

THE ROLE OF SIMULATION IN SUPPORTING LONGER-TERM
LEARNING AND MENTORING WITH TECHNOLOGY

A Thesis Submitted to the College of
Graduate and Postdoctoral Studies
In Partial Fulfillment of the Requirements
For the Degree of Doctor of Philosophy
In the Department of Computer Science
University of Saskatchewan
Saskatoon

By

David Edgar Kiprof Lelei

© Copyright David Edgar Kiprof Lelei, June 2019. All rights reserved.

PERMISION TO USE

In presenting this dissertation in partial fulfillment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this dissertation in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my dissertation work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this dissertation or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my dissertation.

Requests for permission to copy or to make other uses of materials in this dissertation in whole or part should be addressed to:

Head of the Department of Computer Science

176 Thorvaldson Building

110 Science Place

University of Saskatchewan

Saskatoon, Saskatchewan S7N 5C9

Canada

OR

Dean

College of Graduate and Postdoctoral Studies

University of Saskatchewan

116 Thorvaldson Building, 110 Science Place

Saskatoon, Saskatchewan S7N 5C9

Canada

ABSTRACT

Mentoring is an important part of professional development and longer-term learning. The nature of longer-term mentoring contexts means that designing, developing, and testing adaptive learning systems for use in this kind of context would be very costly as it would require substantial amounts of financial, human, and time resources. Simulation is a cheaper and quicker approach for evaluating the impact of various design and development decisions. Within the Artificial Intelligence in Education (AIED) research community, however, surprisingly little attention has been paid to how to design, develop, and use simulations in longer-term learning contexts. The central challenge is that *adaptive learning system designers and educational practitioners have limited guidance on what steps to consider when designing simulations for supporting longer-term mentoring system design and development decisions.*

My research work takes as a starting point VanLehn et al.'s [1] introduction to applications of simulated students and Erickson et al.'s [2] suggested approach to creating simulated learning environments. My dissertation presents four research directions using a real-world longer-term mentoring context, a doctoral program, for illustrative purposes. The first direction outlines a framework for guiding system designers as to what factors to consider when building pedagogical simulations, fundamentally to answer the question: how can a system designer capture a representation of a target learning context in a pedagogical simulation model? To illustrate the feasibility of this framework, this dissertation describes how to build, the SimDoc model, a pedagogical model of a longer-term mentoring learning environment – a doctoral program. The second direction builds on the first, and considers the issue of model fidelity, essentially to answer the question: how can a system designer determine a simulation model's fidelity to the desired granularity level? This dissertation shows how data from a target learning environment, the research literature, and common sense are combined to achieve SimDoc's medium fidelity model. The third research direction explores calibration and validation issues to answer the question: how many simulation runs does it take for a practitioner to have confidence in the simulation model's output? This dissertation describes the steps taken to calibrate and validate the SimDoc model, so its output statistically matches data from the target doctoral program, the one at the university of Saskatchewan. The fourth direction is to demonstrate the applicability of the resulting pedagogical model. This dissertation presents two experiments using SimDoc to illustrate how to explore pedagogical questions concerning personalization strategies and to determine the effectiveness of different mentoring strategies in a target learning context.

Overall, this dissertation shows that simulation is an important tool in the AIED system designers' toolkit as AIED moves towards designing, building, and evaluating AIED systems meant to support learners in longer-term learning and mentoring contexts. Simulation allows a system designer to experiment with various design and implementation decisions in a cost-effective and timely manner before committing to these decisions in the real world.

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my supervisor, Prof. Gordon McCalla, without which this dissertation would not be possible. I am very grateful for all your amazing guidance, insights and suggestion into this research project. Further, I am very thankful for your financial and emotional support. I will never forget your constant concern about the wellbeing of my family. Thank you for giving me a second chance. I feel highly honored to have gotten a chance to work within a research area that you are among its pioneers.

I would also like to thank my committee members for their unwavering support, suggestions, and many questions that made the focus of this work clearer. Thank you to Dr. Jim Greer (we miss you), Dr. Mark Keil, Dr. Regan Mandryk, Dr. Dirk Morrison, and Dr. Julita Vassileva. I am also grateful to Dr. Robin Cohen for her feedback on the final version of my dissertation.

Furthermore, I am thankful to the members of MADMUC lab in 2011 and ARIES lab over the last few years for countless discussions, critiques, and advice around my research topic. I would like to thank the following friends who contributed greatly to my enjoying my doctoral journey: Stephanie Frost, Dr. Bukola Ishola, Dr. Jennifer Seaton, Dr. Terry Peckham, Dr. Mayya Sharipova, Dr. John Iyilade, Dr. Bunmi Adewoyin, Dr. Richard Lomotey, and Dr. Rita Orji.

Also, I would like to recognize and be grateful to other faculty members without which I would not be in this position. I am thankful to Dr. Jim Carter and Dr. Ralph Deters for their roles in my graduate school journey. I am grateful for all the teachers who taught me at various levels of my educational journey.

Finally, I am thankful to God for giving me patience, knowledge, protection, and giving me a loving and patient family. I have made it this far because of the amazing and unconditional support, patience, and encouragement of my family. Special thanks to my Queen, Gladys, for being all I could ask for in a wife and a mother to our children. Thank you to Gloreyas and Ethan for giving me the reason to persist to completion. Thank you to my parents, Edith and Richard, my parents-in-law, Milka and William, for your constant prayers and encouragement. I am grateful to my brother Tonny, my sister-in-law, Hellen, and their children Bethuel and Elizabeth for their encouragement. I am also thankful to my auntie Joanne, my cousins Annie, Raymond, Jeremiah, and Veronica for their support.

This dissertation is dedicated to my wife Gladys, my daughter Gloreyas, my son Ethan, my parents, Edith and Richard, my brother and his family, Tonny, Hellen, Bethuel, and Elizabeth, and my extended family.

TABLE OF CONTENTS

Permission to Use	i
Abstract	ii
Acknowledgments.....	iii
Table of Contents	v
List of Algorithms.....	viii
List of Figures	ix
List of Tables	x
List of Publications of the Author.....	xi
Important Terms.....	xii
Chapter 1 Introduction.....	1
1.1 Problem Statement	2
1.2 Solution Approach.....	3
1.3 Contributions.....	5
1.4 Dissertation Organization.....	6
Chapter 2 Related Work	7
2.1 Educational Computer Systems Research.....	7
2.2 Longer-term Learning: Barriers and Solutions	11
2.2.1 Accessibility and Open Educational Resources.....	14
2.2.2 Personalization and Open Learner Models	16
2.2.3 Student Modeling within AIED Research	17
2.2.4 Evaluation and Simulation.....	19
2.3 Use of Simulation within AIED Research	22
2.3.1 Current Uses of Simulation.....	24
2.3.2 Potential Use of Simulation	25

2.4	Summary	26
Chapter 3	Creating SimDoc: Modeling a Simulated Doctoral Program	28
3.1	Mentorship and Mentoring Systems	28
3.2	Understanding the Doctoral Program.....	30
3.3	Using Simulation to Understand a Longer-Term Mentoring Environment	33
3.4	Modeling a Simulated Doctoral Program: the SimDoc Model	34
3.4.1	A Seven-Step Modeling Framework	34
3.4.2	SimDoc’s Conceptual Model: Key Components and Assumptions	37
3.5	Informing SimDoc’s Model Fidelity: Behavior and Evaluation Functions	49
3.5.1	Informing the SimDoc Model Behavior Functions	50
3.5.2	Informing the SimDoc Evaluation Functions	58
3.6	Summary	62
Chapter 4	Tuning the SimDoc Model: Calibration and Validation.....	64
4.1	Calibrating the SimDoc Model: Matching UofS Learning Outcomes.....	64
4.2	Validating the SimDoc Model: Exploring the Difference in Number of Runs.....	66
4.2.1	Determining an Adequate Number of Simulation Runs: in the Literature	67
4.2.2	Determining an Adequate Number of Simulation Runs: the SimDoc Case Study.	70
4.2.3	Confirmation of Stability and Variability in SimDoc’s Output.....	73
4.3	Summary	78
Chapter 5	Experimenting with SimDoc: Exploring Effects of Personalization	81
5.1	Exploring Personalization Using Simulation	81
5.1.1	Baseline Experimental Set Up	81
5.1.2	Interpreting Simulation Results	83
5.2	Drilling into the Baseline for Personalization Insights	83
5.3	Exploring Hypothetical, ‘what-if’, Questions within the SimDoc Case Study.....	85

5.4	Insights into the Personalization Experimental Findings	95
5.5	Future Experimental Directions	98
5.6	Summary	100
Chapter 6	Discussion and Conclusion	102
6.1	Discussion	102
6.1.1	Research Objectives	102
6.1.2	Simulation for Supporting AIED System Development Process	102
6.2	Summary of the Research	105
6.3	Contributions	108
6.4	Limitations	110
6.5	Future Work	111
6.5.1	Extensions to the Current SimDoc Simulation	111
6.5.2	Directions for Simulation Research	113
References	115
Appendix A – Behavioral Ethics Certificates	132

LIST OF ALGORITHMS

Algorithm 3-1. Simulated Learner's Number of Classes.....	56
Algorithm 4-1. Pseudocode for an Algorithm that Iteratively Calculates Levene, Chi-Square, and ANOVA P-Values	72

LIST OF FIGURES

Figure 3-1. The Four Supervisory Styles and Associated Supervisor Traits	32
Figure 3-2. A Seven-Step Framework for Building a Valid Pedagogical Simulation Model	35
Figure 3-3. SimDoc’s Conceptual Framework	38
Figure 3-4. SimDoc’s Conceptual Model Implemented in AnyLogic™ Programming Platform	40
Figure 3-5. SimDoc’s Learner Model Implemented in AnyLogic™	41
Figure 3-6. SimDoc’s Supervisor Agent Model as Implemented in AnyLogic™	43
Figure 3-7. A Depiction of SimDoc’s Admission Event as Implemented in AnyLogic™ .	46
Figure 3-8. Snippet Showing SimDoc System and Learner Level Events	48
Figure 3-9. Distribution of the Number of Classes Taken per Learner	56
Figure 3-10. Relative Position of the Four Supervisory Styles Identified by Gatfield [151].	58
Figure 4-1. A Chart Showing the Results of SimDoc’s Calibration Process.....	66
Figure 4-2. Comparison Between UofS Students and SimDoc Learners Time-to-Outcome	74
Figure 4-3. Density Plots of 12 Randomly Selected Runs of the Simulation.....	77
Figure 4-4. Variation in the Graduation and Attrition Rates in the 12 Simulation Runs	77
Figure 4-5. Comparing UofS and SimDoc Baseline Learning Outcomes	80
Figure 5-1. SimDoc Baseline Learning Outcomes Differentiated by Meeting Frequency:	84
Figure 5-2. SimDoc Baseline Learning Outcomes Differentiated by Latent Effort:	85
Figure 5-3. SimDoc MFMF Set Up Learning Outcomes.	89
Figure 5-4. Comparing Learning Outcomes: SimDoc Baseline vs MFMF Set Up.	90
Figure 5-5. SimDoc MFMF Set Up Learners’ Time-to-Outcome.....	91
Figure 5-6. Comparing Learners’ Time-to-Outcome: SimDoc Baseline vs MFMF Set Up.	92
Figure 5-7. SimDoc LEMF Set Up Learning Outcomes.	94
Figure 5-8. Comparing Learning Outcomes in the SimDoc Baseline and the LEMF Set Up.	95
Figure 5-9. SimDoc LEMF Set Up Learners’ Time-to-Outcome.....	97
Figure 5-10. Comparing Learners’ Time-to-Outcome: SimDoc Baseline vs LEMF Set Up.	98

LIST OF TABLES

Table 2-1 Major Educational Movements in the Last Two Centuries.....	13
Table 3-1. Learner Agent Attributes and their Parameterization.....	42
Table 3-2. Supervisor Agent Attributes and their Parameterization.....	43
Table 3-3. Research Group Entity Attributes and their Parameterization	44
Table 3-4. Milestone Entity Attributes and their Parameterization	45
Table 3-5. The UofS Dataset Attributes and their Description.....	50
Table 3-6. Expected FTE and Actual Mean Time Taken for Each Milestone	54
Table 3-7. UofS 'Year of Study' vs 'Number of Classes Taken' Program Distribution Matrix	57
Table 3-8. The Frequency of Supervisor-Doctoral Learner Formal Meetings	61
Table 3-9. UofS 2006-2010 Cohorts' Matriculation, Progress, and Attrition Rates	62
Table 3-10. Milestone Completion Rates with Their Derived Attrition Rates	62
Table 4-1. P-Values for Levene, Chi-Square, and ANOVA Tests for the Simulation Runs.....	73
Table 4-2. Frequency Counts of Time-in-Program Between the SimDoc and the UofS Students	73
Table 4-3. Summary of Frequency Count by Outcome and Data Source	76
Table 5-1 The SimDoc Experiments Data Dictionary	82

LIST OF PUBLICATIONS OF THE AUTHOR

- (1) **Lelei, David Edgar K.**, and Gordon McCalla. "How Many Times Should a Pedagogical Agent Simulation Model be Run?" to appear in the *Proceedings of the 20th International Conference on Artificial Intelligence in Education*, Chicago, IL, USA, June 2019, "accepted".
- (2) **Lelei, David Edgar K.**, and Gordon McCalla. "Simulation in Support of Lifelong Learning Design: A Prospectus" to appear in the Proceedings of the *International Workshop on Supporting Lifelong Learning*, to be held in conjunction with the *20th International Conference on Artificial Intelligence in Education*, Chicago, IL, USA, June 2019, "accepted".
- (3) **Lelei, David Edgar K.**, and Gordon McCalla. "How to Use Simulation in the Design and Evaluation of Learning Environments with Self-directed Longer-Term Learners." in the *Proceedings of the 19th International Conference on Artificial Intelligence in Education*, London, UK, June 2018, pp. 253–266.
- (4) **Lelei, David Edgar K.**, and Gordon McCalla. "The Role of Simulation in the Development of Mentoring Technology to Support Longer-Term Learning." in the Proceedings of the *3rd International Workshop on Intelligent Mentoring Systems*, held in conjunction with the *19th International Conference on Artificial Intelligence in Education*, London, UK, June 2018. Available at:
https://imsworkshop.files.wordpress.com/2018/06/ims2018_lelei.pdf
- (5) **Lelei, David Edgar K.**, and Gordon McCalla. "Using Simulation to Explore Reciprocal Help Seeking in a Lifelong Learning Context." *International Journal of Business Process Integration and Management* 7, no. 3 (2015): 228-246. 10.1504/IJBPIIM.2015.071253.
- (6) **Lelei, David Edgar K.**, and Gordon McCalla. "Exploring the Issues in Simulating a Semi-Structured Learning Environment: the SimGrad Doctoral Program Design." in the Proceedings of the *2nd Workshop on Simulated Learners*, held in conjunction with the *17th International Conference on Artificial Intelligence in Education*, Madrid, Spain, June 2015, vol. 5, pp. 11–20. Available at: http://ceur-ws.org/Vol-1432/sl_pap2.pdf
- (7) **Lelei, David Edgar K.** "Supporting Lifelong Learning: Recommending Personalized Sources of Assistance to Graduate Students." in the *Extended Proceedings of 16th International Conference on Artificial Intelligence in Education – Doctoral Colloquium*, Memphis TN, USA, July 2013, pp. 912-915.

IMPORTANT TERMS

In this section, I present definitions of fundamental terms I use throughout this dissertation. A *learner* is a person who attempts to obtain new knowledge or skill by studying or learning from a teacher¹. A *student* refers to a person registered formally in a learning institution for the purposes of pursuing a course of study². In this dissertation, I use the terms *learner* and *student* interchangeably and in a more specific way to mean a person registered formally in a doctoral program. Doctoral students often experience learning difficulties that may slow their progress and elongate their *time-to-completion* – the elapsed time between students’ enrollment and graduation dates [3]. Taking a long time to complete a degree program negatively affects the *completion rate*, that is, the ratio of the number of students who complete a degree program divided by the total number of students who enrolled during the same degree program at the same admission period [4]. Furthermore, slow time to completion because of learning hindrances increases the *attrition rate*, which, refers to the number of students who drop out of a program as compared to the total number of students who were enrolled over a specific period.

I use the term *system* to refer to a group of dynamic or passive elements, interconnected together to form a complex whole. For example, in this dissertation, a doctoral program is a system that is made up of dynamic elements (e.g., students, supervisors, and other stakeholders) and passive and abstract elements, such as courses and departments. A *model* is a simplified representation of an element within a system or a representation of the whole system. *Simulation* is the imitation through computation of aspects of a natural system’s functionality and behavior. A *conceptual model* is a representation of a system to be modeled and simulated; it is comprised of a static description of the composition of a system the model represents [5]. A *computational model* refers to a coded (computer program) version of the conceptual model that is executable and can run on a single or a network of computers to reproduce key behaviors of a system under study. I use the term *model fidelity* to refer to the degree or measure of exactness or similarity a model has to the *natural system* or phenomenon it represents [6]. *Verification* is the processes of determining the accuracy of a computational model to a relevant conceptual model and that the resulting model has no programmatic errors [7]. *Calibration* is the process of adjusting numerical modeling parameters in the computational model for the purpose of improving the match be-

¹ <http://www.learnersdictionary.com/definition/learner> last accessed on February 12, 2019. Note, in AIED a teacher could either refer to a human or a teaching (learning) support system.

² <http://www.thefreedictionary.com/student> last accessed on February 12, 2019

tween simulation output and dataset from the real-world system [8]. I use the term *baseline model* to refer to a calibrated simulation model whose parameters have been tuned such that the output of the model is statistically similar to that of the real-world system. *Validation* involves checking that a computational model's output and behavior are consistent with the data output and behavior of the system requirements under study [9]. Therefore, a *valid* model is a model whose output and behavior are significantly like the output of the real system the model is based on.

SimDoc refers to my model of a doctoral program. *SimDoc* is a simulated doctoral program environment with simulated students, supervisors, classes, and research groups represented as *agents*. An *agent* is a software program designed to represent (act as, play the role of) an element or entity of interest within a domain of interest. An *agent model* refers to the attributes and behaviors captured in the agent about an entity. The notion of an 'agent' in computer science indicates an entity that has some degree of intelligence and the capacity to perform actions autonomously. Similar to how Schroeder, Adesope, and Gilbert [10] define an agent, in this dissertation, I use the term **pedagogical agent** to refer to a computer-based avatar employed in an AIED system's interface to support learning.

CHAPTER 1

INTRODUCTION

Humans have always developed tools to extend their capabilities to solve different types of problems they faced. Today, computer technologies are used to support and extend almost all human capabilities to solve challenges in many walks of life including but not limited to education, communication and transportation, sports and games, business, medicine, building and construction, and social networking. As a result, pervasive and ubiquitous computing technologies are evident in almost every aspect of our lives. These technologies support the collection of data, the storage of data, and the processing of the collected data to support decision making. In the context of education, computer technologies were not the first machines to be used to support learning. Other technologies, such as instructional radio in the 1920s and instructional television in the 1950s [11] were early attempts to use machines to aid learning. Since the 1950s, the use and influence of computer technologies in education have gradually increased [11].

So, what motivated the building and use of educational computer systems? There are at least two main reasons: practical and research motives [12]–[14]. In practical terms, using computers to support learning offers an opportunity to overcome social and economic challenges associated with human tutoring. For example, while individualized human tutoring is very effective [15] and thus it would be ideal to replicate its effectiveness in many learning contexts, it is very costly to hire the right number of human tutors. In this regard, a strong advantage of educational computer systems is the possibility of developing many systems that provide and replicate individualized tutoring that is responsive to learners' individual needs on a par with excellent human tutors [14]. Another motivation for educational computer systems is to provide environments that are more motivating than traditional educational systems, like educational computer games or learning environments with exciting computer bells and whistles. As far as research is concerned, the intersection of three different disciplines including computer science, cognitive science, and education [12] forms the core of educational support systems research. This intersection provides a test-bed for exploring the nature of knowledge and how that knowledge is being learned in different learning contexts and domains. This contributes to the understanding of learning theories [14].

The use of technology to enhance learning has resulted in the emergence of advanced learning technology (ALT) research fields including Artificial Intelligence in Education (AIED). AIED aims to advance “rigorous research and development of interactive and adaptive learning environments for learners of all ages, across all domains”³. Advances in this research area have led to the development of various adaptive learning systems that are helping thousands of learners in numerous learning contexts and domains [16], [17]. It is important to note, however, that most research in these contexts has focused on learning in domain-specific situations taking place over relatively short-time frames often with experiments done in a laboratory setting [18]. Usually, the focus is on certain well-defined domain subject matter, such as mathematics [19], [20], physics, computer literacy, and more [21], [22], whose concepts gradually build on one another and have fairly structured and constrained tasks that learners perform [23].

Recently, though, there has been some effort directed towards modeling and exploring longer-time real-world learning environments [24]–[26]. Mentoring [27], personalization [25], [28], and self-directed learning [29]–[33] are vital parts of professional development and longer-term learning. The key is the one-to-one relationships between learners and their mentors which dictates how the support is given in different styles in accordance with learners’ learning styles and preferred support styles [34], [35].

1.1 Problem Statement

As AIED advances into new areas, such as longer-term mentoring contexts, evaluation of developed learning systems and learners’ learning experiences are critical for ascertaining the success and effectiveness of systems [36], [37] [36], [38]. The nature of longer-term mentoring contexts means that testing of these AIED systems would be very costly as it would require substantial amounts of financial, human, and time resources. Simulation is a cheaper and quicker approach for evaluating the impact of various design and development decisions. However, there is very little research on how to design, develop and use simulations in longer-term learning contexts.

The central challenge is that AIED system designers and educational practitioners have limited guidance on what steps to consider when designing simulations for supporting longer-

³ <http://ijaied.org/about/> last accessed on February 12, 2019

term mentoring system design and development decisions. The fundamental research question underlying my research is:

How does a system designer design, implement, calibrate and validate a simulation to explore adaptive and personalized learning approaches for supporting learners in longer-term mentoring contexts?

My more specific research questions are:

Q1.How can a system designer capture a representation of a target longer-term learning environment in a pedagogical simulation model? To answer this question, I specifically explore how to create a representation of a doctoral program as a pedagogical simulation model.

Q2.How can a system designer inform the resulting pedagogical simulation model to achieve fidelity to a desired level of granularity? Specifically, how does a system designer get and incorporate available data to inform the fidelity of the simulated doctoral pedagogical simulation model?

Q3.How can a system designer know for certain that they have adequately calibrated and validated a simulation model in order to trust its output? Specifically, how many simulation runs does it take for a practitioner to have confidence in the simulated doctoral model's output?

Q4.How does a system designer use the resulting simulation model to ask pedagogical questions concerning the learning environment? Specifically, I explore how different combinations of personalization characteristics in a doctoral program affect learning outcomes as measured by completion rates, attrition rates, and time-to-completion?

1.2 Solution Approach

Simulation enables rapid comparison among the various adaptive and personalization strategies to determine the most effective ones without incurring large costs in terms of financial, time, and human resources. In addition, simulation enables the answering of hypothetical, 'what-if', questions. Using simulation allows for the testing and understanding of the possible impact of various adaptive and personalization measures on a doctoral program before embarking on building an actual system to experiment with real learners. Three central challenges in the design of pedagogical simulations are deciding on key attributes of the target learning domain to model, finding

data to inform the simulation model's fidelity, and determining the best methods to use to validate the simulation model. To this end, my research aims to depict the whole process of design, implementation, calibration, validation, and use of a simulation model for empirical evaluation of various adaptive and personalized support strategies in a longer-term mentoring context using a doctoral program as a case study.

Here is an overview of how I have addressed the concerns of questions raised in section 1.1.

Q1.I have answered this question by building a simulation model of a doctoral program (the SimDoc model) represented as an agent model simulation. In addition, I have used an equation informed by Item Response Theory in modeling and representing learners' knowledge states. Its attributes are derived from a real-world target learning environment, the University of Saskatchewan (UofS) doctoral program. The process of building the SimDoc model depicts a case study of how to build a simulation for any longer-term mentoring context. This includes an illustration of both the software and knowledge engineering approaches used in designing, modeling, and developing the SimDoc model.

Q2.I have answered this question by demonstrating three ways of informing the SimDoc simulation model to have it produce learning outcomes that match the UofS dataset. The three approaches are using reverse engineering techniques on available data about the target environment, the UofS dataset; drawing from data in the research literature about the target environment; deducing important information from 'common sense' assumptions (that must be validated by requiring that a full simulation that incorporates such 'common sense' assumptions matches actual behavior in the target domain).

Q3.I have answered the concerns of this third question by creating an algorithm that determines the number of runs necessary to generate stable outputs with appropriate variability using one-way analysis of variance (ANOVA). Essentially, the approach is to run the simulation iteratively, and after each run, the collected simulation outputs generated to date are compared against real-world data using Chi-Square, Levene, and ANOVA testing methods.

Q4.I have answered the concerns of this fourth question by building a version of the SimDoc model whose features can be parameterized to allow exploration of personalization issues of interest and compute metrics such as the differences in completion rates, attrition rates, and time-to-completion. I have created metrics representing supervisor style

by looking at the literature on supervisory style. Similarly, I found metrics representing learning style by looking at the literature on an allied capability of students. I have run simulations exploring the effects (on the three identified learning metrics) of various algorithms for matching supervisor types against learner types. And, finally, I have run simulations to explore the effects (on the three identified learning outcomes) of varying the number of learners being supervised by each supervisor.

1.3 Contributions

The central contribution of my research is the demonstration of how to build, calibrate and validate, and use a simulation model of a longer-term learning environment to explore various pedagogical issues, including asking hypothetical, ‘what-if’, pedagogical questions. The research makes several contributions to advanced learning technology research and most specifically to artificial intelligence in education (AIED). In my research I have:

- Presented a seven-step framework adopted from [193] for guiding the design and modeling of simulation models. Through a case study based on a doctoral program, I have shown how AIED and other advanced learning technology (ALT) researchers can use it to guide the building, informing, and validating of a pedagogical simulation model for exploring different research issues in longer-term learning and mentoring environments.
- Identified a pedagogical use of simulation not explored very much in AIED or other ALT communities: how simulation could be used to explore various hypothetical, ‘what-if’, pedagogical questions related to understanding issues in longer-term learning and mentoring environments.
- Developed a medium fidelity simulation model, a rarely investigated level of fidelity.
- Illustrated how to inform a simulation model by showing how to combine data from diverse sources related to phenomena of interest.
- Demonstrated how to calibrate a simulation model using a baseline dataset gathered from the target learning environment.
- Showed how to validate a simulation model by providing a pseudo-algorithm that can be used to determine how many replications of a simulation run to perform in order to be confident in the results produced by the simulation.

- Showed how to develop an experimental program to explore specific AIED and advanced learning technology research questions through simulation.
- Illustrated the importance of using simulation in exploring various learning domains where data is not readily available, particularly self-directed, and longer-term learning scenarios. Simulation allows exploration of such domains while also enabling deeper insight into learner models and learning contexts.

1.4 Dissertation Organization

The remainder of this thesis is organized as follows: In Chapter 2, I discuss related work focusing on AIED, longer-term learning, and simulation. In Chapter 3, I introduce a seven-step framework for building a pedagogical simulation and demonstrate how I have used it to guide the creation of the SimDoc model. I next describe SimDoc's conceptual model, enumerating its core components and assumptions. In Chapter 4, I outline how I calibrated and validated the SimDoc model. In Chapter 5, I detail the pedagogical research questions I explored, explaining in more detail the various hypotheses I examined using the SimDoc model. In Chapter 6, I conclude my dissertation by discussing some of the strengths and limitations of simulation, reiterating my contributions, and looking at possible future research directions.

CHAPTER 2

RELATED WORK

This chapter provides a brief description of three main foundational areas for this dissertation research: advanced learning technology research, focused mainly on AIED research; longer-term learning with its challenges and solutions; and the use of simulation within AIED research. The literature discussed in this chapter concerns research that has already been done that I deem related to my research. The concepts discussed are not directly used to specifically inform aspects of the model I created but rather to provide context for my research. Therefore, I will not be pointing out how research discussed here compares and contrasts with my own research. Specific literature that my research critically relies on and that I used to inform specific aspects of my research are discussed in relevant sections of the thesis as the need arises.

2.1 Educational Computer Systems Research

Among the first computer systems to be used to support learning and instruction is PLATO. PLATO was a Computer Assisted Instruction (CAI) centralized mainframe system that was created in 1959 at the University of Illinois [39]. It was designed to offer individualized instruction via terminals to the students [40]. PLATO was among the pioneering early drill-and-practice computer systems used in educational settings in the 1950s through to the 1960s [11]. Due to the rise of popularity of CAI and advancement in technology, many other educational systems were developed in the early 1970s including TICCIT [41], SOPHIE [42] and SCHOLAR [43]. These systems are historically regarded as landmark educational systems as far as Artificial Intelligence in Education (AIED) and Intelligent Tutoring Systems (ITS) research is concerned. TICCIT, a learner-controlled system, had functionalities that were more complex and beyond drill-and-practice systems. TICCIT enabled students to choose their learning strategy component, to decide on when to start or stop, and when to navigate to a previous topic.

SCHOLAR and SOPHIE are early examples of educational systems that incorporated artificial intelligence (AI) techniques. In an educational context, AI is concerned with developing computer systems that support the goals of education, that is to 1) acquire and store knowledge, 2) understand the knowledge, and 3) effectively use the knowledge gained to solve problems and accomplish tasks [44]. The SCHOLAR system [43] was among the first CAI systems developed

based on AI techniques. As such SCHOLAR extended CAI capabilities with the ability to emulate human teachers by offering one-on-one personalized geography tutoring sessions based on Socratic dialogue. The SOPHIE system [42] illustrated how students could learn in an environment that simulated the real-world. Learners could interact with this simulated environment while problem-solving. SOPHIE provided learners with advice as they navigated through their learning tasks. In the SOPHIE system [42], a very detailed representation of an electric circuit from a simple resistor to a complex circuit of complete power supply was modeled to represent the domain knowledge. SOPHIE also modeled diagnostic tactics that allowed it to react to students' misconceptions.

AI techniques are used to develop intelligent tutoring systems that augment learning by helping teachers and learners to maximize learning with the available information [45]. The main goal of any intelligent tutoring system is to provide a platform that allows for interactions with a learner while delivering personalized learning content and feedback. The term Intelligent Tutoring System (ITS) was coined by Sleeman and Brown [46]. ITSs were first developed in the 1970s and 1980s and were sometimes referred to as Intelligent Computer Aided Instruction (ICAI) systems [11]. At first, the inclusion and use of the term 'Intelligent' was not universally accepted leading to some researchers opting for names such as 'Knowledge-Based Tutoring Systems' and others choosing 'Adaptive Tutoring Systems'. Over time, most researchers were content with using the phrase 'Intelligent Tutoring Systems' and its acronym 'ITSs'.

An ITS is an advanced computer-based instructional system that is designed to have characteristics and skills found in human tutors such as the ability to observe students' learning progress, answer questions, and provide relevant feedback and support based on students' misconceptions and learning paths. All these features were made possible with the incorporation of AI techniques. As a result, early ITSs showed some great promise. However, these ITSs still had challenges that were unresolved including concerns over the delivery of effective instruction. As a response to this challenge, Peachey and McCalla [47] proposed an instructional planning framework that was subsequently adopted by practitioners and incorporated in several ITSs. This framework was later augmented in the PEPE system [48], [49] to support content planning involving making decisions on what content to teach and in what order as well as delivery planning which determined how to present the content.

Early forms of ITSs could be grouped into three clusters: curriculum sequencing systems, interactive problem-solving support systems, and systems that intelligently analyze student solutions [50]. These early ITSs were mostly based on cognitive and expert systems principles and therefore were able to help students learn by either answering or asking questions relevant to students' learning tasks [22]. An ITS would achieve its goals by finding misconceptions in each student's knowledge. It would then teach each student according to their knowledge status following a specified set of pedagogical rules [51]. Therefore, it has always been necessary to create models that captured both students' and the domain's knowledge state to facilitate adaptive instruction [52], [53]. Knowledge acquired about a learner is useful in understanding that learner and in recommending personalized learning resources to that learner. To effectively perform its objective, a traditional ITS contains four core components that support its functionality. These four components are (1) the domain knowledge module, (2) the student model, (3) the pedagogical module, and (4) the user interface that facilitates communications between the ITS and human users.

Since the early 1980s, several research communities have sprung up to pursue research into different aspects concerning the use and development of these educational computer systems. Some of these communities include Learning Sciences (LS), Computer Supported Collaborative Learning (CSCL), Educational Data Mining (EDM), Learning Analytics, Technology Enhanced Learning (TEL), Advanced Learning Technology (ALT), Massive Open Online Courses (MOOCs), Artificial Intelligence in Education (AIED), Intelligent Tutoring Systems (ITS) among others. These research communities cover a wide spectrum of ideas and concepts concerning the use of computer technology to aid learning. Note that there is an overlap in the kind of research undertaken by these communities. The main difference is in the level of research depth each community gives to a certain topic(s) of interest.

AIED has three key features that define its nature: openness to new research ideas and areas, willingness to ask hard questions, and keenness to conduct research that is focused on the learner first [54]. The first attribute of AIED is the openness to new research ideas and interdisciplinary themes. This aspect of AIED is very essential because it contributes to the evolution and importance of AIED research [54]–[56]. It is common to find researchers who publish in AIED also publishing in other ALT communities and related technical areas such as User Modeling, Adaptation, and Personalization (UMAP) and Intelligent User Interfaces (IUI). The second

characteristic of AIED research is the willingness to ask hard and deep research questions aimed at providing a precise computational description of what could otherwise be unclear as it pertains to learning [55]–[57]. The third property of AIED is the keenness to focus on the learner-first approach to research with personalization at the center of it [54]. As a result, the personalization of learning environments has been one of AIED’s most important research goals. Studies attest that personalization improves learners’ attainment of learning goals [34].

One of the earliest threads of research that is central to the success of personalization is modeling – the modeling of all key elements in a learning environment [18], [56], [57]. Personalization requires attentive modeling of learning stakeholders (learner, mentors), learning resources (learning objects), and the learning environment. Santos, Kravcik, and Boticario [18] provide a more detailed outline of the various dimensions affecting support for personalization in learning systems; their work focused on research published between the years 2011 and 2016. These dimensions include application scope, interaction and technological devices, and educational domains. In a succinct preamble to the IJAIED 25th Anniversary Issue focusing on AIED research directions for the subsequent 25 years edited by Lane, McCalla, Looi, and Bull [54], it is clear that modeling as a research thread will still be part of AIED future research with an ongoing focus on personalization [58]. Continuous advances in AIED research have led to the development of AIED systems aimed at supporting learning among thousands of learners in numerous learning contexts and domains.

Even with such great success, there is room for improvement. Thus far, most AIED research has focused on learning in domain-specific scenarios with well-defined domain subject matter, such as mathematics [19], [20], physics, computer literacy, etc. [21], [22]. In such domains, concepts gradually build on one another and have fairly structured and constrained tasks that learners perform [23]. Recently, though, there has been some effort towards modeling and exploring longer-time real-world learning environments [24]–[26]. Mentoring [27], personalization [25], [28], and self-directed learning [29]–[33] are important issues in professional development and longer-term learning. From here on, I will use the term AIED system to generally refer to any kind of technology to support learning in any learning context and domain unless I want to refer to a specific kind of learning system, such as an intelligent tutoring system.

2.2 Longer-term Learning: Barriers and Solutions

Longer-term learning tasks can happen in many formal settings that are tailored to take place during the early stages of life where individuals are enrolled in a formal learning institution to pursue education with the aim of obtaining both knowledge and certification [59], [60]. Longer-term learning can also exist in other settings beyond the formal (school) ones, in non-formal contexts. In such contexts, individuals partake in learning on demand with less structure than formal learning settings and can attain professional skills with or without certification [61], [62]. Learning can happen informally in every area of human life where individuals engage in unstructured learning, planned or unplanned experiential learning, that helps them gain knowledge and skills for day-to-day tasks [63]. The need to support longer-term learning is very important as it enables individuals to achieve a sense of self-development, and desired skills.

