learning Content vs. Game Aspect Skills in Bug Game

ing Cocos2d framework which is one of the most important and popular frameworks for iOS development.

### 6.2.2 *Game Process*

First the developed game is non-adaptive;

we appreciated the value of the usability testing protocols in the design process and admit that no set of guidelines alone will guarantee design success.

In this Section, i would introduce the adaptive variables in Bug Game; differentiate between GameAspect variables and LearningContentAspect variables; further we would determine the correlation between these variables and towards children´s attitudes.

To successfully play the "Firefly Getaway" Learning Content skills as well as Game Aspect Skills should be defined in Table 7 on page 113.

A variety of Events can take place during the process of the game as declared in Table 8 on page 114.

The Task Events can have four different states gathered through the parameters of success and correctness. Besides monitoring the percentage of successfully/unsuccessfully, correctly/ incorrectly performed tasks, the percentage of performing one compared to the other is measured, as well as the time taken between the actions. In this

| EVENTS |
| --- |
| Task Event |
| • Squashing bad bugs |
| • Saving good bugs |
| Game Event |
| • Good or bad bug escaped |
| • Good or bad bug squashed |
| • Bad bug doubled |
| • Good bug eaten |
| • Good or bad Bug entering |

Table 8: Events in Bug Game

game it might be additionally interesting to monitor which bugs the actions are being performed with: whether the child is reacting equally to each number or for example prefers simple numbers like one or two; or whether a child is only reacting to numbers that have a bigger difference to each other (e.g. if the good bugs are those with three dots on their back: Is he/she only squashing bad bugs with one or six dots, because they are easier to distinguish? That might indicate that he/she is using recognition rather than counting.) But that again goes beyond the scope of this thesis. The Game Events correlate with the Task Events. Some are directly induced by a Task Event, others are consequential. A high amount of good bugs eaten and bad bugs escaped by chance will show a negative correlation with the number of bugs squashed. A high amount of good bugs escaped by chance will show a positive correlation.

## 6.3 ADAPTIVITY MODEL

### 6.3.1 *Adaptive Approach*

[54] The game structure offers the option to include adaptivity on different levels. The use of adaptivity on the macro level was skipped in order to include a higher level of control [40]. This hides the linear structure of the game and gives the player the feeling that his/her

| ADAPTABLE VARIABLES | ADAPTABLE PROPERTIES |
|---|---|
| Number Sign | {1, 2} |
| Bugs | {Art, Speed, Color} |
| Entrance of bugs | {Time Intervall, Percentage of Good/Bad} |
| Digits on bugs | {Dice-like, Random} |
| Encounter of bad bugs | Duplicate {yes, no} |
| Escape of bad bugs | Exit {yes, no} |

Table 9: Adaptable variables & their properties in Bug Game

actions have an impact on the game world. It creates the feeling of playing the game the way he/she wants, with a sense of agency. This increases the flow experience (Sweetser & Wyeth, 2005). The idea of adaptivity through Aptitude-Treatment interaction approach by dynamically generating the feedback given by the main characters in the game (Emil and Pauline) would be a fairly time- and cost-intensive process. The necessity of a complex feedback depends strongly on the task. Tasks which require a complex knowledge and understanding of the subject require detailed feedback while tasks which primarily serve for repetitive training – when failure happens mostly due to a lack of practice – do not require such detailed feedback. Since the main focus of the game lies in practicing skills learned at school rather than teaching new content, the need for complex feedback is relatively low and therefore was skipped for the time being. The same applies to the Constructivistic-Collaborative adaptive approach. If practical use of the game shows the need for it, adaptive feedback can relatively easily be added retrospectively. [40]

The current implementation of the game focuses on the implementation of adaptivity on the micro level. On the level of the particular task (the mini-game) the difficulty is adjusted to the progress the child makes while playing. Additionally game data is stored and used for the setup of the next mini-game. Dynamic difficulty adjustment can be achieved through two different approaches:

- Game Aspects: Game Aspects include all aspects which are not directly linked with the learning content itself but rather with attributes mainly influencing the gaming experience: parameters such as speed or number and size of the elements presented.
- Learning Content Aspects: Learning Content Aspects are the actual learning objectives, the exercises presented, such as how many tokens to count, or what numbers to calculate. The concept of easier and harder Learning Contents and the influence of their

order of presentation is based on the Knowledge Space Theory. [40]

The Knowledge Space Theory first introduced by Doignon and Falmagne (1985) is a structural approach towards human knowledge gaining. Knowledge Spaces are based on competencies and are somehow connected with each other and dependent on each other. The theory categorizes problems, learning objects and skills. [44] Learning objects are assigned to a set of skills which are relevant for solving the problems. Skill functions can be formulated based on the subset of skills sufficient for a particular problem. Some skills are prerequisites for others. This leads to a structure of competence, where each skill is assigned to a collection of subsets of skills. Acquiring one skill requires mastering of all skills of at least one of those 'underlying' subsets of skills. [3] To sort the Learning Content by difficulty the Bavarian first grade curriculum was used as a point of reference (Staatsinstitut für Schulqualität und Bildungsforschung München, 2000).

### 6.3.2 *Model*

*DIAGRAMS MISSING WITH DESCRIPTION TEXT*

- Reasoning engine?
- Modeling Adaptivity (Events, Tasks, Skills, LCA, GA)
- Adaptive Events in the Bug Game
  - TaskEvents (+): Task finished successfull/correct
    - (+): Good Bug saved
    - (+): Bad Bug destroyed
  - (+): Task finished unsuccessfull/incorrect
    - (-): Good Bug killed (tried to kill)
    - (-): Bad Bug saved GameEvents
    - (+): Good Bug saved
    - (+): Bad Bug destroyed
    - (-): Good Bug killed (tried to kill)
    - (-): Bad Bug saved
  - Consequential Events
    - (+): Good Bug went out
    - (-): Bad Bug went out
    - (-): Bad Bug doubled
    - (-) : Good Bug eaten by bad bug
  - GameAspectAdaptEvents

    * (-/+) decreasing/ increasing of the speed of the fireflies (GA)
    * (-/+) decreasing/ increasing the size of the next appearing firefly (GA)
- LearningContentAspectAdaptEvents
    * choice of the number of dots on the next appearing firefly (LCA)
    * Bigger difference between 'good' and 'bad' flies (LCA)
    * More 'good' or more 'bad' flies (LCA)
    * change of order of dots on the next appearing firefly (LCA)

# 7

# EVALUATION STUDY: BUG GAME

Since we would evaluate a game prototype, it is unrealistic to aim to determine the learning effectiveness through the game. And since we are at a phase, when we would change the game from non-adaptive to adaptive & individual in order to increase the learning experience and to make the game events more challenging.

We would focus more specific on the effect of micro adaptive interventions such as cognitive hints related to the skill assessment of the learner and motivational hints related to the motivational assessment of the learner.

- Guided by information gathered in part 3 of the literature study, we will test the validity of the guidelines formulated in phase 2 with a series of experiments. These will primarily involve unstructured interviews and observation of children using appropriate software packages. Video and audio recordings will support data gathered through observation. Each session will focus on a subset of the guidelines and reactions to design elements related to these guidelines will be observed and recorded. We will organise the information according to the guidelines it relates to. Evidence relating to each guideline will be analysed to decide whether it should be accepted, rejected or adjusted.
- Impact of the dynamic Scaffolding on Children.
- Immediate and long term impact
- Study with psychological and pedagogical assessment
- The present study emphasize the application of digital games in preschool education by evaluating the potentials of games in facilitating the learning of math concepts and skills that are required by core curriculum content standards.
- Why have I choosed the Qualitative or the Quantative Approach?
  - Thus, whether we should emphasize qualitative or quantitative research methods may depend on the conditions and purposes of our inquiry.
    Qualitative methods may be more suitable when flexibility is required to study a new phenomenon about which we know very little, or when we seek to gain insight into the subjective meanings of complex phenomena to advance our conceptualization of them and build theory that can be

tested in future studies.
Qualitative research thus can sometimes pave the way for quantitative studies of the same subject. Other times, qualitative methods produce results that are sufficient in themselves.
In sum, you do not need to choose one camp or the other. Each approach is useful and legitimate. Each makes its unique contribution to inquiry. Each has its own advantages and disadvantages. Each is a set of tools, not an ideology. Researchers need to match the tools they use with the research questions and conditions they face using quantitative methods for some studies, qualitative methods for others, and both methods in combination for still others. Rubin and Babbie [84]

- Aims: Deeper understandings, Describing contexts, Generating hypotheses, Discovery, ...
- Structure: Flexible procedures evolve as data are gathered, ...
- Setting for data gathering: Natural environment of Research participants, ...
- Theoretical approach most commonly employed: inductive
- Sample size likely or preferred: Smaller
- Most likely timing in investigating phenomena: early, to gain familiarity with phenomenon, more time-consuming, ...
- Subjective
- Nature of data emphasized: Words
- Depth and generalizability of findings: Deeper but less generalizable
- Richness of detail and context: Rich descriptions with more contextual detail
- Nature of data-gathering methods emphasized: Lengthier and less structured observations and interviews, Open-ended items and interviews with probes
- Case studies, Focus groups
- Data Analysis Process: Search for patterns and meanings in narratives, not numbers

- Evaluation
  A fourth purpose of social work research is to evaluate social policies, programs, and interventions. The evaluative purpose of social work research actually encompasses all three of the preceding purposes:

  - exploration,
  - description, and
  - explanation.