Support for the recognition and promotion of longer-term learning has been ongoing for some time. Early on, Cropley [62] advocated for an academic system that formally recognizes longer-term learning as a part of learning. According to Cropley [62] and Bagnall [64], longer-term learning is purposeful learning that happens throughout a person's life involving all the three domains (formal, non-formal and informal) leading to an individual's development in every aspect of their life [65]. As such, longer-term learning is not constrained by learning settings or restricted to any age group; it is an opportunity for learners to take advantage of the available resources to achieve longer-term goals. Moreover, it encompasses learning that takes place throughout life that is meant to improve learners' knowledge, skills, competence, and sense of enjoyment [66]. Other scholars have synonymized longer-term learning with continuing education [67], adult learning [68], and higher education including both undergraduate and graduate studies [69]–[71]. Further, there is a plethora of literature on longer-term learning, especially dealing with learning after a person enters the workforce [72] that is focused mostly on non-formal learning.

Support for the importance of longer-term learning is not universal. Some researchers have differing opinions concerning longer-term learning viewing it as a form of control, compulsion, and a way of diverging systemic failure to individuals. They claim the involvement of government agencies in efforts to encourage participation in longer-term learning among citizens creates a new form of moral authoritarianism [73] and social control [74]. Tight [75] postulates that

the perceived notion of longer-term learning as an extension of a career without which some people's careers may stall makes longer-term learning a compulsion rather than an opportunity to grow. According to Coffield [76], some central tenets of longer-term learning are flawed as they aim at diverting the failure of systemic policies from the system to individuals. The extent to which these opposing views have shaped research depends partly on the sentiments of the authors.

While some might argue against supporting longer-term learning, their arguments are weak. If you are worried about unwanted government control, it is important to consider the government's responsibility in preparing its citizens to be competitive in the ever-changing global workplace. The emergence of globalization and advances in technology have led to changes in the nature of innovative tasks [77]. Further, the opposition might contend that citizens are under compulsion to participate in longer-term learning. However, rapid information and knowledge evolution within many domains mean mastery of context-specific information will continue to be a challenging, but necessary, objective for individuals [78]. Research has shown that longer-term learning has the potential to enhance social cohesion by providing individuals of communities the same opportunities to learn, be informed, be able to form networks of peers and experts and be able to solve life's challenges [64], [68], [78]. This social view is seen by the EU-Commission and the OECD as a means to enable people fighting social exclusion caused by the rise of globalization, missing out on early education, and overcoming the disappearance of low-skilled jobs [65], [79], [80]. Finally, there is a humanitarian motivation supported by UNESCO that longer-term learning can affect personal life satisfaction by empowering people, as well as influencing job satisfaction, communal participation and personal health and well-being [64], [65], [81]. Clearly, there are many reasons why individuals should participate in longer-term learning that are motivated by a variety of economic, social, and personal satisfaction reasons.

This concept of longer-term learning is evidently not new; however, what is relatively new is the need for educational systems that support longer-term learning in all formal, non-formal, and informal settings. Educational systems not only provide and maintain knowledge but also prepare and urge individuals to update their knowledge and skills. The emergence of longer-term learning as a new research topic for AIED [82] underscores this need. AIED research covers learning happening in formal settings, for example, classrooms as well as learning happening in

non-formal contexts, such as workplaces, and has the potential to offer personalized, flexible, and effective longer-term learning experiences [45].

The success of longer-term learning lies in the identification of its challenges and solutions to these challenges. Longworth [83] highlights major movements in the nature of education in the last two centuries and identifies the actions that need to be taken in order to fully realize them (see Table 2-1).

Table 2-1 Major Educational Movements in the Last Two Centuries (based on Figure 6.1 from [83])

Education and training C20th	Lifelong learning C21st	Action for change
Educational decision making is rooted in a 20th-century mass education and training paradigm	Decisions are made on individual learning needs, demands and styling of all citizens of all ages, aptitudes and abilities	Find the barriers to learning and dismantle them. Develop and market a strategy based on lifelong and lifewide learning for all

The first of these actions is to find and deal with barriers to learning. The second one is to develop a strategy that supports longer-term learning. So, what are some of the barriers to longer-term learning? The one size fits all concept mostly used in formal learning settings is not perceived to prepare individuals for longer-term learning as it often leads to poor learning culture as revealed in a Glasgow City survey [83]. Further, because formal learning domains are created to address the educational needs of the masses, individual students' learning styles and the need to address individual needs are often ignored [84]. In addition, lack of financial resources at both societal and individual levels is a big barrier to participation in longer-term learning [85]. Lack of finances negatively affects people's ability to travel to educational centers and/or buy facilities that would enable them to study at home. It is crucial to develop solutions to address the barriers to longer-term learning which include lack of personalization, time, and access to learning resources.

The time to use technology to support longer-term learning is ripe. The ubiquitous nature of technology means there is a great opportunity to study the role of technology in supporting longer-term learning more than ever before [86]. Dorf [87] claims that applying technology to education provides an opportunity to realize both longer-term learning, on one hand, and chances

of experiencing social consequences, on the other. In yet another study, Stahl [88] has argued for the need for computers to support longer-term learning in a collaborative learning environment for both design tasks and sharing of information. In addition, Sharples in [89] postulates that mobile phones already have the capacity to support longer-term learning by enabling learners to access learning material from any location at the same time facilitating communication between learners and their peers or their instructors (mentors). There is no question technology can support learners of all ages to participate in the longer-term learning process and acquire knowledge that would be needed in various contexts. The impact of technology on longer-term learning has been an area of research since the 1960s although not within AIED research. Most AIED research and most educational institutes' use of technology to support learning have focused on short learning episodes [85]. In the next section, I briefly discuss some of the research directions that have recently emerged as researchers attempt to use technology to support longer-term learning and mitigate challenges experienced by longer-term learners.

2.2.1 Accessibility and Open Educational Resources

A portion of the population does not attend higher education institutions because of lack of funds. Other times individuals make personal choices not to participate in higher education because they feel there are not made for it. Open Education Resources (OER) is a timely development in the last two decades that helps decrease the financial and accessibility barriers [85]. OER initiatives have seen learning institutions such as MIT publish learning resources for free. The concepts of Massive Open Online Courses (MOOCs) are related to OER; however, MOOCs are mostly concerned with knowledge transmission while OER is concerned with networked learning. Learners who decide on different learning paths need support to identify different yet fulfilling learning paths. Lifelong Learning in London for ALL (L4All) [86] is a web-based system developed to support such longer-term learners in exploring and planning both career and educational choices. The target learners are at least 16 years old and have not attended a higher education institution after their high school education and live around London in the UK. The support mechanism is based on the concept of allowing learners to create learning paths of linked learning content (text and/or images) as a basis for future learning opportunities as well as a record of learning. The possibility of sharing the resulting learning paths with peers forms an important

integrated social feature in the system that creates an opportunity as well as supports collaborative longer-term learning.

Another challenge that longer-term learners face is the need to be in the right place, not just physically but also online. Depending on the learning domain, there are probably plenty of learning resources already available on the Internet thanks to OER and MOOC initiatives, as well as other stakeholders that keep generating more learning content on a continual basis. The main challenge learners face is being aware of where to look and having the capabilities to access available and relevant learning resources that would meet their needs. In an effort to address these challenges, Sharples in [89] proposes an architecture for developing mobile technology for supporting longer-term learning. Sharples used this framework to guide his work in identifying the design requirements for the software, the hardware, and the user interface for a handheld device, HandLeR. HandLeR is designed to play the role of a mentor with the following objectives: support learners in the capacity of a learning companion; support learning by playing the role of a case and concept map archive to suggest the best learning paths and organization of learning resources; and support learners acting as a communication device. HandLeR is designed to be portable, individual and adaptable, available, and persistent.

Longer-term learners also face the barrier of accessing relevant learning resources in a timely manner. Searching for learning material on the Internet has become an integral part of longer-term learners' to-do-lists in their quest to acquire knowledge. The main goal of the CROKODIL project is to support longer-term learners in their search for on-the-job learning materials, more specifically, on collaborative learning endeavors based on learning material available on the Internet [90]. CROKODIL utilizes semantic tagging of learning material and social networking features to encourage collaborative learning among longer-term learners. Social networks allow learners to collaborate easily by facilitating interactions where individuals can ask questions, participate in discussions, and recommend and/or share learning resources. Therefore, social networks not only support communication but also provide avenues for flexible collaborative learning. However, there are challenges associated with this kind of learning that stem from a poor organization of the learning materials. Often, available learning materials are not structured or created following a recognized standard and therefore there is a need to assess the quality of the learning materials. The goal of the CROKODIL project is to support learners to overcome these challenges.

2.2.2 Personalization and Open Learner Models

Adaptation and personalization require tailoring of learning content or system behavior based on a learner's personal characteristics captured in a learner model. Learner modeling is at the center of the adaptation and personalization of educational systems [57]. Since longer-term learners often change their learning goals, location, environments, and learning technologies, modeling and recognizing all relevant learner activities into a unified learner model while considering the different contexts is the main challenge. Consider a typical day of a longer-term learner named John.

John starts his day by reading news articles on his smartphone while having breakfast, as he commutes to work, he reads work-related articles on his tablet, later at work he participates in project discussions taking notes on a work tablet, and in the evening, he joins an online class offering guitar lessons, a long-held hobby on his home computer.

These short learning episodes depict longer-term learning. Longer-term learners are often actively learning in different learning contexts where resources are not known at the design time but, there is constant change in both the learning resource and context. To successfully adapt the vast amounts of learning materials that are readily available in different repositories including the Internet to support longer-term learners, it is critical that information found in different pervasive and ubiquitous devices about a target learner is transformed into a longer-term learner model [91], [92]. Kay [92] proposes an approach for linking and aggregating learner models from different learning contexts into a longer-term open learner model. For example, a longer-term learner model of John would aggregate all learner models about John found in different independent and domain-specific (learning) systems John has used. The idea of a longer-term open learner model is not only helpful in supporting personalization for both short-term and longer-term learning goals but also for transparency and scrutiny. Learning content can be adapted to a learner based on his/her prior knowledge, interests, preference, and learning goals. When considering short-term learning goals, contextual and environmental information could carry more weight in the decision making on what resources to recommend to the learner. A longer-term learner model can be reused to facilitate memory augmentation through external memory [93]. The concept of an open learner model addresses the adaptation barrier in longer-term learning and gives longer-term learners the ability to take control by scrutinizing the representation of their knowledge state in the open learner model.

Supporting longer-term learners to plan and determine their future learning paths is important. A similarity concept based on searching for ‘people like me’ is used in the MyPlan project [94], [95] to extend the functionality of the L4All system by determining personalized learning paths to recommend to a target longer-term learner. The idea is based on a learner-driven three-level process of determining ‘people like me (target learner)’. At the first stage, a learner chooses the features out of their profile to be matched. At the second stage, the learner specifies what part of the historical timeline to consider. Finally, the learner chooses the similarity measure to use based on a provided classification. After these filters have been applied, the system provides a ranked list of candidate timelines from which the learner can make a final decision. Beyond supporting longer-term learners to identify learning paths that are relevant to their learning goals, it is important to support learners in providing flexible learning paths that take into account planning and costs [59].

Apart from empowering longer-term learners to search for their learning resources as done in the MyPlan project, a recommender system can be used. Numerous educational recommender systems have been used to support learners in finding novel learning resources, finding peer learners, and finding relevant learning paths [96]. Unlike in a commercial transaction where a recommendation is often an event that happens once, rarely does learning happen just once; instead, learning happens over time as learners attain different levels of competency. Therefore, educational recommenders not only have to consider learners’ learning goals but also learners’ prior knowledge and learning tasks.

The Learning Networks project supports longer-term learners by connecting distributed learners in highly flexible and learner-centric learning networks [97]. A learning network incorporates many learning resources and different learning activities offered by different stakeholders with the ability to contribute to the learning resource pool either by editing, deleting, evaluating, or creating new learning resources. Longer-term learners benefit from accessing readily available, updated, and evaluated learning resources. Further, learners get to participate in a virtual community of learners where collaboration is made possible.

2.2.3 Student Modeling within AIED Research

Modeling of a precise representation of a student’s knowledge state has long been an essential goal of AIED systems research [98]. This representation of a student’s current knowledge state

was early on referred to as a ‘student model’ but since 2000s the term *learner model* has been more widely used⁴. Among the pioneering student modeling methods are doing model tracing [98], constructing a library of misconceptions called bug libraries [99], representing knowledge using techniques such as Bayesian networks and fuzzy set modeling [13], [99], and constraint-based modeling [100] just to list a few of the techniques. These student modeling approaches diagnose students’ knowledge states through inference about how students use AIED systems.

Self [101] proposed the concept of involving students in the student modeling process. This approach to student modeling led to the emergence of the concept of *Open Learner Modeling* (OLM) [102]. Generally, OLM is about opening up the learner model to the student it represents for scrutiny with various degrees of access restrictions that may or may not allow access to learner models of peers [103]. There are several advantages of allowing students to access their learner models including correcting misdiagnosis of knowledge states, helping learners understand and reflect on their learning, and supporting effective help-seeking and collaboration especially after examining models of peers [102]–[104].

In 2003 Dimitrova [104] introduced a framework for supporting *Interactive Open Learner Modelling* (IOLM), an architecture that allowed students to inspect, discuss, and potentially alter their learner models while collaborating with an AIED system. Evaluation studies have revealed that involving students in learner modeling leads to high-quality learner models and improved student learning. Another framework designed to help explain, compare and assess OLMs is *Student Models that Invite the Learner In* (SMILI) introduced by Bull and Kay [105]. Adaptation in AIED systems, which relies on correct learner models, is important in fostering effective, efficient, and satisfactory learning [105]. Effective modeling approaches continue to be a crucial part of AIED system development; in fact, in a succinct preamble to the IJAIED 25th Anniversary Issue in 2016 focusing on AIED research directions for the subsequent 25 years, Lane et al. [54] make it clear that modeling as a research thread will still be part of AIED future research with more focus on personalization, a sentiment also reflected and supported by Baker in [58].

Model tracing and its generalization knowledge tracing have been among the most widely used approaches to student modeling in the development of AIED systems. A key feature to both

⁴ In this dissertation I am using the terms student modeling and learner modeling indistinguishably from one another. In the AIED research field the term "student" modeling was commonly used starting with Self’s seminal paper in 1974 [241]; but in the 2000s learner modeling became preferred because the word "learner" implied a broader idea of learning than just being a student in a course.

model and knowledge tracing is the need to capture detailed information on the required steps for a student to generate a solution for a given problem [98]. While model and knowledge tracing has been successfully used in many AIED systems, the need for a model that faithfully imitates details of student problem-solving has been a drawback. Another downside relates to the computation power required to analyze recorded student actions against a detailed student model. Further, model and knowledge tracing approaches are costly to develop as it requires extensive modeling of student misconceptions. Nevertheless, depending on the complexity of the problem space, model and knowledge tracing can provide targeted and appropriate feedback [106].

In an effort to overcome the challenges of needing a very specific formal knowledge representation such as that underlying model and knowledge tracing approaches, Ohlsson [100] proposed the idea of using constraint-based modeling (CBM). CBM theory aims to use abstraction to overcome the over-specificity of the student models. CBM captures the state of a student's knowledge in the form of constraints s/he violates or not while solving a given learning challenge [100]. Violations of constraints indicate a lack of adequate knowledge by a student on a specific topic. CBM has been used in numerous AIED systems including the first constraint-based tutor, the SQL-Tutor developed in the late 1990s [99]. The SQL-Tutor is used for tutoring SQL, the database query language. Empirical tests of the SQL-Tutor reveal that students find it easy to use the tutor and that in a controlled study, students who used SQL-Tutor outperformed those who did not use the tutor [99], [107]. In a comparative study with model-tracing, Kodaganallur and Weitz [106] concluded that CBM is best suited to an information-rich problem domain.

2.2.4 Evaluation and Simulation

In the early 1990s Self [52] emphasized the importance of evaluation as part of the AIED system development process. The evaluation contexts and evaluation approaches are crucial considerations in the development of AIED systems to support longer-term learning. Depending on the circumstances, an AIED system designer can decide to perform either or both of the following two forms of AIED system evaluation: formative and summative [52], [108]. Formative evaluation is the process of examining the architecture and behavior of an AIED system being developed to detect potential problems (using, for example, pilot testing or expert-based assessment) to help guide any required modifications. Because of the complexity of an AIED system, forma-

tive evaluation helps in the assessment of various AIED system components including a domain knowledge component, teaching elements, a communication module, a student model, and learning and control components. Summative evaluation, on the other hand, is a rigorous assessment process that is performed to support formal claims about the educational benefits or behavior of a completely developed AIED system.

Empirical evaluation is necessary to examine the effectiveness of an AIED system on students. Among the measures that can be evaluated are an AIED system's affective impact (motivating or uninspiring) and its educational effectiveness (retention, completion rate, learning times, skills transfer) [38], [108]. Greer and Mark in [108] highlight and describe some of the popular ITS evaluation methods used in the early 1990s that are still in use, such as the use of learning curves, kappa scores, simulated students. In [38] they present relatively new approaches that include the use of crowdsourcing, educational data mining, and propensity score matching. In brief:

- Learning curves plot students' number of mistakes against their estimated knowledge level [109].
- Kappa score is often used in a situation where human experts are involved in evaluating an AIED system. Kappa score measures the rate of agreement among the participating experts [110].
- Software systems in the form of simulated learners can be designed and developed to mimic student behavior and therefore be used for evaluation purposes. Simulated learners can be used to study various pedagogical issues including evaluating theory, experimenting with a teaching approach, and validating AIED systems [1].
- Crowdsourcing takes advantage of the wisdom of the crowd concept where through incentives a high number of evaluation participants can be attracted [111].
- Educational data mining involves the utilization of machine learning techniques to identify meaningful patterns from large often fine-grained data sets to make sense of student activities [112].
- Propensity scores aim at removing selection bias during the participant recruitment process. Participants with comparable attribute values are placed in the same treatment groups. The results of similar students are then used for the evaluation process [113].

Although the importance of evaluation of AIED systems was identified and advocated from the early 1990s [52], [108], there was minimal attention given to it at that time as more research focus was on developing the actual AIED systems and their components [108]. This is confirmed by the findings of Roll and Wylie [114] (published in 2016) who analyzed 47 papers published in the AIED journal in the years (1994, 2004, 2014). Roll and Wylie chose these papers to represent AIED's early, mid, and latest (as of 2016) research foci and trends respectively. Roll and Wylie show that only 5% of papers published in the year 1994 included evaluation in their description choosing more to focus on domain and learner modeling. In contrast, in 2004 the percentage of published papers including a system description and evaluation had jumped to 62% and further increased to 71% in the year 2014. Another interesting trend reflected in the Roll and Wylie analysis is that with more publications including empirical evaluations, there seems to have been a shift towards STEM (*Science Technology Engineering Maths*) domains.

As shown above, several learning system evaluation methods are provided in the literature, but the diversity of these approaches gives rise to another challenge that the literature does not address. There is no clear guidance for system designers on which approach to use when considering a longer-term learning domain. Further, most of these evaluation methods are empirical which does not allow generalization and often excludes factors that might affect the end results [37]. When considering which evaluation methods to use, two factors need to be well-thought-out. The first is to consider the focus of the evaluation: is it on the whole system or on a component of the learning system [52]? The second consideration is the type of evaluation to run: is it an experimental evaluation requiring systemic variation of variables or an exploratory study requiring a deep understanding of interacting factors [37]?

The complex nature of AIED systems means that the process of developing and evaluating AIED systems is costly [115], [116]. This cost comes in different forms. For example, the development of a cognitive tutor was costly because it required deep cognitive modeling of skills and bugs, which takes a lot of time and many iterations of system building [116]. Other factors that contribute to the cost of developing AIED systems include the need to consider intercultural dimensions of potential users (learners), their privacy concerns, the need to support the collaboration of thousands of learners, and the need to explore diverse learning contexts [117]. Further, it is costly to run a closely controlled classroom and other real-world experiments [118]. Other

learning goals might require learners to use the system for a longer time, probably months, in order to gather a reliable dataset for evaluation [119].

Generally, AIED systems are evaluated with randomized studies; however, these evaluations are costly and time consuming often requiring experimental designs that require approvals by Ethics boards and recruiting and paying of study participants [37], [120]. When considering longer-term learning domains, the nature of the learning environment means that the financial cost would be much higher, and it would be almost impossible to use human subjects because it would require too much time. In addition, there are feature design decisions that have to be made, for example, deciding to include forgetfulness in the open learner model [92], [121]. This is crucial for representing accurate knowledge of a learner as knowledge and skills decay over time especially because learning in this context happens longer-term, measured in years. Moreover, as I illustrated earlier using John’s learning scenario, longer-term learning happens episodically and in multi-institutional settings. Most evaluation settings do not recreate these important phenomena [85].

Given these evaluation challenges, it is desirable to advocate for a cheaper, faster, and flexible evaluation approach for validating and evaluating design and development decisions underlying an AIED system for supporting longer-term learning. A promising approach is the use of simulation, simulating important features of the learning environment. Analogous to the use of wind tunnels to evaluate the aerodynamics of various airplane components [52], [122], simulation provides an opportunity to experiment with various hypothetical, ‘what-if’, design decisions and learner decisions in a more cost-effective and faster approach than using human learners.

2.3 Use of Simulation within AIED Research

In the mid-1990s VanLehn, Ohlsson, and Nason [1] asserted that technological advances had made it possible to create simulated agents that could exhibit human-like behavior. Therefore, a novel idea of using artificial learners (agent technology) either as co-learners, learning companions, or collaborators was introduced to improve learning [123], [124]. The development of simulated learning environments and simulated learners as agents within AIED systems can be traced back to the creation of intelligent tutoring systems in the early 1970s [35]. In their pioneering article on the use of simulated agents within AIED research, Kurt VanLehn, Stellan Ohlsson and Rod Nason [1] identified three main uses of simulation in learning environments.

These uses are (i) to provide an environment for human teachers to practice their teaching approaches; (ii) to provide an environment for testing different pedagogical instructional design efforts; (iii) to provide simulated learners who can act as companions for human learners.

The third of these uses have had, by a wide margin, greater follow-up research in the interceding years. There have been numerous simulation systems where simulated humans can play an explicit part in the learning environment, for instance as learning companions [125], or as pedagogical agents [124], or as teachable agents [126], [127], or even as mentors [128]. Another example is Vizcaíno's [129] proposed simulated student architecture designed for a collaborative environment with the aim of assisting human learners to avoid situations that would lead to less learning while collaborating with other human learners. Vassileva, McCalla, and Greer [130] provide a succinct review on the impact of the PHelpS system, a system that was based in part on using personalized companion agents to help students to select a helper. There has been research on animated interface agents, knowledge-based learning environments, visual embodiment, and dialogue systems in the early 1990s that spearheaded the change in the functionality of intelligent tutoring systems by enabling the introduction of pedagogical agents [131][132]. These agents can connect and interact with human students.

Similarly, the first use has also received substantial research work where simulated learners interact with human teachers or mentors. The goal is to provide humans the opportunity to gain valuable training by teaching a simulated learner, for example, the tutor-able physics student [133] and SimStudent [134]. SimStudent⁵ is a customizable computational model that enables research in three major research directions: theory building – examining factors that affect learning by observing how SimStudent simulated as a student learns [135]; intelligent authoring – SimStudent helps improve Cognitive Tutoring Systems by modeling the cognitive skills from examples [134]; and teachable peer agent – SimStudent can play the role of a peer learner. As such, SimStudent allows human tutors to learn by teaching it [118], [136].

In contrast, it is surprising that the second role has not only received the least follow-up research among the three roles, but it took a while before research on the use of simulation to test systems design issues started to gather traction. Champaign in [137] used a very simple abstraction of learners and learning objects, a low-fidelity simulation, to design a peer-based ITS aimed at helping simulated learners by suggesting appropriate learning objects. These suggestions are

⁵ <http://www.simstudent.org/> last accessed on February 12, 2019

based on those learners' past performances and the performances of like-minded peers, personalized allocation of learning objects and matching students' needs with beneficial annotations on learning objects. Similarly, Frost and McCalla [138] utilized simple learner agents to examine the impact of peers on learning. They compared the learning outcomes in a situation where there are two types of peer effects: 1) a learner reacts negatively to the success of their peers, 2) a learner responds positively to the success of their peers. Dorça in [34] showed how to use simulation to test three approaches for detecting learners' learning styles more efficiently while saving on time, financial, and human resources. Another 'second use' project is Erickson et al. [2] who modeled a specific learning architecture with simulated learners and learning objects. Aspects of this simulation were informed with data gathered in two different and unrelated empirical studies about human behavior.

2.3.1 Current Uses of Simulation

With the introduction of pedagogical agents in the late 1980s and early 1990s [139], the possible pedagogical strategies of adaptive learning systems became even more complex and at times even involved reversed roles such as having human learners take on the role of tutors and computer applications taking on the role of students [140]. The initial goal of using learning companions was to make AIED systems more interactive and potentially to make them more impactful as learning tools [124].

There are two popular types of AIED systems based on agent technology discussed in the literature: conversational pedagogical agent systems and learning by teaching systems that use conversational agents and teachable agents respectively [35]. As the names suggest, conversational agents are systems that can converse with human learners and teachable agents are artifacts that human learners can teach to complete their learning tasks. Simulated pedagogical agents are used to serve a variety of pedagogical purposes in AIED systems including producing realistic simulations; promoting engagement; improving students' learning and performance; facilitating help-seeking; providing learning companionship; and adapting to learners' sociocultural needs.

Simulated learning environments and pedagogical agents are increasingly used within game-based learning environments and intelligent tutoring systems. However, research results suggest that learning benefits resulting from the use of these teachable and conversation agents

vary based on agent characteristics (e.g., facial representation and gender) and learner features (e.g., demographic background and prior knowledge) [118], [136]. Conversational agents are useful in engaging students and providing a sense of learning companionship as well as facilitating personalized tutoring in different learning domains. Some conversational agents are designed with the ability to monitor learners' learning gains based on the responses provided by the learners. Teachable agents not only give learners an opportunity to learn by teaching but also give researchers a platform to explore various learning theories. Simulated game-based learning environments enable learners to learn by immersion, experimenting with virtual environments and its features without fear of making catastrophic mistakes.

The quality of a student model that an AIED system uses to make its adaptive decisions affects the effectiveness of that AIED system [119]. In an effort to create effective student models, SimStudent is used to develop expert models representing students' cognitive skills in problem-solving tasks [134]. In another study, Dorça [34] showed how to use simulation to test three approaches for detecting learners' learning styles more efficiently while saving on time, financial, and human resources. Another issue is evaluating systems targeted at a special education population, where in addition to a lack of sufficient study participants, there is the difficulty of meeting the ethics requirements associated with studying students with special needs [141].

2.3.2 Potential Use of Simulation

There are at least two challenges associated with the AIED system development process. First, during the design and prototyping stage, it is not often possible to design for all anticipated student misconceptions. These misconceptions occur when students are interacting with an AIED system. To fully evaluate the results of such misunderstandings at the initial stages of an AIED system development process is difficult because of the high financial cost and length of time required [51]. Secondly, while AIED systems are meant to provide personalized instruction to students, the kind of individualization provided by AIED systems is primarily performance-based. As a result, it is not easy for a human learner to establish any form of relationship with an AIED system [142].

As AIED research trends towards supporting learners in longer-term learning contexts, these challenges are magnified. Longer-term learning consists of potentially many learning episodes from different domains, so understanding the longer-term impact of a system design deci-

sion is crucial. In addition, personalized social interactions and recommendations of learning resources are important to achieving longer-term learning goals. Use of simulation and simulated learners for validating such systems is a promising alternative. There has been little use of simulation, however, in the education domain, with the goal of informing educational system designers about how to create environments that support learners and teachers.

Simulation gives a researcher the ability to conduct experiments and shed light on real systems that are otherwise impractical to investigate because of the nature of the environment or the length of the investigation in real time that is required [34], [129], [143]. Simulation can be used to replicate a real-world situation by modeling the target domain's key characteristics and behavior over a span of time. Simulation has been considered as an important decision support tool since the 1950s and has been used in numerous areas including healthcare, the military, crowd behavior, and market or customer behavior [144]–[147]. It is the primary tenet of this dissertation that the time has now come for simulation to be used much more in educational domains, and in particular to become a key element used by the developers of advanced learning technology to help in the design and testing of their systems.

2.4 Summary

AIED research into the use of technology to support learners of all ages and across all domains has led to the development of various AIED systems that are helping thousands of learners in numerous learning contexts and domains. As a result, many learners have had enhanced learning experiences. It is important to note, however, that over time, the ambitious approach to address learning challenges in an increasingly large number of domains in different contexts exposes AIED research to at least two challenges.

First, most AIED research has focused on learning in domain-specific scenarios taking place over relatively short-time frames or experiments done in a laboratory setting. Usually, the focus is on a certain well-defined subject matter, such as mathematics, physics, computer literacy, and other similar domains whose concepts gradually build on one another and have structured and constrained tasks that learners perform. Focusing primarily on these kinds of learning contexts limits the positive impact of learning technologies from reaching a wider population of learners who might want to engage in longer-term learning. Second, as AIED ventures into new learning domains, there is a need for a deep understanding of learning system design challenges

especially concerning learner modeling and evaluating the potential impact of using technology. Evaluation is critical for ascertaining the success and effectiveness of systems. However, when a new domain is first being explored, for example, an ill-defined domain, or in a longer-term learning context, the minimal research focus AIED has placed on such environments means that there is inadequate knowledge as to what approaches to use for evaluating such AIED systems. Further, the nature of these learning environments means that designing, developing, and testing an AIED system would be very costly as it would require substantial amounts of financial, human, and time resources.

Recently, more and more researchers have started to tackle these challenges. Technological advances over the last decade have contributed to the development of better systems and enabled the deployment of these systems in large-scale contexts, for example, the deployment of MOOCs. Moreover, as I discussed in the longer-term learning section above, there have been some efforts towards modeling and exploring longer-duration real-world learning environments. As more AIED systems are developed to support learners engaged in longer-term learning, it is desirable to advocate for a cheaper and faster approach for validating and evaluating the design and development decisions underlying an AIED system. Indeed, a cheaper and promising approach for exploring research issues in such dynamic domains is using simulation.

Simulation model fidelity is an issue that arises when using simulation to study real-world phenomena. Different researchers have demonstrated that it is possible to use different levels of model fidelity to gain insight into various pedagogical research issues within educational technology research. For example, while Champaign [148] used a very low fidelity model, Matsuda et al. [118] used a model with high cognitive fidelity to reach compelling conclusions about the use of ITSs to personalize student learning experiences. Erickson et al. [2] also demonstrated that it is possible to use a medium simulation model fidelity to uncover interesting results. Pedagogical simulation models can be based on real [118] or fictitious [2], [148], [149] learning environments. In choosing the appropriate fidelity of the simulation model it is important to consider both the objectives of the study and the research questions a researcher hopes to explore.

CHAPTER 3

CREATING SIMDOC: MODELING A SIMULATED DOCTORAL PROGRAM

In this chapter I introduce SimDoc, a simulated learning environment meant to capture aspects of a university doctoral program. First, I present a discussion on mentorship and mentoring systems. Next, I give an overview of the doctoral program as an example of a learning environment with largely self-directed learners who are active over the longer term (usually years) and are mentored by supervisors. I then show how I modeled a doctoral program demonstrating how an AIED system designer can model learning environments s/he wishes to explore. In particular, I describe the basic simulation model architecture and its agent models (representing students and supervisors). I conclude this chapter with a demonstration of how I use a combination of available target environment data, data from the literature, and ‘common sense’ assumptions to inform the simulation model.

3.1 Mentorship and Mentoring Systems

Mentorship is a deliberate process where more knowledgeable and often mature individuals in a specific field or profession encourage and teach novice individuals how to acquire relevant knowledge and develop appropriate skills [150]. Mentoring exists in all walks of life. For example, the relationship between Ph.D. students and their academic supervisors [151], or the relationship between apprentices and master craftsmen, can be considered as instances of longer-term mentoring relationships. Short-term mentoring relationships include cases such as the peer review process where there exists a relationship between authors of scholarly work and reviewers [152] or the relationship that exists amongst students who help each other out or collaborate (peer mentorship) in a project or course [153], [154].

Minor [155] argues that group peer-mentoring is the most cost-effective approach because it takes advantage of perceived trust among peers and willingness to receive feedback from individuals considered to be at the same level of knowledge. However, for an effective peer mentoring process, there is a need for a more knowledgeable and experienced individual in the mix and formal structures can be useful [153]. For example, in a learning institution, the participation of both faculty and students is paramount for the success of group peer mentoring. Other factors that affect the quality of the mentoring relationship and hence the effectiveness of mentoring in-

clude differences in expectation, miscommunication, lack of appreciation of each party's circumstances, and trust [156]. These are not dissimilar to factors affecting doctoral supervision, especially where trust and communication (through feedback on written drafts and other issues) between doctoral students and their supervisors is important [157], [158].

Mentoring can take different forms including one-to-one, one-to-many, many-to-one, peer group, and many-to-many [159]. In addition, mentoring can either be classified as informal or formal [160]. Friendships and professional acquaintances form the basis of informal mentoring where a more knowledgeable (expert) individual agrees with no formal arrangement to mentor a novice; hence, there is self-selection of mentors and mentees [161]. This spontaneous kind of arrangement is longer-term and is based on an extension of an existing relationship between two individuals. Further, the goals and outcomes of the mentoring process are not time-bound. On the other hand, formal mentoring relationships are established in the context of an organization in a matchup between a mentor (expert) and a novice [160], [161]. In addition, most formal mentoring relationships are short-term, often lasting less than a year (although some formal mentoring relationships can last many years, as in doctoral supervision). As such, objectives and outcomes are specified and are time-bound. Another difference between formal and informal mentoring has to do with effectiveness. According to [160], [161], informal mentoring is much more effective than formal mentoring because novices receive greater benefits and overall satisfaction in an informal relationship. Some of the reasons for this phenomenon include lack of self-motivation and the short-term nature of the formal mentoring relationship [161]. Therefore, it is important for professional organizations to incorporate aspects of informal mentoring into their formal mentoring processes.

Several systems have been developed that support mentoring including [162]–[164]. Adewoyin and Vassileva in [162] introduce a mentorship framework for serving short-term mentoring that exists between peer reviewers (mentors) and scholarly authors (mentees). *myPAL* [163] is a virtual mentor serving in a longer-term mentorship scenario – a 5-year medical undergraduate course. The structure of the program is based more on professional values found in clinical settings and less on a course-based curriculum. Therefore, in such contexts, it is important to focus on aspects geared towards fostering learners to achieve and complete major milestones as opposed to fine-grained course-work level curriculum details. This is also true for my work on the supervisor-student mentoring relationship in a doctoral program. AutoMentor [164] is a web-

based ecological mentoring game that supports students learning about urban planning and how to handle various stakeholders interested in and competing for land use. The goal of AutoMentor is to mentor students to respond to inquiries in a professional manner.

3.2 Understanding the Doctoral Program

Doctoral programs are complex and dynamic social environments made up of many heterogeneous stakeholders: students, faculty, administrators, and government, who interconnect in a myriad of ways. These connections matter to these stakeholders and affect how they act and react, therefore impacting how the whole doctoral learning environment operates. In addition, they each also have different roles, for example, government acts as a funding body, faculty members act as supervisors and/or instructors, and students as learners but also sometimes as tutors and teachers too.

Doctoral supervision like most other supervision in formal and informal settings is widely viewed as a unidirectional endeavor where the supervisor actively passes knowledge to a supervisee. However, the supervisory process actually provides an opportunity for the supervisor and supervisee to learn from each other [165]. As such, supervision forms the basis for reciprocal learning. Just as in other professional development where training to acquire relevant skills takes years, doctoral supervision is an example of a mentoring process in a formal setting that happens over a relatively long period of time.

Attrition rates, completion rates, and time-to-completion are important factors influencing the perception and experience of a doctoral program by interested stakeholders [3], [166]. Long time-to-completion and high attrition rates are costly financially to the funding bodies and the learning institutions and costly, timewise, to the individuals involved: student(s) and supervisor(s). Research on doctoral program attrition and time-to-completion indicates that over time both the attrition rates and time-to-completion have continued to increase. On average, the attrition rate is currently reported to be between 50% and 60% [166]–[168]. Literature shows that various factors have an influence on doctoral attrition rates, completion rates, and time-to-completion. These factors include: learner’s gender, nationality, ethnicity, age, marital status, time management skills, writing strategy, supervisory style, availability and size of funding, discipline of study, and sense of learning community [3], [151], [166], [169]–[178].

Almost three decades ago Baird [179] explored the effect of the discipline of study on time-to-degree and showed that the average time-to-completion amongst doctoral students varied across different disciplines. Most recent research [175], [180], [181] agrees with Baird's finding that the discipline of study impacts doctoral students' time-to-completion and attrition rates. While these studies, [175], [179]–[181], focused on discipline features such as social integration, a disciplines' job prospects, funding opportunities, supervision ratio and support, others such as [182], [183] focused on the factors affecting doctoral students' persistence at various stages of their doctoral degree.

Studies, [166], [182], [183], have shown that attrition mostly occurs when students have finished their coursework, during the years of study when students are working on their thesis research and when isolation is a prevalent phenomenon. Others, [184]–[186], have asserted that it is not the stage at which the student is at, but rather the lack of both academic and social integration that affects doctoral students' decisions to either persist with their studies or not to persist.

Supervisory style is another factor that affects doctoral students and their propensity to persist to completion that has received substantial research focus, [151], [158], [169], [177], [178], [183], [187], [188]. Heath [178] attributes most of the doctoral students' success or lack thereof on the supervisor-student relationship. His research indicates that, even though in most Ph.D. programs a doctoral student is supervised by 1 or 2 supervisor(s), the number of supervisors involved with a student makes no difference in that student's satisfaction regarding the doctoral program experience. Nevertheless, a supervisor can have far-reaching impact on a doctoral student's progress towards their doctoral degree because s/he plays a crucial role in supervising, guiding, supporting (financial & otherwise), bringing appropriate expertise to the table, fostering a student's learning skills, and assessing the student [177], [178], [189].

Gatfield analyzed 60 items found in the literature associated with supervisor-student relationships and Ph.D. completion to identify four supervisory styles. These four styles are *Laissez-Faire*, *Pastoral*, *Directorial*, and *Contractual*. See Figure 3-1 for a short outline of supervisor traits for each supervisory style as identified by Gatfield. These descriptors suggest a preferred mode of supervision style by a supervisor and not a fixed form of operation. Supervisors may often be required to adapt their supervisory styles based on the needs of students they are interacting with and the stage of progress of the students.

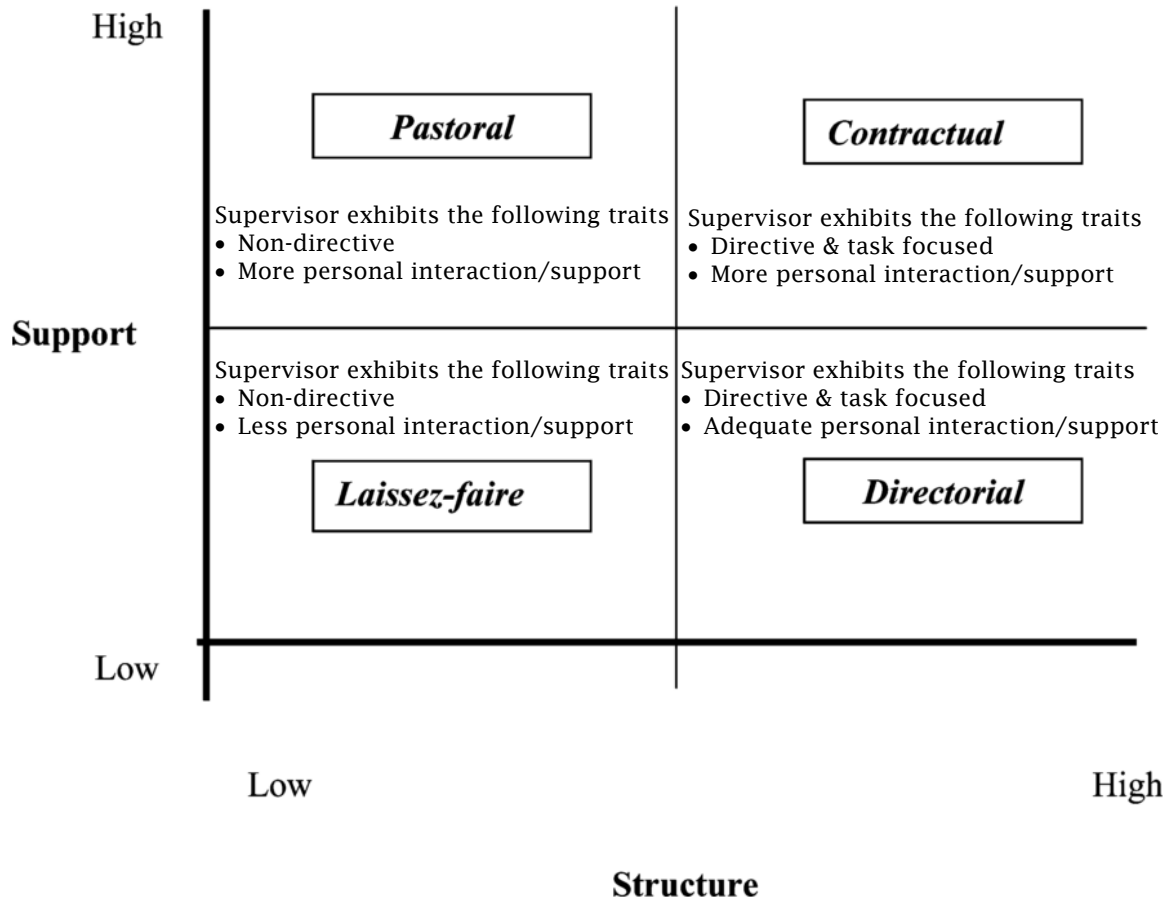


Figure 3-1. The Four Supervisory Styles and Associated Supervisor Traits Based on Supervisory Management Grid [151].

Another factor that is related to supervision that affects the progress of doctoral students is the frequency of meetings that occur between doctoral supervisors and their doctoral students. Regular supervisor-student meetings have been shown to play major roles in the satisfaction of students leading to successful completion of the program [177], [178], [190], [191]. The frequency and the setting for supervisor-student meetings is a balancing act between creating a sense of community and providing individualized one-on-one interaction. For some students interactions that take place in a group-setting create a sense of community and therefore alleviate feelings of isolation, while other doctoral students desire to work independently and require personalized feedback instead [190].

Although several individual and institutional factors have been shown to affect a doctoral student's decision to persist or drop out, research on doctoral students, such as [172], [173], provides only a few recommendations on how institutions can help students overcome the challenges that might lead to long time-to-completion or the desire eventually to drop out. Even with these findings, there is still much to learn especially in a scenario where an adaptive AIED system is introduced. To explore and gain insight into how such adaptation and personalization affects students' progress through their doctoral program, it is important to address at least four major challenges.

First, because of the nature of the doctoral program, it is difficult to determine what personalization data, learning environment data, and learner data, to collect and use for supporting an AIED system's design and personalization decisions [143]. Second, in a doctoral program, learning goals are achieved over longer-term periods. Hence, it is difficult to quickly evaluate the impact of a design decision. Third, where such personal data are collected, there are privacy concerns over who gets to access what subset of the data and how such data are distributed [117], [130], [192]. Finally, educational data mining and learning analytics techniques used for evaluation currently do not allow researchers and system designers to manipulate existing data such that various hypothetical, 'what-if', research questions could easily be asked for purposes of gaining insights into new personalized learning strategies, or new learning systems design decisions [38].

3.3 Using Simulation to Understand a Longer-Term Mentoring Environment

As I have mentioned earlier in section 2.2.4, there are two types of system evaluation: formative and summative. A system designer performs a formative evaluation early on during the system design and development process. It is performed to identify and correct design and development issues including misconceptions about the potential impact of the designed system. By contrast, summative evaluation is performed when the system development has been completed. It is aimed at providing evidence that formally supports the behavior and effects of using the implemented system.

Designing and developing systems to support longer-term mentoring is tasking and it takes a long time to evaluate them. The designers not only need to know what system features to develop, but also must understand the learning context in which the system will be used. The ques-

tion then is how can longer-term mentoring system designers explore and test the impact of various system design decisions and components in an effective and timely manner? I illustrate how this is possible using the SimDoc model: a complete simulated doctoral environment with both mentors and mentees. Such simulation allows a system designer to evaluate many features and their impact in a reasonable time frame without subjecting human mentors and mentees to unfavorable conditions.

In a doctoral program, attrition rates, completion rates, and time-to-completion are important factors that impact students' learning experiences and satisfaction [3]. High attrition rates and lengthy time-to-degree are costly to both the learning institutions and students. Studies show that among those factors that affect time-to-degree and attrition rates are supervisory style and a sense that progress is being made in learning [151], features analogous to peer mentoring and expert-novice mentoring. Developing a mentoring system to support learning in a doctoral program requires an understanding of its functionalities and its stakeholder behaviors.

So, in my explorations of the doctoral program, I am concerned especially with the effects of various policies on time-to-degree and attrition rates. Specifically, I am interested in exploring the following hypothetical, 'what-if', questions:

Q1.How effective would each of the four (4) supervisory styles identified by Gatfield [151] and categorized based on desired meeting frequency be among students with the four different desired meeting frequency levels based on Heath's [178] distribution?

Q2.How effective would each of the four (4) supervisory styles identified by Gatfield [151] and categorized based on desired meeting frequency be among students grouped by the three different levels of latent effort⁶?

3.4 Modeling a Simulated Doctoral Program: the SimDoc Model

In this section, I demonstrate the steps I took to model SimDoc, a model of a doctoral program.

3.4.1 A Seven-Step Modeling Framework

I use a seven-step framework depicted in Figure 3-2 to guide the SimDoc modeling process. The framework is adapted from a social science seven-step architecture for conducting simulation

⁶ A numeric simulation model attribute that indicate how much a student is willing to work on their doctoral work.

studies described by Law [193]. Having such a framework is helpful in building valid simulation models, which in turn a researcher can use to conduct pedagogical simulation studies. It is important to note that iteration within the seven steps is necessary. In the next section, I briefly discuss what happens in each step and how this framework differs from that presented by Law.

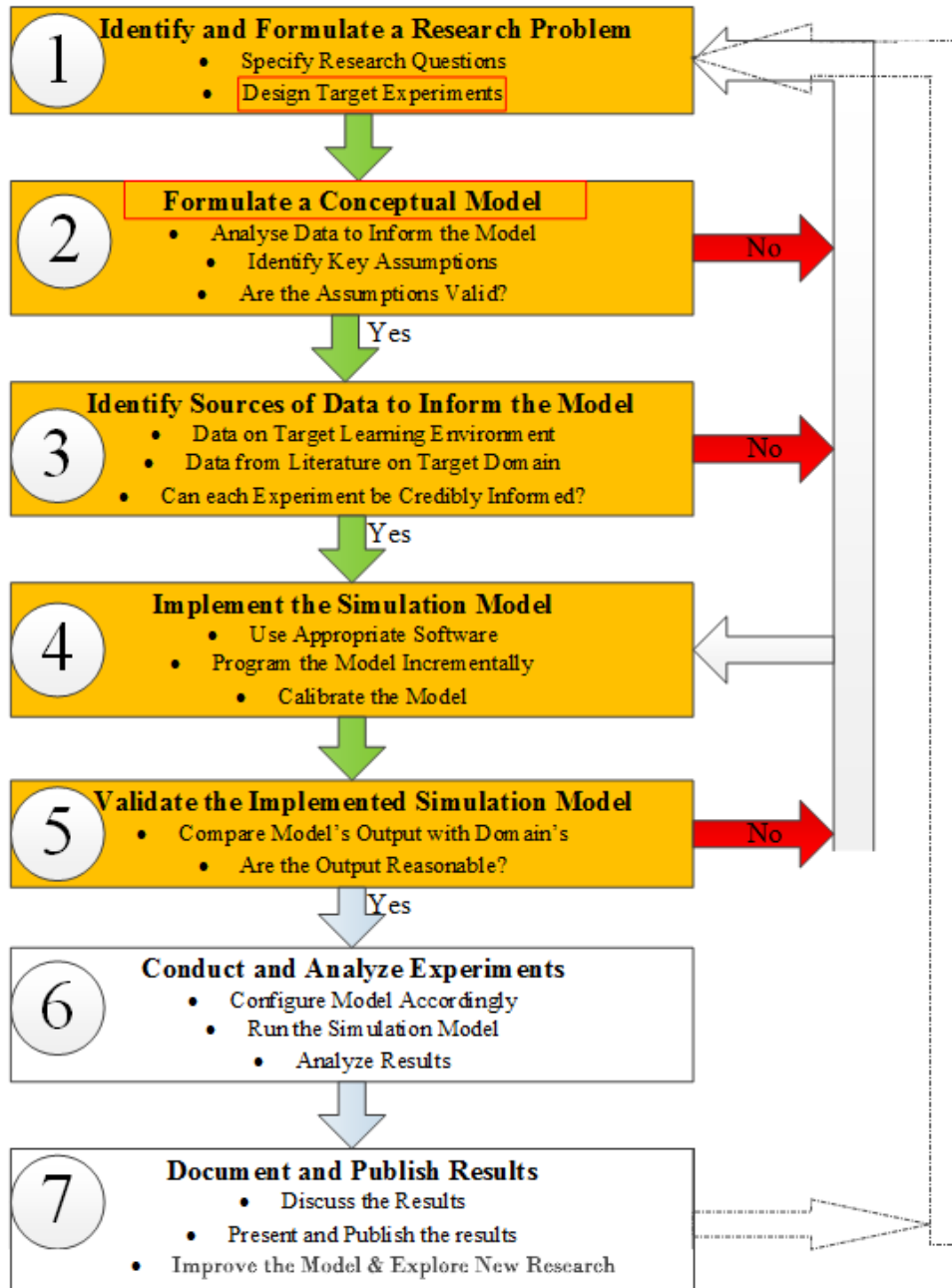


Figure 3-2. A Seven-Step Framework for Building a Valid Pedagogical Simulation Model

Step 1. Identify and Formulate a Research Problem

A researcher/designer can use simulation to explore pedagogical research issues without incurring the high expenses often needed to perform human-based studies. To do this successfully, a researcher/system designer must have clear and specific research questions concerning the phenomenon s/he is interested in exploring. Further, a designer should have preliminary designs of expected experiments s/he is interested in studying. This is one sub-step in my framework that is different from Law's approach [194]. According to Law's architecture, the experimental design comes later in the 6th step. I believe this sub-step best serves the designer and the model building process if performed early on because it helps guide the next two steps: formulation of a conceptual model and data collection. In addition, it helps the designer to think critically about the research problem s/he is interested in exploring.

Step 2. Formulate a Conceptual Model

One of the keys to creating an appropriate simulation model is the understanding of how the organizational structure of the target learning context contributes to how various stakeholders behave. This step involves describing an abstract model of the target learning environment. In addition, where applicable, a designer can incorporate insights found in the literature concerning aspects of interest in the target learning environment to strengthen the conceptual model and improve its validity. It is essential in this step to make sure the conceptual model's assumptions are valid before moving on.

Step 3. Identify Sources of Data to Inform the Model

The third step involves identifying, collecting, and analyzing raw data about the target learning environment to inform the various model assumptions and system behavior. A designer should try to get access to data about the target learning environment and its stakeholders. A designer can then use data from the literature to fill the gaps where raw data are missing. Both sets of data are helpful in informing model parameters, key assumptions, and algorithms, that form the theoretical foundation for the model. It is important to establish that the assumptions of the experiments designed in step 1 can be credibly informed. Otherwise, a revision of step 1 is necessary.

Step 4. Implement the Simulation Model

After formulating a conceptual model, a designer can then program the simulation model using simulation software of choice. It is important that the model is built incrementally, and that the designer calibrate and verify each programmed component of the model against the conceptual model.

Step 5. Validate the Implemented Simulation Model

Where there is existing data about the target learning environment, to ensure that it is valid a designer can compare the simulation model's output with comparable real-world data collected in step 3. In circumstances where real data do not exist, a designer must judge the validity of the model output preferably with help from a panel of experts in the target context and domain. To achieve the desired output, a designer might have to iterate between steps 4 and 5 more than once. With a valid model implemented, a designer can then move to the next steps, conducting and analyzing experiments and finally reporting on insights discovered.

Step 6 & 7. Experimentation and Publishing of Results

After validating a simulation model, a researcher can then use the model to explore various issues and to test various hypothetical, 'what-if', research questions of interest. These two steps are important phases that are useful in reporting the insights gained using the resulting simulation model. Another feature that differentiate this framework from that of Law is using the insight from the simulation results to explore ways of improving the simulation model.

3.4.2 SimDoc's Conceptual Model: Key Components and Assumptions

In this section, I present a practical example of steps 2, 3, and 4 being used to guide the development of the SimDoc model. I have already identified the research questions of interest (step 1) in section 3.3 and will illustrate the validation of the model (step 5) in the next chapter before demonstrating steps 6 & 7 in chapter 5. Step 2 is about developing SimDoc's conceptual model. This conceptual model includes agents and their attributes, and the behavior and evaluation functions needed to model a simulated doctoral program. I have designed SimDoc's conceptual model to have five key components: agents, normative rules, dialogic rules, events, and scenes based on features for building an electronic institution proposed by Esteva et al. [195]. See Figure 3-3

for a depiction of SimDoc’s conceptual model. I model SimDoc’s entities following the agent-based modeling (ABM) [196] technique (and see Figure 3-4, later in the section, for a depiction of SimDoc’s conceptual model as implemented in the AnyLogic™ programming platform⁷). Use of ABM is applicable because it enables me to model entities of interest as agents with various characteristics. Note that I use the term learner model when discussing what attributes, I modeled at a conceptual level, and the term learner agent when discussing the agent that instantiates this model within SimDoc.

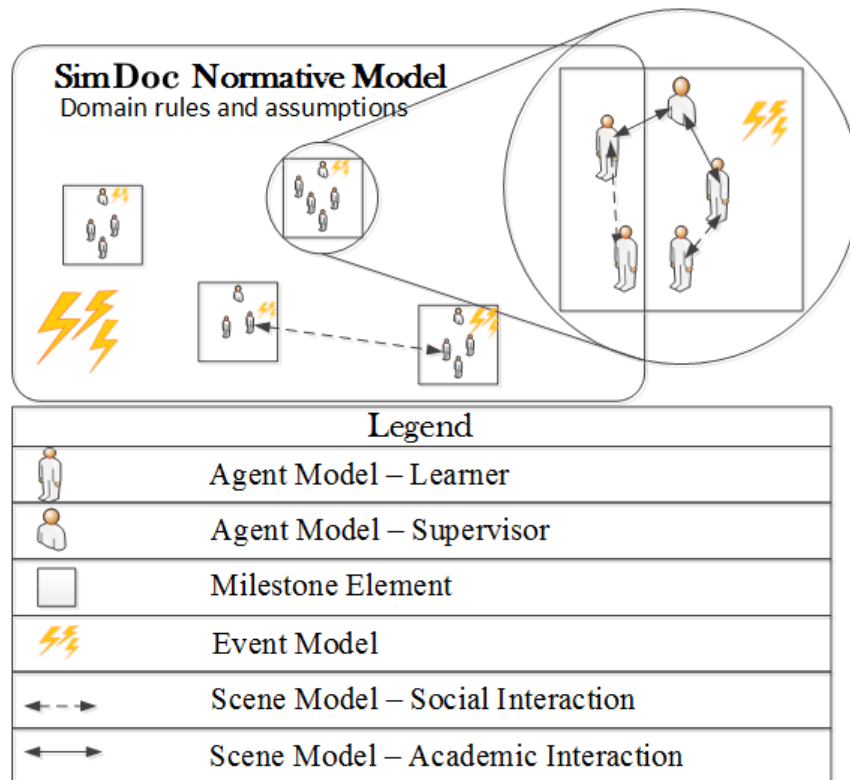


Figure 3-3. SimDoc’s Conceptual Framework - Its Three Element Types, and their Interaction Patterns

Agent Models

I use the notion of agents to represent stakeholders. As stated earlier, different stakeholders play several roles in the doctoral program learning environment. The concept of a role is fundamental in understanding and modeling the activities taking place in a given learning environ-

⁷ I implemented SimDoc model in AnyLogic™, a Java-based platform for modeling and simulation. <https://www.anylogic.com/> last accessed February 12, 2019

ment. A role is defined as a conceptual representation of a stakeholder's collection of established behavior and functions that have observable features [195], [197], [198]. Specific to SimDoc's elements, I model two types of roles: learner and supervisor. As such, I model two types of agents to represent doctoral students (who we will refer to as learners in our simulation) and supervisors. A key element in any simulation in AIED is the agent model underlying the simulated learners. Numerous probabilistic methods have been used in building such agent models that model learner data based on probability distribution functions, as in [199]. The Bayesian Knowledge Tracing technique is one of the most popular approaches for modeling learners' knowledge [200]. A common approach to modeling learner knowledge is combining Bayesian Knowledge Tracing with other methods, for example, Bayesian Knowledge Tracing and Item Response Theory [201]. Still another Bayesian related approach is Bayesian Decision Theory [202].

Another main approach is to use simulation strategies based on cognitive theories in which cognitive problem-solving processes are modeled to produce rich simulated learners. Faghihi, Fournier-Viger, and Nkambou introduced a technique based on neuroscientific theories that combine several human learning capabilities including emotional, episodic, causal and procedural learning to capture a learner model for a cognitive tutoring agent [203]. Rule-Based Modeling often augments the cognitive approach by representing learners' knowledge using procedural and production rules as in the cognitive tutoring systems [135]. These approaches attempt to create high fidelity agent models based on historical interactions between an ITS and learners.

Alternatively, some researchers have chosen to build low fidelity agent models that typically don't need to draw on a large amount of human learners' usage data. For instance, Frost and McCalla [204] model very simple abstractions of learning objects and learners. Each simulated learner is modeled to have an attribute called aptitude-of-learner whose value is between (0 and 1). Similarly Champaign and Cohen [205] have used low fidelity models consisting of learners and learning objects. Each learner is modeled to have a level of knowledge, a value between (0 and 1). These low fidelity models are useful for exploring parameter interactions and general pedagogical tendencies in the simulated learning environment. In this dissertation, I use agent models with medium fidelity. The learner agent model captures the following key attributes: preferred meeting frequency, class workload, class requirements, stage of program, research group,

actual frequency of meeting with supervisor, latent effort (how much they are willing to work on their doctoral tasks), weekly effort-of-learner (how much they actually work on their doctoral tasks), and satisfaction (with the program). See Figure 3-5 for an exhaustive list of attributes implemented in AnyLogic™. In Table 3-1, I provide brief descriptions that demonstrate how each learner agent’s attribute is initialized and changed.

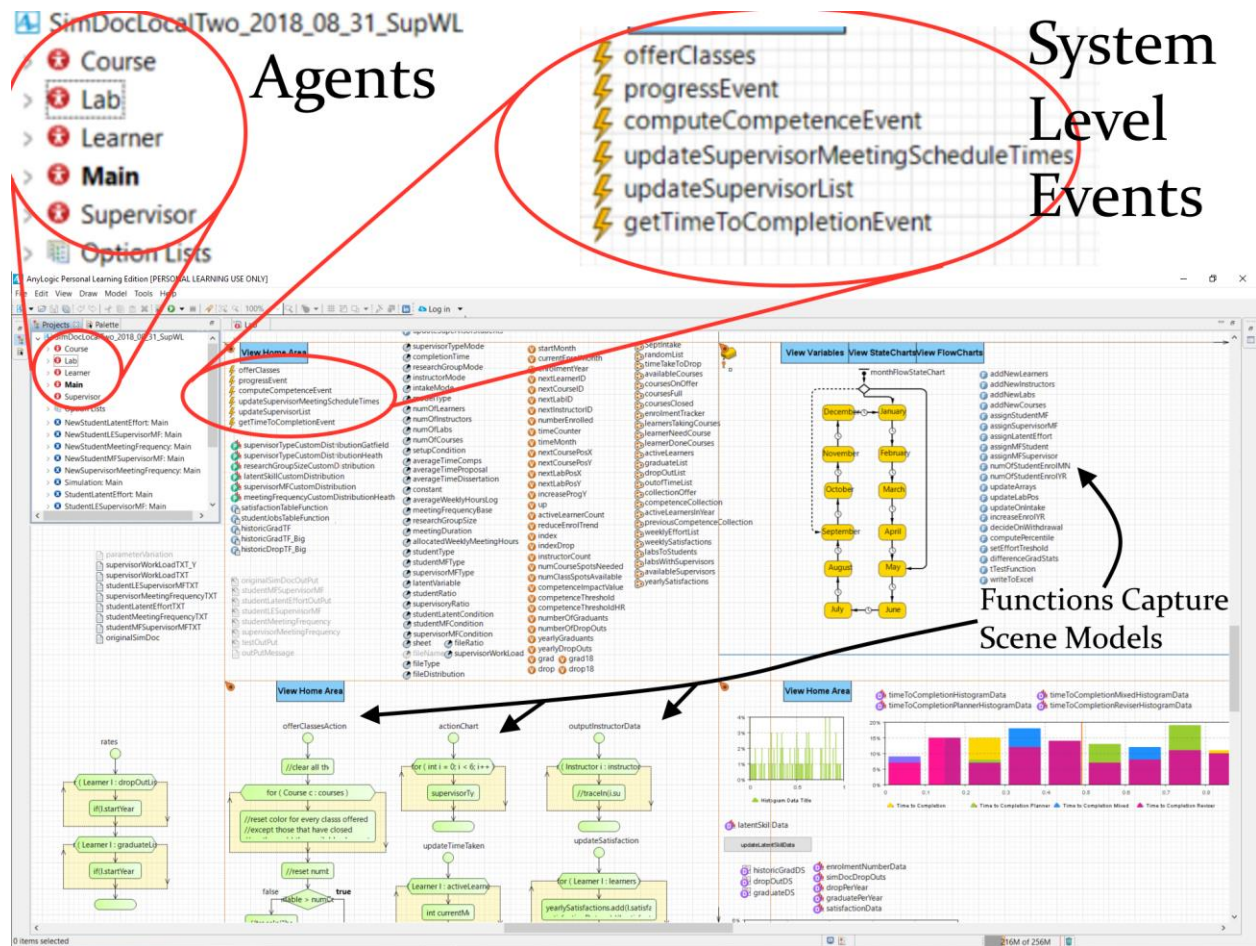


Figure 3-4. SimDoc’s Conceptual Model Implemented in AnyLogic™ Programming Platform

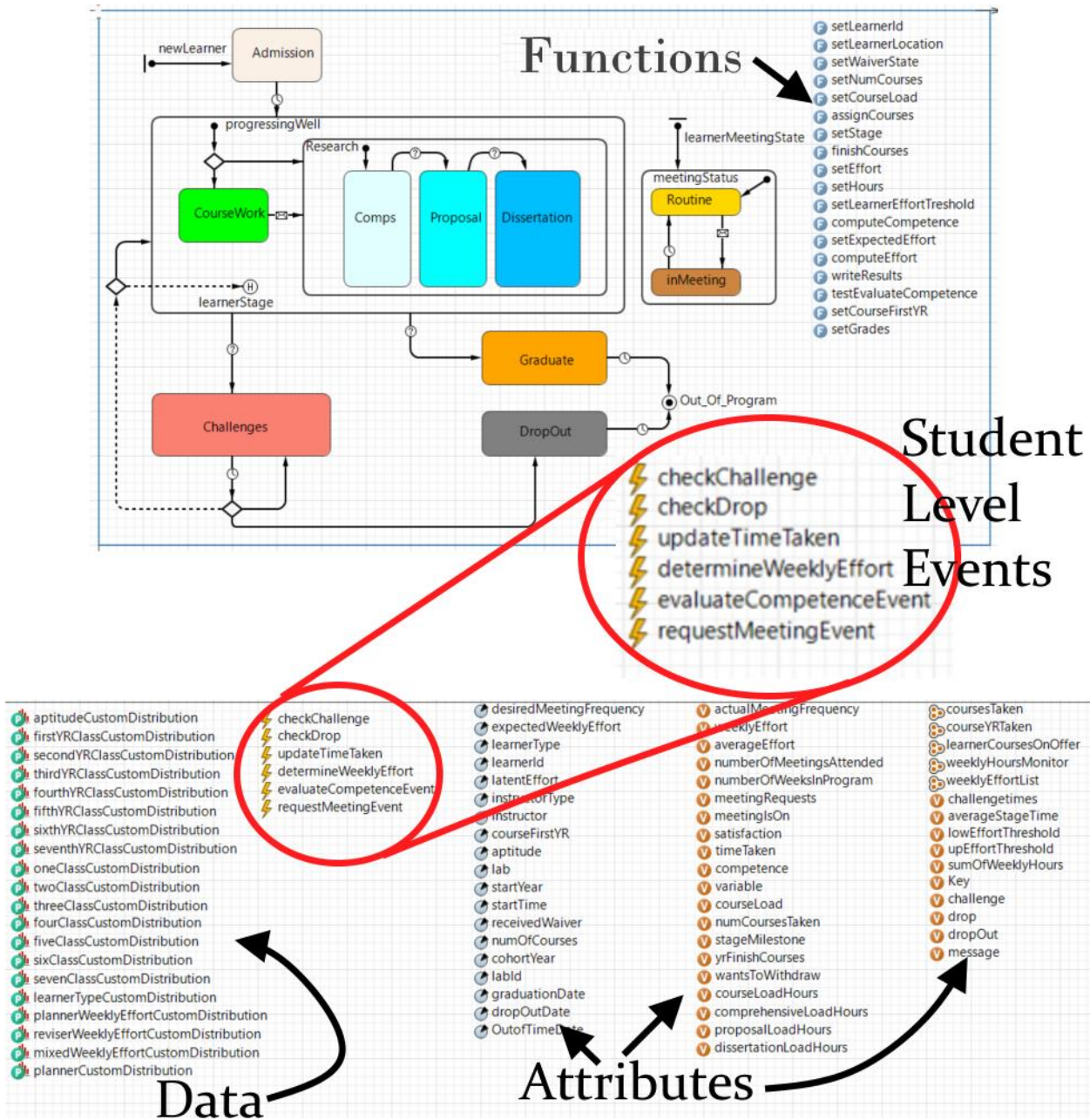


Figure 3-5. SimDoc's Learner Model Implemented in AnyLogic™

Table 3-1. Learner Agent Attributes and their Parameterization

<i>Attribute Description</i>	<i>Data type (value range)</i>	<i>How its value is assigned How its value changes</i>
Learner Agent Id	<i>Integer (>0)</i>	Its value is generated sequentially as agents are added to the simulation <i>Once set, its value does not change</i>
Start Date	<i>Date</i>	Its value is assigned when a learner is added <i>Once set, its value does not change</i>
Stage <i>This attribute indicates the milestone a learner is working on at a given time</i>	<i>String (coursework, comprehensive, proposal, dissertation)</i>	If a learner has class requirements to complete, a learner is first assigned to the coursework stage. Otherwise, a learner is first assigned to the comprehensive stage. <i>A learner agent progresses to the next stage once it has completed the requirements of the previous stage</i>
Research Group <i>Indicates research group a given learner agent is assigned to upon admission</i>	<i>Integer (research group id)</i>	Its value is assigned based on the research group sizes derived from UofS CS department data – indirect assignment of a supervisor since every research group is led by a supervisor. <i>Once set, its value does not change</i>
Class Requirements <i>The number of classes a learner is assigned to take in the course of their program</i>	<i>Integer (0,8)</i>	Its value is determined and allocated based on a distribution derived from UofS dataset <i>Once set, its value does not change</i>
Class Workload <i>Indicates the number of classes a learner chooses to take in one academic year</i>	<i>Integer (0,8)</i>	Its value is determined and allocated based on a distribution derived from UofS dataset <i>Its value changes every academic year</i>
Preferred Meeting Frequency <i>A number indicating the number of interactions a learner desires to have with their supervisor per week</i>	<i>Numeric (0,2)</i>	Its value is assigned randomly based on Heath's [178] findings <i>Once set, its value does not change</i>
Weekly Effort-of-Learner <i>Indicates the number of hours a learner is devoting to learning per week</i>	<i>Integer (0,70)</i>	Assigned based on values deduced from UofS dataset student performance and UofS requirements <i>Its value changes weekly</i>
Actual Meeting Frequency <i>A number indicating the number of interactions a learner has with their supervisor per week</i>	<i>Numeric (0,2)</i>	Its value is generated during simulation runtime and is based on learner's and supervisor's availability <i>This value changes every week</i>
Latent Effort <i>A number that indicates how much a learner is willing to work on their doctoral tasks.</i>	<i>Numeric (1.16,1.32)</i>	Derived from the UofS dataset. <i>Once set, its value does not change.</i>
Satisfaction <i>Measures academic satisfaction.</i>	<i>Numeric (0,1)</i>	Derived based on Actual Meeting Frequency and Desired Meeting Frequency. <i>Its value is calculated weekly.</i>

I model supervisor agents to have the following key attributes: research group, workload, meeting frequency, and preferred supervisory style. See Figure 3-6 for a depiction of a supervisor agent model as implemented in AnyLogic™.

Table 3-2, I provide brief descriptions that demonstrate how each supervisor agent attribute is initialized and changed.

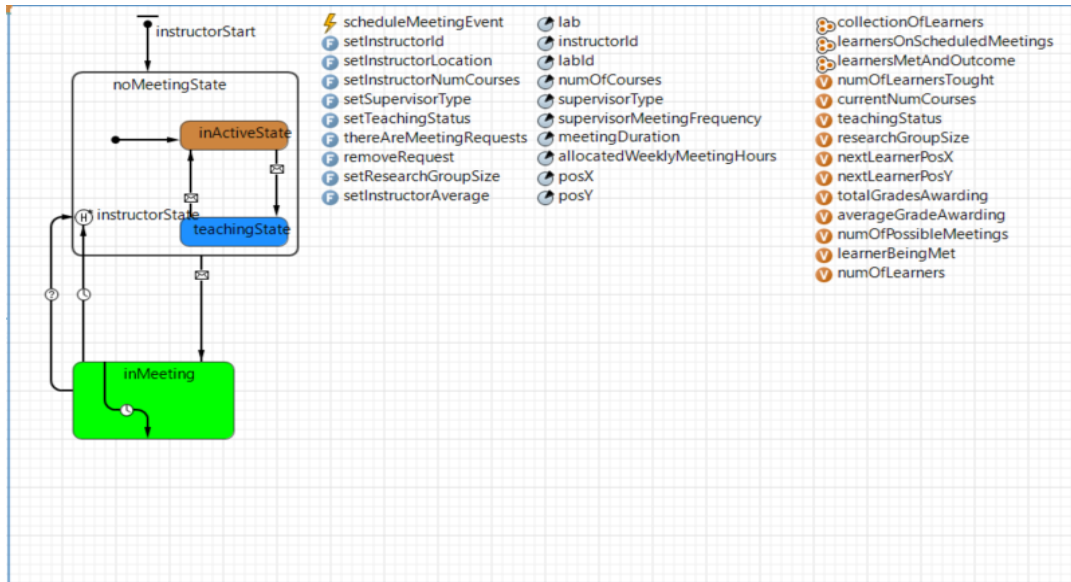


Figure 3-6. SimDoc's Supervisor Agent Model as Implemented in AnyLogic™

Table 3-2. Supervisor Agent Attributes and their Parameterization

Attribute Description	Data type (value range)	How its value is assigned How its value changes
Supervisor Agent Id	Integer (>0)	Its value is generated sequentially as agents are added to the simulation Once set, its value does not change
Workload Indicates the total number of hours a supervisor is committing to their weekly work including preparation, teaching, and meeting time	Integer (0,40)	Its value is assigned based on the number of classes taught in a semester and supervisor-learner meeting commitments Its value changes weekly
Research Group Indicates the ID of a research group headed by a supervisor	Integer (research group id)	Its value is assigned randomly – indirect assignment of learners since every research group is led by a supervisor. Once set, its value does not change
Preferred Supervisory Style A supervisor's preferred supervisory style	Integer (0,2)	Assigned based on distribution statistics derived from the Gatfield [151] model Once assigned, it does not change
Preferred Meeting Frequency A number indicating the number of interactions a supervisor desires to have with their supervisor per week	Numeric (0,2)	Its value is assigned based on Heath's [178] findings and supervisory style Once set, its value does not change

Normative Model

The notion of a normative model allows me to abstract from the complex doctoral program the key requirements and constraints that affect learner agents in the aspects that are important to the questions I am interested in exploring. The normative model represents the environment (programs, research groups, classes, and milestones) and therefore determines the consequences of actions a learner agent might undertake. Specific to SimDoc’s normative model, I model a doctoral program in a multi-level hierarchical organizational structure in the following order: doctoral program, research groups, and stakeholders (supervisors and learners). I view the doctoral program through a lens of four major milestones: the coursework, comprehensive, proposal, and dissertation. In Table 3-3, I provide brief descriptions that demonstrate how each research group’s attribute is initialized and changed.