- For example, we might conduct open-ended exploratory interviews with community residents as a first step toward evaluating what services they need. We might conduct a descriptive community survey to evaluate the problems residents report having and the services they say they need. A descriptive study might also evaluate whether services are being implemented as intended. We might conduct an explanatory analysis to evaluate whether factors such as ethnicity or acculturation explain why some residents are more likely than others to utilize services.

- Evaluative studies also might ask whether social policies, programs, or services are effective in achieving their stated goals. Evaluations of goal achievement can be done in an exploratory, descriptive, or explanatory way. For example, if we simply ask practitioners in an open-ended fashion to recall techniques they have employed that seemed to be the most or least effective in achieving treatment goals, we would be conducting an exploratory evaluation to generate tentative insights as to what ways of intervening might be worth evaluating further. Suppose we evaluate the proportion of service recipients who achieve treatment goals, such as whether they graduate from high school as opposed to dropping out. That would be a descriptive evaluation. We should not call it explanatory unless our study design enables us to determine whether it was really our service, and not some other factor, that explained why the goal was achieved. Perhaps the students who were the most motivated to succeed were more likely to seek our services than those who were least motivated. If, however, we assess such alternative factors, then we will have an explanatory evaluation one which enables us to determine whether it was really our services that caused the desired outcome.

## 7.1 RIGOR OF THE STUDY

- The rigor of the study is important in order to generate credible and trustworthy results (Creswell, 1998; Strauss & Corbin, 1998; Yin, 1994). [87]
  The rigor in this study is achieved in the following ways:

  1. We followed the standard procedures of Usability Testing (Software Engineering Part). In addition, we incorporated the guidelines for ...
  2. ...

3. ...

### 7.1.1 *Restatement of the objectives and Research Questions of the Study*

#### 7.1.1.1 *Objectives*

- Our Study doesn´t aim at contributing to the understanding and measuring whether digital games can be considered suitable and effective tools for enhancing learning. Then this has been already in early chapters discussed and ensured through the contributions of other researchers on Serious Games in the last decade.
- This research is not focused on the exploration of the learning effects of Emil&Pauline Game, however, since this has already been performed in a previous studies regarding such an educational game.
- The evaluation of games' outcomes in this study should be a comprehensive measurement of cognitive, metacognitive, and **motivational aspects**.
- Gone from the emotions of the child, which are observed (effect), we search for the reason/trunk (cause) of success and/or failure. The cause can be part either of the GameAspect, or of the LearningContentAspect; which should be adapted accordingly: "Backwards Adjustment of Micro-Adaptivity".

### 7.1.2 *Researcher´s situated perspective*

- Methodology Review

### 7.1.3 *Aim*

### 7.1.4 *Evaluability Assessment*

- a study to determine whether or not a program or project can be evaluated.
- Evaluability assessment (EA), which is given the most extensive treatment here, assesses the extent to which programs are ready for future evaluation and helps key stakeholders come to agreement on realistic program goals, evaluation criteria, and intended uses of evaluation information. Wholey et al. [98]
- A process in which a program's or project's goals, objectives, activities, expected outputs and outcomes and client group are articulated and examined to assess the program's plausibility, the

evaluability of its goals and its readiness for evaluation. Usually depicted pictorially.

- http://www.socialresearchmethods.net/kb/intreval.htm
  - Formative evaluation includes several evaluation types:
    - needs assessment determines who needs the program, how great the need is, and what might work to meet the need
    - evaluability assessment determines whether an evaluation is feasible and how stakeholders can help shape its usefulness
    - structured conceptualization helps stakeholders define the program or technology, the target population, and the possible outcomes
    - implementation evaluation monitors the fidelity of the program or technology delivery
    - process evaluation investigates the process of delivering the program or technology, including alternative delivery procedures
  - Summative evaluation can also be subdivided:
    - outcome evaluations investigate whether the program or technology caused demonstrable effects on specifically defined target outcomes
    - impact evaluation is broader and assesses the overall or net effects – intended or unintended – of the program or technology as a whole
    - cost-effectiveness and cost-benefit analysis address questions of efficiency by standardizing outcomes in terms of their dollar costs and values
    - secondary analysis reexamines existing data to address new questions or use methods not previously employed
    - meta-analysis integrates the outcome estimates from multiple studies to arrive at an overall or summary judgement on an evaluation question

### 7.1.5 *Logic Model*

- My Logic Model; Bug Game Adaptive Variables and their claims, cause - effect, ...
- Input, Output; Impact?

### 7.1.6 *Evaluation Design*

- Single-Group Design: The simplest but least satisfactory evaluation design is the posttest only design, symbolized, X O (where X is the program intervention and O is a post-program observation).

### 7.1.7 *Site Selector*

- We have chosen to visit children at their home, play yard and even at kindergarten.

### 7.1.8 *Participants Selection*

- "Fifteen 4th–5th grade students were enrolled in the summer math program and participated in this research project. They were 10–13 years old, with five being socio-economic disadvantaged, 10 being girls, and all being white.2 Participants' pre-program school grades were collected. Their math abilities were classified into four levels – advanced, proficient, basic, and below basic based on their performances in the prior Pennsylvania System of School Assessment test. Out of the 15 students, four were advanced, six were proficient, and five were basic or below basic in math achievement. Participants were questioned on their prior gaming experience and if necessary, trained to know basic computer skills, such as using a mouse to click buttons on the computer screen. At the beginning of the summer math program, all participants took one orientation session to familiarize them with the gaming environment and were trained to do think-aloud, a strategy in which participants verbalize aloud while interacting with computer games, thus modeling the cognitive and affective processes of game-playing." [48]

- "Participants were selected using a random number generator. Overall, 33 out of 57 students attending the fifth grade were selected for the evaluation of the PwN application. Each student played the game individually in the peaceful environment of the school computer lab. Throughout the experimental procedure the students were observed by at least one researcher, who was responsible for the collection of the experimental data." [70]

- For this study, we focused on students from different kindergarten schools in Munich. The full study sample was a group of fifteen preschool children aged between 4 and 8 years.

### 7.1.9 *Research Methodology*

- The focus of this methodology is:
  1. to observe children playing the game without disturbing them,
  2. to send the collected videomaterial to experts for analysis,
  3. to integrate the proposed variables from the experts´Evaluation Report in our Adaptivity Framework / Remote Tool so adjusting the first adaptive version of the game,
  4. to conduct successive observation sessions, this time manipulating each Session one variable, in order to define easier the causality of each variable,
  5. to send the collected videomaterial to experts for re-analysis,
  6. last but not least, to adjust the Framework as suggested from the pedagogues.
- As Wacheux (1996) emphasizes, any qualitative study is specific. 'If the authors generally clarify their question and their results, rarer is research where the methodology is correctly presented. Nevertheless, the researcher must be capable of justifying his/her choice, and of defining the general framework in which the study will be deployed' (p.95). We have resolved to emphasize the research methodology for it is in the adoption of the **interpretativist perspective** that resides **the originality of this study**.
  1. First we are going to explain the reasons for the adoption of an **interpretativist perspective** based on the case study technique.
  2. Second we are going to analyze the collection and the treatment of data. Leclercq et al. [58]

### 7.1.10 *Research Design*

- This study will be as a quasi-experiment conducted in that there was no random assignment of children to condition.
- An experimental design with mixed method was used to conduct the research. Experts were separately asked after having explained the scenarios and requirements of the game as the Storytelling

### 7.1.11 *Data Collection instruments*

- As Plato already said "*Watch a man at play for an hour and you can learn more about him than in talking to him for a year*" (Plato,

427-348 v.Chr.). The best way to collect information about the player (e.g. a child playin a serious game) is to observe him playing. Ismailović [40]

- The following subsections explain the factors related to observation of players necessary for a serious game to be adaptable. Ismailović [40]
  - First of all different levels of evidence are explained that are necessary to apply adaptivity
  - Followed by a description of how this data can be used to recognize skills of a player.
  - Observation Mode: is necessary as a tool for teachers and pedagoges for beeing able to observe the child playing directly.
- Observational studies with statistical controls
  Widespread use, especially among academics
  Advantages:
  Doesn't require an experimental design
  Allows greater understanding of the pathways to change (tests conceptual framework)
  Limitations:
  Difficult to rule out confounding factors
- Method, procedure
- Questionnaires, Interviews, Field Notes, Analytic Memos, Observation, ...
- "**Observation**: The researcher closely observed participants' behaviors, verbal and nonverbal, and facial expressions when they interacted with the computer program, peers, and the external environment. A semi-structured observation protocol was developed to guide the researcher's attention during observation, though the actual observation was open to any situational changes." [48]
- All observed Children were recorded during the Gameplay.

## 7.2 QUALITATIVE STUDY: EXPERTS´ INTERVIEWING

The aim of the interviews with experts is to gain insight into their opinion about how freely could we adapt the game, the settings that are needed for each child and for each game situation. The interviews with experts had an open character. However, a semi-structered interview was formulated. *See appendix C: Interviews with experts.*

- I used the qualitative approach of open-ended, probing interviews to generate hypotheses on:

- how would Experts adapt the game individually to Children of different Ages variying between 4 and 8 years.
  - How would they recognize the varying states of the children playing the Game.
  - ...
- The qualitative analysis of the interview data suggested the following tentative factors that might influence how experts adjust to racial identity problems: parental attitudes, sibling and peer relationships, role-model availability, extended family factors, racial composition of school and community, and experience with racism and discrimination.

### 7.2.1 *Interviews*

We enlisted the help of several different kinds of experts to acquire and set up the levels settings for the adaptive mode that assure a reasonnable amount of difficulty or motivation in the Game.