Table 3-3. Research Group Entity Attributes and their Parameterization

<i>Attribute Description</i>	<i>Data type (value range)</i>	<i>How its value is assigned How its value changes</i>
Research Group Id <i>Research group identifier</i>	<i>Integer (>0)</i>	Its value is assigned sequentially <i>Once set, its value is not changed</i>
Supervisor <i>Leads a research group</i>	<i>Integer (supervisor id)</i>	Assigned randomly <i>Once assigned, no change is expected</i>
Learners <i>Learners assigned to a group</i>	<i>Array [learner id, status]</i>	List of learners are assigned randomly <i>A learners list is updated every admission period</i>
Max Intake Size <i>Research group yearly intake</i>	<i>Integer (0,10)</i>	Its value is determined based on UofS distribution <i>This value is re-assigned every admission period</i>

I model each milestone as an agent. Each milestone contains only four attributes: milestone id, type, required hours of study, and a list of learners who have engaged with it. In Table 3-4, I provide brief descriptions that demonstrate how each milestone’s attribute is initialized and changed. As far as time is concerned, I model the following time granularities: hour – I model the smallest time unit to be in hours to reflect time learners allocate to their studies; week – I assume that as learners work on each milestone, the hours they each allocate to their research work are allocated and tracked on a weekly basis; semester – I model each semester to be 16 weeks, this is the length of time learners will work on a course; year – I model each year to have 52 weeks. Also, I measure time-to-completion in year time units.

Table 3-4. Milestone Entity Attributes and their Parameterization

<i>Attribute Description</i>	<i>Data type (value range)</i>	<i>How its value is assigned How its value changes</i>
Milestone Id <i>Milestone identifier</i>	<i>Integer (>0)</i>	Its value is assigned sequentially <i>Once set, its value is not changed</i>
Milestone Type <i>The Type of a Milestone representing the stage of the doctoral learner is currently in</i>	<i>String [class, coursework, comprehensive, proposal, dissertation]</i>	Its value is assigned on creation <i>Once set, its value is not changed</i>
Hours <i>The number of hours that a learner must engage the milestone to pass</i>	<i>Integer (>0)</i>	Its value is assigned based on the type of milestone <i>Once set, its value is not changed</i>
Learners <i>List of learner agents that have registered in a given milestone</i>	<i>Array [learner id]</i>	Its values are augmented every time a learner enters <i>Where applicable, updates occur (every start of the semester (year)for classes) that a given milestone is assigned to learners</i>

In the SimDoc model, the normative model controls events concerning the admission period, the dissertation period, and the coursework period. See Figure 3-7 for a depiction of this admission process. The admission process has functions that define the admission numbers at each admission period and gives new learners identifier numbers, waiver status, and research group assignment. The SimDoc model is programmed to handle both single-admission and multi-admission period(s) per year. The default setting is a single-admission period – admission happens once a year in September. Everything starts with an admission event where a number of learners are admitted. Once learners enroll, the SimDoc model assigns them to different research groups. Each research group is led by a supervisor who also acts as a supervisor to learners assigned to that research group. At admission, the SimDoc model also determines whether a learner receives a class waiver or not. Learners who receive a waiver straight away enter the comprehensive stage while the rest start in the coursework stage.

If a learner does not experience learning challenges, such a learner continues with their learning endeavors through the doctoral program milestones. However, some learners may experience learning challenges (e.g. feeling of inadequacy, time management issues, and lack of social support, to name a few) and require more support from their supervisors. Such support happens during a supervisor-learner meeting. Reduced numbers of supervisor-learner meetings as well as lack of desired progress, increases the probability of failure – dropout or long time-to-completion. I model learner-supervisor interaction as a contributing factor to the learning out-

comes of the doctoral program. The impacts of such interactions are indirect; for example, a positive outcome increases a learner’s motivation, which in turn causes an increase in effort resulting in timely completion of learning goals.

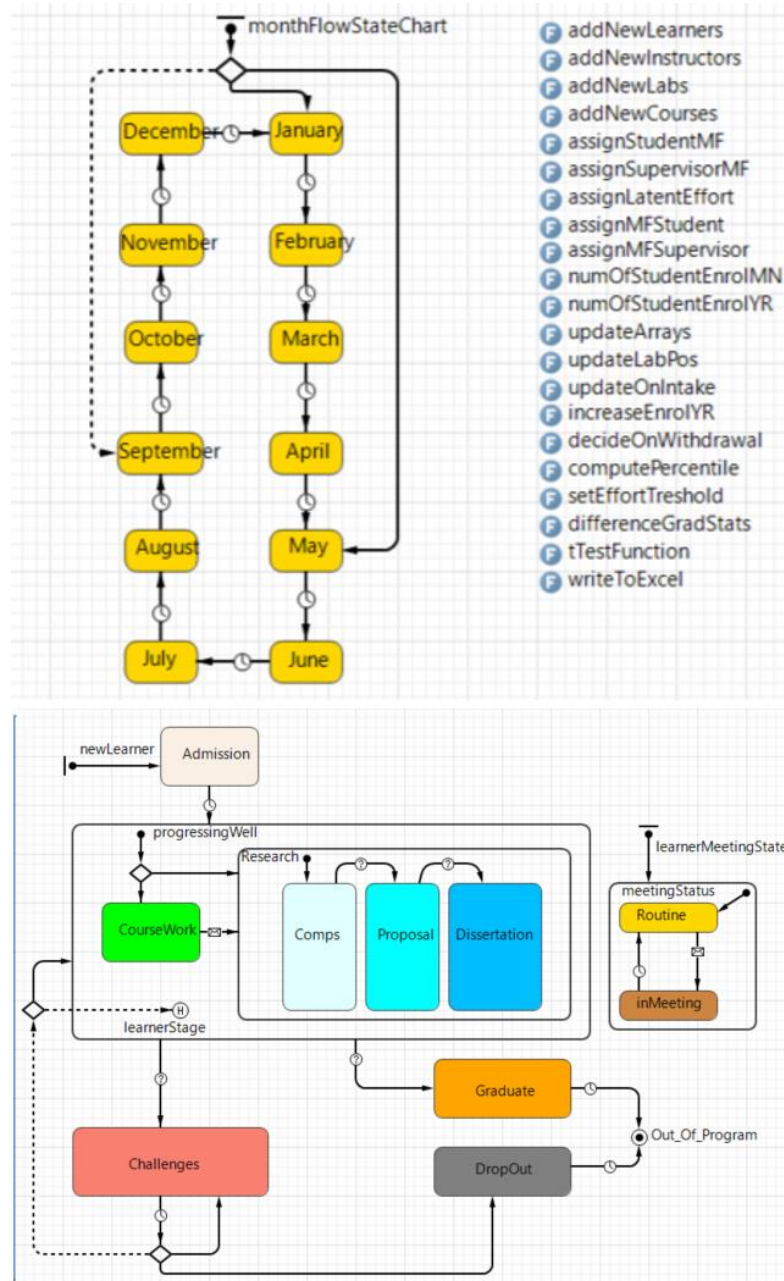


Figure 3-7. A Depiction of SimDoc’s Admission Event as Implemented in AnyLogic™. The SimDoc model is programmed to handle both single-admission and multi-admission period(s) per year. The default setting is a single-admission period – admission happens once a year in September.

Dialogic Model

The dialogic model refers to interaction strategies and communication mechanisms that the agents use within the simulated learning environment. Specific to the dialogic model, I use a message passing mechanism [206] to facilitate communication between learners and their supervisors. I use the message mechanism to trigger a meeting event between a learner agent and its supervisor agent. In case a meeting event is triggered, the agents do not interact or discuss anything. Instead, the SimDoc model determines the outcome of the meeting and updates the learner agent's attributes – which could affect a learner in a positive or a negative way going forward.

Event Model

An event refers to a happening, a period, a change, or a stimulus within a model environment that triggers (re)action by other agents. I model two types of events: system level and agent level, see Figure 3-8 for an illustration of these implementations. I model agents to react to events by initiating appropriate actions. Each event has different enabling conditions associated with it. Further, each event leads to different outcome(s), including other event(s). The normative model manages system level events by scheduling new ones, triggering due ones, and monitoring ongoing ones. One example of a system level event is the learner admission and enrollment process. The admission process in a real doctoral program is complex and involves a lot of stakeholders and processes, but I have simplified it in the SimDoc model, although the goal is still to determine and model the year-to-year enrolment patterns, so they exhibit behavior similar to those shown by UofS data. At UofS there are three admission periods in a year: September, January, and May, but for simplicity, I have chosen to model admissions to happen once a year.

Any agent type can initiate an interaction which I refer to as an agent level event. For example, a learner could send a request for a meeting with its supervisor. When a supervisor receives a request for a meeting, its response is based on its workload and supervisory style. A supervisor agent can accept a request for a meeting with a learner only when they have free time and they have completed assigned tasks (workload and scheduled meeting(s)), all measured in terms of hours. In the SimDoc model, if a supervisor wishes to initiate a meeting, a supervisor would broadcast their availability as a message to learner agents representing all its supervisees. The supervisor can then schedule meetings based on learners' responses.

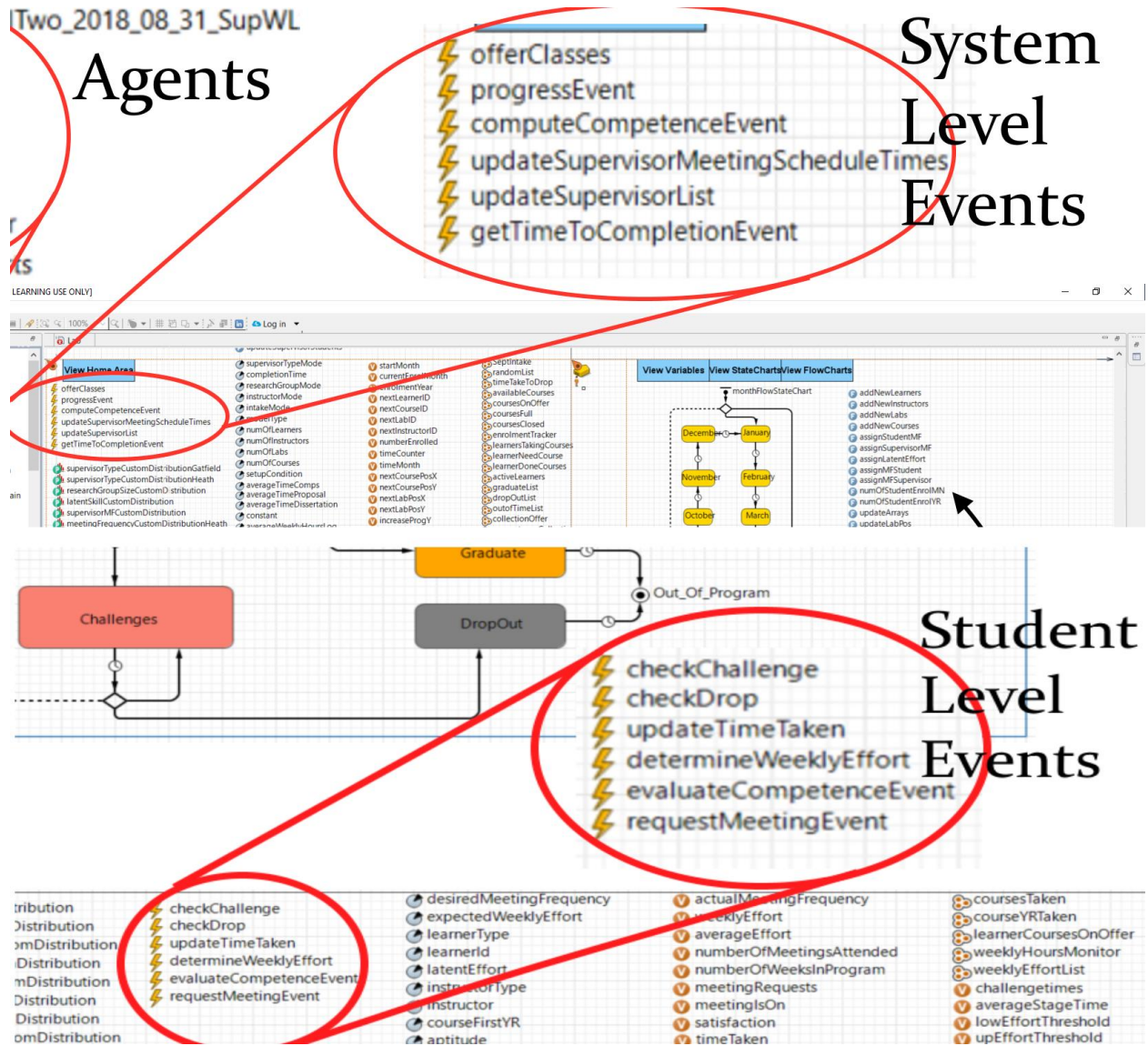


Figure 3-8. Snippet Showing SimDoc System and Learner Level Events

Scene Model

I use the notion of a scene to describe a single interactive session between two agents. Once an interaction event has been initiated, the scene model takes over from the event model. The scene model captures and monitors all initiated events, all responses, potential start and end times of the interaction events, and the outcomes of each interaction event. In a scenario where a learner requests a meeting with their supervisor, the scene would capture all the subsequent events and potential outcomes. For example, if on one hand, the response is favorable to the learner, a meeting between the learner and its supervisor happens. Such an interaction might be the catalyst that

provides an opportunity for each learner agent to go on and achieve their learning goals. On the other hand, if the response is not favorable, a learner's production and progress may be reduced in the coming weeks until it meets its supervisor or withdraws.

3.5 Informing SimDoc's Model Fidelity: Behavior and Evaluation Functions

I have chosen to model a medium-level fidelity simulation model primarily based on a real-world doctoral program, the University of Saskatchewan's doctoral program (UofS). I have created a dataset (the UofS dataset) based on data about the UofS doctoral programs kept in the University Data Warehouse (UDW⁸). For those attributes whose information is not available from this dataset but is necessary in the model in order to explore the research questions elaborated in step 1 of the framework, I get data from other sources. These sources include different departmental web pages and various studies in the literature, specifically, information from [151], [169], [174], [178], [188]. The diverse nature of these studies and learners involved allows for the capturing of a broad spectrum of doctoral learners' behavior.

To gain better insights about any phenomenon using simulation, it is necessary to model two types of important functions as suggested by Erickson et al. [2]. These functions are behavior functions and evaluation functions. Behavior functions inform the decision making of an active agent and dictate interaction patterns between them, other modeled elements, and the learning environment. Evaluation functions determine the outcomes of the various interactions between different agents and/or between learner agents and the learning environment (e.g., determining the outcome of an interaction between a learner and a supervisor).

In the subsequent sub-sections, I focus on describing the steps I take to inform SimDoc's key base-line behavior and evaluation functions both using the UofS dataset as well as 'secondary' data from other sources. Note, based on the research questions I have identified in step 1, the key aspects of the SimDoc model that I deem necessary to inform with data include: the distribution of supervisor types based on supervisory styles; learner types based on desired meeting frequency; and information on doctoral learner-supervisor interaction and learning outcomes. This information is not available in the UofS dataset, so I will return to how I get it later, after a more detailed description of how I informed attributes of the simulation that was available through the UofS dataset.

⁸ I got this data with permission, and my research has been approved by the U of S Behavioral Ethics Review Board

3.5.1 Informing the SimDoc Model Behavior Functions

Step 3 of the seven-step framework is to inform the simulation. As much as possible this should be from data collected in the target environment, in this case data captured about the UofS doctoral program. The UofS dataset contains information on students' class performances, their persistent registration, graduation, and drop out information as described in Table 3-5. Therefore, I can derive important learning outcomes of doctoral students at UofS measured in terms of attrition rates, completion rates, and time-to-completion. Information from the UofS graduate programs webpage⁹ shows that as of 2015, UofS had 86 graduate programs. Of these, 54 programs offered doctoral degree programs. This number is in line with the number of doctoral programs indicated in the UofS dataset I received.

Table 3-5. The UofS Dataset Attributes and their Description Provided by the UofS Data Warehouse

<i>Attribute</i>	<i>Description</i>
Academic Year	The academic year in which a student was registered
Student ID	An obfuscated version of the student's identification number
Major Department Code	An obfuscated version of the department of the student's major
Class Department Code	An obfuscated version of the department offering a class
Class ID	An obfuscated version of the class identification number
Instructor ID	An obfuscated version of the identity of a class' instructor
Supervisor ID	An obfuscated version of the identity of a student's supervisor
Class Count	The number of classes taught by the instructor/Ph.D. student
Grade Range	A banded version of the grade of a student in a numeric class
Grade Mode Description	The type of grade entered for a class; numeric/percentage
PHD Student Instructor	Number of classes (including lab/tutorials) taught by a student

As a first step towards informing the SimDoc model behavior and evaluation functions, I analyze the UofS dataset. This dataset contains information on doctoral students registered for a period of 10 years (2005-2014). Within this period, there were 2291 doctoral students with 52850 data points on class registration. The year 2005 registration includes students who had joined the program earlier than the year 2005. This group of students would add noise to any patterns derived from the UofS dataset. Therefore, I only consider students whose registration started from the year 2006 onwards. Taking this step reduces the population size of students to 1962. From the students who matriculated between the years 2006 and 2014, I use different matricula-

⁹ <http://grad.usask.ca/programs/find-a-program.php> last accessed on February 12, 2019

tion cohort ranges to inform different SimDoc attributes. For example, to inform the length of time students take in their program I consider 2006-2010 cohorts and the 349 students from these cohorts who had graduated. I use these cohorts because I need to be able to trace the last cohort entering in the year 2010 through to their expected completion year, 2014. I decided on this range because according to the UofS guidelines students are expected to complete their doctoral program in 4 years. Even though students are ideally expected to have graduated in 4 years, analysis of UofS data shows that there would still be many students who came in the year 2010 who would still be in the program beyond the year 2014.

Admission pattern

Admission is an important part of a doctoral program that contributes to its dynamic nature. I take values for each of the admissions months for the years 2006-2015 and combine those of the same academic year. By doing so, I obtain a distribution of enrolment numbers for each of the 10 years. I then use the resulting distribution to generate a scatter plot of admission numbers. A sigmoid pattern emerges. Next, I perform a non-linear curve fitting to the scatter plot so that the admission function can be represented in the form shown in Equation 3-1, where N represents the total number of learners who will register in each year, t , and a , c , d , and e are variables necessary for creating a sigmoid pattern that matches the UofS pattern. Subsequently, I run a regression analysis to find values of each of these variables: $a = 0.08$, $c = 1.2$, $d = 3.9$, and $e = 1.4$. Equipped with this equation, I model SimDoc's admission patterns over a period of ten years. I use a two-tailed paired t-test statistical measure to examine the difference between the SimDoc and UofS admission patterns. At 95% confidence level, there is no significant difference ($p\text{-value} = 0.4876$) of the means of the two admission patterns. The maximum difference of the mean can be as low as -0.3411372 and as high as 0.3411372 . Therefore, this result shows that SimDoc's admission patterns are statistically like those observed in the UofS dataset.

Equation 3-1: Informing the SimDoc Model Enrolment Number based on UofS Yearly Intakes

$$N = at + \frac{\sin(ct)}{d} + e$$

Length of Time for the Doctoral Program

The UofS provides guidance on the expected time-to-completion for every doctoral student in the doctoral program guidelines web pages¹⁰. According to these guidelines, it is expected that doctoral students spend at least 40 hours per week on their research work¹¹; when it comes to coursework, students are expected to spend 15-20 hours per class per week¹². Overall, in an ideal situation, doctoral students are expected to complete their degree in 4 years. Time-to-completion analysis of the UofS dataset does not reflect this ideal situation for every student. The data reveals that some students were able to complete their degrees sooner while others took way longer than the 4 years. This variation in time-to-completion could be attributed to many factors. One factor that I consider for the purposes of informing the SimDoc model is a learner's time management skills and allocation of effort to research work.

Rodwell and Neumann [175] say that calculating a student's time-to-completion by just simply measuring the elapsed time from the start (admission) to finish (graduation) is misleading. Instead, they suggest using the concept of full-time equivalent (FTE) workloads. The FTE approach determines a more accurate workload time by weighting and comparing elapsed time for both full-time and part-time engagement as suggested and used by [175], [207]. In an analysis of approximately 20,000 students, Rodwell and Neumann's findings suggest that enrolment choice (choosing between part-time and full-time) is significant in determining a student's time-to-completion and that part-timers finish faster than full-timers when their time to degree is measured in terms of FTE. Note that I am not interested in modeling different types of enrolments in the SimDoc model but, I find the concept of FTE to be very interesting and useful in modeling the time to completion in the SimDoc model. To estimate the total number of hours a doctoral learner in the SimDoc simulation needs to complete their program, I assume that the total number of FTE hours a learner should invest in their doctoral program is determined as shown in Equation 3-2, where t_d is the time in hours it takes to finish a doctoral program, $y = 4$ (ideal number of years), $m = 12$ (months in a year), $w \approx 4$ (weeks in a month), and $h_r = 40$ (expected hours allocated to research per week). As such, an estimated value for t_d would be 7680 hours.

¹⁰ <https://grad.usask.ca/programs/find-a-program.php> last accessed on February 12, 2019

¹¹ <http://artsandscience.usask.ca/psychology/department/gradteaching.php> last accessed on February 12, 2019

¹² <https://www.grad.ubc.ca/campus-community/life-grad-student-ubc/what-expect> last accessed on February 12, 2019

Equation 3-2. Estimated Full-Time Equivalent Hours Required of Learners to Finished Their Doctoral Program in SimDoc

$$t_d \approx ymwh_r$$

Length of Time for each Milestone

Students' time management skills and ability to allocate time to the milestones affect the average length of time required to complete each of the four doctoral milestones and hence affect the overall time to completion. I deduce an ideal duration for each milestone based on UofS program information provided in the program web pages¹³. For example, doctoral students are expected to complete their coursework requirements within the first academic year (i.e. 2 semesters) and their comprehensive exam within the first 18 months¹⁴, whereas, nothing is explicitly mentioned as to when doctoral students are expected to complete their proposal and dissertation milestones.

To determine the length of time a learner should be working on coursework in the SimDoc simulation, I examine the total number of classes taken by all student in the 2006-2010 time-frame and perform the operation shown in Equation 3-3, where: t_{cw} is length of time dedicated to course in years; c is the number of classes; $m \approx 4$ (months - length of a semester); $w \approx 4$ (weeks per month); $h_{cw} \approx 20$ (ideal maximum number of hours per week a student is expected to allocate to their coursework), and t_{cw} FTE ≈ 1920 . When measured in years, this is equivalent to 0.52 years.

Equation 3-3. Average Full-Time Equivalent for Taking Coursework Workload in the SimDoc Simulation

$$t_{cw} = \frac{cmwh_{cw}}{FTE}$$

Since there is no explicit information in the UofS dataset that I can use to derive the average time students need to complete the other two milestones, I use the basic idea of measurement of uncertainty based on standard deviation calculations. This concept suggests that a combination

¹³ <https://grad.usask.ca/programs/find-a-program.php> last accessed on February 12, 2019

¹⁴ <https://www.cs.usask.ca/students/grad-programs/doctoral/index.php> last accessed on February 12, 2019

of any independent normal random variables results in another normal random variable. In this case concerning learner times in milestones, the result should be equal to the average distribution of time spent on each milestone by learners who have graduated. That is, the sum of the expected mean values and variance values for each milestone should be equal to the overall mean and variance values 5.2 and 1.96 (SD^2) derived earlier from the 349 students who persisted to completion. Consider the following computations and assumptions:

- the average length of time for the coursework milestone is 0.52 (0.33);
- the total length of time to finish the program is 5.2 (1.4);
- an assumption that comprehensive milestone time is approximately equal to that of proposal milestone;
- a further assumption that the dissertation takes approximately twice the time required for the proposal.

Then, the resulting length of time expected for each milestone is shown in Table 3-6. The factors affecting agents progressing through the milestones should, ideally, lead to variance similar to UofS in the simulation runs.

Table 3-6. Expected FTE and Actual Mean Time Taken for Each Milestone Derived from Both the UofS Ph.D. Program Description Webpages and Provided UofS Dataset Respectively

Milestone stage	Mean time taken (in years)	
	Expected FTE	Actual mean time taken
Overall	4	5.20
Coursework	0.52	0.52
Comprehensive exam	0.88	1.18
Proposal	0.96	1.30
Dissertation	1.64	2.20

The expected FTE and actual mean time for coursework are the same because this information is available and gleaned from the raw UofS dataset. However, I must compute the expected FTE for the comprehensive exam, the proposal, and the dissertation milestones based on the actual overall average time taken to finish the program (5.2 years), the mean time taken to finish coursework, the actual mean time taken for each of the other three milestones, and the ideal expected overall time to degree (4 years). Therefore, I inform each milestone's length of time using a derived formula as shown in Equation 3-4. For example, while completing coursework is dependent on the number of classes taken by a learner, FTE hours dedicated to a proposal on a

weekly basis should sum up to $(0.96 * 1920 \approx 1843)$ for the same learner to complete the proposal milestone.

Equation 3-4. Computing the Expected FTE for Each of the Three Milestones (Comprehensive Exam, Proposal, and Dissertation)

$$FTE_{milestone} \approx \frac{actual\ mean\ time_{milestone}}{(overall\ mean\ time - 0.52)} (overall\ FTE - 0.52)$$

Learner's Number of Classes

The UofS program website¹⁵ shows that, currently, the minimum number of classes required of doctoral students varies from program to program; it ranges between 2 and 7. Nevertheless, there are situations where this requirement is waived for some students depending on a given student's academic background, qualifications, and achievements. This phenomenon is apparent in the dataset provided by the UDW where 12% of the doctoral student population had not taken any classes. An analysis of the UofS dataset indicates that of those doctoral students who took classes, the number of classes ranges between 1-8. To inform the number of classes each SimDoc learner who does not receive a waiver will be taking, I determine the frequency of each range value to get a distribution of the number of classes taken. Using the resulting distribution (see Figure 3-9, graph *a*), I derive a class load probability distribution function (PDF), $f(x)$, see graph *b*. From this $f(x)$ I obtain a cumulative distribution density function (CDF, see graph *c*), $F(x)$, that I use to inform the number of classes taken by each simulated learner as shown in Algorithm 3-1.

¹⁵ <http://www.usask.ca/programs/colleges-schools/grad-studies/programs/index.php> last accessed on February 12, 2019

For all newly created learners
 Normalize the x -axis (grade points) to become PDF, $f(x)$
 From $f(x)$, obtain a CDF, $F(x)$
For each learner agent which does not receive a waiver
 Draw a uniform random number, x , (bounded by the $F(x)$ upper limit)
End for
 Learner's number of classes = $F^{-1}(x)$
End for

Algorithm 3-1. Simulated Learner's Number of Classes Assignment Algorithm

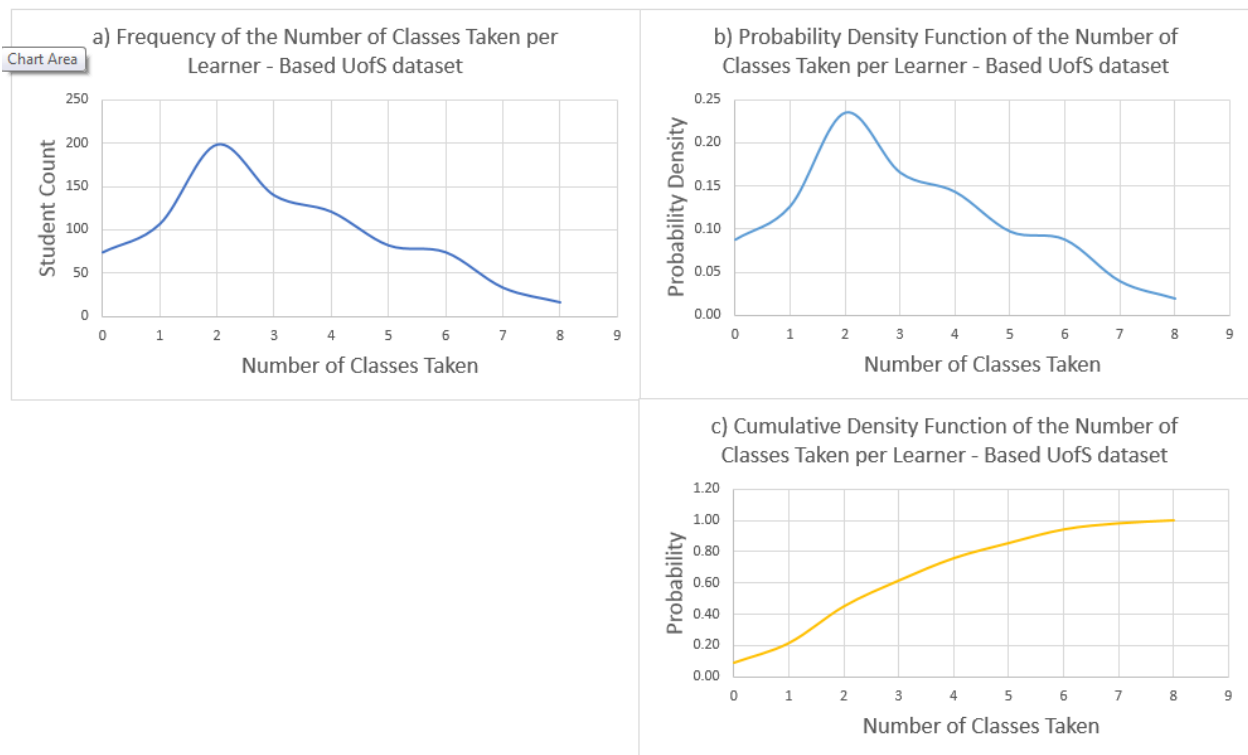


Figure 3-9. Distribution of the Number of Classes Taken per Learner Derived from UofS Dataset

Learner's Coursework Load per Semester

To inform a learner's coursework load – the number of classes a learner is going to take in a given semester - I first consider if that learner received a waiver during admission. Secondly, I check if a learner has completed their course requirements. If not, then I consider the 'year of study' vs the 'number of classes taken' program distribution matrix generated from the UofS dataset as shown in the matrix in Table 3-7. This matrix is based on the class taking frequency of

the entire 2006-2010 cohort. I use this matrix to inform the average FTE taken by simulated learners in their coursework milestone as follows: I consider the year of study of a learner and use it to obtain the corresponding column from the matrix. I then use the resulting values to derive a class count probability distribution function (PDF), $f(x)$. From this $f(x)$, I obtain a cumulative distribution density function (CDF), $F(x)$, that I use to generate the potential number of class(es) to assign to a learner. Finally, I verify if this potential number is greater than the class requirement in the agent model. If so, the learner is assigned a coursework load equivalent to the class requirement and its class requirement attribute is updated to ‘complete’. If not, the learner is assigned coursework load equivalent to the potential number of class(es) and its class requirements attribute is updated with a difference between the class requirements value and the potential number of classes.

Table 3-7. UofS 'Year of Study' vs 'Number of Classes Taken' Program Distribution Matrix

		Year of Study							
		1	2	3	4	5	6	7	8
Number of Classes Taken	1	203	229	116	43	14	1	2	
	2	393	249	62	14	4		2	
	3	362	132	24	6				
	4	270	91	4	4				
	5	235	10						
	6	168	18						
	7	42							
	8	9							

Supervisor Type

I use Gatfield’s doctoral student supervisory model [151] to inform SimDoc’s supervisor types. Gatfield’s model is based on two dimensions: structure and support, which are viewed as axes in a two-dimensional space to identify four supervisory styles. These four styles are *Laissez-Faire*, *Pastoral*, *Directorial*, and *Contractual*. These descriptors suggest a mode of supervision style that a supervisor would naturally adapt which may vary depending on the needs of students they are supervising. However, in the SimDoc simulation, I model supervisory styles to be fixed with the percent composition of each supervisor type based on the ratio derived for an analysis of supervisor characteristics in the Gatfield study ($n = 12$). See the relative position of the 12 supervisors in Figure 3-10.

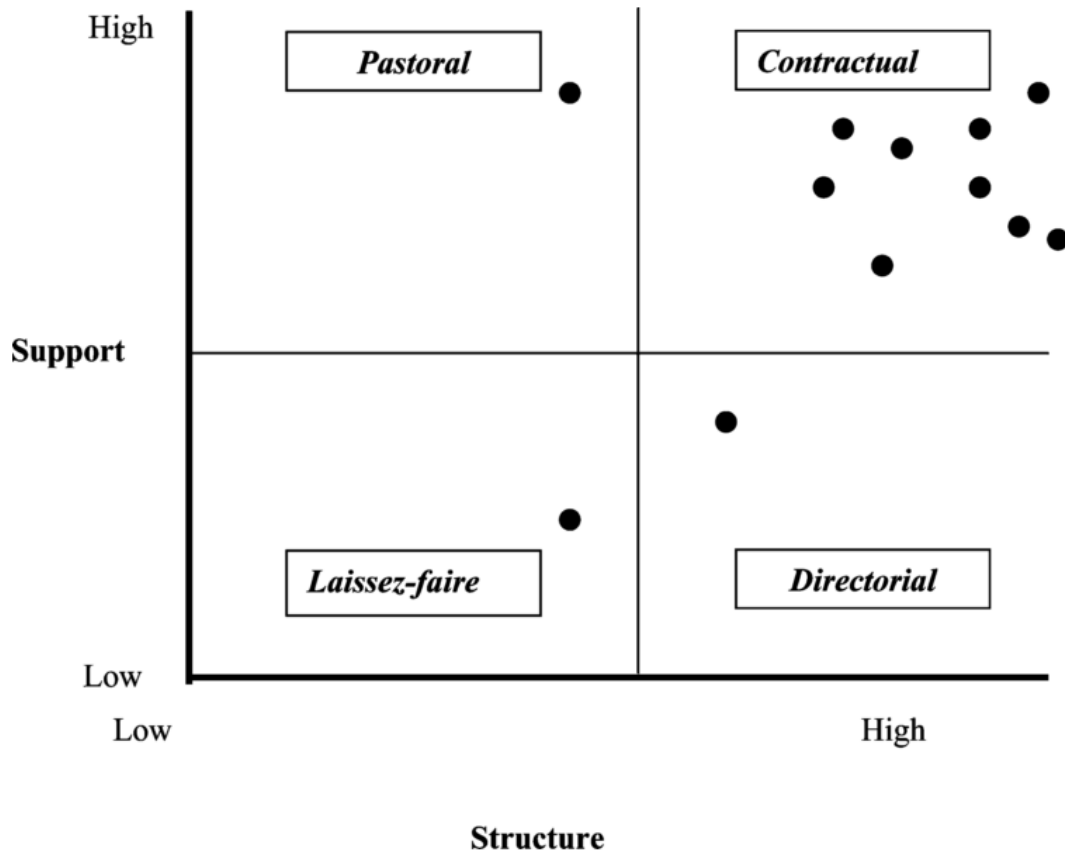


Figure 3-10. Relative Position of the Four Supervisory Styles Identified by Gatfield [151].

3.5.2 Informing the SimDoc Evaluation Functions

It is important to capture how well the simulated doctoral learners are progressing through their program: that is, to design relevant evaluation functions. In a teaching and learning episode, a psychometrics assessment plays an important role in tracking and measuring students' progress in acquiring targeted knowledge and skills. An assessment also reveals the current states of students' knowledge which are helpful in adapting instruction to students as per their learning needs and challenges [208]. In the SimDoc simulation, simulated learners are evaluated based on their time management skills and frequency of student-supervisor meetings. I model these factors to affect students' satisfaction and therefore their progress: withdrawal or persistence, and time-to-completion in case of the latter outcome.