Such interviews might not be high on a hierarchy designed to determine objectively the effects of the intervention of Adaptivity, but would be high on a hierarchy designed for generating in-depth insights about children's perceptions, feelings and attitudes.

*MUST BE REWRITTEN*

- Was ist adaptiearbar im BugGame?
- Beim start des Spiels:
  - totalNumberOfBugs = 60;
    - totalNumberOfGoodBugs = 30;
    - goodbugs = 5, 2;
      Welche 2 Nummern sind Good-bugs?
    - Random in order to assess the rate of right answers (Learning effectiveness)
  - Spiel:
    - badBugExits = NO;
      Sollen Bad-Bugs raus gehen können?
    - destroy = NO;
      Kann man Good-Bugs killen?
    - maxLimitDuplication = 8;
    - maxLimitEntrance = 8;
    - goodBugsRatio = 0.5;
      Verhältniss zwischen Good und Bad enteringRate = 6.0;
      alle 6 Sec kommen neue bugs
    - isAdvanced_BadBugsExit = NO;

| NAME | SECTOR |
|---|---|
| Dr. Pohl (Psy) | Psychology |
| Frau Bartl (GS) | Game Story |
| Guido (Ped) | Pedagogy |
| Damir (GD) | Game Design |

Table 10: Intervieews

    * isAdvanced_BadBugsDuplicate = YES;
    * isAdvanced_Speed = NO;
      Hier haben alle Bugs die gleiche geschwindigkeit
    * speed_factor = 1;
  - Aufgabe:
    * isAdvanced_NumbersOnly = YES;
      If NO: Then BadBugs are red and Good are green
    * isAdvanced_RandomNumbers = NO;
      Zufällige Anordnung der Zahlen

### 7.2.2 *Interviewees*

The interviewees ranged from Psychologues, Pedagogues, Math Teachers to game designers. All of the interviewees had a clear and detailed understanding of the scenarios of our game.

### 7.2.3 *Interview Structure*

- Conduct Personal Interviews
- In-depth interviews with experts
- semi-structured interviews

### 7.2.4 *Interview Procedure*

We contacted the experts one week earlier, send them the scenarios and requirements of the game being evaluated and the game itself, so that they could try and generate first impressions and ideas.

Before the interview started the purpose of this research was explained. Subsequently, consent was asked for recording the interview and each of the experts have had the game in front of him to simulate the situations discussed.

- How would experts judge and estimate the child´s skills in order to give him the right game situation?

  - if **underestimated**, beneath his/her skills, Both events and tasks should be more complex and harder. (Boredom & Annoying State)
  - if **overestimated**, over his/her skills, Both events and tasks should be more basic and easier. (Challenging & motivated State; Confusion & Frustration)

### 7.2.5 *Interview Results*

The pedagogue, Mrs. Bartl, was consulted to identify elements and properties, which should be adaptable. She specified six elements which should be adaptable: The quantity of number-signs held by Emil and Pauline, which indicate the 'good' fireflies, can vary between one and two. The consequence of two 'bad' fireflies hitting each other (encounter of 'bad' fireflies) is either none or they multiply. Whether 'bad' fireflies can leave the cave on their own (escape of 'bad' fireflies) is adaptable. The fireflies can be adjusted in regard to their speed and color. Which firefly is to appear (appearance of fireflies) when can be adapted. The amount of fireflies currently on the screen, and the percentage of 'good' and 'bad' fireflies being introduced are the influencing factors. The digits on fireflies can be either arranged as they are on a dice or have a random arrangement.

- **Level 1 (3-5)**
  - Number-sign: One good number
  - Encounter of bad bugs: no multiplication
  - Escape of bad bugs: no autonomous escape
  - Fireflies: all have the same speed (slow); good bugs are green, bad bugs are red
  - Appearance of bugs: maximum of five bugs on the screen; equal amount of good and bad bugs appear
  - Digits on fireflies: dice-like
- **Level 2 (4-6) or LevelUp**
  - Number-sign: One good number
  - Encounter of bad bugs: no multiplication Escape of bad bugs: no autonomous escape
  - Fireflies: all have the same speed (slow); random color
  - Appearance of bugs: maximum of five bugs on the screen; equal amount of good and bad bugs appear
  - Digits on bugs: dice-like
- **Level 3 (5-7) or LevelUp**

| ADAPT EVENTS | INFLUENCE ON |
|---|---|
| Set quantity of number-signs | LCA |
| Set encounter of bad bugs | GA |
| Set escape of bad bugs | GA |
| Set speed of bugs | LCA & GA |
| Set color of bugs | LCA |
| Set number of bugs on screen | LCA & GA |
| Set percentage of good/bad bugs on screen | GA |
| Set order of digits | LCA |

Table 11: Adapt Events & their influence on the game process in Bug Game

  - Number-sign: two good numbers
  - Encounter of bad bugs: multiplication
  - Escape of bad bugs: autonomous escape sbugs: bugs with higher numbers are faster; random color
  - Appearance of bugs: maximum of eight bugs on the screen; equal amount of good and bad bugs appear
  - Digits on bugs: dice-like

- **Level 4 (6-8) or LevelUp**
  - Number-sign: two good numbers
  - Encounter of bad bugs: multiplication
  - Escape of bad bugs: autonomous escape
  - Fireflies: all bugs are faster; bugs with higher numbers are faster; random color
  - Appearance of bugs: maximum of ten fireflies on the screen; more 'bad' fireflies
  - Digits on bugs: random

Correlation between Adapt Events & Adaptable variables.

The following Adapt Events Table 11 on page 130 can be identified, to manipulate the described Adaptable variables. Some of the Adapt Events manipulate the difficulty of Learning Content Aspects others the difficulty of Game Aspects.

These Adapt Events can be applied independently and customized, to meet the individual child's standard.

### 7.2.6 *Validity*

### 7.2.7 *Reliability*

## 7.3 PARTICIPANT OBSERVATION STUDY

The functional requirements are the business functions that the system is to perform. These requirements should also be met after integrating the Adaptivity Framework, but this needs a lot of Adjustment, since the adaptive game would adapt himself individual to his player.

From the Usability Testing Sessions i conducted during the development of the Game, i learned that, most of the unclear questions that a Game Designer/Developer ask himself about his Software solution could be answered by observing children playing the game.

So give the child an iPad, let him play, record him, analyse yourself and let the experts tell you their opinions.

### 7.3.1 *Video Material Permission*

- Please note that for studies with audio or video recordings, participants must be told:
  - (a) that the interviews or sessions will be audio or videotaped;
  - (b) that the cassettes will be coded so that no personally identifying information is visible on them;
  - (c) that they will be kept in a secure place (e.g., a locked file cabinet in the investigator's office);
  - (d) that they will be heard or viewed only for research purposes by the investigator and his or her associates; and
  - (e) that they will be erased after they are transcribed or coded.
- If you wish to keep the recordings because of the requirements of your professional organization with respect to data or because you may wish to review them for additional analyses at a later time, the statement about erasing them should be omitted and you should state that they will be retained for possible future analysis.
- If you wish to present the recordings at a convention or to use them for other educational purposes, you should get special permission to do so by adding, after the signature lines on the consent form, the following statement, "We may wish to present some of the tapes from this study at scientific conventions or as demonstrations in classrooms. Please sign below if you are

willing to allow us to do so with the tape of your performance." And add another signature line prefaced by, "I hereby give permission for the video (audio) tape made for this research study to be also used for educational purposes." This procedure makes it possible for a participant to agree to being taped for research purposes and to maintain the confidentiality of the information on that tape.

### 7.3.2 *Participant Observation*

### 7.3.3 *Interviews Post-test*

Hast du das Spiel gemocht?
Was hast du im Spiel gemocht?
Was hättest du noch gerne im Spiel?
Welche Game Situation / Level war dir zu leicht (langweilig)?
Welche Game Situation / Level war dir zu schwer (überfordert)?
Welche Game Situation / Level passt dir am besten?
(Schwierigkeitsgrad? Geschwindigkeit? Überwachen des Ausganges? ...)
Könntest du durch das Spiel die Zahlen besser erkennen?

- Did you like the game?
- Was it challenging?
- What was your favorite level?
- Was the game fun?
- Was the game too easy?
- Was the game too hard?
- Was the game easy to learn?
- Are the controls intuitive?
- Is the interface clear and easy to navigate?
- Rate yourself on a scale of one to ten.
- Rate how fun playing was.

The answers may make the game producers feel good and may indicate that the players will at least play the game, but they **do not assess the behavioral change** that takes place, and they **do not tell you that the players have learned the skills**.

In order to answer these questions, playtesting observes what users did in the game, how well the game supported their agency, and how users chose to solve gameplay challenges.

## 7.4 RESEARCH PRODUCT

- The time is ripe to create educational gaming environments since technology is now available to integrate assessments and implement data collection and adaptivity in near real-time. We have developed a set of capabilities to design and integrate assessments into games, visualize and analyze data, and individually adapt games. Additionally, we are continuing to research and develop methods for making inferences about behavior and learning in gaming environments. Embedded assessments that are truly integrated into the design of the game provide learners with the opportunity to immediately learn from their mistakes. In order to take advantage of this idea, **we must design games to integrate opportunities for learners to display the skills they are intended to learn, allowing the learner to enjoy and experiment with trial and error without interrupting gameplay**. Underwood et al. [93]

# Part V

# EVALUATION REPORT

Data Analysis; Results; Discussion; Conclusion; Contributions to adjust the Adaptivity Framework; Recommendations for future work in Evaluation.
Diplomarbeit: "Evaluation of the Impact of Micro-Adaptivity in Serious Games."
Studie entwerfen mit Evaluability Assessment Diagramm, Logic Model Diagramm, Interviews, Observation Experiment with different Game Situations.