Learner's Weekly Time Management

To determine how many hours a learner allocates to their research, I use an evaluation function that computes the numbers of hours on a weekly basis, as illustrated in Equation 3-5. Based on the outcomes of this evaluation function, the SimDoc model directly updates the values of two attributes: the weekly effort of a learner and the learner's satisfaction level. This update indirectly affects learners' progress and the choice to either persist or dropout.

This evaluation function assumes a linear relationship between a learner's latent effort and a logarithm of time to complete the whole program. This evaluation function is inspired by item response theory (IRT) [201] which has mainly been used to model students' probability of providing correct answers in tests. Other researchers have also used IRT to model students' problem-solving times [209]. To determine the probability of a student providing a correct answer in a test item, IRT considers a student's latent ability, the basic difficulty of an item, a discrimination factor, and a random factor.

Similarly, SimDoc's evaluation function primarily assumes that a learner's weekly effort can be defined in terms of time-to-completion of a milestone unit, their latent effort θ , the difficulty of the current milestone unit (measured in time) c_{at} , the number of classes taken $e^{\frac{x}{8}}$, the desired supervisor-learner meeting frequency f^{mf} , and a stochastic factor z . Therefore, a learner's weekly time spent on their doctoral studies is given by c_{tl} based on an equation adapted from a similar IRT equation, see Equation 3-5. In this equation $f^{mf} \approx$ impact factor based on the exponential meeting frequency base f which is a variable and exponential power mf which is the meeting frequency $\frac{\text{number of meetings so far}}{\text{number of weeks in program by student}}$, $e^{\frac{x}{8}} \approx$ exponential of x which is the number of classes taken in a semester divided by 8 (the maximum number of classes a learner can take). This evaluation function does a weekly calculation of effort-of-learner. The weekly effort-of-learner can vary widely as various factors change over time, so the overall time in the program can only be determined by simulating the learner's progress through the entire program.

Equation 3-5 Determining the Value Learner Agent's Weekly Effort-of-Learner Attribute

$$c_{tl} = \theta \cdot c_{at} \cdot e^{\frac{x}{8}} \cdot z \cdot f^{mf}$$

The Frequency of Learner-Supervisor Meetings

Another factor that the literature shows to affect doctoral students' satisfaction and progress in their doctoral program is the frequency of meetings between students and their supervisors. Seagram, Gould, and Pyke [174] used regression analysis to examine 10 independent variables, and their findings suggest that one of the important factors affecting students' progress is the frequency of meetings with their supervisors. They showed that the more frequently the meetings occur, the faster completion time is. Heath [178] goes a step further to actually examine the frequency of these meetings and their impact on students. Heath's findings show that on average 80% of students who met at least once biweekly expressed satisfaction with their supervision.

Heath's findings also show that up to a third of students who have at least one meeting a week desire to have more frequent meetings during the early and later stages of their research and don't desire as many in the middle stage, see Table 3-8. To provide some fidelity to the determination of supervisor meeting frequencies, I draw on data from studies carried out by Heath [178] and Gatfield [151]. Even with the known benefits of frequent meetings between a student and their supervisor, convening a meeting can be challenging at times because the whole process involves at least two persons with converging and diverging interests [158], [187], [190], [210].

Another contributing factor is the conflict between the supervisory style of a supervisor and the preferred supervisory style based on the desired-support level of a doctoral student. Similar to the inherent nature in supervisors to change their supervisory style preference based on the circumstances at hand, according to Lavelle and Bushrow [188] students do not rigidly adhere to one learning approach but are influenced by input from their learning environment. However, in the SimDoc simulation, I model supervisory style and student type to be unchangeable once assigned. Lavelle and Bushrow further suggest that when students receive clear guidelines, they are likely to choose a productive approach. In contrast, when students receive unclear suggestions, they tend to choose a less productive approach.

I inform simulated learners' desired-support level based on the averages of the frequency of meetings derived from the Heath [178]. I then use insight derived from Seagram et al.'s [174] observations, that suggest that a student who meets their supervisor more frequently tends to complete their program quicker, so the actual number of meetings that occur between a student and their supervisor impact on the student's progress.

Table 3-8. The Frequency of Supervisor-Doctoral Learner Formal Meetings over Time from Heath [178]

Stage of candidature:	Frequency				
	One week or less	Two weeks	One month	2–3 months	6 months or more
Early	38	29	19	9	4
Middle	24	28	27	15	7
Late	36	26	18	10	9

Learner’s Decision to Persist or Leave the Program

Tinto [184] notes that a student’s decision to drop out does not happen all at once; instead their satisfaction waxes and wanes. Therefore, I use a progress evaluation function that computes the kind of influences, positive, neutral, or negative, students experience on a weekly basis. Based on the outcomes of the evaluation function, the SimDoc model updates a learner’s satisfaction value appropriately. The evaluation function considers the new learner’s satisfaction value and compares it with a dropout threshold derived from the UofS yearly attrition rate and milestone completion rates. To inform this dropout threshold, I glean the yearly attrition rates from the UofS dataset. Table 3-9 shows the UofS attrition values for students matriculating in the years 2006 to 2010 included. Table 3-10 shows all milestones’ overall attrition rates. I compute these attrition rates for each milestone based on the average time it takes for students to go through the various milestones as illustrated earlier with Equation 3-3 and Table 3-6. Note that the sum of each column in Table 3-10 matches the results depicted in Table 3-9.

Table 3-9. UofS 2006-2010 Cohorts' Matriculation, Progress, and Attrition Rates

Year - Of study	Enrollment, progress, and drop out numbers										Attrition rates in %	
	2006		2007		2008		2009		2010		Cumulative	Yearly
	Enrolled	drop	Enrolled	drop	Enrolled	drop	Enrolled	drop	Enrolled	drop		
1	154	9	133	7	154	9	200	12	205	10	6%	6%
2	145	6	126	7	145	11	188	7	195	14	11%	6%
3	139	1	119	5	134	5	181	6	181	13	14%	4%
4	132	4	113	4	124	4	169	7	162		17%	3%
5	112	3	95	3	100	8	143		152		20%	5%
6	74	1	63	2	67		111				21%	2%
7	46	3	37		43						23%	7%
8	21		20								0%	0%

Table 3-10. Milestone Completion Rates with Their Derived Attrition Rates for Different Year of Study Periods

Year of study	Coursework		Comprehensive		Proposal		Dissertation	
	Completion %	Attrition	Completion %	Attrition	Completion %	Attrition	Completion %	Attrition
1	100	6	33	2				
2			63	4	20	1		
3			4		59	2	3	
4					20	1	20	1
5					1		38	2
6							29	1
7							9	1
8							1	
Sum		6%		6%		4%		4%

3.6 Summary

In this chapter, I have addressed two of my four dissertation goals. I have provided answers to the following two research questions: (i) how can an AIED system designer create a simulation model for a longer-term learning and mentoring environment and (ii) how can the same AIED designer inform the resulting simulation model in order to improve its fidelity. To answer the first question, I introduced a seven-step framework (adapted from a 7 step process suggested by Law [193]) for building pedagogical simulations. The seven steps are: identifying and formulating a research problem, formulating a conceptual model, identifying sources of data to inform the model, implementing the simulation model, calibrating and validating the resulting simulation model, experimenting with the resulting model, and finally reporting the results of the simulation and exploring ways of improving the simulation model in future cycles of iterative design. I demonstrated how system designers can use this framework to design their pedagogical simula-

tion models by using it to guide the creation of the SimDoc model – a simulation model of a doctoral program. I next described the SimDoc conceptual model, enumerating its core components and assumptions. These components include an agent model, normative model, dialogic model, event model, and scene model. To answer the second question, I outlined how I gleaned appropriate attributes and their values from the UofS dataset and determined the supervisor style and student style categories from studies in the literature.

CHAPTER 4

TUNING THE SIMDOC MODEL: CALIBRATION AND VALIDATION

In this chapter, I continue with a practical example of how to carry out the fifth step in the seven-step framework, that is, I focus on the question: how can a system designer know for certain that they have adequately informed a simulation model in order to trust its output? Often when building a simulation model and is the case with the SimDoc model, reliable information concerning every parameter is not always available. Therefore, it is common to have aggregate data on emergent behavior that relate to a combination of attributes. Sometimes data on a few parameters are available but might not have the detailed breakdown required to inform target parameter values. To address this challenge, calibrating and validating a simulation model is necessary.

Calibration is the process of adjusting selected numeric parameters whose values are not well-known in the computational model for the purpose of improving the match between simulation output and the dataset from the real-world system [8]. The result of the calibration process is to produce a baseline simulation model, the one whose parameter tunings best match the real world system. Validation involves checking that a calibrated computational model's output and behavior are statistically similar to the data output and behavior for the system elements under study [9].

To illustrate how a system designer can calibrate and validate a simulation model, I focus on what calibration and validation approaches a system designer would use and how they would use them to calibrate and validate their model. In the SimDoc case study, the main concept of calibration and validation is to ascertain the validity of the output resulting from running the SimDoc model by matching its output with the UofS dataset in three key aspects: completion rates, attrition rates, and time to degree.

4.1 Calibrating the SimDoc Model: Matching UofS Learning Outcomes

Thus far, the SimDoc model has been equipped with behavior functions and evaluation functions that have been informed with real-world data (some obtained from the UofS dataset and other data drawn from the research literature pertaining to relevant attributes to my research questions of interest) and assumptions. However, before using the SimDoc simulation to run experiments to explore hypothetical, 'what-if', scenarios, it is important to make sure that I start with a ver-

sion of the SimDoc model, the *baseline model*, that as closely as possible matches the real world data. This tuning process is called *calibration*.

Calibrating the SimDoc model involves adjusting values of parameters (attributes) whose values are not known through other means, so that SimDoc's resulting behavior better matches the UofS data. These three attributes are stochastic factor z , meeting frequency base f in Equation 3-5, and supervisor-learner meeting duration. The goal is to get the AnyLogic™ application software to answer the question: 'What must be the values for each of these (unknown) parameters that would make sure the simulation output matches real-world system dataset?'. Calibration reveals whether the current SimDoc model can actually be tuned to reproduce a dataset with similar patterns (of important attributes) to the UofS dataset. In addition, where there is a reasonable number of datasets that cannot be directly used to parameterize a model, calibration enables system designers to take advantage of the available dataset indirectly. In AnyLogic™, calibration uses a (global) optimization algorithm to try to adjust unknown parameter values to find the model attribute values that correspond to the simulation output that best fits the real-world dataset. The optimization algorithm will run the model many times to find the best match for a real-world dataset. Therefore, it is important to know what outcome to match, what parameters to vary and over what range to vary them, and what payoff function (objective function) to use.

The process of calibrating the SimDoc model so its outputs match the UofS dataset consists of 500 simulation runs where I vary systematically the values of the three attributes: z , f , supervisor-learner meeting duration. The "final" values of the three attributes in the best SimDoc model are determined from their values in the simulation run whose graduation and attrition rates most closely match the UofS data set with the highest confidence level. The 132nd run out of the 500 calibration simulation runs is the best one, with a 93% confidence level as shown in Figure 4-1 below.

In Figure 4-1, the upper graph shows the objective function numbers that capture the difference between the simulation output and historical data – the lower the value, the better the confidence level. This objective function is only applicable to the calibration process. The light blue line captures the value for each simulation run while the dark blue line captures the value for the current lowest (best and feasible) objective function value. If the best value was not feasible, the brown line would be depicted on the graph.

The lower graph depicts the graduation numbers each year. The black line represents the data from the real-world system and that is why it is labeled ‘Historic’. The brown line represents the output of the latest simulation run and hence we call it ‘Current’. Finally, the light red line represents the closest match so far between the simulation run results and actual real-world data, so we labeled it ‘Best feasible’. This calibration process leads to a better parameterized computational model with dataset output that best matches relevant data of the real-world target system.

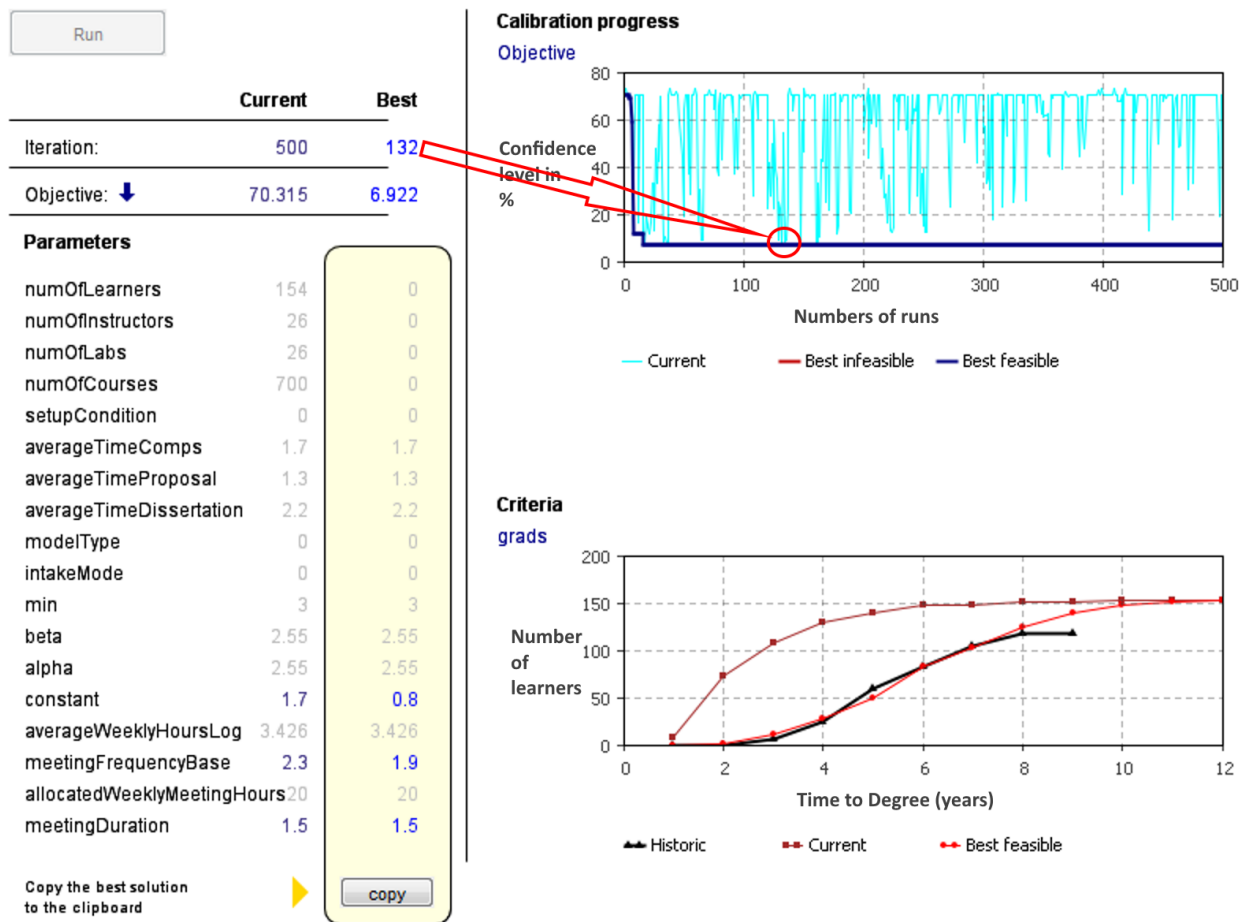


Figure 4-1. A Chart Showing the Results of SimDoc’s Calibration Process

4.2 Validating the SimDoc Model: Exploring the Difference in Number of Runs

The calibration process has been used to tune the parameters of the simulation model whose values are not well known. The resulting calibrated model, however, must still be validated. Since

stochastic elements are often part of a simulation model (as in SimDoc), such validation requires many runs of the simulation model to ensure the model is behaving appropriately. Validation involves checking to what degree a computational model's output and behavior are consistent with the data output and behavior of the system elements under study [9]. This process necessitates prudent experimentation in order to ascertain that the model works as expected [34]. The objective of such experimentation is to determine whether the resulting model is correctly implemented and to show that its output accurately reflects the output of the real-world system of interest [9]. For system practitioners and designers to fully gain the benefits that simulation offers, they need to know how many times it is necessary to run a simulation model in order to validate it. Knowing this would allow practitioners to not only effectively explore and test various hypotheses but to also accept or reject them with confidence.

The baseline model should have been validated against the target real-world environment (the UofS dataset in this case) and the validation results show that the simulation outcomes statistically match the real-world outcomes along measures of interest (yearly graduation, attrition rates, and time to degree in this case). That is, the simulation must be run enough times to be sure that the phenomena of interest are properly explored and have generated results that are stable enough to allow reliable answers to the hypothetical, 'what-if', research questions of interest. It is also important to be sure that the baseline model has not been overfitted, and displays appropriate statistical variability from run to run. This is to give confidence that the SimDoc model in fact accurately captures the relevant characteristics of the UofS doctoral program along the dimensions affecting the issues being explored as determined in step 1 of the process.

The main question I aim to answer in this section is thus how many replications of a simulation run does it take to have confidence in SimDoc's output? I use several statistical measures to determine the answer to this question. I illustrate my discussion by using the results of simulation runs I perform when evaluating the SimDoc model [215].

4.2.1 Determining an Adequate Number of Simulation Runs: in the Literature

Replication Runs within Simulation Research

When using simulation modeling to explore pedagogical phenomena, there are several issues a designer/practitioner should consider. One of the most important decisions has to do with deter-

mining how many replications of a simulation run to perform in order to be confident in the results produced by the simulation [216]. With a deterministic model, a single simulation run is adequate. This issue becomes more challenging when part of the simulation model is based on stochastic elements. One of the solutions that have been used to address this challenge in other research communities is the use of Monte Carlo simulation [217]. Generally, the answer to the question of the number of runs to make in a simulation depends on the question at hand and project-specific constraints. The number may range from 25 to 800 when using Monte Carlo approximation [217], [218]. Monte Carlo methods are used to explore the behavior of statistical measures under controlled situations. Usually, in any simulation study, a summary statistic is calculated after a finite number of replications of a simulation run have been performed. Often there is a between-run variability within the simulation results that depends on experimental set up and the number of replications performed. Thus, determining the number of replication runs is critical.

One approach uses standard deviation and the confidence interval convergence rate to determine the stopping point as described in [216]. This approach has the advantage of minimizing the waste of simulation runs that would otherwise have been performed if too many replications were specified a priori. A similar approach that recalculates sample standard deviation and mean when a new replication run is added until a stopping condition is achieved is proposed by Truong, Sarvi, Currie, and Garoni in [219]. Yet, other methods may consider confidence intervals of measures of performance [219]. The domains for which these methods have been explored tend to be simpler and more predictable than in AIED, where simulation often involves many more variables, a range of statistical sub-models, and pedagogical agents.

Replication Runs in AIED Research

Within the AIED research community, however, this question of how many times should a pedagogical simulation model be run to produce predictions in which the designer can have confidence has received surprisingly little attention. In AIED there is no clear guidance on how to determine the number of runs a practitioner should use to evaluate their simulation model output. Unfortunately, the number of runs practitioners have used is rarely reported, with more general descriptions of the model and/or results being the focus of discussion (as in [220], [221] and [222] for example). In the few papers that have reported on the number of iteration runs used, the number varies greatly, ranging from as low as 2 (see [223]) to as high as 1000 (see [224]) runs.

Sometimes the number of simulation runs is justified based on pedagogical or theoretical grounds. So, in an experiment to determine how students learn composite concepts, Liu in [223] used Bayesian Networks to represent student models as they are a popular way of capturing the relationship between students' competence and their performance. Liu indicates that the simulation needed at least two runs given that the number of concepts being explored is also two. Desmarais and Pu in [225] used Bayesian methods to model a new approach to Computer Adaptive Testing (CAT) based on a theory of knowledge spaces and item graphs with no hidden nodes called POKS (Partial Order Knowledge Structure). CAT systems are used to administer adaptive tests that are used to determine if the examinee is a master or a non-master using the least number of test items. In evaluating the performance of POKS, an average of 9 simulation runs was used.

Most often, though, the number of runs seems to have been arbitrarily chosen. A simulation-based physics tutor, BEETLE II [224], was developed to encourage effective self-explanation using adaptive feedback. The BEETLE II tutor expected students to provide explanations for experiments using natural language in the form of sentences as input. An important statistical significance test that can be done is the F-Score [226]. The F-Score for BEETLE II was evaluated using the approximate randomization significance test with 1000 simulation runs. The evaluation was used to determine whether the system made a correct decision on either accepting or rejecting a student answer. In a proof of concept study exploring a medium fidelity simulation of a multi-agent pedagogical environment, Erickson et al. [2] 100 simulation runs are performed to experiment with three learning approaches for simulated learners. The goal was to determine which learning condition is most desirable between unstructured, semi-structured, and structured approaches to assigning learning objects. In another study to explore the impact of an instructional planner that employed collaborative filtering based on learning sequences, Frost and McCalla [204] used 25 simulation runs to show how different groups of learners would perform. In yet another study, StudyWise [227], researchers used simulated learners to test an application meant to help students memorize collections of basic techniques required for an effective scheduling algorithm. The researchers performed 100 simulation runs to evaluate the pedagogical effectiveness of their system.

4.2.2 Determining an Adequate Number of Simulation Runs: the SimDoc Case Study

Characteristics of a Set of Simulation Run Outputs from a Valid Simulation Model

A system designer can have confidence in the simulation outputs when the following two conditions are met:

- i. The collected replication runs are *stable*; the average of aggregate outputs of simulation runs statistically match the outputs of the target real-world system under study.
- ii. The collected replication runs have enough statistical *variability* from run to run; the model has not been overfitted to the data used to inform the simulation model.

Testing for the Characteristics of a Valid Simulation Model

To test the stability of a simulation model, it is vital to consider the difference between the expected frequencies (average of aggregate outputs of the attributes of interest in the simulation runs – henceforth I refer to this average output as ‘consolidated dataset’) and the observed frequency from the real-world system under study along the same attributes of interest. A stable model should have a consolidated dataset that is statistically similar to that of a real-world scenario, especially, in cases where the simulation model attributes are informed with data from a real-world scenario. To determine whether these frequencies are distributed in a statistically identical manner, Chi-Square test of homogeneity could be used to check for consistency among the consolidated dataset and a Fisher’s exact test of independence could be used to provide an exact test of independence among the overall total frequency count per outcome.

To test for appropriate variability in the consolidated dataset, a system designer can either use a graphical approach, a statistical method, or both approaches. A system designer could choose to examine the characteristics of the results produced by each simulation run or a subset of the total number of simulation runs. Demonstrating variability in a simulation model’s output is important to show that there is no overfitting in the parameterization of the model. Graphically, a system designer could use density plots or box plots for example. To demonstrate whether the variability observed is statistically significant, a system designer could use one-way analysis of variance (ANOVA) hypothesis testing methods. ANOVA is an extension of independent two-sample t-test that is used to analyze data organized in groups (must be at least have three (3)). ANOVA allows exploration of the variance in means of each of the iteration runs. Before per-

forming an ANOVA test, it is important to establish that the three main ANOVA assumptions are met. These are the independence of observations, homogeneity of variance, and the dependent variable is normally distributed. Note, before an ANOVA test can be performed, a Levene's test can be used to check for homogeneity of variances. If the Levene's test is positive ($p < 0.05$) then an ANOVA test could then be carried out.

Determining an Adequate Number of Runs for the SimDoc Simulation

To the best of my knowledge, this is the first attempt (at least within AIED research) to explore the appropriate number of runs there needs to be of a simulation model to get results about which the experimenter can be confident with the outputs of their simulation model. As identified by Ritter et al. [228], many authors fail to report the number of runs used in testing a simulation model. Even when the number is included, the reason behind choosing a given number of runs is barely mentioned. My approach is based on defining characteristics necessary of the simulation output, namely that the simulation runs, collectively, meet statistical standards of stability and variability when measured against comparable real-world data. Essentially, the approach is to run the simulation iteratively, and after each run, take the average of aggregate simulation outputs generated (consolidated dataset) and compare them against real-world data using Chi-Square, Levene, and ANOVA testing methods. Algorithm 4-1 summarizes the approach

I will now illustrate this approach for validating the SimDoc model, produced through the calibration process described in section 4.1. Before performing the ANOVA tests, I establish that the three main ANOVA assumptions are met: independence of observations, homogeneity of variances, and the dependent variable is normally distributed. To do this I perform a Chi-Square test of independence, a normality check, and a Levene's test. Since ANOVA also requires at least three (3) groups of data to analyze, the initial number of runs of the simulation is set to 2. Subsequently, this number is systematically increased until a stopping condition is matched in one of two ways. Either all the statistical conditions of Chi-Square, Levene's test, and ANOVA are satisfied demonstrating that the simulation model produces stable outputs with appropriate variability. Or the stopping condition set by a system designer is met meaning the consolidated dataset could not converge, and thus that the simulation model structure and parameterization need to be improved.

Algorithm 4-1. Pseudocode for an Algorithm that Iteratively Calculates Levene, Chi-Square, and ANOVA P-Values

```
run simulation twice generating two sets of simulation data
set stopping condition
iteration = 2
  consolidate simulation runs outputs
  compare consolidated dataset against a real-world system dataset
  until p-values of Chi-Square and Levene's Test are >0.05 and p-value of ANOVA is < 0.05
    if stopping condition is met
      end until
    endif
    iteration = iteration + 1
    run simulation generating the next set of simulation data
    consolidate simulation runs outputs
    compare consolidated dataset against a real-world dataset
  end until
output iteration
```

Applying Algorithm 4-1 to the best calibrated SimDoc model, I find that when comparing SimDoc's consolidated dataset against the UofS dataset, the statistical conditions of Chi-Square, Levene's test, and ANOVA are satisfied when the input number of runs is 100. The results of running this algorithm on SimDoc's consolidated dataset show how ANOVA's values slowly converge between the runs 96-100, as summarized in Table 4-1. It can be observed that the Chi-Square p-value for each run is greater than 0.05, therefore the ANOVA requirement of independence is met. Similarly, the Levene Test's p-value for each run is also greater than 0.05, so the ANOVA assumption of homogeneity is met. Since all the statistical conditions of Chi-Square, Levene's test, and ANOVA are satisfied in the 100th iteration run, when the ANOVA p-value <0.05, more runs are not necessary. So, in the case of the SimDoc simulation, appropriate statistical significance on the relevant measures is achieved with 100 simulation runs, and fewer runs won't give us this significance.

Table 4-1. P-Values for Levene, Chi-Square, and ANOVA Tests for the Simulation Runs (96 to 100)

<i>run</i>	<i>Levene</i>	<i>Chi-Square</i>	<i>ANOVA</i>	<i>run</i>	<i>Levene</i>	<i>Chi-Square</i>	<i>ANOVA</i>
100	0.17	0.28	0.04	95	0.15	0.28	0.08
99	0.2	0.28	0.06	94	0.17	0.28	0.07
98	0.18	0.28	0.06	93	0.16	0.28	0.08
97	0.18	0.28	0.05	92	0.16	0.28	0.12
96	0.16	0.28	0.06	91	0.15	0.28	0.11

4.2.3 Confirmation of Stability and Variability in SimDoc’s Output

Confirmation of Stability

In this section, I explain in more detail how I test for SimDoc’s stability by elaborating on the results achieved by running Algorithm 4-1. The results show that 100 runs of the SimDoc simulation are adequate to gain simulation output that is statistically similar to the UofS dataset. For clarity, I consider only one output measure from the simulation: learners’ time-in-program. I compare the time-in-program in the consolidated dataset (averaged over 100 runs of the simulation) against the observed time-in-program in the UofS dataset. Table 4-2. shows the time-in-program frequency counts of the UofS students and the time-in-program frequency counts of the consolidated dataset (divided into students who graduated, those who withdrew, and the cumulative total of these two figures). These comparisons are depicted graphically in Figure 4-2.

Table 4-2. Frequency Counts of Time-in-Program Between the SimDoc and the UofS Students

<i>Time(years)</i>	<i>Graduated</i>		<i>Withdrew</i>		<i>Cumulative</i>	
	Real World	Simulation	Real World	Simulation	Real World	Simulation
0	0	0	9	10	9	10
1	0	0	6	2	6	2
2	0	0	1	2	1	2
3	6	6	4	6	10	12
4	16	21	3	3	19	24
5	35	29	1	5	36	34
6	27	27	3	8	30	35
7	22	19	4	4	26	23
8	11	7	1	0	12	7
9	1	0	4	5	5	5

To determine whether these frequencies are distributed in a statistically similar manner, I conduct a Chi-Square test of homogeneity to check for consistency among the yearly distributions, separately for each column: Cumulative, Graduated, and Withdrew. Since the simulation model was informed and calibrated based on the UofS dataset, I expect that the consolidated dataset is statistically similar to the UofS dataset. Therefore, my null hypothesis is that the frequency counts of the UofS dataset and the consolidated dataset are equally distributed. Thus, the alternative hypothesis is that there is a difference between the distributions of the frequency counts. For this analysis, the significance level I use is 0.05. I then apply the Chi-Square test of homogeneity to the cumulative contingency table and compute the degree of freedom, the Chi-Square test statistic, and p-value. Since the p-value is more than the significance level (0.05), I accept the null hypothesis that the frequency counts are statistically consistent between the UofS dataset and the consolidated dataset, χ^2 (df=9) = 5.0904, p = 0.8264.

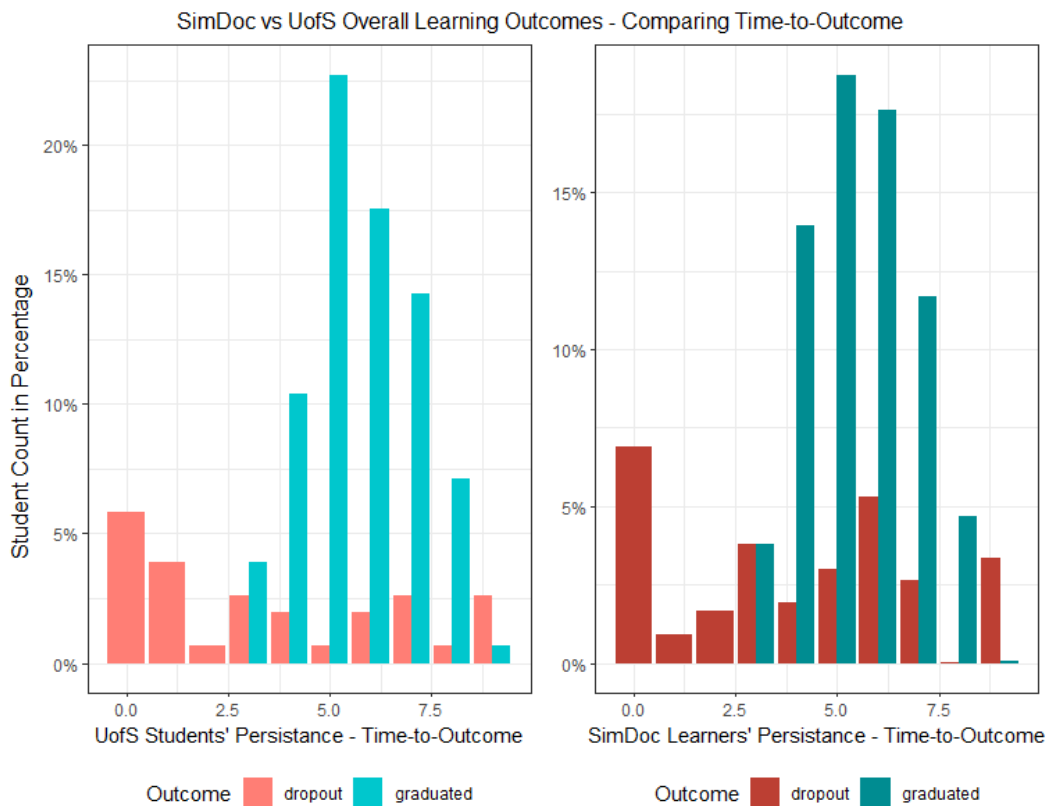


Figure 4-2. Comparison Between UofS Students and SimDoc Learners Time-to-Outcome

The second analysis is to determine if the distribution of frequency counts in time-in-program among the graduated learners are similar. Given that the distributions of the cumulative frequencies are statistically similar, I expect that the real-world graduated frequency counts are like the simulated graduated frequency counts per year. Thus, my null hypothesis is that the frequency counts of the UofS (graduated) dataset and the consolidated (graduated) dataset are equally distributed. As such, the alternative hypothesis is that there is a difference between the frequency counts between these distributions. As in the first analysis, I choose a significance level at 0.05. I then conduct the Chi-Square test for homogeneity and the results show that we can accept the original hypothesis since the p-value is greater than the significance level, and thus the frequency counts are statistically consistent between the UofS (graduated) dataset and the consolidated (graduated) dataset, χ^2 (df=6) = 2.9945, p = 0.8095.

The third analysis is to assess whether the distribution of frequency counts in time-in-program among the learners who withdrew were similar between the real-world dataset and simulated dataset. Given that the distributions of the cumulative datasets were statistically similar, my null hypothesis is that the UofS dataset withdrawal frequency counts are similar to the consolidated dataset withdrawal frequency counts per year. As such, the alternative hypothesis is that there is a difference between the frequency counts between these distributions. As in the first analysis, I choose a significance level of 0.05. I then conduct the Chi-Square test for homogeneity and the results show that there is no significant difference in the distribution of frequency counts per year between the UofS (withdrew) dataset and the consolidated (withdrew) dataset, since the p-value is greater than the significance level, χ^2 (df=9) = 7.9344, p = 0.5408.

Table 4-3. shows the overall total frequency count per outcome for the UofS and the consolidated SimDoc datasets. Since the resulting contingency table is small (2x2), to test whether the proportions for one nominal variable are different from other nominal variables, the Chi-Square test of homogeneity is not recommended but instead, it is advisable to use a Fisher's exact test. In this analysis, I am exploring if the frequency counts per outcome between the UofS and the consolidated datasets differ. My hypothesis is that the proportions of the outcome variables are not the same between the real-world and the simulation datasets. Therefore, the alternative hypothesis is that there is no difference in the proportion of the frequency counts in the outcome variables. I then conduct a Fisher's exact test which yields a result with p-value = 0.3005,

indicating that I can accept the hypothesis that there is no significant difference in frequency counts per outcome between the UofS and the consolidated datasets.

Table 4-3. Summary of Frequency Count by Outcome and Data Source

	<i>Real World</i>	<i>Simulation</i>
<i>Graduated</i>	118	109
<i>Withdrew</i>	36	45

Confirmation of Variability

In this section, I describe in more detail how I test for appropriate variability in SimDoc’s simulation output. As in the previous section, I examine the characteristics of the results paying attention to variability in time-in-program, produced by 100 iteration runs. Before looking at all 100 runs, however, I would first like to get a sense of whether there is variation. Thus, I will randomly select 12 out of the 100 runs to examine them graphically for insight into the variance among them. Figure 4-3 depicts graphically the results for the 12 randomly selected iterations in the form of density plots. This figure shows that there is evidence of variation in the graduation and withdrawal rates between the runs.

A box plot sheds more insight into the nature of the simulation results as shown in Figure 4-4. This box plot shows that indeed there are variations among the different runs and in fact, a few outliers exist. To check for homogeneity of variances, I run a Levene’s test for the 100 simulation runs against the UofS dataset, the test reveals a p-value (0.1707) > 0.05, as such, equal variance can be assumed. Thus far, I have shown that there is variability among the results of the 100 simulation runs. In addition, I have demonstrated that there is no difference in the means among the 100 simulation iterations and the UofS dataset.

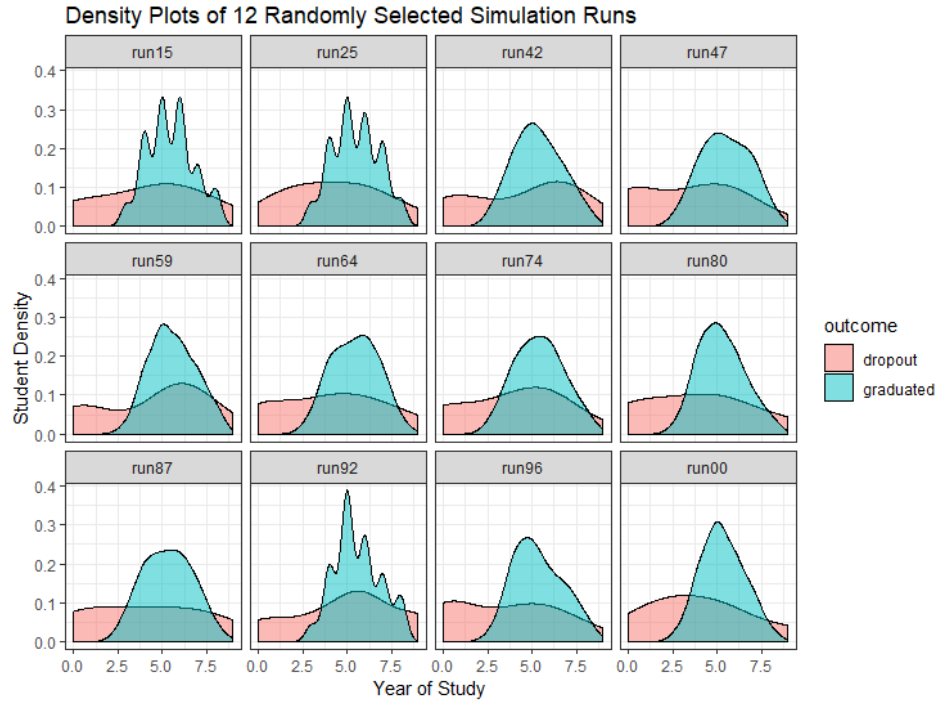


Figure 4-3. Density Plots of 12 Randomly Selected Runs of the Simulation

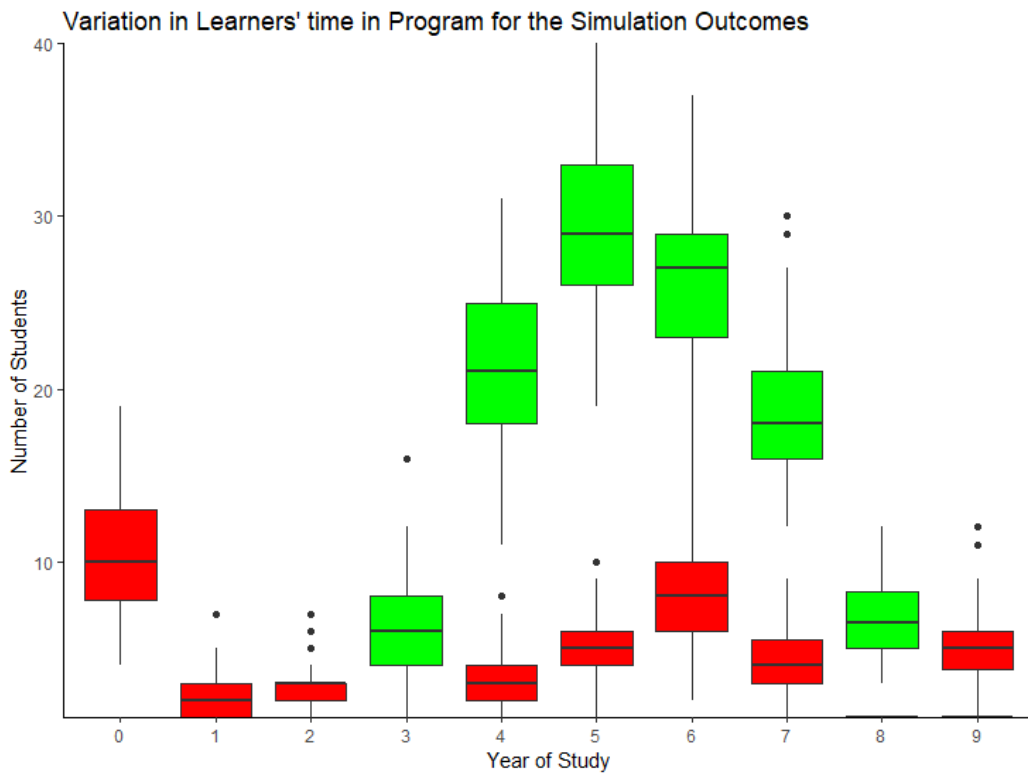


Figure 4-4. Variation in the Graduation and Attrition Rates in the 12 Simulation Runs

However, are these variations statistically significant? Are there any significant differences among the 100 iterations of the simulation? To answer these questions, I use a one-way analysis of variance (ANOVA). With ANOVA I explore the variance in means of each of the runs between the distribution of student counts in the program per year in the UofS dataset and SimDoc's output (each of the 100 runs). In this case, there are 101 groups: 100 groups representing the 100 runs of the simulation and 1 group representing the student graduation and withdrawal counts gleaned from the UofS dataset.

I am interested in confirming that there are significant differences in the average mean time learners were in the program either leading to completion of their degree or withdrawal from the program among the simulation's 100 runs and the UofS dataset. Since I have shown that the ANOVA assumptions are met, I am thus able to conduct a one-factor ANOVA to compare the difference in learners' time-in-program among SimDoc's 100 runs and the UofS dataset. The ANOVA results show that there is a statistically significant difference in the average time-in-program [$F(100,15453) = 1.272, p = 0.0352$] among the 100 iterations of the simulation and the UofS dataset. Therefore, I reject the null hypothesis and thus accept the hypothesis that there is statistical evidence to suggest that there is a difference in the means among SimDoc's 100 runs and the UofS dataset. This result shows that there is a difference between at least one or more pairings.

Whenever the null hypothesis is rejected in ANOVA, all that is known is that at least 2 groups differ from each other. ANOVA cannot tell us which of these groups are different. Therefore, to explore how the mean for each of the 100 iterations compared to that of the real-world dataset, I perform a post hoc test using the Tukey's Honest Significant Difference test at $p < .05$. The results show that there is no significant difference between the UofS dataset when compared to each of the 100 iterations when considering the time-to-program measure. The difference therefore exists within the 100 iterations, thus ensuring appropriate variability among the simulation runs.

4.3 Summary

In this chapter, I addressed the fourth goal of my dissertation: to demonstrate how an AIED system designer can calibrate and validate the resulting simulation model. I first performed calibration experiments and then validation experiments. The calibration process systematically assigns

values to under-determined variables looking for the assignment of values that yields outputs that most closely match the real world dataset on outcome variables of interest. This best matching run defines the baseline simulation model. 500 runs were used to find the SimDoc baseline. It is important to make sure that the calibration process has not gone too far and overfitted the SimDoc simulation model to the UofS dataset. Therefore, I subsequently performed a validation process on the calibrated SimDoc model.

I provided an algorithm that can be used to determine when an appropriate number of runs has been carried out. This is achieved by running the best calibrated model until it is clear that this model meets appropriate stability and variability criteria. The validation analysis for the calibrated SimDoc model turned out to converge at the 100th run in producing a consolidated dataset that is like the UofS dataset. The methods used here are not specific to the SimDoc simulation and should generalize to any simulation. Knowing when the simulation has been run an appropriate number of times should allow system designers to be confident in the results of subsequent experimentation with the simulation model and should avoid them having to needlessly make extra simulation runs. This is especially important for medium and high fidelity simulations that can take a long time to run.

The result of the calibration and validation process is a validated baseline simulation (called the ‘SimDoc baseline’ in the SimDoc case study) that produces learning outcomes that match reality, which can then be used to explore the effects of making various hypotheses about changing some of the characteristics of the simulated doctoral program. Essentially, step 5 of the seven-step framework is complete. I will explore step 6 of the framework in the next chapter: how to use the baseline model as the foundation for a set of experiments designed to explore interesting hypotheses of interest, again using SimDoc to illustrate the process. I show in Figure 4-5 the completion rate produced by the SimDoc baseline model compared to the UofS completion rate. Generally, this figure shows a slight difference in these completion rates. Is this difference significant? I answer this question in the next chapter as I address the 6th and 7th steps of the seven-step framework.

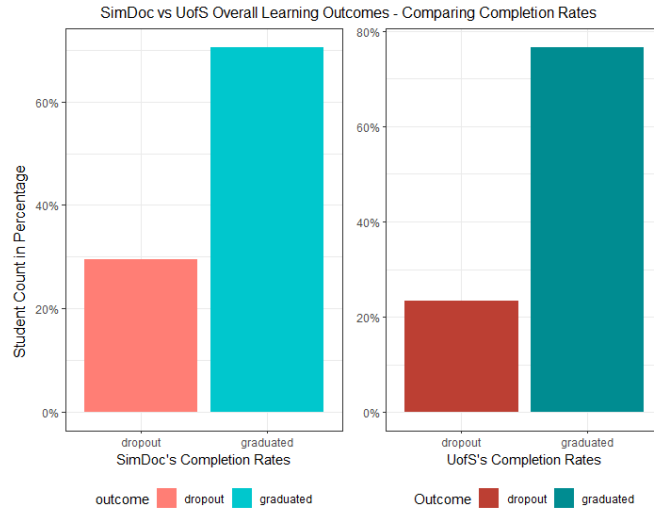


Figure 4-5. Comparing UofS and SimDoc Baseline Learning Outcomes

CHAPTER 5

EXPERIMENTING WITH SIMDOC: EXPLORING EFFECTS OF PERSONALIZATION

In this chapter, I use the SimDoc baseline model to illustrate how to carry out the sixth and seventh steps of the seven-step framework. That is, I show how a system designer can run simulation experiments, analyze the results, discuss the results in readiness for presentation and publication, and explore ways of improving the simulation in preparation for subsequent design iterations and new research directions.

5.1 Exploring Personalization Using Simulation

My research aim is to demonstrate how a system designer can build a simulation of a longer-term learning environment and use the resulting simulation model to ask hypothetical, ‘what-if’, pedagogical questions for better understanding of the dynamics of the modeled longer-term learning and mentoring environment. Before I use SimDoc to explore hypothetical, ‘what-if’, research questions concerning personalization, it is important to have a validated baseline model for referencing and comparisons.

5.1.1 Baseline Experimental Set Up

The simulation model I use for illustration purposes, the SimDoc model, is a medium-fidelity validated model involving simple abstractions of learners, supervisors, and milestones. The milestones consist of coursework, comprehensive, proposal, and dissertation. Each of these milestones has a prescribed time requirement. In each experimental set up, a new simulated doctoral program structure is created that does not necessarily have an existing real-world parallel to compare to. I choose to have a single type of learner or supervisor in each of the experimental set ups in order to exaggerate the differences among conditions so as to see clearly the differences in effects. To achieve variability and stability in the simulation results, I run 100 iterations for each simulation condition as informed by the results of the validation process as to the number of runs to achieve an appropriate dataset of simulation results (see chapter 4 for details). For each of these 100 runs, a new random seed is selected to ensure variability in simulation results.

Table 5-1 The SimDoc Experiments Data Dictionary

Learner		Supervisor
<i>Latent Effort:</i>	<i>DMF</i>	<i>Supervisor Type (desired meeting frequency)</i>
Low (1.0 – 1.1)	Weekly	Laissez-Faire (Quarterly)
Medium (1.1 – 1.23)	Biweekly	Pastoral (Monthly)
High (1.23 – 1.34)	Monthly	Directorial (Biweekly)
	Quarterly	Contractual (Weekly)

The key supervisor attributes I am mainly focused on here include:

- *supervisorType* – one of these types: Laissez-Faire, Pastoral, Directorial, or Contractual, based on Gatfield’s [151] study.
- *desiredMeetingFrequency* – represents how often a supervisor is willing and available to meet learners. Its value is informed based on supervisor type at a rate of once every week for Contractual, once every two weeks for Directorial, once every three weeks for Pastoral, or once every six weeks for Laissez-Faire. Rate values are assigned based on results from Heath’s [178] study.
- *allocatedWeeklyMeetingHours* – represents the time allocated for total meeting time each week (currently set to 5hrs). This value is set based on the SimDoc calibration results. A meeting event is made up of one or more meeting sessions depending on the number of learners that have requested for a meeting. A meeting session refers to a one to one meeting between a supervisor and a learner.
- *meetingDuration* – represents the time allocated for each meeting session (currently set to 1.5hrs). This value is set based on the SimDoc calibration results.

The main learner attributes we are focused on include:

- *latentEffort* – a number between (1.0, 1.34) representing a learner’s basic capability and allows learners to be divided into groups: low (1.0 – 1.1), medium (1.1 – 1.23) and high effort (1.23 – 1.34). Its value is assigned inversely based on UofS time-to-completion distributions. Therefore, shorter times indicate higher latentEffort.
- *desiredMeetingFrequency* – represents how often a learner ask for a meeting per week. Its value is assigned based on Heath’s [178] distribution of a rate of once every week for 38%, once every two weeks for 29%, once every three weeks for 19%, or once every six weeks for 13% of the learner population.

- *actualMeetingFrequency* – represents the actual meeting frequency per week. Its value is calculated weekly as follows: the number of meetings attended / number of weeks in the program.
- *satisfaction* – a value representing how satisfied a learner is in the program based on the frequency of meetings attended. It is calculated as follows: *actualMeetingFrequency* / *desiredMeetingFrequency*.
- *receivedWaiver* – an attribute that indicates whether a learner received a waiver not to take classes or not. Its value is derived from the UofS dataset.
- *weeklyEffort* – a value representing how many hours a learner is allocating to their studies on a weekly basis as they go through the program and interacts with her/his supervisor. The is based is computed based on the IRT-based Equation 3-5

5.1.2 Interpreting Simulation Results

Simulation modeling is a useful tool in research for a number of reasons. A simulation can be used to drill down more deeply into various variables that are hidden in the real world, and their interactions. A simulated model can be used for testing various hypothetical scenarios. Further, the process of creating the simulation model representing a target real-world system of study often leads to a better understanding of the real-world system. Depending on the objective of a simulation study, using a simulation might help a system designer identify potential areas for improvement in the real-world system including changing the structure of the current system or adding new components. Where two or more solutions exist to address a challenge, simulation allows all competing solutions to be simulated and experimental results to be compared to find the best option. It is important to note that any discoveries made using simulation are actually predictions for the real-world system and not necessarily reflective of reality.

5.2 Drilling into the Baseline for Personalization Insights

In this section, I drill down into the baseline simulation results, looking for patterns based on supervisory type and learner type. One of the advantages of medium and high fidelity simulations is that system designers are afforded an ability to drill down into the model details to discover potentially interesting patterns. The baseline model can be used to explore ideas around potential

questions of interest. For the SimDoc case study, I am interested in exploring the effects of matching different supervisors (distinguishing by their desired meeting frequency) and learner types (distinguishing first by their desired meeting frequency and second their latent effort level) as discussed in detailed experimental set ups in section 5.3. Before I get to the experiments, I am interested in discovering the predictions in the baseline simulation. Figure 5-1 reveals potential predictions of various learning outcomes based on a combination of different supervisory types and different types of learners, at least as both are distinguished by desired meeting frequencies. A more detail comparison with SimDoc case study results is provided later in section 5.3. It is not surprising to observe that SimDoc baseline forecasts that learners who desire weekly meeting would only perform well under Contractual supervision. It is, however, surprising to see that when considering the different types of learners supervised by Contractual supervisors, learners who desire weekly meeting perform the poorest.

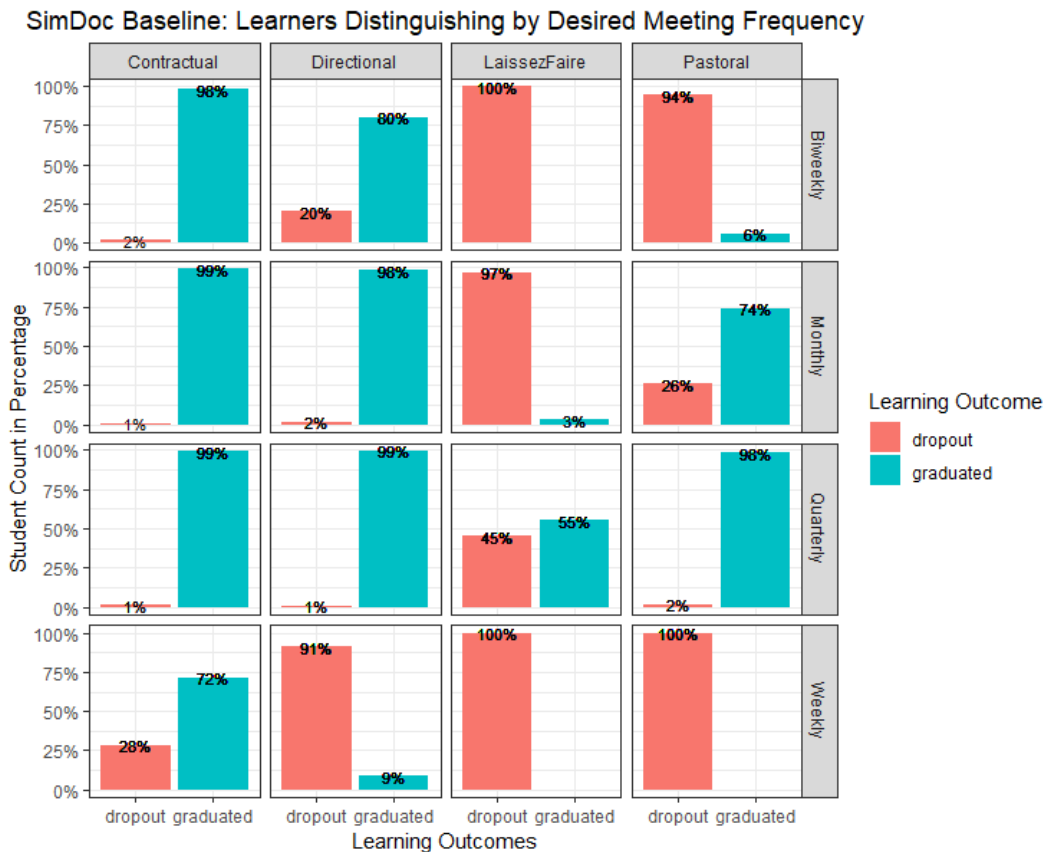


Figure 5-1. SimDoc Baseline Learning Outcomes Differentiated by Meeting Frequency: when Learners are Distinguished by Their Desired Meeting Frequency and Assigned Supervisory Type.

Figure 5-2 show SimDoc Baseline learning outcomes based on supervisor type (differentiated by their desired meeting frequency) and learner type (distinguished by their learner effort level). As with Figure 5-1, a more descriptive comparison with SimDoc experimental results is provided later in section 5.3.

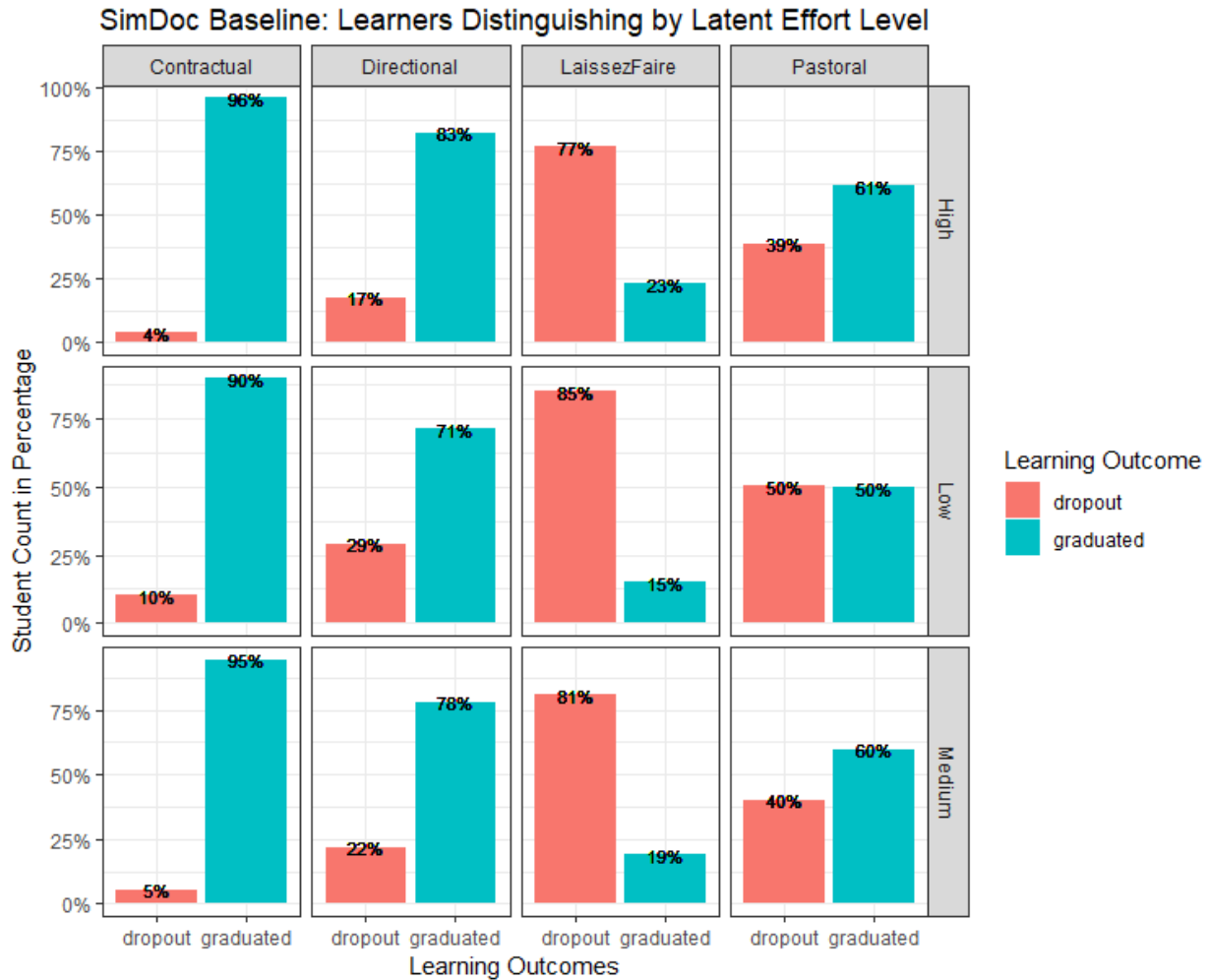


Figure 5-2. SimDoc Baseline Learning Outcomes Differentiated by Latent Effort: when Learners are Grouped by Their Latent Effort Level and Assigned Supervisory Type.

5.3 Exploring Hypothetical, ‘what-if’, Questions within the SimDoc Case Study

Personalization of learning support has been one of AIED’s most central research objectives. Studies attest that personalization improves learners’ attainment of learning goals. As such, personalization [25], [28] is important in promoting effective and efficient learning in these longer-term learning domains. The key is the one-to-one relationships between learners (mentees) and

the source of their support (mentors); more specifically how the support styles are in accordance with learners' learning styles and preferred support styles [34]. In the SimDoc case study, I examine the effectiveness of personalized supervisory styles. The specific hypothetical, 'what-if', questions I focus on here are:

E1. How effective would each of the four (4) supervisory styles identified by Gatfield [151] and categorized (with a major distinguishing feature being desired meeting frequency) be among learners with the four different desired meeting frequencies based on Heath's [178] distribution?

E2. How effective would each of the four (4) supervisory styles identified by Gatfield [151] and categorized (with a major distinguishing feature being desired meeting frequency) be among learners grouped by the three different latent effort values?

E1. Exploring the Effect of Matching Different Types of Learner and Supervisor based on Desired Meeting Frequencies

In this first experiment I would like to explore the effect of matching different types of learners and supervisors based on their desired meeting frequencies. The goal of this experimental set up is to explore what would happen in a doctoral program that had only one type of supervisor and one type of learner as determined by their desire for meeting frequencies. To achieve this objective, I set up an experimental with 16 different conditions: four (4) types of supervisory desired meeting frequencies multiplied by four (4) groupings of learners' desired meeting frequencies. Since I am experimenting with desired meeting frequencies (MF) for both supervisors and learners, I named this experimental set up – the MFMF set up.

For this experimental set up, I start with SimDoc's baseline simulation model, that is the version of SimDoc tuned and validated to match the UofS dataset. From the SimDoc baseline, I only change the way I assign values for two attributes: supervisor's *desiredMeetingFrequency* and learner's *desiredMeetingFrequency*. That is, for each MFMF experimental set up, I assign all supervisors and all learners to have one supervisor desired meeting frequency type and learner desired meeting frequency type respectively. That is the only difference between conditions; every other aspect of the SimDoc simulation continues to operate in the baseline mode. I run each experimental set up 100 times. This value was determined in the validation stage to provide results with stability and appropriate variability. For the outcomes of the simulation runs, I take the average of the aggregate simulation output for each of the 100 runs in each condition and plot

them. Note that it would have been possible to drill down and look separately at each of the 100 runs in each configuration of supervisors and students, and interesting differences between runs may have been observed and possibly shed some light on certain issues, but for these initial experiments averages over the 100 runs seemed to be more useful measures.

E1. Experimental Results: MFMF set up

It is important to keep in mind that results yielded by these experiments are actually predictions for what might happen in the real-world if the conditions were similar to that in the SimDoc MFMF model. Figure 5-3 summarizes the experimental results of the 16 SimDoc MFMF experimental set ups. In total this experiment consisted of 1600 iterations of the simulated doctoral program involving hundreds of simulated learners and supervisors.

As expected, the results show a difference in the effectiveness of each supervisory style on each learner type. It is very surprising that the MFMF SimDoc set up predicts that in a doctoral program with a single supervisory type, learners who desire weekly meeting frequencies underperform all the other types of learners. A similar pattern is seen in the simulation outputs of the SimDoc baseline model as depicted in Figure 5-4 (charts for both learning outcomes produced by the SimDoc baseline and the SimDoc MFMF set up are included to help with comparison). However, the results predict unsurprisingly that learners supervised by Laissez-Faire supervisor will have a worse completion rate. On the other hand, the MFMF SimDoc model envisions that the highest average of aggregate completion rate would be achieved in a condition where supervisors of a Contractual type are involved. A comparable outcome is seen in the simulation outputs of the SimDoc baseline model.

Further, when considering time-to-outcome (persistence to graduation or drop out) it is unsurprising to see the forecast that learners supervised by a Contractual supervisor – those willing to meet weekly and give clear direction – would have a high chance of graduating and doing so in a relatively good time-to-degree as illustrated in Figure 5-5. It is very interesting to observe that of the learner types working with Contractual supervision, learners desiring quarterly meeting frequencies had better time-to-degree; there was no learner graduating or dropping out beyond the 7th year. Even more surprising is the forecast by the MFMF SimDoc model that no learner will be able to persist beyond the 2nd year when learners with the desire for weekly meetings are paired up with a supervisor that prefers the Laissez-Faire approach to supervising.

Visually, Figure 5-4 and Figure 5-6 show that there are some similarities but also substantial differences between the learning outcomes produced by the SimDoc baseline and the SimDoc MFMF set up. To ascertain whether these results are statistically similar, I perform a statistical test. I conduct an independent-samples t-test to compare the learning outcome in SimDoc's baseline set up and experimental results in the MFMF SimDoc set up. I conducted two sets of t-tests: one for a positive outcome (persistence to completion) and the other for a negative outcome (withdrawal). The results show that among the learners who persisted to completion there was no significant difference in the completion rates for SimDoc's baseline ($M=55.375$) and SimDoc's MFMF set ups ($M=45.375$); $t = -0.048264$, $df = 29.999$, $p\text{-value} = 0.9618$. Similarly, the results among learners who withdrew indicate that there is no significant difference in the completion rates for SimDoc's baseline ($M=44.625$) and SimDoc's MFMF set ups ($M=54.625$); $t = -0.048264$, $df = 29.999$, $p\text{-value} = 0.9618$. These t-test results indicate that the two (2) groups are not different statistically.

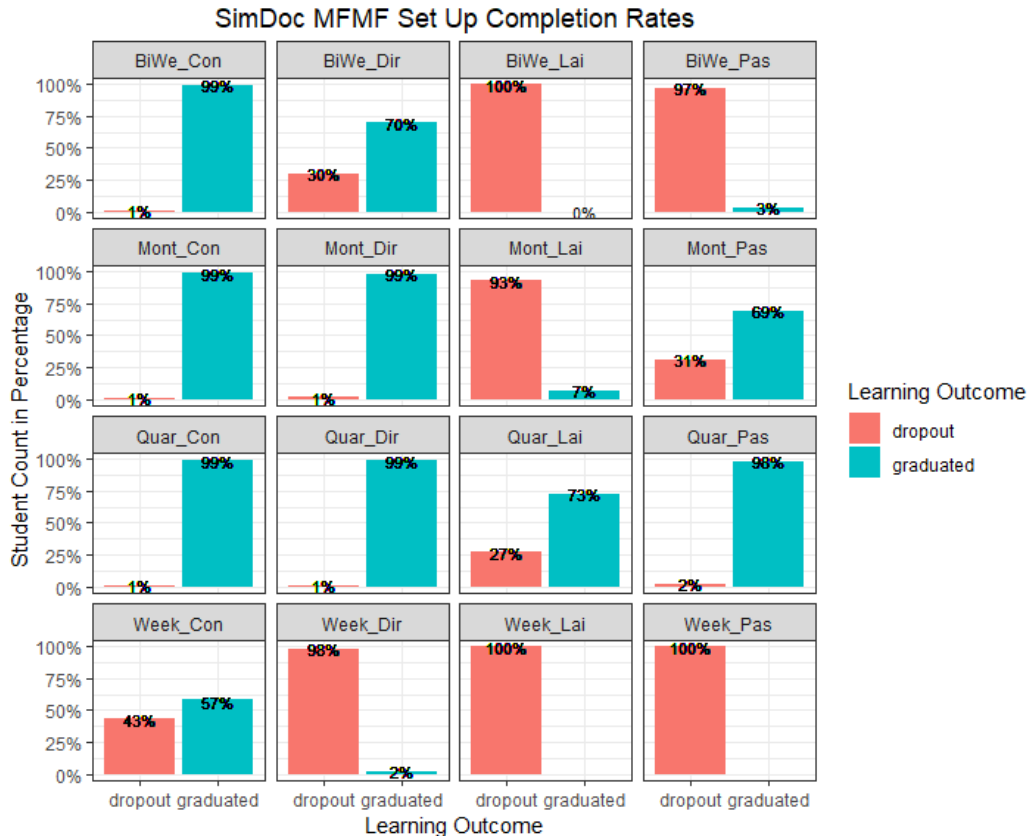


Figure 5-3. SimDoc MFMF Set Up Learning Outcomes. A Chart Depicting Learning Outcomes of Learners Under Different Conditions Distinguished by Their Desired Meeting Frequency and Assigned Supervisory Type.

The following is the key to notations used for each experimental condition:

- BiWe_Con – represents an interaction between **BiWeekly** learners and **Contractual** supervisors
- BiWe_Dir – represents an interaction between **BiWeekly** learners and **Directorial** supervisors
- BiWe_Lai – represents an interaction between **BiWeekly** learners and **Laissez-Faire** supervisors
- BiWe_Pas – represents an interaction between **BiWeekly** learners and **Pastoral** supervisors
- Mont_Con – represents an interaction between **Monthly** learners and **Contractual** supervisors
- Mont_Dir – represents an interaction between **Monthly** learners and **Directorial** supervisors
- Mont_Lai – represents an interaction between **Monthly** learners and **Laissez-Faire** supervisors
- Mont_Pas – represents an interaction between **Monthly** learners and **Pastoral** supervisors
- Quar_Con – represents an interaction between **Quarterly** learners and **Contractual** supervisors
- Quar_Dir – represents an interaction between **Quarterly** learners and **Directorial** supervisors
- Quar_Lai – represents an interaction between **Quarterly** learners and **Laissez-Faire** supervisors
- Quar_Pas – represents an interaction between **Quarterly** learners and **Pastoral** supervisors
- Week_Con – represents an interaction between **Weekly** learners and **Contractual** supervisors
- Week_Dir – represents an interaction between **Weekly** learners and **Directorial** supervisors
- Week_Lai – represents an interaction between **Weekly** learners and **Laissez-Faire** supervisors
- Week_Pas – represents an interaction between **Weekly** learners and **Pastoral** supervisors



Figure 5-4. Comparing Learning Outcomes: SimDoc Baseline vs MFMF Set Up. Left: the Chart Presented in Figure 5-1. Right: the Graph Presented in Figure 5-3.



Figure 5-5. SimDoc MFMF Set Up Learners' Time-to-Outcome. A Graph Showing Time-to-Outcome of Learners Under Different Conditions Distinguished by Their Desired Meeting Frequency and Assigned Supervisory Type.

The following is the key to notations used for each experimental condition:

BiWe_Con – represents an interaction between **BiWeekly** learners and **Contractual** supervisors

BiWe_Dir – represents an interaction between **BiWeekly** learners and **Directorial** supervisors

BiWe_Lai – represents an interaction between **BiWeekly** learners and **Laissez-Faire** supervisors

BiWe_Pas – represents an interaction between **BiWeekly** learners and **Pastoral** supervisors

Mont_Con – represents an interaction between **Monthly** learners and **Contractual** supervisors

Mont_Dir – represents an interaction between **Monthly** learners and **Directorial** supervisors

Mont_Lai – represents an interaction between **Monthly** learners and **Laissez-Faire** supervisors

Mont_Pas – represents an interaction between **Monthly** learners and **Pastoral** supervisors

Quar_Con – represents an interaction between **Quarterly** learners and **Contractual** supervisors

Quar_Dir – represents an interaction between **Quarterly** learners and **Directorial** supervisors

Quar_Lai – represents an interaction between **Quarterly** learners and **Laissez-Faire** supervisors

Quar_Pas – represents an interaction between **Quarterly** learners and **Pastoral** supervisors

Week_Con – represents an interaction between **Weekly** learners and **Contractual** supervisors

Week_Dir – represents an interaction between **Weekly** learners and **Directorial** supervisors

Week_Lai – represents an interaction between **Weekly** learners and **Laissez-Faire** supervisors

Week_Pas – represents an interaction between **Weekly** learners and **Pastoral** supervisors

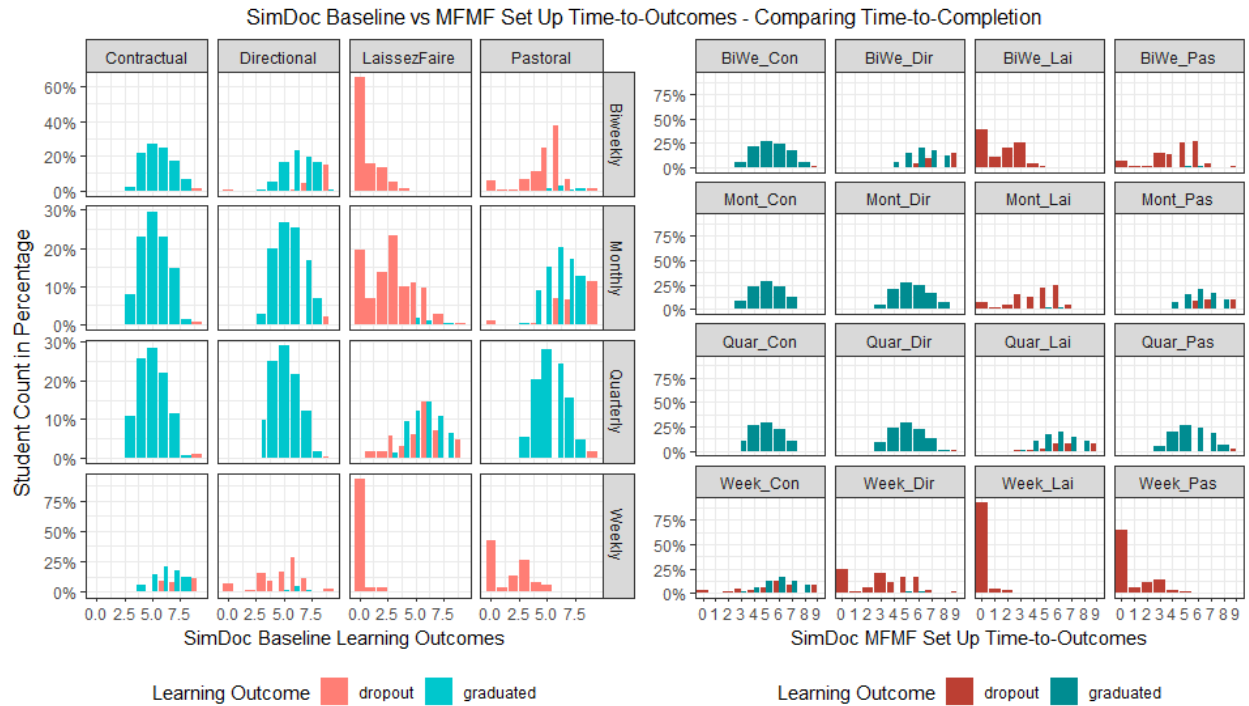


Figure 5-6. Comparing Learners’ Time-to-Outcome: SimDoc Baseline vs MFMF Set Up. Left: A Graph Showing Time-to-Outcome of Learners in SimDoc Baseline Under Different Conditions Distinguished by Their Desired Meeting Frequency and Assigned Supervisory Type. Right: the Chart Presented in Figure 5-5.