# 8

# DATA ANALYSIS & DISCUSSION

## 8.1 SUMMARY OF THE METHODOLOGY

bla bla from chapter 7

## 8.2 PRESENT EVIDENCE AND IDEAS FROM SOURCES

The idea comes from this paper[1].

- ZPD Theory
- Interpret the Unproductive Patterns Wu [101]

The Learner ZPD Evaluator model is constructed based on the ZPD theory, which was firstly introduced in .... The model depicted in Figure 1.1 helps us understand under what circumstances the unproductive activities are likely to happen for a regular learner.
In the retrospect pattern, the learner tries to complete a task with inadequately developed skills. As a result, this learner ends up in the zone of confusion (the grey circle area in Figure 5.4) and cannot perform efficient learning.

After a number of times of failure, the learner diagnoses his/her problem with the current mission task is because of the relatively low-level skills. The strategy of this self-regulated learner uses is to increase his/her skill level. After he/she has improved the skill level to a certain degree and selects to play the mission task again, he/she is likely to shift the position into the ZPD (the while circle area in Figure 5.3). Of course, they cannot always come back to the ZPD zone. Chances are that after practicing the irrelevant pages or exercises with learning materials their skill level is still below the required level to win the challenge. And this shows the necessity of pedagogical support for this pattern of behavior.

Learners who behavior in the Onrush pattern have the ambition to complete all the mission tasks as soon as possible. It is possible it is because they gain confidence in the simple scenes with less challenge levels compared with their skill level. But once the mission scenes raise the challenge level to a degree higher than what he can get over with the current skill level, this learner will be shifted to the zone of confusion (the white circle area in Figure 5.5). He can of course lower the challenge level of the mission task by abusing the hint facilities.

1 REDUCING UNPRODUCTIVE LEARNING ACTIVITIES IN SERIOUS GAMES P23-108-109-110 Wu [101]

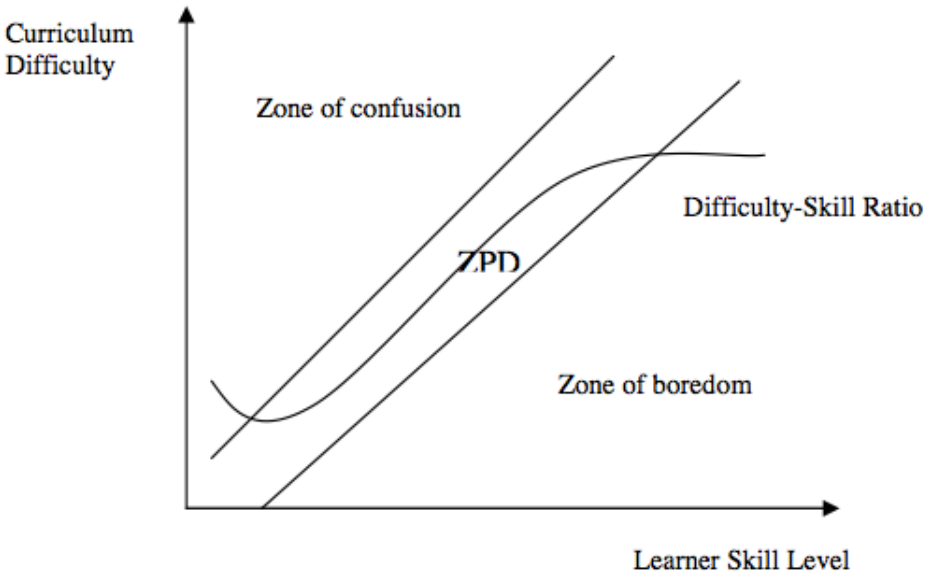



Figure 1.1 Learner's ZPD Relative to the Curriculum Difficulty

Figure 13: REDUCING UNPRODUCTIVE LEARNING ACTIVITIES IN SERIOUS GAMES P23

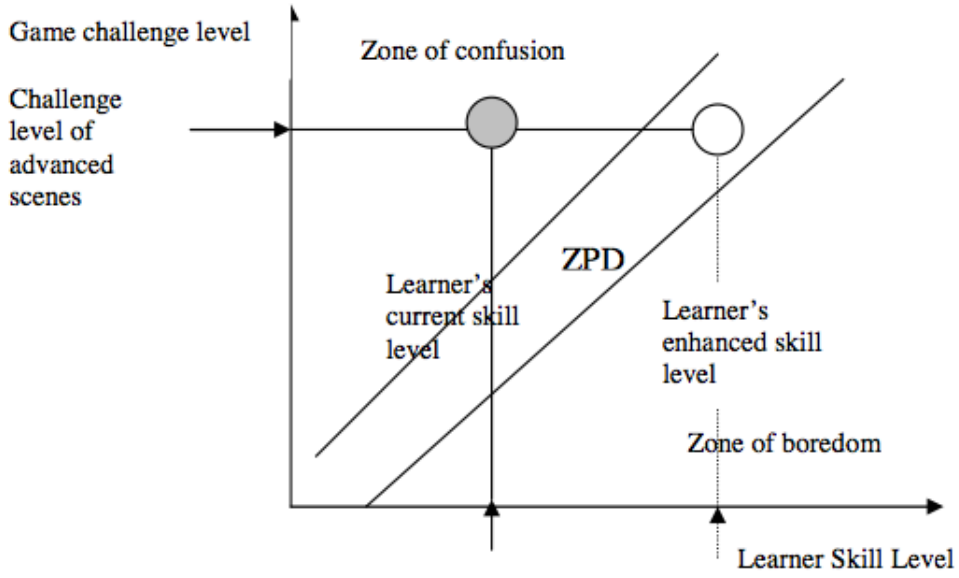



Figure 5.4 The Shift of Learner's Position to ZPD in the Retrospect Pattern

Figure 14: REDUCING UNPRODUCTIVE LEARNING ACTIVITIES IN SERIOUS GAMES P108

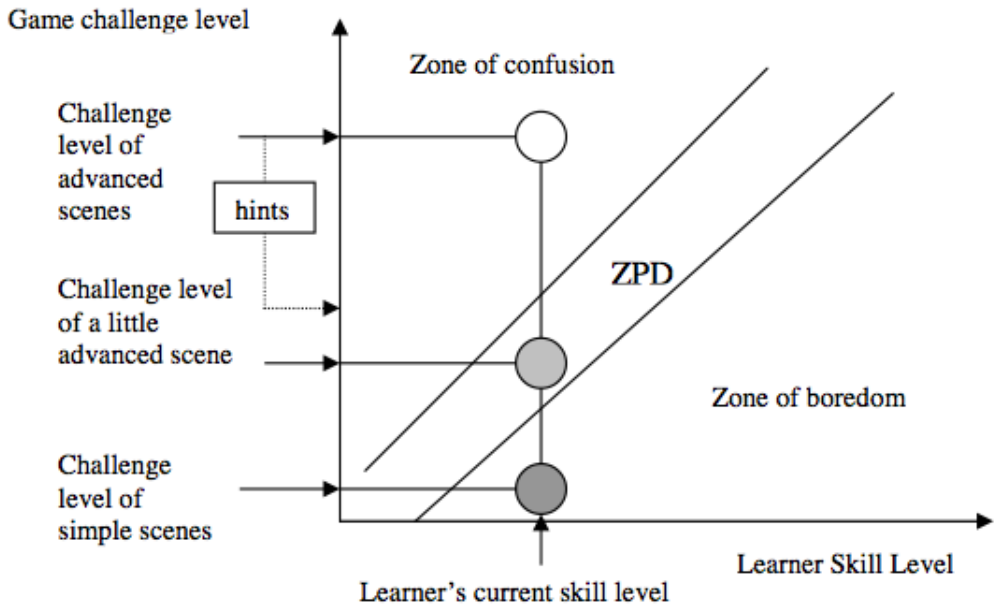



Figure 5.5 The Shift of Learner's Position to ZPD in the Onrush Pattern

Figure 15: REDUCING UNPRODUCTIVE LEARNING ACTIVITIES IN SERIOUS GAMES P109

And if he succeeds, he can bring himself back the ZPD. We have some examples showing that learners complete the mission takes under the assistance of hints. The strategy is acceptable as the learner can learn from hints to advance their skill level. However, in the situation that they fail to lower the challenge level to shift their position into the ZPD zone, the potential unproductive behavior will be turned into an actual unproductive one.

Compared with the previous two patterns, the HelpRefusal pattern least benefits the skill development of learner. Because the learners' inaction to acquire required skills and or to access the learning materials or help facilities, the position of this learner, once in the zone of confusion, is difficult to change (Figure 5.6). Without tutor intervention or self-diagnosis, this learner will remain in the zone of confusion, which is claimed to be the area where low efficient learning occurs.