E2. Investigating the Effect of Matching Different Types of Learners and Supervisors based on Latent Effort and Desired Meeting Frequency Respectively

In this second experiment, I would like to investigate the effect of matching different types of learners and supervisors based on latent effort and desired meeting frequency respectively. The goal of each experimental set up is to discover what would possibly happen in a doctoral program that had only one type of supervisor and one type of learner as determined by their desire for meeting frequencies and latent effort respectively. To achieve this objective, I set up an experiment with 12 different conditions: four (4) types of supervisory desired meeting frequencies multiplied by three (3) types of learners’ latent efforts. As with the previous experiment, I name this experimental set up based on the attributes being experimented with (learners’ latent efforts (LE) and supervisors’ desired meeting frequencies (MF)) – thus the LEMF set up.

In total this experiment consisted of 1200 iterations of a simulated doctoral program involving hundreds of simulated learners and supervisors. As with experiment E1, I start with the

SimDoc validated baseline simulation model, and only change attributes directly involved in the hypothetical question being explored. In this case, I change the way I assign values for only two attributes: supervisor's *desiredMeetingFrequency* and learner's *latentEffort*. That is, I assign all supervisors to have one supervisor desired meeting frequency type and assign and all learners to have one learner latent effort type for each experimental set up. As with E1, each experimental set up consists of 100 simulation runs. For the outcomes of the simulation runs, as in E1 I take the average of the aggregate simulation output for each of the 100 runs in each condition and plot them. Figure 5-2 reveals a prediction of potential learning outcomes under each of the 12 experimental set ups exploring various E2 hypotheses.

E2. Experimental Results: LEMF set up

Once again, the discoveries of the SimDoc LEMF experimental set up are actually predictions for what might happen in the real-world if the conditions were similar to that in the SimDoc LEMF model. Figure 5-7 summarizes the experimental results of the 12 SimDoc LEMF experimental set ups. In total this experiment consisted of 1200 iterations of the simulated doctoral program involving hundreds of simulated learners and supervisors. Figure 5-7 depicts a prediction by the LEMF experiment that learners under Laissez-Faire supervision would generally have a very high attrition rate. The opposite is true when I examine learners under the similar condition in the SimDoc baseline model as depicted Figure 5-2. This pattern of observing opposite outcomes can also be seen when considering learning outcomes under the Contractual supervisory type. In the SimDoc baseline model, learners under Contractual supervision have high completion rates while the LEMF SimDoc model predicts a scenario where learners, regardless of type, would have the highest attrition rates among their group.

Visually, Figure 5-8 shows that the differences that exist between the two graphs is surprisingly very high. To establish whether the results depicted in Figure 5-8 are statistically different, I conducted an independent-samples t-test to compare the learning outcome in SimDoc's baseline set up and experimental results in SimDoc's LEMF set up. As with E1, I conducted two sets of t-tests: one for a positive outcome (graduate) and the other for a negative outcome (withdrawal). The results show that among those learners who persisted to completion there was no significant difference in the completion rates for SimDoc's baseline ($M=61.81$) and SimDoc's LEMF set ups ($M=55.15$); $t = 0.66764$, $df = 13.134$, $p\text{-value} = 0.5159$. Comparably, the results among learners who withdrew indicate that there is no significant difference in the completion rates for

SimDoc’s baseline (M=38.18) and SimDoc’s LEMF set ups (M=44.84); $t = 0.66764$, $df = 13.134$, $p\text{-value} = 0.5159$.

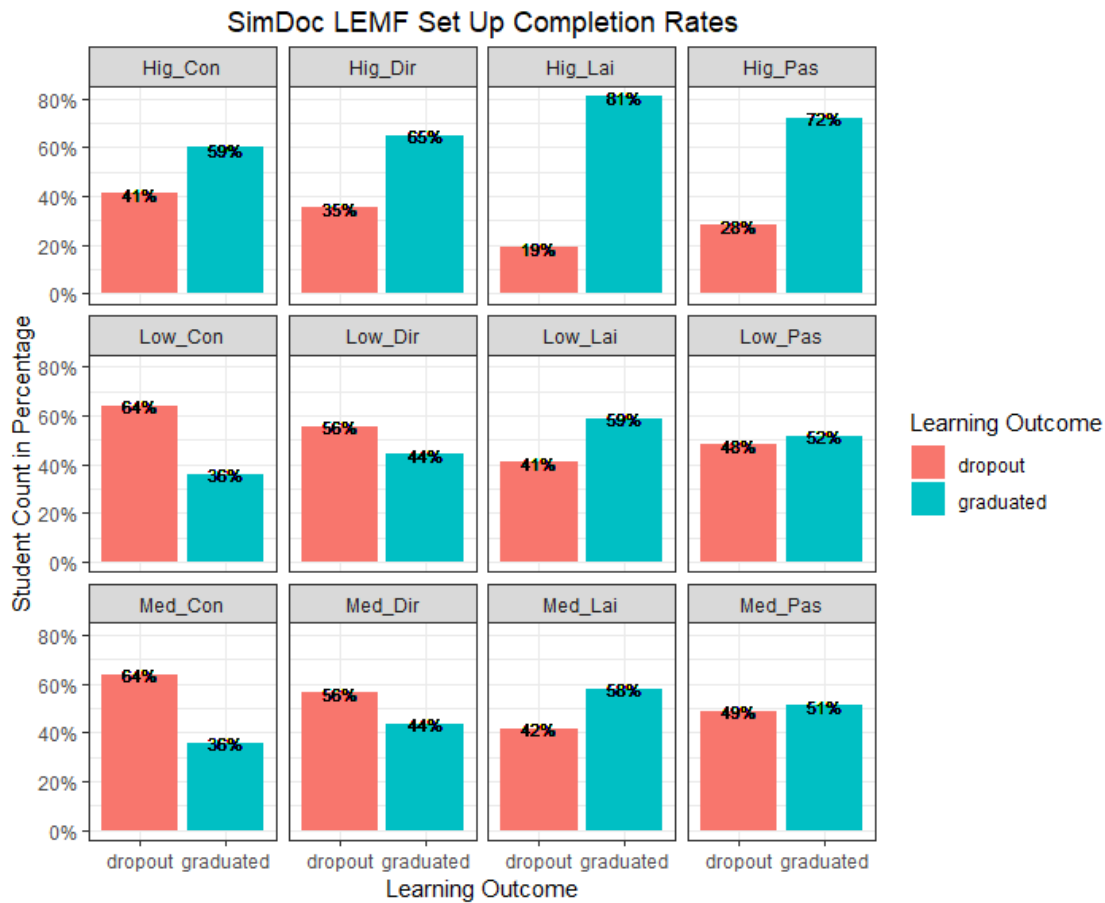


Figure 5-7. SimDoc LEMF Set Up Learning Outcomes. A Chart Depicting Learning Outcomes of Learners Under Different Conditions Distinguished by Their Latent Effort Level and Assigned Supervisory Type.

The following is the key to notations used for each experimental condition:

Hig_Con – represents an interaction between **BiWeekly** learners and **Contractual** supervisors

Hig_Dir – represents an interaction between **BiWeekly** learners and **Directorial** supervisors

Hig_Lai – represents an interaction between **BiWeekly** learners and **Laissez-Faire** supervisors

Hig_Pas – represents an interaction between **BiWeekly** learners and **Pastoral** supervisors

Low_Con – represents an interaction between **Monthly** learners and **Contractual** supervisors

Low_Dir – represents an interaction between **Monthly** learners and **Directorial** supervisors

Low_Lai – represents an interaction between **Monthly** learners and **Laissez-Faire** supervisors

Low_Pas – represents an interaction between **Monthly** learners and **Pastoral** supervisors

Med_Con – represents an interaction between **Quarterly** learners and **Contractual** supervisors

Med_Dir – represents an interaction between **Quarterly** learners and **Directorial** supervisors

Med_Lai – represents an interaction between **Quarterly** learners and **Laissez-Faire** supervisors

Med_Pas – represents an interaction between **Quarterly** learners and **Pastoral** supervisors

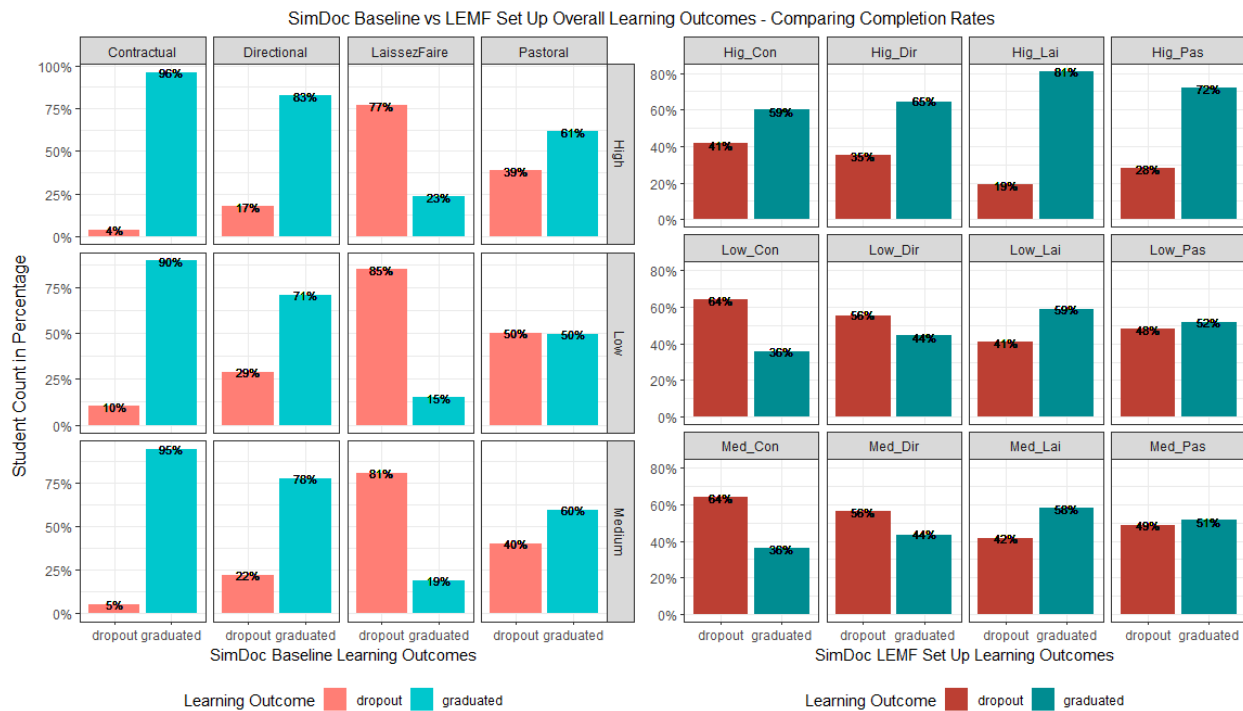


Figure 5-8. Comparing Learning Outcomes in the SimDoc Baseline and the LEMF Set Up. Left: the Graph Presented in Figure 5-2. Right: the Chart Presented in Figure 5-7.

5.4 Insights into the Personalization Experimental Findings

The tasks done in this chapter help illustrate how to achieve the requirements of step 6 and 7 of the seven-step framework. I have experimented with the SimDoc model and analyzed the results. The results of the simulations should not be taken literally. However, they are predictions of what could actually happen in a real-world system if all the conditions were similar. The 7th step of the framework suggests providing a discussion of the results and identifying potential ways of improving the model and hence new research directions.

It is very interesting to observe that in a doctoral program scenario with only Contractual supervisors and only learners of one learner type (those who desire weekly meetings), SimDoc predicts that the outcome is worse in this SimDoc’s MFMF set up than in SimDoc’s baseline set up. This discovery is counterintuitive because one would expect that matching highly productive supervisors with high performing high maintenance learners would result in better outcomes. However, probably the envisioned high attrition among learners in this condition is caused by high demand for supervisor meeting time. This phenomenon is magnified by very high attrition

rates among this type of learner and the other three types of supervisors who do not have the time to meet with learners regularly with the worst matchup being with Laissez-Faire type supervisors. Simulation results show a prediction that almost all learners in this situation will drop out in their first year – see Figure 5-5. Furthermore, the simulation results forecast that learners with quarterly desired meeting frequency tend to complete their program under all supervisor types. My interpretation of this observation is that the most important thing that leads to a prediction of high completion rates is keeping learners satisfied with their progress and meeting frequencies.

Furthermore, I find it fascinating to discover a prediction that when there is only a single learner type based on latent effort, the overall completion rates are worse in a one-supervisor type (SimDoc's LEMF) condition as compared to SimDoc's baseline set up except for the Laissez-Faire supervisory condition. This suggests that it is not desirable to have a single type of learner in a doctoral program. I observe another interesting phenomenon in SimDoc's LEMF set up (see Figure 5-9, showing completion rates by year in the LEMF set up). Results depicted in this figure predict that, overall, learners stay in the program longer as compared to outcomes shown in SimDoc's baseline set up (see Figure 5-10). Further, there is no expected condition in SimDoc's LEMF set up that leads to 100% attrition rate as compared to SimDoc's MFMF conditions.



Figure 5-9. SimDoc LEMF Set Up Learners' Time-to-Outcome. A Chart Depicting Time-to-Outcome of Learners Under Different Conditions Distinguished by Their Latent Effort Level and Assigned Supervisory Type

The following is the key to notations used for each experimental condition:

Hig_Con – represents an interaction between **BiWeekly** learners and **Contractual** supervisors

Hig_Dir – represents an interaction between **BiWeekly** learners and **Directorial** supervisors

Hig_Lai – represents an interaction between **BiWeekly** learners and **Laissez-Faire** supervisors

Hig_Pas – represents an interaction between **BiWeekly** learners and **Pastoral** supervisors

Low_Con – represents an interaction between **Monthly** learners and **Contractual** supervisors

Low_Dir – represents an interaction between **Monthly** learners and **Directorial** supervisors

Low_Lai – represents an interaction between **Monthly** learners and **Laissez-Faire** supervisors

Low_Pas – represents an interaction between **Monthly** learners and **Pastoral** supervisors

Med_Con – represents an interaction between **Quarterly** learners and **Contractual** supervisors

Med_Dir – represents an interaction between **Quarterly** learners and **Directorial** supervisors

Med_Lai – represents an interaction between **Quarterly** learners and **Laissez-Faire** supervisors

Med_Pas – represents an interaction between **Quarterly** learners and **Pastoral** supervisors

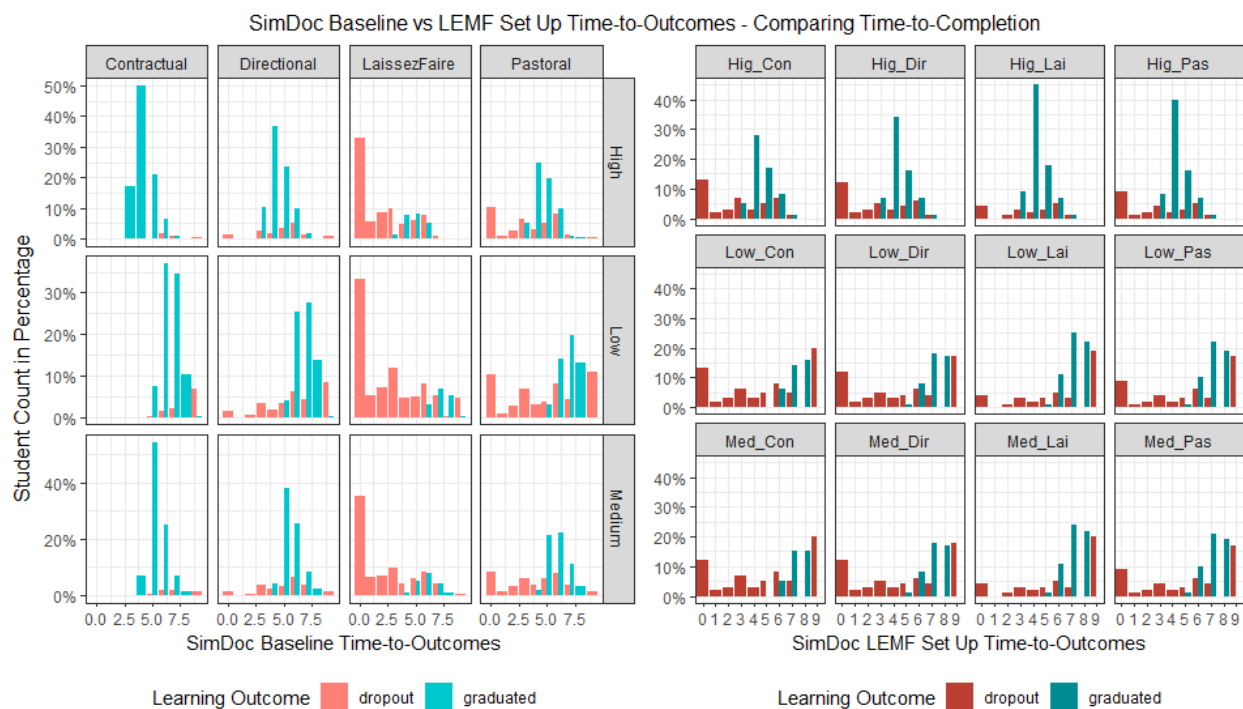


Figure 5-10. Comparing Learners' Time-to-Outcome: SimDoc Baseline vs LEMF Set Up. Left: A Graph Showing Time-to-Outcome of Learners in SimDoc Baseline Under Different Conditions Distinguished by Their Latent Effort Level and Assigned Supervisory Type. Right: the Chart Presented in Figure 5-9.

5.5 Future Experimental Directions

E1&E2 Future Directions

Results from E1 and E2 shed some light on interesting predictions on the effect of using personalized approaches to selecting and allocating supervisors to doctoral learners. Drawing from the lessons learned in these first set of experiments, here are some future directions that could be explored in future experimentation and application. A system designer could develop an algorithm that can match doctoral learners to supervisors, based on characteristics of the learners and the supervisors derived from the current version of SimDoc's experimental results that offers the best chance of learner success. Success, in this case, is measured in having a short time-to-completion with high completion rates. The algorithm could be modified to enable learners to benefit the most from getting an appropriate match with their supervisors. In these follow up sets of experiments, it would particularly be interesting to examine the contrast between the new al-

location algorithm, the baseline as informed by UofS data, and random assignment of supervisors to learners to determine which allocation strategy works best.

Studying the Impact of Different Ratios of Learner Types

A study by Heath [178] shows that the most important factor in the success of doctoral mentorship is not the format (one-to-one or many-to-one) but rather the quality of the supervisor-learner (mentor-mentee) relationship, with the supervisor having a far more crucial role to play in fostering, encouraging, and supporting the learner through regular supervisor-learner meetings [177], [178]. However, one major challenge is maintaining a regular or even a frequent enough meeting schedule [178], [187]. Thus, it will be interesting to explore the impact of a combination of various meeting frequencies between supervisors and learners and different ratios of the learner type. The aim of this experimental set up would be to gain insight into the impact different learning environment set ups have on different types of learners. It would be interesting to run follow-up experiments that examine how the different types of learners perform in an environment with a different combination of learners. Results from such experiments could lead to the development of an algorithm that would be used to assign learners to different research groups based on that learner's and research group's attributes, that is, a personalized allocation of learners to research groups.

Investigating the Effect of Research Group Sizes/Supervisor Workload

Learners take on learning endeavors for a myriad of reasons. Relatedness, a sense of belonging to a community, affects such motivations and individual performances [229]. These studies [182], [230] show that in different learning environments learners that feel supported and respected by their teachers and their family are more likely to demonstrate natural inquisitive tendencies and desire to learn new skills. In a doctoral program, the number of other learners supervised by the same supervisor could indirectly affect a given learner's relatedness to the rest of the group because of demand for shared supervisor's time.

Supervisor's workload is one factor amongst others that affect the availability of a supervisor to meet regularly with the learner(s) s/he supervises. One of the contributing factors to a supervisor's workload is the number of learners s/he supervises and her/his preferred supervisory style. It would be interesting to explore the following hypothetical question: how would the different sizes of the research groups affect the learning outcomes of a doctoral program? This

question could be answered by setting up an experiment where the simulation is run under ten different conditions: from a scenario where each supervisor has only one learner through a situation where each supervisor has ten learners. The various conditions could then be compared to learning outcomes.

The goal of this experimental set up would be to explore the effect of both the research group size and the supervisory workload on doctoral learners with characteristics and behaviors captured in the SimDoc model. Another intention of such an experiment would be to figure out an optimum research group size based on the attributes of supervisors and learners. In this experimental set up, no outcome need be compared with SimDoc baseline outputs. The findings would shed light on personalization and could lead to recommendations being made to supervisors as to whether they should take on more learners or not. In addition, insights would be gained on the impact of supervisor decisions on their supervisees and how new learners can get personalized recommendation on which supervisor to choose based on supervisors' workload.

Moreover, going beyond the current SimDoc model, new learner model attributes and event models could be devised that took into account the effects (both beneficial and detrimental) of a student's interactions within their research group on their progress through the doctoral program. This would require another design iteration through the 7-step framework, with new questions leading to revised SimDoc models, a new round of calibration and validation, and a new experimental program. Such iterative improvement and deepening of the simulation model would, over time, slowly enhance the model's fidelity and the realism of its predictions.

5.6 Summary

In this chapter, I addressed the fourth goal of my dissertation, that is to demonstrate how an AIED system designer can use a simulation model of a longer-term learning and mentoring environment to better understand the environment, its characteristics, learning and teaching needs, and possible tools to support these needs. I illustrated this concept by showing how an AIED system designer can experiment with the resulting simulation model (step 6 of the process in Figure 3-2). I carried out two experiments aimed at understanding more about personalization issues in the learner-supervisor relationship. I further identified potential research directions that could be carried out relatively easily using the simulation (and certainly much more easily and quickly

than in the real world). I also looked briefly forward to future experiments and subsequent cycles of iterative design.

CHAPTER 6

DISCUSSION AND CONCLUSION

In this chapter, I conclude my thesis. First, I revisit the main underlying objective of my research. Next, I describe the need for simulation to support the AIED system development process, especially when creating systems for supporting longer-term learning and mentoring environments. Subsequently, I summarize how I used the seven-step framework to guide the modeling of a doctoral program as an example of a longer-term mentoring and learning environment. Finally, I present a discussion on some potential future research directions.

6.1 Discussion

6.1.1 Research Objectives

My overriding research goal is to demonstrate how to build, calibrate and validate, and use a simulation model of a longer-term learning environment to explore pedagogical issues, including being able to pose hypothetical, ‘what-if’, research questions. This demonstration is made concrete through the development of a simulation for a real world longer-term learning environment, a doctoral program, focused on the interaction between different types of mentors (supervisors) and different kinds of mentees (doctoral students) in a longer-term mentoring environment (doctoral program). Insight from my research provides a template for other AIED researchers who desire to design AIED systems for supporting learning in other longer-term learning environments. This research also sets the stage for future research related to many factors that affect doctoral students’ persistence to degree.

6.1.2 Simulation for Supporting AIED System Development Process

Simulation has been considered as an important decision support tool since the 1950s and has been used in numerous areas including healthcare, the military, crowd behavior analysis, commerce, as well as education [144]–[147], [231], [232]. The idea of using simulation within AIED research was introduced more than two decades ago VanLehn et al. [1]. Simulated learners in the form of pedagogical agents have been used within various AIED systems to play significant roles, for example, playing the role of learning companions, acting as conversational agents, or

taking the role of teachable agents. Research results show that learning gains from the use of these virtual agents are dependent on agent and learner characteristics. The use of simulation is of increasing importance within AIED research as the field trends towards supporting longer-term learning.

There are several reasons why AIED system designers should have simulation as part of their toolkit. Using simulation as an evaluation tool enables rapid comparison among the various adaptive and personalization strategies to determine the most effective ones. Simulation allows for the testing and understanding of the possible impact of various adaptive and personalization measures in a learning environment before embarking on building an actual system to experiment with real learners. Moreover, the time taken to develop a simulation model would be relatively shorter than the time required to develop a functional real system prototype for evaluation purposes. Furthermore, simulation gives a researcher the ability to conduct experiments that shed light on real-world systems that are otherwise impractical to investigate because of the nature of the environment or the length of the investigation in real time that is required [34], [129], [143]. Finally, simulation can be used to replicate a real-world situation by modeling key characteristics of the target domain and learner behavior over a span of time enabling system designers to explore interactions among variables that are hidden in the real world as well as to answer hypothetical, ‘what-if’, design decision questions.

Simulation can be used at different stages of the AIED system building process and for different purposes. For instance, a simulation can be built to be used in the early stages of the design process for formative evaluation of proposed designs which might require developing a standalone simulation, for example, exploring how to create an effective sequence of learning material for an AIED system [205]. Or simulation can be used to help in the design of AIED system components as in the use of SimStudent for the authoring of student models for a cognitive tutor [233]. Or simulated learners can be developed as part of AIED systems in the capacity of pedagogical agents created to act as peer learners [234] or help in peer assessment [235]. Or a pedagogical agent can be developed and iteratively redesigned based on a complex simulated learner that is useful in studying theories of learning such as learning by teaching [236] or the design of effective pedagogical conversations [237].

Simulation can be used to address the lack of an adequate amount of data to allow exploratory study. For example, in an interactive learning game situation, generating interactive narra-

tive plans is crucial for gameplay experiences. Data-driven approaches are often used to inform policies for creating new plans; however, these data-driven approaches require huge datasets that are mostly not available. As a solution, researchers in [238] proposed the use of simulation, specifically the simulation of high-fidelity players. High-level fidelity simulated players which mimicked human player behaviors were generated and used to examine the effectiveness of reinforcement learning on narrative policy planning. Experimental results showed that the use of reinforcement learning led to narrative plans that are generalizable.

Some researchers might argue against the use of simulation to study pedagogical issues and AIED system design concerns. They could question whether it is possible to capture the right level of model fidelity to represent a target complex learning context being studied. As a result, they might raise concerns about the validity of the simulation's outputs and how well those outputs match real-world outcomes. Likewise, they may question the appropriateness of using simulation or simulated learners in the AIED system development process, especially where there is a lack of good data to either inform the simulation model or verify the simulation model outputs. Furthermore, they might even argue that all that needs to be known about system design issues can be gathered by engaging stakeholders (users and experts), not through simulation. Further, they may contend that with pervasive and ubiquitous technology there are loads of data about learners.

These are genuine concerns. However, as AIED research trends towards supporting learners in longer-term learning and mentoring contexts where many operations and the roles of various stakeholders are interconnected with innate variability and complexity, simulation is an essential aid to experimentation considering costs, time and the need for control of experimental set ups. Further, having loads of data is not enough by itself to explore design issues in AIED, especially when designing an AIED system to serve longer-term learners. Merely having lots of data doesn't help directly with 'what-if' questions - simulation is needed for this, although the data can be useful for informing the simulation models and functions. In addition, the concern about the validity of the model could be mitigated by clearly describing the level of model fidelity to be used in a given research.

Simulation model fidelity is an issue that arises when using simulation to study real-world phenomena because it dictates the level of reality captured in the model. Different researchers have demonstrated that it is possible to use different levels of model fidelity to gain insight into

various pedagogical research issues within educational technology research. The level of cognitive fidelity required is dictated by the nature of the research issues being explored. For example, while Champaign [148] used a very low fidelity model to explore the interactions of various parameters in a learning system, Matsuda et al. [118] used a model with high cognitive fidelity to reach compelling conclusions about the use of AIED systems to personalize student learning experiences. Erickson et al. [2] demonstrated that it is possible to use a medium fidelity simulation model to uncover interesting results. In choosing the appropriate fidelity of the simulation model it is important to consider the objectives of the research and the research questions researchers hope to explore [122]. Evaluation of AIED systems for supporting learning in longer-term learning contexts can be achieved by using simulation since it provides an economically cheaper, and ethically safer alternative to using human learners. Whatever the use of simulation, an important challenge is to understand how to design and develop pedagogical agents and simulated learning environments with the appropriate level of simulation model fidelity and including the participation of instructors and students in the design and development decisions.

6.2 Summary of the Research

Many AIED systems have been developed that assist learners in single learning sessions whose duration is hours or at most months. There is a relative paucity of research that provides a framework for exploring challenges facing designers of AIED systems for supporting mentoring of learners over longer-term durations, such as the mentoring of doctoral students. Designing mentoring systems to support longer-term learning and mentoring such as a doctoral program requires a large financial investment to pay for study participants as well as to pay for the cost of prototyping. It also takes a long time to seek ethics approval even for initial prototypes. Many human resources are required for a successful system development. Therefore, an important question is how can a system designer evaluate designs for a longer-term mentoring system while mitigating the cost and testing the system in a timely manner? The obvious answer is the use of simulation. This then leads to another important question: how can a system designer build a simulation of a longer-term learning context, especially one that would allow the system designer to ask hypothetical, ‘what-if’, research questions about the learning environment and make informed design decisions? Answering this question was the focus of my dissertation.

In my dissertation, I have explored the least examined role of simulation within AIED research: how to use simulation to support the AIED system design process. Specifically, I examined how to use simulation as a tool to support AIED system design for longer-term learning and mentoring environments. I presented a seven-step framework, adopted from [193], that can be used to guide the modeling and simulation of longer-term learning environments. These steps are: identification and formulation of research questions of interest around the target real-world system; designing a conceptual simulation model based on the research questions; identification of sources of real-world data to inform the simulation model; building a simulation model; calibrating and validating the resulting simulation model; performing experimentation with the resulting model, analyzing and interpreting the results of the experimentation; and presenting the results and suggesting improvements to the model as a foundation for future research directions.

To apply this seven-step framework to model building, I presented a case study where I demonstrated how to build and use a simulation of a doctoral program to help to understand the interactions of hidden variables of interest and to answer hypothetical, ‘what-if’, pedagogical questions concerning the doctoral program. The simulation model I developed for this case study is called SimDoc. At the outset, it was unclear how such a simulation model should be developed, what features it should have, and how these attributes should be informed. I was able to access a dataset from the target learning environment, the UofS doctoral program, to inform various attributes in order to explore initial research questions. But the dataset did not contain all the information necessary to inform the simulation.

The iterative process in the seven-step framework was very useful in helping overcome this difficulty. Over the analysis of what data was available I was able to refine the research questions. This allowed for a significant reduction in the number of stakeholders and attributes to model. I designed the SimDoc conceptual model, with five key components: agents, normative rules, dialogic rules, events, and scenes. I used the notion of agents to represent stakeholders according to their role. I only modeled two types of agents to represent doctoral students (who we will refer to as learners in our simulation) and supervisors. I chose to create a medium-level fidelity simulation model. Though the model was already simplified, I found out that the UofS dataset still did not contain enough information to inform all the agent attributes necessary to enable the exploration of the identified research questions.

Knowing that the UofS dataset was insufficient, I considered other sources of information to augment the UofS dataset. These sources included different departmental web pages and various studies in the literature. The diverse nature of these studies and students involved allowed for the capturing of a broad spectrum of doctoral students' behavior. I used the data from these sources to inform characteristics of two important types of functions in SimDoc: behavior functions and evaluation functions. Behavior functions inform the decision making of an active agent. Evaluation functions determine the outcomes of the various interactions between different agents. This process of seeking additional data led to the need to refine the SimDoc conceptual model.

After refining the SimDoc model by establishing essential elements, functions, and characteristics and identifying relevant sources of data to inform it, I implemented the SimDoc model in AnyLogicTM, a Java-based platform for modeling and simulation¹⁶. I modeled SimDoc's entities following the agent-based modeling (ABM) technique. Using ABM, I was able to model all entities of interest as agents with their various characteristics. After implementing the SimDoc simulation model, it was important to ascertain the validity of its output by comparing the simulation output against the real-world dataset along key measures of interest. In the UofS dataset, I had aggregate data on expected emergent behavior but not necessarily data to specifically inform exact parameter values for the SimDoc model so that SimDoc reproduces the same emergent behavior.

Identifying that it was vital to tune parameter values for those attributes whose parameter values were not well-known, I calibrated and validated the SimDoc model. The calibration process consisted of 500 simulation runs. The best run had a 93% match between the UofS dataset and the SimDoc dataset. I then validated the SimDoc model. I provided a pseudo-algorithm that can be used to determine the number of runs necessary for a model to produce results with stability and appropriate variability to ensure that the model had not been overfitted. The overall outcome of the validation analysis showed that 100 runs were the appropriate number of times SimDoc should be run in order to be confident in the results and avoid having to needlessly make extra simulation runs. With 100 simulation runs, the SimDoc model reproduced a consolidated dataset that is statistically similar to the UofS dataset on variables of interest.

¹⁶ <https://www.anylogic.com/> last accessed February 12, 2019

Having calibrated and validated the SimDoc model, I used the resulting SimDoc model to illustrate how to explore interactions of "hidden" variables and to ask hypothetical, 'what-if', pedagogical issues concerning personalization within a doctoral program. This resulted in discoveries that constitute predictions for the real-world system. For example, simulating (exaggerated) hypothetical situations in which there is a supervisor of only a single type reveals that learners who prefer weekly meeting frequency will be subject to very high attrition rates, regardless of the supervisory type. However, having learners who preferred quarterly meeting frequency leads to high completion rates across all the four supervisory styles.

6.3 Contributions

The central contribution of this dissertation is the demonstration of how to build, calibrate and validate, and use a simulation model of a longer-term learning environment to explore various aspects of the environment, and in particular to hypothetical, 'what-if', research questions.

This dissertation also presented several contributions to advanced learning technology research and most specifically to artificial intelligence in education. In my research I have:

- Presented an extended seven-step framework (based on [193]) for guiding the design and modeling of simulation models. In a case study, I showed how AIED and other advanced learning technology researchers could use this seven-step framework to guide the building, informing, and validating of a pedagogical simulation model for exploring different research issues in longer-term learning and mentoring environments.
- Identified a pedagogical use of simulation not explored very much in AIED or other advanced learning technology communities: how simulation could be used to explore various hypothetical, 'what-if', pedagogical questions related to understanding issues in longer-term learning and mentoring environments.
- Developed a medium fidelity simulation model, a rarely investigated level of fidelity. In this dissertation, I have described the steps that I took to design and build SimDoc, a simulation of doctoral program, as an example of a longer-term learning environment. Other designers with aspirations to build simulations of learning environment, especially longer-term environment, can benefit greatly by following the steps I followed in designing and developing SimDoc.