## 8.3 DISCUSSION

- How to motivate? Motivation Challenges
  - Provide optimally challenging activities
  - Change sensory conditions to arouse curiosity
  - State goals or allow goals to emerge

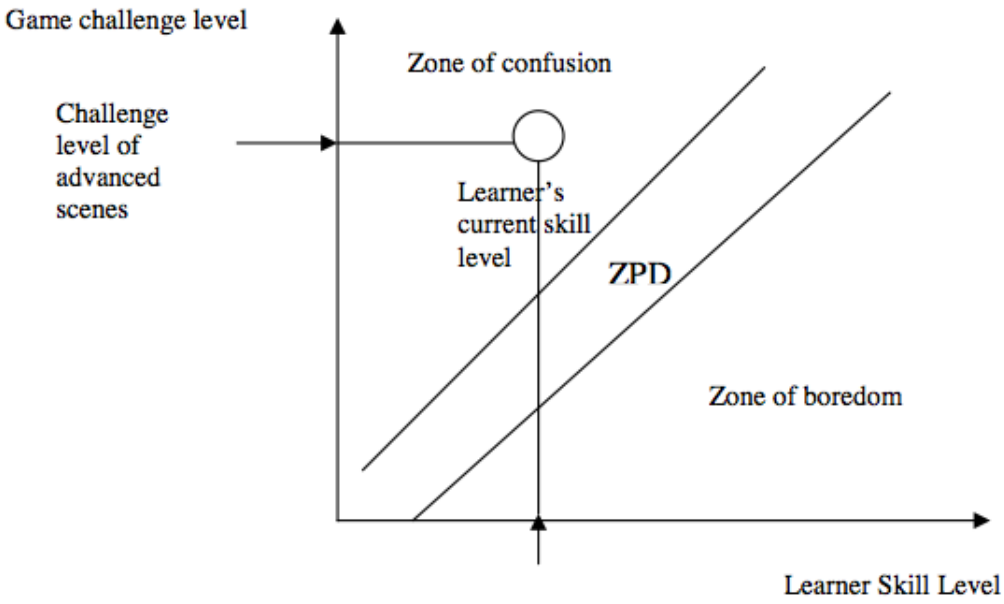



Figure 5.6 The Shift of Learner's Position to ZPD in the HelpRefusal Pattern

Figure 16: REDUCING UNPRODUCTIVE LEARNING ACTIVITIES IN SERIOUS GAMES P110

  - Provide an optimal level of challenge
  - Provide performance feedback.
  - Provide control over the learning environment
  - Use fantasy to help the student experience power, success, fame, and fortune. Also helps learners relate new learning to a past experience

- In my Diagramms i should present The Relations Among the Adaptive Variables and their effect on the motivation of children as shown in 17
  Motivation, a dependent variable, was considered as heart and outcome of the cycle

### 8.3.1 *Bug Game: Game Situation I*

- Interaction, dependence or coherence among adaptive variables
- Adaptive Variables Bounds on a game flow chart
- observed effect on children
- If we differentiate the color of ´bugs, the speed should be relatively higher.

| ADAPTIVE VARIABLES | VALUE PRESENTED | IMPACT WITHDRAWN |
|---|---|---|
| Number of Bugs (Non-Adaptive) | 60 | |
| Good/Bad Ratio (goodBugsRatio) | 0.7 | |
| Entering Rate () | 7 | |
| Speed Rate (Speedfactor) | 0.75 | |
| Speed Bugs (isAdvanced_Speed) | NO | |
| Collision - Duplication (isAdvanced_BadBugsDuplicate ) | NO | |
| Entering Bugs: 4 (PREDEFINED Small iPad Screen) | 4 | |
| maxLimitEntrance: 8 (PREDEFINED Small iPad Screen) | 8 | |
| maxLimitDuplication: 8 (PREDEFINED Small iPad Screen) | 8 | |
| Color (isAdvanced_NumbersOnly) | YES | |
| Exit (isAdvanced_BadBugsExit) | NO | |
| Random (MISSING) | NO | |
| Target Numbers (E&P) (MISSING) | NO | |
| OnlyOneTarget (MISSING) | YES | |

Table 12: Game Situation 1 (3-5 years); Starter Level

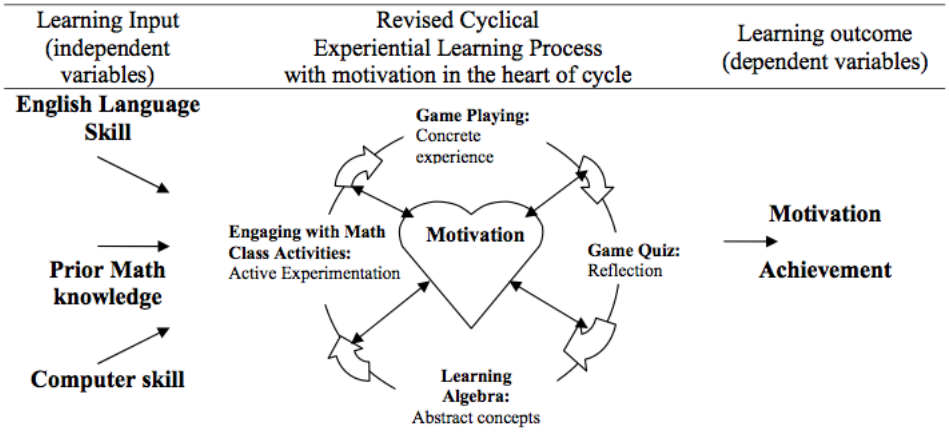



Figure 6: The Relations Among the Variables of Language Background, Prior Knowledge, Computer Experience, Motivation and Achievement with the Cyclical Learning Process of Experiential Learning Theory Extended from a Kolb's (1984) Circular Learning

Figure 17: The Relations Among the Adaptive Variables and their effect on the motivation of children

### 8.3.2 *Bug Game: Game Situation II*

### 8.3.3 *Bug Game: Game Situation III*

### 8.3.4 *Bug Game: Game Situation IV*

## 8.4 FINDINGS IN CHART FORM

- Groupings may be organized according to your Research questions.
- Factor Duplication, factor Speed and factor Collision correlate together; So if we adjust one of them, we should be aware of the others, in order to mainatain the flow.

## 8.5 EXPERIENCE AND PERCEPTION

- Causes of conflict;
- How to resolve the conflict?;
- Learning experiences;
- Personal commitments;
- Impact of social action; ...

| ADAPTIVE VARIABLES | VALUE PRESENTED | IMPACT WITHDRAWN |
|---|---|---|
| Number of Bugs (Non-Adaptive) | 60 | |
| Good/Bad Ratio (goodBugsRatio) | 0.6 | |
| Entering Rate () | 6 | |
| Speed Rate (Speedfactor) | 1 | |
| Speed Bugs (isAdvanced_Speed) | YES | |
| Collision - Duplication (isAdvanced_BadBugsDuplicate ) | NO | |
| Entering Bugs: 4 (PREDEFINED Small iPad Screen) | 4 | |
| maxLimitEntrance: 8 (PREDEFINED Small iPad Screen) | 8 | |
| maxLimitDuplication: 8 (PREDEFINED Small iPad Screen) | 8 | |
| Color (isAdvanced_NumbersOnly) | NO | |
| Exit (isAdvanced_BadBugsExit) | NO | |
| Random (MISSING) | NO | |
| Target Numbers (E&P) (MISSING) | NO | |
| OnlyOneTarget (MISSING) | NO | |

Table 13: Game Situation 2 (4-6 years)

| ADAPTIVE VARIABLES | VALUE PRESENTED | IMPACT WITHDRAWN |
|---|---|---|
| Number of Bugs (Non-Adaptive) | 60 | |
| Good/Bad Ratio (goodBugsRatio) | 0.5 | |
| Entering Rate () | 5 | |
| Speed Rate (Speedfactor) | 1 | |
| Speed Bugs (isAdvanced_Speed) | YES | |
| Collision - Duplication (isAdvanced_BadBugsDuplicate ) | YES | |
| Entering Bugs: 4 (PREDEFINED Small iPad Screen) | 4 | |
| maxLimitEntrance: 8 (PREDEFINED Small iPad Screen) | 8 | |
| maxLimitDuplication: 8 (PREDEFINED Small iPad Screen) | 8 | |
| Color (isAdvanced_NumbersOnly) | NO | |
| Exit (isAdvanced_BadBugsExit) | NO | |
| Random (MISSING) | YES | |
| Target Numbers (E&P) (MISSING) | YES | |
| OnlyOneTarget (MISSING) | NO | |

Table 14: Game Situation 3 (5-7 years)

| ADAPTIVE VARIABLES | VALUE PRESENTED | IMPACT WITHDRAWN |
|---|---|---|
| Number of Bugs (Non-Adaptive) | 60 | |
| Good/Bad Ratio (goodBugsRatio) | 0.4 | |
| Entering Rate () | 4 | |
| Speed Rate (Speedfactor) | 1 | |
| Speed Bugs (isAdvanced_Speed) | YES | |
| Collision - Duplication (isAdvanced_BadBugsDuplicate ) | YES | |
| Entering Bugs: 4 (PREDEFINED Small iPad Screen) | 4 | |
| maxLimitEntrance: 8 (PREDEFINED Small iPad Screen) | 8 | |
| maxLimitDuplication: 8 (PREDEFINED Small iPad Screen) | 8 | |
| Color (isAdvanced_NumbersOnly) | NO | |
| Exit (isAdvanced_BadBugsExit) | YES | |
| Random (MISSING) | YES | |
| Target Numbers (E&P) (MISSING) | YES | |
| OnlyOneTarget (MISSING) | NO | |

Table 15: Game Situation 4 (6-8 years)

# 9

# CONCLUSION & RECOMMENDATIONS FOR FUTURE WORK

## 9.1 RESTATEMENT OF THE RESEARCH QUESTION

The intention from this evaluation is neither to gather data such as video material or experts interviews in order for further Analysis with statistical analysis software or to develop charts mentioning the success of the developed game. The intention, due to the evaluation period (Alpha Phase), is to develop theories from the hypotheses that mess the real correlation between the adaptive variables.

These results would average the adaptive variables, bound them to assure a pedagogical and psychological terrain to better integrate the Adaptivity Framework. (make the game adaptable)

Keeping in mind the theories of Gameplay, Learning through play, Game Flow, Learning through mistakes, Vygotsky's Theories of individual ZPD & Scaffolding; if we would adapt, we should be aware of all of these theories, better said, we would adapt individually to the player playing the game at the moment.

## 9.2 SUMMARY THAT SYSTEMATICALLY ANSWER YOUR RESEARCH QUESTIONS

## 9.3 POSSIBILITIES & CONSTRAINTS

We experimented with few samples of children, have defined general game situations settings that would ensure the game flow depending on the right association skill level and child/gamer; But even if the game situation at the moment presents a suitable difficulty for the child playing it, after some hours enjoying the game, it would be annoying, if we don´t switch to another game situation more complex;

The question here is when should we switch? and should we make it more difficult or more easier? All depending on who´s playing? What skills does he/she have?

How would we switch individually?

## 9.4 RECOMMENDATION FOR FUTURE RESEARCH

As a recommendation for future evaluation and in order to let experts interact live with the players into the game although without interrupting or letting the gamer feel that some one is changing the game artefacts while he/she plays, we imagined a test solution, called "RemoteEqualizerTool" consisting of two iPads, facilitating us the remote control of the adaptive variables wirelessly while playing. The interest relies in tasks that have been characterized as 'remote help giving' (Tolmie et al., 2004) where one of the collaborators has the task knowledge and one of the collaborators manipulates the task artefacts.

As far as is known that "Learning is often characterized in terms of the relationship between Instructor and Learner with the Instructor either passing on knowledge or creating an environment for the Learner that is rich for self-discovery (e.g. Vygotsky's zone of proximal development, Vygotsky, 1978; Bransford et al., 2000)." [52], during instruction, the expert should be able to define the limits of understanding of the Learner, they must successfully pass on knowledge and they must be able to competently assess that the Learner has understood.

The need from this method is to investigate the impact of varying adaptable variables on the quality of both gameplay and learning.

## 9.5 CONTRIBUTIONS

*Qualitative Research*

In the first step we need to answer the two research questions RQ1 [1] & RQ2 [2]

Therefore a qualitative research will be applied by interviewing experts and by observing children while playing in several iterations.

For answering this two questions we will carry out interviews with experts (pedagogists, psychoslogists, professional game-developers, serious game researchers, students and players). We will iteratively develop a serious game with those experts that is able to provide examples to the given questions, and that helps experts to reflect about given answers. We will use the designed serious game to observe children while playing, to give experts more data. Finaly we will execute retrospective interviews with experts based on the given case study.

This study will provide following results:

1 RQ1: What can be observed when a learner play a serious game and how can this information be used to characterize the individual learner?

2 RQ2: What are changes that need to be applied to a serious game (while playing) to provide help?

- A detailed description of data that can be observed in one example serious games for children with one specific topic.
- Detailed description of how experts caracterize the child in the given example.
- Description about what are possible adaptable properties in the given example that could be used to provide help to children.

This exploratory study will help us to better comprehend the nature of adaptivity.

*Quasi-Experimental Study with experts*

In the next step we will setup a quasi-experimental study with experts based on the results of RQ1 and RQ2 to answer the research question RQ3 [3]. In this study we will use the designed game to find out how experts would apply non-invasive help. This study will provide following results:

- Description about how the adaptable properties in the given example can be used to provide help to children.

3 RQ3: How would experts apply non-invasive help to a serious game by just observing learners? How can this task be formalized?

# Part VI

# APPENDIX

# A

# APPENDIX TEST

## A.1 APPENDIX SECTION TEST

Video Material Parents Permission
Interview
Observation protocol
Questionnaire Post-Observation

## A.2 ANOTHER APPENDIX SECTION TEST

## A.3 ANOTHER ANOTHER APPENDIX TEST

# BIBLIOGRAPHY


[1] Michael J. Albers and Brian Still. *Usability of Complex Information Systems Evaluation of User Interaction*. 2011. ISBN 9781439828946.

[2] Dietrich Albert, Cord Hockemeyer, Michael D Kickmeier-rust, Neil Peirce, and Owen Conlan. Microadaptivity within Complex Learning Situations a Personalized Approach based on Competence Structures and Problem Spaces. 2008.

[3] K Barbara. *Domain Independent Framework for Story Driven Game Based Learning*. PhD thesis, 2011.

[4] Carol M. Barnum. *Praise for Usability Testing Essentials*. 2011. ISBN 9780123750921.

[5] Bartl. Emil & Pauline in der Höhle. Technical report, 2011.

[6] F. Bellotti, R. Berta, a. De Gloria, and L. Primavera. Enhancing the educational value of video games. *Computers in Entertainment*, 7(2):1, June 2009. ISSN 15443574. doi: 10.1145/1541895.1541903. URL http://portal.acm.org/citation.cfm?doid=1541895.1541903.

[7] M. Bopp. Immersive Didaktik: Verdeckte Lernhilfen und Framingprozesse in Computerspielen. *kommunikation@ gesellschaft*, 6: 1–18, 2000. URL http://en.scientificcommons.org/23415593.

[8] Matthias Bopp. Digital Games Research. *Game Studies*, (May), 2007.

[9] Susan Bowers. Assessing Young Children: Whats Old, Whats New, and Where Are We Headed?, 2007. URL http://www.earlychildhoodnews.com/earlychildhood/article_print.aspx?ArticleId=210Seite.

[10] Johanna Bragge and Jan Storgå rds. Profiling academic research on digital games using text mining tools. 15:714–729, 2007. URL http://www.digra.org/dl/db/07311.26413.pdf.

[11] Dunphy Bruce C and L Stacey. Assisted Performance and the Zone of Proximal Development ZPD; a Potential Framework for Providing Surgical Education. 3:48 –58, 2003.

[12] Canada. How to Design an Evaluation from Scratch. *Power Analysis Inc. Experts in Program Evaluation, Survey Research and Data Analysis*.

[13] John M Carroll and Mary Beth Rosson. A Case Library for Teaching Usability Engineering: Design Rationale, Development, and Classroom Experience. *ACM Journal on Educational Resources in Computing*, 5(1):1–22, 2005. doi: ACM1531-4278/05/0300-ART3\$5.00.

[14] Sande Chen and David Michael. Proof of Learning: Assessment in Serious Games. *Media*, 2005.

[15] Jennifer L Chiu, Douglas B Clark, Brian C Nelson, Cynthia M D Angelo, Kent Slack, and Mario Martinez. Promoting Links and Developing Students Criteria for Visualizations by Prompting Judgments of Fidelity. *ICLS2010*, 2:230–231, 2010.

[16] Douglas B Clark, Brian C Nelson, Cynthia A D Angelo, Kent Slack, and Mario Martinez-garza. SURGE: Integrating Vygotskys Spontaneous and Instructed Concepts in a Digital Game? Research Goals and Theoretical Framework. 2:384–385, 2010.

[17] Genevieve Clarke and Michelle Treagust. Gaming for reading. (June), 2010.

[18] Karen Collins, Kamen Kanev, and Bill Kapralos. Using Games as a Method of Evaluation of Usability and User Experience in Human Computer Interaction Design. pages 5–10, 2010.

[19] Thomas Connolly and Mark Stansfield. *2nd European Conference on Games Based Learning*. Published by Academic Publishing Limited, 2008. ISBN 9781906638191. URL www.academic-publishing.org.

[20] Ulrike Cress, Ursula Fischer, Korbinian Moeller, Claudia Sauter, and Hans-Christoph Nuerk. The use of a digital dance mat for training kindergarten children in a magnitude comparison task. *ICLS*, 1(1994):105–112, 2010.

[21] Aristides Dasso and Ana Funes. *Verification, Validation and Testing in Software Engineering*. ISBN 1591408512.

[22] Felipe Soares de Oliveira, Tatiana Aires Tavares, Andrew Anderson Chagas, Aquiles Burlamaqui, Edna Gusmão Brennand, and Guido Lemos de Souza Filho. Experiences from the use of a shared multimedia space for eLearning in Brazil primary schools. *Proceedings of the 3rd international conference on Digital Interactive Media in Entertainment and Arts DIMEA 08*, page 91, 2008. doi: 10.1145/1413634.1413655. URL http://portal.acm.org/citation.cfm?doid=1413634.1413655.

[23] Claire Dormann and Robert Biddle. Understanding game design for affective learning. *Proceedings of the 2008 Conference on Future*

*Play Research, Play, Share Future Play 08*, page 41, 2008. doi: 10.1145/1496984.1496992. URL http://portal.acm.org/citation.cfm?doid=1496984.1496992.

[24] Tammie Hutto Egloff. Edutainment: A Case Study of Interactive CDROM Playsets. *ACM Computers in Entertainment*, 2(1):1–22, 2004. doi: ACM1544-3574/04/0100-ART04\$5.00.

[25] Gerald D Everett and Raymond McLeod. *Software Testing Testing Across the Entire Software Development Life Cycle*. 2007. ISBN 9780471793717.

[26] Lindsay Evett. Advanced Frontiers of eAccessibility Serious Games and Interactive Devices, 2011.

[27] Magued Michel Farah George. *Bug Game: Learning Math in a Serious Game for iPad*. PhD thesis, 2011.

[28] Jorge Ferreira Franco, Sandra Regina da Cruz, and Roseli de Deus Lopes. Computer Graphics, Interactive Technologies and Collaborative Learning Synergy Supporting Individuals Skills Development. *ACM SIGGRAPH 2006 Educators program on SIGGRAPH 06*, page 10, 2006. doi: 10.1145/1179295.1179338. URL http://portal.acm.org/citation.cfm?doid=1179295.1179338.

[29] Josef Froschauer, Ingo Seidel, Gärtner Markus, Helmut Berger, and Dieter Merkl. Design and Evaluation of a Serious Game for Immersive Cultural Training.

[30] A.S.V. Games. Learning Language through Video Games: A Theoretical Framework, an Evaluation of Game Genres and Questions for Future Research Jonathan deHaan.

[31] R. Garris, R. Ahlers, and J. E. Driskell. Games, Motivation, and Learning: A Research and Practice Model. *Simulation & Gaming*, 33(4):441–467, December 2002. ISSN 1046-8781. doi: 10.1177/1046878102238607. URL http://sag.sagepub.com/cgi/doi/10.1177/1046878102238607.

[32] Franca Garzotto. Investigating the Educational Effectiveness of Multiplayer Online Games for Children. *Proceedings of the 6th international conference on Interaction design and children IDC 07*, page 29, 2007. doi: 10.1145/1297277.1297284. URL http://portal.acm.org/citation.cfm?doid=1297277.1297284.

[33] Helene Gelderblom. Designing software for young children: theoretically grounded guidelines. *IDC 2004*, June1-3(College Park, Maryland, USA):121–122, 2004. doi: ACM1-58113-791-5/04/0006\ldots\$5.00.

[34] Stefan Göbel. Serious Games DEG. 2010.

[35] Begoña Gros. Digital games in education: The design of games based learning environments. *Journal of Research on Technology in Education*, 40(1):23, 2007. URL http://www.mrgibbs.com/tu/research/articles/gros_Game_design.pdf.

[36] Casper Harteveld. Triadic Game Evaluation: A Framework for Assessing Games with a Serious Purpose.

[37] Koen Hendrix, Robert van Herk, Janneke Verhaegh, and Panos Markopoulos. Increasing Childrens Social Competence Through Games, an Exploratory Study. *Proceedings of the 8th International Conference on Interaction Design and Children IDC 09*, page 182, 2009. doi: 10.1145/1551788.1551823. URL http://portal.acm.org/citation.cfm?doid=1551788.1551823.

[38] Layla Husain. Getting Serious about Math: Serious Game Design Framework and an Example of a Math Educational Game. 2011.

[39] Katherine Isbister, Mary Flanagan, and Chelsea Hash. Designing games for learning: insights from conversations with designers. In *Proceedings of the 28th international conference on Human factors in computing systems*, pages 2041–2044. ACM, 2010. ISBN 9781605589299. URL http://portal.acm.org/citation.cfm?id=1753637.

[40] Damir Ismailović. *Adaptivity in Serious Games*. PhD thesis, 2011.

[41] Damir Ismailović and Dennis Pagano. Barwin on the Beagle. 2008.

[42] Damir Ismailović and Dennis Pagano. weMakeWords - a collaborative game based learning system for the iPhone. 2009.

[43] Damir Ismailović, Johannes Krugel, and Christopher Schulz. On the Application of different Research Methods in Software Engineering. Technical report, 2009.

[44] Damir Ismailović, Dennis Pagano, and Bernd Brügge. WEMAKEWORDS AN ADAPTIVE AND COLLABORATIVE SERIOUS GAME FOR LITERACY. *IEEE Transactions on Learning Technologies*, 2011.

[45] Nick Iuppa and Terry Borst. *EndtoEnd Game Development Creating Independent Serious Games and Simulations from Start to Finish*. Focal Press, Elsevier, Inc., 2010. ISBN 9780240811796.

[46] Victor Kaptelinin and Michael Cole. Individual and Collective Activities in Educational Computer Game Playing. *Proceedings*

*of the 2nd international conference on Computer support for collaborative learning CSCL 97*, pages 142–147, 1997. doi: 10.3115/1599773.1599790. URL http://portal.acm.org/citation.cfm?doid=1599773.1599790.

[47] Nakamura Kazuo. On the Concept of Cultural Age in Vygotskys Cultural Historical Theory For the purpose of a clearer understanding of the concept ZPD. *The Report of Tokyo University of Fisheries*, pages 1–6, 2003.

[48] F Ke. A case study of computer gaming for math: Engaged learning from gameplay? *Computers & Education 51 2008 Elsevier Ltd*, 51(4):1609–1620, December 2008. ISSN 03601315. doi: 10.1016/j.compedu.2008.03.003. URL http://linkinghub.elsevier.com/retrieve/pii/S0360131508000523.

[49] MANSUREH KEBRITCHI. *EFFECTS OF A COMPUTER GAME ON MATHEMATICS ACHIEVEMENT AND CLASS MOTIVATION: AN EXPERIMENTAL STUDY*. PhD thesis, 2008.

[50] Michael D Kickmeier-rust. Bibliography on Digital Game Based Learning. *Journal of Information Technology*, 53(July):86–107, 2010.

[51] Michael D. Kickmeier-Rust, Birgit Marte, Stephanie Linek, Tiphaine Lalonde, and Dietrich Albert. The effects of individualized feedback in digital educational games, 2008.

[52] David S. Kirk and Danaë Stanton Fraser. The Effects of Remote Gesturing on Distance Instruction. *Proceedings of the 2005 conference on Computer support for collaborative learning learning 2005: the next 10 years CSCL 05*, pages 301–310, 2005. doi: 10.3115/1149293.1149332. URL http://portal.acm.org/citation.cfm?doid=1149293.1149332.

[53] Kenneth R Koedinger, Elizabeth A Mclaughlin, Neil T Heffernan, and Kenneth Koedinger. A QuasiExperimental Evaluation of an Online Formative Assessment and Tutoring System. *Journal of Educational Computing Research*, 43:489–510, 2010.

[54] Christina Kolb. *Evaluation of a serious game*. PhD thesis, 2011.

[55] Effie Lai-chong Law and Michael Rust-kickmeier. 80Days: Immersive Digital Educational Games with Adaptive Storytelling.

[56] Effie Lai-chong Law, Michael Kickmeier, Dietrich Albert, and Andreas Holzinger. Challenges in the Development and Evaluation of Immersive Digital Educational Games. *HCI and Usability for Education and Work*, 2008.

[57] T.L. Leacock and J.C. Nesbit. A framework for evaluating the quality of multimedia learning resources. *Subscription Prices and Ordering Information*, 10(2):44–59, 2007. URL http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.119.7249&rep=rep1&type=pdf#page=49.

[58] Aurélie Leclercq, Pierre Romelaer, and Henri Isaac. The Perceptual Evaluation of Information Systems Using the Construct of User Satisfaction: Case Study of a Large French Group. *Data Base For Advances In Information Systems*, 38(2):27–60, 2007.

[59] C.S. Loh, Arnond Anantachai, Jaehwan Byun, and Joe Lenox. ASSESSING WHAT PLAYERS LEARNED IN SERIOUS GAMES: IN SITU DATA COLLECTION, INFORMATION TRAILS, AND QUANTITATIVE ANALYSIS. *Computer Games: AI, Animation, Mobile, Educational & Serious Games (CGAMES 2007), Louisville, KY*, 2007. URL http://www.csloh.com/research/pdf/cgames2007_loh_patent.pdf.

[60] Rosemary Luckin, Benedict D U Boulay, Nicola Yuill, Darren Pearce, and Amanda Harris. Using Software Scaffolding to increase Metacognitive skills amongst young learners. *Psychologie Schweizerische Zeitschrift Für Psychologie Und Ihre Andwendungen*, pages 1–3.

[61] Javier Marco, Eva Cerezo, Emanuela Mazzone, and Janet C Read. Bringing Tabletop Technologies to Kindergarten Children. *People and Computers*, pages 103–111, 2009.

[62] Dominic W Massaro. Just in Time Learning: Implementing Principles of Multimodal Processing and Learning for Education.

[63] Maic Masuch. Digitale Lernspiele. *Fachhochschule Trier*.

[64] Patrick McCue. The Crucial Role of Animated Childrens Educational Games. *ACM SIGGRAPH 2005 Educators program on SIGGRAPH 05*, page 6, 2005. doi: 10.1145/1187358.1187366. URL http://portal.acm.org/citation.cfm?doid=1187358.1187366.

[65] Bente Meyer. Proceedings of the 4th European Conference on Games Based Learning. (October), 2010.

[66] Chris Mills and Barney Dalgarno. A conceptual model for game based intelligent tutoring systems. *Proceedings ascilite Singapore 2007*, pages 692–702, 2007. URL http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.95.1113&rep=rep1&type=pdf.

[67] Tom Murray and Ivon Arroyo. Toward Measuring and Maintaining the Zone of Proximal Development in Adaptive Instructional Systems. *Learning*, (1):1–8, 2002.

[68] Lennart Nacke, Anders Drachen, and Stefan Göbel. Methods for Evaluating Gameplay Experience in a Serious Gaming Context. pages 1–12, 2000.

[69] Staffan Nählinder and Per-anders Oskarsson. Learning from Computer Games. *Defence Research Agency Command and Control Systems*, (December 2007).

[70] Chris T. Panagiotakopoulos. Applying a Conceptual Mini Game for Supporting Simple Mathematical Calculation Skills: Students Perceptions and Considerations. *World Journal of Education*, 1(1):3–14, April 2011. ISSN 1925-0754. doi: 10.5430/wje.v1n1p3. URL http://www.sciedu.ca/journal/index.php/wje/article/view/150www.sciedu.ca/wje.

[71] Dimitra Pappa. GameBased Learning for Knowledge Sharing and Transfer: the eVITA project.

[72] Kai Pata, Tago Sarapuu, and Raymond Archee. Collaborative Scaffolding in Synchronous Environment: Congruity and Antagonism of TutorStudent Facilitation Acts. *Interpretation A Journal Of Bible And Theology*, (2001), 2004.

[73] Albert-Jan Pomper, Bruno Jordaan, and Johan Ravesloot. The Added Value of Serious Games in Management Development Programs The Slowesa Case. *PMI Global Congress Proceedings Amsterdam, Netherlands*, pages 1–10, 2009.

[74] David Mw Powers, Richard Leibbrandt, Darius Pfitzner, Martin Luerssen, Trent Lewis, Arman Abrahamyan, and Kate Stevens. Language Teaching in a Mixed Reality Games Environment. *Proceedings of the 1st ACM international conference on PErvasive Technologies Related to Assistive Environments PETRA 08*, 2008. doi: 10.1145/1389586.1389668. URL http://portal.acm.org/citation.cfm?doid=1389586.1389668.

[75] Marc Prensky. The motivation of gameplay: The real twenty first century learning revolution. *On the Horizon*, 10(1):5–11, 2002. ISSN 1074-8121. doi: 10.1108/10748120210431349. URL http://www.emeraldinsight.com/10.1108/10748120210431349.

[76] Clark N. Quinn. *Designing MLearning: Tapping Into the Mobile Revolution for Organizational Performance*. Pfeiffer, An Imprint of Wiley, 2011. ISBN 978-0-470-94593-3. URL www.pfeiffer.comwww.designingmlearning.com.

[77] Hayes Raffle, Rafael Ballagas, Glenda Revelle, Hiroshi Horii, Sean Follmer, Janet Go, Emily Reardon, Koichi Mori, Joseph Jofish Kaye, and Mirjana Spasojevic. Family Story Play: Reading with Young Children and Elmo Over a Distance. *CHI 2010*, pages 1583–1592, 2010.

[78] Elaine M. Raybourn and Nathan Bos. Design and evaluation challenges of serious games. *CHI 05 extended abstracts on Human factors in computing systems*, page 2049, 2005. doi: 10.1145/1056808.1057094. URL http://portal.acm.org/citation.cfm?doid=1056808.1057094.

[79] G. Rebolledo-Mendez, Katerina Avramides, S. de Freitas, and K. Memarzia. Societal impact of a serious game on raising public awareness: the case of FloodSim. In *Proceedings of the 2009 ACM SIGGRAPH Symposium on Video Games*, pages 15–22. ACM, 2009. ISBN 9781605585147. URL http://portal.acm.org/citation.cfm?id=1581076.

[80] Thomas Reeves. Can Educational Research Be Both Rigorous and Relevant? *keynote address at the Educational Design Research*, 1:1–24, 2010. URL http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Can+Educational+Research+Be+Both+Rigorous+and+Relevant?#4.

[81] Brian J. Reiser. Why Scaffolding Should sometimes Make Tasks More Difficult for Learners. *Proceedings of the Conference on Computer Support for Collaborative Learning Foundations for a CSCL Community CSCL 02*, page 255, 2002. doi: 10.3115/1658616.1658652. URL http://portal.acm.org/citation.cfm?doid=1658616.1658652.

[82] S.M. Robertson and M. Hart. Video Games as an Education Tool. *cs.cmu.edu*. URL http://www.cs.cmu.edu/afs/cs.cmu.edu/misc/mosaic/common/omega/Web/People/smrobert/tucson/video_draft.doc.

[83] Guido Rößling, Ari Korhonen, Rainer Oechsle, J. Ángel Velázquez Iturbide, Mike Joy, Andrés Moreno, Atanas Radenski, Lauri Malmi, Andreas Kerren, Thomas Naps, Rockford J. Ross, and Michael Clancy. Enhancing learning management systems to better support computer science education. *ACM SIGCSE Bulletin*, 40(4):142, November 2008. ISSN 00978418. doi: 10.1145/1473195.1473239. URL http://portal.acm.org/citation.cfm?doid=1473195.1473239.

[84] A. Rubin and E.R. Babbie. *Research Methods for Social Work*. Brooks/Cole Pub Co, 2008. ISBN 9780495811718. URL http://books.google.com/books?hl=en&lr=&id=

xUlfMaY4RskC&oi=fnd&pg=PR4&dq=Research+Methods+for+Social+Work&ots=0xGLLU947J&sig=H2aN5TDCpZRdLagMY7DNWdnDkpk.

[85] Paula Savioja, Leena Norros, and Leena Salo. Evaluation of systems usability. *Proceedings of the 15th European conference on Cognitive ergonomics the ergonomics of cool interaction ECCE 08*, page 1, 2008. doi: 10.1145/1473018.1473051. URL http://portal.acm.org/citation.cfm?doid=1473018.1473051.

[86] J Schell. *The Art of Game Design*. 2008. ISBN 9780123694966. URL http://books.google.com/books?hl=en&lr=&id=LP5xOYMjQKQC&oi=fnd&pg=PP1&dq=The+Art+of+Game+Design&ots=Y87SqDO3tJ&sig=80iFD1Gof0toeZwV88-z_5lFPUI.

[87] Hong Sheng, Keng Siau, and Fiona Fui-Hoon Nah. Understanding the Values of Mobile Technology in Education: A Value Focused Thinking Approach. *The DATA BASE for Advances in Information Systems*, 41(2):25–44, 2010.

[88] Blagina Simeonova. *An Adaptive Serious Game for Preschool Mathematics on Mobile End Devices*. PhD thesis, TUM, 2011.

[89] Elizabeth Sklar and Debbie Richards. The Use of Agents in Human Learning Systems. *Proceedings of the fifth international joint conference on Autonomous agents and multiagent systems AAMAS 06*, page 767, 2006. doi: 10.1145/1160633.1160768. URL http://portal.acm.org/citation.cfm?doid=1160633.1160768.

[90] A.J. Stapleton. Serious games: Serious opportunities. In *Health Care*, pages 1–6, 2004. URL http://andrewstapleton.com/wp-content/uploads/2006/12/serious_games_agdc2004.pdf.

[91] Edward Tse, Richard Beckwith, Johannes Schöning, Yvonne Rogers, Jochen Huber, Max Mühlhäuser, and Lynn Marentette. Child Computer Interaction: Workshop on UI Technologies and Educational Pedagogy. *CHI 2011*, pages 2445–2448, 2011.

[92] Mary Ulicsak. Games in Education: Serious Games: A Futurelab Literature Review, 2010.

[93] Jody S. Underwood, Stacy Kruse, and Peter Jakl. Moving to the Next Level: Designing Embedded Assessments into Educational Games. *Design and Implementation of Educational Games: Theoretical and Practical Perspectives*, pages 126–140, 2010.

[94] R. Van Eck. Digital Game Based Learning: It is Not Just the Digital Natives Who Are Restless. *EDUCAUSE review*, 41(2):16, 2006. URL http://www.itu.dk/people/jrbe/DMOK/Artikler/digitalgamebasedlearning2006.pdf.

[95] Irina Verenikina. Understanding Scaffolding and the ZPD in Educational Research. *Faculty of Education, University of Wollongong, NSW, Australia*, 2004.

[96] Irina Verenikina, Pauline Harris, and Pauline Lysaght. Childs Play: Computer Games, Theories of Play and Childrens Development. *IFIP Working Group 3.5 Conference: Young Children and Learning Technologies, held at UWS Parramatta in July 2003*, Australian:99–106, 2003.

[97] Lev Vygotsky. Literary Creativity in School Age Children. *Imagination and Creativity in Childhood*, pages 41–47.

[98] Joseph S. Wholey, Harry P. Hatry, and Kathryn E. Newcomer. *HANDBOOK OF PRACTICAL PROGRAM EVALUATION*. 2010. ISBN 978-0-470-52247-9.

[99] Bradford B Wiles. Toward a theory of how young children learn to read in the ZPD: Implications for research and practice. 2008.

[100] Wee Ling Wong, Cuihua Shen, Luciano Nocera, Eduardo Carriazo, Fei Tang, Shiyamvar Bugga, Harishkumar Narayanan, Hua Wang, and Ute Ritterfeld. Serious video game effectiveness. *Proceedings of the international conference on Advances in computer entertainment technology ACE 07*, page 49, 2007. doi: 10.1145/1255047.1255057. URL http://portal.acm.org/citation.cfm?doid=1255047.1255057.

[101] Shumin Wu. *Reducing Unproductive Learning Activities in Serious Games for Second Language Acquisition*. PhD thesis, 2008.

[102] Pat Youngblood. Evaluating Serious Games Confusing Terminology.

## COLOPHON

Technische Universität München
Institut für Informatik / I1
Lehrstul für Angewandte Softwaretechnik
Boltzmannstrasse 3
D-85748 Garching bei München
Germany

Taha Dhiaeddine Amdouni; amdouni@gmx.de

[ December 21, 2011 at 14:08 ]

# DECLARATION

I assure the single handed composition of this thesis only supported by declared resources.
*Garching bei München, December 2011*

Taha Dhiaeddine Amdouni