- Illustrated how to inform a simulation model by showing how to combine data from diverse sources related to phenomena of interest. AIED system designers creating simulation models to be used to explore issues other than those in the doctoral program can follow the steps I took to inform the SimDoc model, since these steps are generalizable to other domains and contexts. The first step is all about identifying the experimental question to be explored using simulation. To successfully use simulation, it is important to have clear and specific experimental questions. Doing so helps in the formulation of a conceptual model and data collection. In addition, it helps the designer to think critically about the problem he or she is interested in exploring. The next step involves identifying, collecting, and analyzing data about the target learning environment to inform the various behavior and evaluation functions. There are important data and behaviour functions of the doctoral program that the UofS data didn't provide any insight on. I have shown how to inform the model from either the real-world environment, from results in the research literature, or both.
- Demonstrated how to calibrate a simulation model using a baseline dataset gathered from the target learning environment. With data from different sources, it was important to combine and synthesize them to make sure they still reproduce learning outcomes that match the target learning environment. This is where the calibration process comes in. Calibration also has a further advantage of helping determine the values of as yet unsigned parameters and attributes. I showed how to use the calibration process available in AnyLogic™, a Java-Based simulation platform to find a very good baseline model in SimDoc. Designers of other simulations can follow my example to calibrate their own models.
- Showed how to validate a simulation model by providing a pseudo-algorithm that can be used to determine how many replications of a simulation run to perform in order to be confident in the results produced by the simulation. With the best calibrated model, it is important to establish that the model is stable (i.e. that it matches the real-world system over time) but also that the model's attribute values are not overfitted (i.e. that it displays appropriate variability over time). This is where validation comes in. The pseudo-algorithm that I have provided is generalizable and a system designer can use it by following the steps I took in validating SimDoc and validate the model they have built.

Moreover, the validation pseudo-algorithm terminates after a certain number of simulation runs, when the appropriate stability and variability has been achieved. This number can then be used to determine the number of runs to use in actual experiments with the simulation. This pseudo-algorithm is based on statistical measures that generalize to any simulation context.

- Showed how to develop an experimental program to explore specific AIED and advanced learning technology research questions through simulation. In particular I explored some interesting patterns in supervisor-student interactions, as they affected time-in-program, and graduation and attrition rates. While the particular experiments were SimDoc specific, others can learn from them even in different learning contexts, especially in the use of ‘what if’ scenarios that allow exploration of situations that don’t occur in the real world, but nevertheless might reveal interesting patterns.
- Illustrated the importance of using simulation in exploring various learning domains where data is not readily available, particularly self-directed, and longer-term learning scenarios. Simulation allows exploration of such domains while also enabling deeper insight into student models and learning contexts.

6.4 Limitations

The research I conducted in this thesis has limitations. One of the issues concerns the availability of enough data on graduate students. The amount of detailed information about graduate students that the university of Saskatchewan was willing to provide was limited. To address this concern, I had to glean data from other sources including the research literature and doctoral program webpages, but using such derived data is certainly not as informative as having all the data from a single source – the UofS doctoral program – the target environment in this case.

Another issue is that the SimDoc model as currently constructed is a major simplification of the doctoral program given the social complexity of the doctoral environment. This is a valid concern; however, I informed, calibrated and validated the SimDoc model around the research issues of interest. In order to experiment with other issues concerning the doctoral program, the SimDoc model must be revised, informed, calibrated, and validated accordingly.

In addition, the actual experimental results are possibly not reliable. However, at the very least they constitute predictions of what could possibly happen in a doctoral program if the con-

ditions were at least somewhat similar to the assumptions underlying the simulation. These findings can form the basis of new research questions that can be explored either with a modified simulation model or with a real-world experimental set up.

Some researchers have expressed reservations concerning the use of simulation to evaluate learning systems¹⁷. The argument is that results derived from running simulations might not reflect the reality and accuracy of learning outcomes. This concern is justified particularly if a system designer wants a simulation to handle all the complexities of a learning environment. However, a simulation can still be informed well enough to answer specific questions about the learning environment, even at low and medium fidelities. Moreover, even with limitations, simulation is the only possible way of exploring issues in some specific contexts, especially longer-term learning environments. Constraints of research time and the nature of the research context necessitate the use of simulation. Additionally, simulation enables the exploration of issues in learning contexts that would otherwise be difficult to study. The fidelity and validity of the models play key roles in the accuracy or the trustworthiness of the simulation predictive output.

6.5 Future Work

It would be nice to see follow up research directions resulting from this dissertation work. One of these directions is better informing SimDoc model following a few directions I have briefly discussed in section 5.5 and a few more I suggest in section 6.5.1. These directions focus on particular experiments that could be run following the SimDoc experiments performed in this dissertation. Another research direction is to have follow up research into the art/science of building longer-term pedagogical simulations.

6.5.1 Extensions to the Current SimDoc Simulation

In the future, there are several extensions to the SimDoc baseline simulation model that could be improved on. Some of these directions might require changes in the composition of SimDoc's conceptual model while others might not.

Better Informing SimDoc

¹⁷ At the Artificial Intelligence in Education 2013 conference, for example.

Although there are many possible improvements that could be done to improve SimDoc and its predictive power, the most consequential one in my view is informing the whole model with data from one source – the target learning environment to be investigated. Future studies might explore the impact of informing all aspects of the SimDoc model with data from the same doctoral program as opposed to using only aggregate data. This might also require the reproduction of similar research performed in the referenced literature in the context of the target learning environment. This research can be performed without the need for changing SimDoc’s conceptual model. The only step I foresee to be necessary before experimenting with such an updated SimDoc model would be calibration and validation. Such an undertaking would shed further light on the use of simulation in longer-term learning environments that are not captured in this thesis research.

Impact of Informed Meeting Frequencies

To follow up the SimDoc experiments described in this thesis, a specific fidelity issue is to explore the impact of informing desired meeting frequencies with data from the target learning environment. This would require an historical dataset on student-supervisor meetings – frequencies, durations, and learning outcomes. Such a dataset could be collected through a detailed quantitative study. With this change, it would be interesting to see how the research findings will compare with the predictive findings provided by the current set up.

Exploring Impact of Departmental Factors

Future studies might consider improving the SimDoc simulation model by actually limiting the scope of the model from the doctoral program as a whole to modeling a single department’s doctoral program. In this thesis research, the SimDoc model is informed by aggregated data from different departments of the University of Saskatchewan doctoral program and data from the literature that were initially collected and used for different purposes. There is a need for research on whether and to what extent do different departmental factors influence doctoral learners’ persistence to degree. Studying various departmental types and incorporating their factors into the model will lead to a better understanding of factors impacting persistence in different departments and programs.

6.5.2 Directions for Simulation Research

The following are more general ideas for research that might be done in the future to explore the use of simulation to investigate issues within longer-term learning environments.

Integrating a Simulation Model into an Actual Learning Environment

It would be very interesting to explore ways of integrating a simulation model into an actual learning environment, much as SimStudent has done. This would probably lead to the simulation model getting more sophisticated over time as the actual learning environment produces more data. At best there could be a positive feedback loop here, where simulation leads to a better real-world learning environment which creates data that can lead to a better simulation, and so on. This iterative design-use cycle would be highly informative for a simulation, as well as leading to immediate real world applications at each iteration.

Deepening the Cognitive Fidelity of a Simulation Model

Further, it would be fascinating to investigate different ways of deepening a simulation model's cognitive fidelity, gradually going from medium fidelity to higher fidelity. This would involve an increase in the number of attributes and functions in the simulation over time, perhaps informed as indicated in my previous point through interleaving the simulation with the use in the real world of actual learning systems spinning off from the simulation. Eventually so many attributes and behaviours could be captured that the fidelity becomes high, rather than medium.

Better Informing Simulations of Other Longer-Term Learning Environments

It would be interesting to explore better informing simulations of other longer-term learning environments. The pervasive and ubiquitous nature of technology has enabled many learners to engage in learning beyond the limits of learning institutions. More and more people are using learning environments such as MOOCs to improve their knowledge. While this makes access to learning material easy, the completion rates of learners using such platforms is very low and would benefit greatly with improved mentoring services within the platform. I would suggest using simulation to explore how such mentoring services could be integrated into the platform of interest. This would include informing a simulation model by taking advantage of the large datasets that are increasingly becoming available in well used MOOCs. This step would require exploring ways of better informing a simulation (beyond what would be done in the immediate future for

SimDoc). This would also include finding ways of integrating other scientific results, as I have explored in this dissertation work. This future direction would involve accessing a much bigger scientific literature.

Despite the limitations of this research and the need for future research to overcome them, this thesis makes a strong case for the importance of simulation in designing learning environments, especially in longer-term domains. It also confirms the possibility of actually building such a simulation of a longer-term learning environment through a case study demonstrating how to build a "medium fidelity" simulated doctoral program, the SimDoc model. Specific lessons are drawn at every stage of the creation and experimentation with SimDoc that can inform other advanced learning technology researchers who wish to use simulation in their own research.

REFERENCES

- [1] K. VanLehn, S. Ohlsson, and R. Nason, "Applications of Simulated Students: An Exploration," *Journal of Artificial Intelligence in Education*, vol. 5, no. 2, pp. 1–42, 1994.
- [2] G. Erickson, S. Frost, S. Bateman, and G. I. McCalla, "Using the Ecological Approach to Create Simulations of Learning Environments," in the *Proceedings of the 16th International Conference on Artificial Intelligence in Education*, 2013, pp. 411–420.
- [3] H. Groenvynck, K. Vandevelde, and R. Van Rossem, "The Ph.D. Track: Who Succeeds, Who Drops Out?," *Research Evaluation*, vol. 22, no. 4, pp. 199–209, 2013.
- [4] S. P. Kerlin, "Pursuit of the Ph.D.: 'Survival of the Fittest,' Or is it Time for a New Approach?," *Education Policy Analysis Archives*, vol. 3, no. 16, pp. 1–28, 1995.
- [5] F. Landriscina, *Simulation and Learning: A Model-Centered Approach*. Springer Science & Business Media, 2013.
- [6] J. M. Beaubien and D. P. Baker, "The Use of Simulation for Training Teamwork Skills in Health Care: How Low Can You Go?," *Quality and Safety in Health Care*, vol. 13, no. suppl_1, pp. i51–i56, Oct. 2004.
- [7] J. Banks, "Introduction to Simulation," in the *Proceedings of the 31st Conference on Winter Simulation*, 1999, vol. 1, pp. 7–13.
- [8] D. S. A Vrijdag and T. van Terwisga, "Systematic modelling , verification , calibration and validation of a ship propulsion simulation model," *Journal of Marine Engineering & Technology*, vol. 8, no. 3, pp. 3–20, 2014.
- [9] B. L. Heath and R. R. Hill, "Final Report Developing an Agent-Based Modeling Verification and Validation Approach for Improving Air Force Analytical Support," Dayton, OH, 2010.
- [10] N. L. Schroeder, O. O. Adesope, and R. B. Gilbert, "How Effective are Pedagogical Agents for Learning? A Meta-Analytic Review," *Journal of Educational Computing Research*, vol. 49, no. 1, pp. 1–39, 2013.
- [11] S. Aslan and C. M. Reigeluth, "A Trip to the Past and Future of Educational Computing: Understanding Its Evolution," *Contemporary Educational Technology*, vol. 2, no. 1, pp. 1–17, 2011.
- [12] H. Nwana, "Intelligent Tutoring Systems: An Overview," *Artificial Intelligence Review*, vol. 4, no. 4, pp. 251–277, 1990.
- [13] J. R. Anderson, C. F. Boyle, A. T. Corbett, and M. W. Lewis, "Cognitive Modeling and Intelligent Tutoring," *Artificial Intelligence*, vol. 42, no. 1, pp. 7–49, 1990.
- [14] J. R. Anderson, C. F. Boyle, and B. J. Reiser, "Intelligent Tutoring Systems," *Science*, vol. 228, no. 4698, pp. 456–462, 1984.
- [15] J. Ireson, "Private Tutoring: How Prevalent and Effective is it?," *London Review of Education*, vol. 2, no. 2, pp. 109–122, 2004.
- [16] N. Manouselis, H. Drachsler, R. Vuorikari, and H. Hummel, "A Sneak Preview to the Chapter 'Recommender Systems in Technology Enhanced Learning,'" in the *Proceedings*

- of the *3rd International Workshop on Social Information Retrieval for TEL*, 2009, pp. 1–25.
- [17] E. G. Blanchard, “Socio-Cultural Imbalances in AIED Research: Investigations, Implications and Opportunities,” *International Journal of Artificial Intelligence in Education*, vol. 25, no. 2, pp. 204–228, 2015.
 - [18] O. C. Santos, M. Kravcik, and J. G. Boticario, “Preface to Special Issue on User Modelling to Support Personalization in Enhanced Educational Settings,” *International Journal of Artificial Intelligence in Education*, vol. 26, no. 3, pp. 809–820, 2016.
 - [19] T. Käser, A. G. Busetto, B. Solenthaler, G. Baschera, J. Kohn, K. Kucian M. Aster, and M. Gross, “Modelling and Optimizing Mathematics Learning in Children,” *International Journal of Artificial Intelligence in Education*, vol. 23, no. 1–4, pp. 115–135, 2013.
 - [20] D. B. Lenat and P. J. Durlach, “Reinforcing Math Knowledge by Immersing Students in a Simulated Learning-By-Teaching Experience,” *International Journal of Artificial Intelligence in Education*, vol. 24, no. 3, pp. 216–250, 2014.
 - [21] K. VanLehn, C. Lynch, K. Schulze, J. A. Shapiro, R. Shelby, L. Taylor, D. Treacy, A. Weinstein, and M. Wintersgill, “The Andes Physics Tutoring System: Lessons Learned,” *International Journal of Artificial Intelligence in Education*, vol. 15, no. 3, pp. 147–204, 2005.
 - [22] B. D. Nye, A. C. Graesser, and X. Hu, “Autotutor and Family: A Review of 17 Years of Natural Language Tutoring,” *International Journal of Artificial Intelligence in Education*, vol. 24, no. 4, pp. 427–469, 2014.
 - [23] K. VanLehn, “Reflections on Andes’ Goal-free User Interface,” *International Journal of Artificial Intelligence in Education*, vol. 26, no. 1, pp. 82–90, 2016.
 - [24] J. Kay, “Lifelong Learner Modeling for Lifelong Personalized Pervasive Learning,” *Transactions on Learning Technologies, IEEE*, vol. 1, no. 4, pp. 215–228, 2008.
 - [25] L. M. Tang and K. Yacef, “Personalising E-Learning Systems: Lessons Learned From a Vocational Education Case Study,” in the Proceedings of the *5th Workshop on Personalization Approaches for Learning Environments*, held in conjunction with the *23rd International Conference on User Modeling, Adaptation and Personalization*, 2015, pp. 24–30.
 - [26] M. Chacón-Rivas, O. C. Santos, and J. G. Boticario, “Modeling Learner Information within an Integrated Model on Standard-Based Representations,” in the Proceedings of the *5th Workshop on Personalization Approaches for Learning Environments (PALE 2015)*, 2015, pp. 31–38.
 - [27] C. P. Leão and A. C. Ferreira, “Talking about Mentoring Relationships from the Perspectives of PhD Students: A Conceptual Model Development,” in the *Proceedings of the 2015 International Conference on Interactive Collaborative Learning, ICL 2015*, 2015, pp. 1209–1215.
 - [28] O. C. Santos, “Training the Body: The Potential of AIED to Support Personalized Motor Skills Learning,” *International Journal of Artificial Intelligence in Education*, vol. 26, no. 2, pp. 730–755, 2016.
 - [29] M. Hosseini Bidokht and A. Assareh, “Life-Long Learners Through Problem-Based and

- Self Directed Learning,” *Procedia Computer Science*, vol. 3, pp. 1446–1453, Jan. 2011.
- [30] V. Aleven, I. Roll, B. M. McLaren, and K. R. Koedinger, “Help Helps, but only so Much: Research on Help Seeking with Intelligent Tutoring Systems,” *International Journal of Artificial Intelligence in Education*, vol. 26, no. 1, pp. 205–223, 2016.
- [31] P. H. M. Cremers, A. E. J. Wals, R. Wesselink, N. Nieveen, and M. Mulder, “Self-directed lifelong learning in hybrid learning configurations,” *International Journal of Lifelong Education*, vol. 33, no. 2, pp. 207–232, 2014.
- [32] J. L. Sabourin, L. R. Shores, B. W. Mott, and J. C. Lester, “Understanding and Predicting Student Self-Regulated Learning Strategies in Game-Based Learning Environments,” *International Journal of Artificial Intelligence in Education*, vol. 23, no. 1–4, pp. 94–114, 2013.
- [33] K. Leelawong and G. Biswas, “Designing Learning by Teaching Agents: The Betty’s Brain System,” *International Journal of Artificial Intelligence in Education*, vol. 18, no. 3, pp. 181–208, 2008.
- [34] F. Dorça, “Implementation and use of Simulated Students for Test and Validation of new Adaptive Educational Systems: a Practical Insight,” *International Journal of Artificial Intelligence in Education*, vol. 25, no. 3, pp. 319–345, 2015.
- [35] G. Veletsianos and G. S. Russell, “Pedagogical Agents,” in *Handbook of Research on Educational Communications and Technology*, J. M. Spector, M. D. Merrill, J. Elen, and M. J. Bishop, Eds. Springer New York, 2014, pp. 759–769.
- [36] S. K. D’Mello, “Giving Eyesight to the Blind: Towards Attention-Aware AIED,” *International Journal of Artificial Intelligence in Education*, vol. 26, no. 2, pp. 645–659, 2016.
- [37] A. Iqbal, R. Oppermann, A. Patel, and Kinshuk, “A Classification of Evaluation Methods for Intelligent Tutoring Systems,” in *Software Ergonomie*, 1999, vol. 181, pp. 169–181.
- [38] J. E. Greer and M. A. Mark, “Evaluation Methods for Intelligent Tutoring Systems Revisited,” *International Journal of Artificial Intelligence in Education*, vol. 26, no. 1, pp. 387–392, 2016.
- [39] B. P. Woolf, *Building Intelligent Interactive Tutors: Student-Centered Strategies for Revolutionizing E-Learning*, Revised. Morgan Kaufmann Publisher, 2010.
- [40] R. S. Hart, “The Illinois Plato Foreign Languages Project,” *Computer Assisted Language Instruction Consortium Journal*, vol. 12, no. 4, pp. 15–37, 1995.
- [41] M. Urban-Lurain, “Intelligent Tutoring Systems: An Historic Review in the Context of the Development of Artificial Intelligence and Educational Psychology,” *Intelligent Tutoring Systems*, 1996. [Online]. Available: <http://www.cse.msu.edu/rgroups/cse101/ITS/its.htm>. [Accessed: 20-Aug-2018].
- [42] J. S. Brown, R. R. Burton, and A. G. Bell, “SOPHIE: A Step Toward Creating a Reactive Learning Environment,” *International Journal of Man-Machine Studies*, vol. 7, no. 5, pp. 675–696, 1975.
- [43] J. R. Carbonell, “AI in CAI: An Artificial-Intelligence Approach to Computer-Assisted Instruction,” *IEEE Transactions on Man-Machine Systems*, vol. 11, no. 4, pp. 190–202,

- 1970.
- [44] D. Moursund, *Brief Introduction to Educational Implications of Artificial Intelligence*. University of Oregon Technical Report, 2006.
 - [45] J. Underwood and R. Luckin, “What is AIED and Why Does Education Need It?,” Teaching and Learning Research Programme Technical Report, UK, 2011.
 - [46] D. Sleeman and J. S. Brown, *Intelligent Tutoring Systems*. London: Academic Press, 1982.
 - [47] D. R. Peachey and G. I. McCalla, “Using Planning Techniques in Intelligent Tutoring Systems,” *International Journal of Man-Machine Studies*, vol. 24, no. 1, pp. 77–98, 1986.
 - [48] J. Vassileva and B. Wasson, “Instructional Planning Approaches: From Tutoring Towards Free Learning,” in the *Proceedings of the European Conference on Artificial Intelligence in Education*, 1996, pp. 1–8.
 - [49] B. Brecht, G. I. McCalla, J. E. Greer, and M. L. Jones, “Planning the Content of Instruction,” in the *Proceedings of the 3rd International Conference on Artificial Intelligence and Education*, 1989, no. May, pp. 32–41.
 - [50] P. Brusilovsky, E. W. Schwarz, and G. Weber, “ELM-ART: An Intelligent Tutoring System on World Wide Web,” in the *Proceedings of the 3rd International Conference on Intelligent Tutoring Systems*, 1996, pp. 261–269.
 - [51] O. C. Santos, J. G. Boticario, S. Baldiris, G. Moreno, D. Huerva, and R. Fabregat, “Recommender Systems For Lifelong Learning Inclusive Scenarios,” in the Proceedings of the *Workshop on Recommender Systems*, held in conjunction with the *17th Biennial European Conference on Artificial Intelligence*, 2008, pp. 45–58.
 - [52] J. Self, “Theoretical Foundations for Intelligent Tutoring Systems,” *Journal of Artificial Intelligence in Education*, vol. 1, no. 4, pp. 3–14, 1990.
 - [53] M. J. Baker, “The Negotiation of Meaning in Epistemic Situations,” *International Journal of Artificial Intelligence in Education*, vol. 26, no. 1, pp. 133–149, 2016.
 - [54] H. C. Lane, G. I. McCalla, C.-K. Looi, and S. Bull, “The Next 25 Years: How Advanced Interactive Learning Technologies will Change the World,” *International Journal of Artificial Intelligence in Education*, vol. 26, no. 1, pp. 539–543, 2016.
 - [55] G. Cumming and A. McDougall, “Mainstreaming AIED into Education?,” *International Journal of Artificial Intelligence in Education*, vol. 11, no. 2, pp. 197–207, 2000.
 - [56] J. Self, “The Birth of IJAIED,” *International Journal of Artificial Intelligence in Education*, vol. 26, no. 1, pp. 4–12, 2016.
 - [57] J. Self, “The Defining Characteristics of Intelligent Tutoring Systems Research: ITSS Care, Precisely,” *International Journal of Artificial Intelligence in Education*, vol. 10, no. 3, pp. 350–364, 1999.
 - [58] R. S. J. D. Baker, “Stupid Tutoring Systems, Intelligent Humans,” *International Journal of Artificial Intelligence in Education*, vol. 26, no. 2, pp. 600–614, 2016.
 - [59] J. Janssen, A. J. Berlanga, and R. Koper, “Evaluation of the Learning Path Specification: Lifelong Learners’ Information Needs,” *Journal of Educational Technology & Society*, vol. 14, no. 3, pp. 1–7, 2011.

- [60] R. Kagia, T. Linden, H. A. Patrinos, and J. Salmi, “Lifelong Learning in the Global Knowledge Economy: Challenges for Developing Countries,” Technical Report. Washington, D.C, 2003.
- [61] G. Fischer, “Lifelong Learning and its Support with New Media,” *International Encyclopedia of Social and Behavioral Sciences*, vol. 13, no. 41, pp. 1–7, 2001.
- [62] A. J. Cropley, “Some Guidelines for the Reform of School Curricula in the Perspective of Lifelong Education,” *International Review of Education*, vol. 24, no. 1, pp. 21–33, 1978.
- [63] L. Hall, S. Tazzyman, C. Hume, B. Endrass, M. Lim, G. Hofstede, A. Paiva, E. Andre, A. Kappas, and R. Aylett, “Learning to Overcome Cultural Conflict through Engaging with Intelligent Agents in Synthetic Cultures,” *International Journal of Artificial Intelligence in Education*, vol. 25, no. 2, pp. 291–317, 2015.
- [64] R. G. Bagnall, “Lifelong Learning as an Idea for Our Times: An Extended Review,” *International Journal of Lifelong Education*, vol. 28, no. 2, pp. 277–285, Mar. 2009.
- [65] M. O. Oketch, “Perspectives on Lifelong Learning in Africa,” in *Second International Handbook of Lifelong Learning*, D. N. Aspin, J. Chapman, K. Evans, and R. Bagnall, Eds. Dordrecht: Springer Netherlands, 2012, pp. 375–397.
- [66] P. C. Candy, *Self-Direction for Lifelong Learning. A Comprehensive Guide to Theory and Practice*. San Francisco, CA: Jossey-Bass, 1991.
- [67] D. M. Gaymer, “Continuing Education and Lifelong Learning Trends,” *Encyclopedia of Management*. MM., Helms (Ed.), pp. 128–131, 2006.
- [68] Canadian Council on Learning, “State of Learning in Canada: Toward a Learning Future,” Technical Report. Canadian Council on Learning, Ottawa, Ontario, 2008.
- [69] P. Jarvis, “Global Trends in Lifelong Learning and the Response of the Universities,” *Comparative Education*, vol. 35, no. 2, pp. 249–257, 1999.
- [70] R. Brooks and G. Everett, “The Impact of Higher Education on Lifelong Learning,” *International Journal of Lifelong Education*, vol. 27, no. 3, pp. 239–254, May 2008.
- [71] P. C. Candy, “Reaffirming a Proud Tradition Universities and Lifelong Learning,” *Active Learning in Higher Education*, vol. 1, no. 2, pp. 101–125, 2000.
- [72] A. Parkinson, “Developing the Attribute of Lifelong Learning,” in the *Proceedings of the 29th Annual Conference on Frontiers in Education*, 1999, pp. 16–20.
- [73] K. Ecclestone, “Care or Control?: Defining Learners’ Needs for Lifelong Learning,” *The British Journal of Educational Studies*, vol. 47, no. 4, pp. 332–347, 1999.
- [74] J. Crowther, “‘In and Against’ Lifelong Learning: Flexibility and the Corrosion of Character,” *International Journal of Lifelong Education*, vol. 23, no. 2, pp. 125–136, Mar. 2004.
- [75] M. Tight, “Lifelong Learning: Opportunity or Compulsion?,” *British Journal of Educational Studies*, vol. 46, no. 3, pp. 251–263, Sep. 1998.
- [76] F. Coffield, “Breaking the Consensus: Lifelong Learning as Social Control,” *British Educational Research Journal*, vol. 25, no. 4, pp. 479–499, Sep. 1999.
- [77] B. L. McCombs, “Motivation and Lifelong Learning,” *Educational Psychologist*, vol. 26,

- no. 2, pp. 117–127, 1991.
- [78] R. Klamma, M. A. Chatti, E. Duval, H. Hummel, E. T. Hvannberg, M. Kravcik, E. Law, A. Naeve, and P. Scott, “Social Software for Life-long Learning,” *Educational Technology & Society*, vol. 10, no. 3, pp. 72–83, 2007.
- [79] K. Larsen, “Lifelong Learning for All,” *OECD Observer*, 2001. [Online]. Available: http://oecdobserver.org/news/archivestory.php/aid/432/Lifelong_learning_for_all.html. [Accessed: 21-Aug-2018].
- [80] D. N. Aspin, J. Chapman, K. Evans, and R. Bagnall, “Introduction and Overview,” in *Second International Handbook of Lifelong Learning*, 2012, pp. 45–84.
- [81] J. Field, “Lifelong Learning, Welfare and Mental Well-being into Older Age: Trends and Policies in Europe,” in *Active Ageing, Active Learning*, G. Boulton-Lewis and M. Tam, Eds. Dordrecht: Springer Netherlands, 2012, pp. 11–20.
- [82] J. Underwood and R. Luckin, “Themes and Trends in AIED Research, 2000 to 2010,” Teaching and Learning Research Programme, UK, 2011.
- [83] N. Longworth, *Lifelong Learning in Action: Transforming Education in the 21st Century*. London and Sterling, VA: Taylor & Francis or Routledge, 2013.
- [84] R. Koper and C. Tattersall, “New Directions for Lifelong Learning Using Network Technologies,” *British Journal of Educational Technology*, vol. 35, no. 6, pp. 689–700, 2004.
- [85] M. Kalz, “Lifelong Learning and its Support with New Technologies,” *International Encyclopedia of the Social & Behavioral Sciences*. Oxford: Pergamon, 2014.
- [86] S. de Freitas, I. Harrison, G. Magoulas, A. Mee, F. Mohamad, M. Oliver, G. Papamarkos, and A. Poulouvasilis, “The Development of a System for Supporting the Lifelong Learner,” *British Journal of Educational Technology*, vol. 37, no. 6, pp. 867–880, Nov. 2006.
- [87] R. C. Dorf, “Short Papers and Classroom Notes,” in *IEEE Transactions on Education*, 1969, vol. 1962, no. March, pp. 50–52.
- [88] G. Stahl, “Collaborative Information Environments to Support Knowledge Construction by Communities,” *AI & Society, Journal of Knowledge, Culture and Communication*, vol. 14, no. 1, pp. 71–97, 2000.
- [89] M. Sharples, “The Design of Personal Mobile Technologies for Lifelong Learning,” *Computers & Education*, vol. 34, no. 3–4, pp. 177–193, Apr. 2000.
- [90] M. Anjorin, C. Rensing, K. Bischoff, C. Bogner, L. Lehmann, A. L. Reger, N. Faltin, A. Steinacker, A. Lüdemann, and R. D. García, “CROKODIL - A Platform for Collaborative Resource-Based Learning,” in the *Proceedings of the 6th European Conference on Technology Enhanced Learning*, 2011, pp. 29–42.
- [91] J. Kay and B. Kummerfeld, “Lifelong Learner Modeling,” in *Adaptive Technologies for Training and Education*, P. J. Durlach and A. Lesgold, Eds. New York, USA: Cambridge University Press, 2012, pp. 140–164.
- [92] J. Kay, “Lifelong Learner Modeling for Lifelong Personalized Pervasive Learning,” *Transactions on Learning Technologies, IEEE*, vol. 1, no. 4, pp. 215–228, 2008.

- [93] J. Kay, “Life-Long Learning, Learner Models and Augmented Cognition,” in the *Proceedings of the 9th International Conference on Intelligent Tutoring Systems*, 2008, pp. 3–5.
- [94] N. Van Labeke, G. D. Magoulas, and A. Poulouvasilis, “Searching for ‘People Like Me’ in a Lifelong Learning System,” in the *Proceedings of the 4th European Conference on Technology Enhanced Learning*, 2009, pp. 106–111.
- [95] N. Van Labeke, A. Poulouvasilis, and G. Magoulas, “Using Similarity Metrics for Matching Lifelong Learners,” in the *Proceedings of the 9th International Conference on Intelligent Tutoring Systems*, 2008, pp. 142–151.
- [96] N. Manouselis, H. Drachler, R. Vuorikari, H. Hummel, and R. Koper, “Recommender Systems in Technology Enhanced Learning,” in *Recommender Systems Handbook*, F. Ricci, L. Rokach, B. Shapira, and P. B. Kantor, Eds. Boston, MA: Springer US, 2011, pp. 387–415.
- [97] R. Koper, “Building Learning Networks for Lifelong Learners: Challenges, Models, Technologies and Standards,” in the *Proceedings of the IEEE International Conference on Advanced Learning Technologies*, 2006, pp. 1098–1099.
- [98] V. Aleven, “Rule-Based Cognitive Modeling for Intelligent Tutoring Systems,” in *Studies in Computational Intelligence: Advances in Intelligent Tutoring Systems*, vol. 308, R. Nkambou, J. Bourdeau, and R. Mizoguchi, Eds. Springer Berlin Heidelberg, 2010, pp. 33–62.
- [99] A. Mitrovic and S. Ohlsson, “Evaluation of a Constraint-Based Tutor for a Database Language,” *International Journal of Artificial Intelligence in Education*, vol. 10, no. 3, pp. 238–256, 1999.
- [100] S. Ohlsson, “Constraint-Based Student Modeling,” *Journal of Artificial Intelligence in Education*, vol. 3, no. 4, pp. 429–451, 1993.
- [101] J. Self, “Bypassing the Intractable Problem of Student Modelling,” in *Intelligent Tutoring Systems: At the Crossroad of Artificial Intelligence and Education*, C. Frasson and G. Gauthier, Eds. 107-23, Norwood, N.J.: Ablex, 1990, pp. 107–123.
- [102] S. Bull and T. Nghiem, “Helping Learners to Understand Themselves with a Learner Model Open to Students, Peers and Instructors,” in the *Proceedings of the Workshop on Individual and Group Modelling Methods that Help Learners Understand Themselves*, held in conjunction with the *International Conference on Intelligent Tutoring Systems*, 2002, pp. 5–13.
- [103] V. Dimitrova, G. I. McCalla, and S. Bull, “‘Open Learner Models: Future Research Directions’ Special Issue of the IJAIED (Part 2),” *International Journal of Artificial Intelligence in Education*, vol. 17, no. 3, pp. 217–226, 2007.
- [104] V. Dimitrova, “STyLE-OLM: Interactive Open Learner Modelling,,” *International Journal of Artificial Intelligence in Education*, vol. 13, no. 1, pp. 35–78, 2003.
- [105] S. Bull and J. Kay, “SMILI: A Framework for Interfaces to Learning Data in Open Learner Models, Learning Analytics and Related Fields,” *International Journal of Artificial Intelligence in Education*, vol. 26, no. 1, pp. 293–331, 2016.
- [106] V. Kodaganallur and R. R. Weitz, “A Comparison of Model-Tracing and Constraint-

- Based Intelligent Tutoring Paradigms,” *International Journal of Artificial Intelligence in Education*, vol. 15, no. 2, pp. 117–144, 2005.
- [107] A. Mitrovic and S. Ohlsson, “Implementing CBM: SQL-Tutor after Fifteen Years,” *International Journal of Artificial Intelligence in Education*, vol. 26, no. 1, pp. 150–159, 2016.
- [108] M. A. Mark and J. E. Greer, “Evaluation Methodologies for Intelligent Tutoring Systems,” *Journal of Artificial Intelligence in Education*, vol. 4, no. 306, p. 129, 1993.
- [109] B. Martin, A. Mitrovic, K. R. Koedinger, and S. Mathan, “Evaluating and Improving Adaptive Educational Systems with Learning Curves,” *User Modeling and User-Adapted Interaction*, vol. 21, no. 3, pp. 249–283, 2011.
- [110] S. E. Stemler, “A Comparison of Consensus, Consistency, and Measurement Approaches to Estimating Interrater Reliability,” *Practical Assessment Research Evaluation*, vol. 9, no. 4, pp. 1–19, 2004.
- [111] A. Kittur, E. H. Chi, and B. Suh, “Crowdsourcing User Studies With Mechanical Turk,” in the *Proceedings of the Special Interest Group on Computer-Human Interaction Conference on Human Factors in Computing Systems*, 2008, pp. 453–456.
- [112] R. S. J. D. Baker, “Data Mining for Education,” in *International Encyclopedia of Education*, 3rd ed., vol. 7, 2010, pp. 112–118.
- [113] P. C. Austin, “An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies,” *Multivariate Behavioral Research*, vol. 46, no. 3, pp. 399–424, 2011.
- [114] I. Roll and R. Wylie, “Evolution and Revolution in Artificial Intelligence in Education,” *International Journal of Artificial Intelligence in Education*, vol. 26, no. 2, pp. 582–599, 2016.
- [115] T. Murray, “Authoring Intelligent Tutoring Systems: An Analysis of the State of the Art,” *International Journal of Artificial Intelligence in Education*, vol. 10, pp. 98–129, 1999.
- [116] N. T. Heffernan and C. L. Heffernan, “The ASSISTments Ecosystem: Building a Platform that Brings Scientists and Teachers Together for Minimally Invasive Research on Human Learning and Teaching,” *International Journal of Artificial Intelligence in Education*, vol. 24, no. 4, pp. 470–497, 2014.
- [117] N. Pinkwart, “Another 25 Years of AIED? Challenges and Opportunities for Intelligent Educational Technologies of the Future,” *International Journal of Artificial Intelligence in Education*, vol. 26, no. 2, pp. 771–783, 2016.
- [118] N. Matsuda, W. W. Cohen, K. R. Koedinger, V. Keiser, R. Raizada, E. Yarzebinski, S. P. Watson, and G. Stylianides, “Studying the Effect of Tutor Learning Using a Teachable Agent that Asks the Student Tutor for Explanations,” in the *Proceedings of the 4th International Conference on Digital Game and Intelligent Toy Enhanced Learning*, 2012, pp. 25–32.
- [119] S. Lallé, J. Mostow, V. Luengo, and N. Guin, “Comparing Student Models in Different Formalisms by Predicting Their Impact on Help Success,” in the *Proceedings of the 16th International Conference on Artificial Intelligence in Education*, 2013, pp. 161–170.

- [120] J. P. González-Brenes and Y. Huang, “Using Data from Real and Simulated Learners to Evaluate Adaptive Tutoring Systems,” in the *Proceedings of the 2nd Workshop on Simulated Learners*, held in conjunction with the *17th International Conference on Artificial Intelligence in Education*, 2015, pp. 31–34.
- [121] P. I. Pavlik and J. R. Anderson, “Practice and Forgetting Effects on Vocabulary Memory: An Activation-Based Model of the Spacing Effect,” *Cognitive Science*, vol. 29, no. 4, pp. 559–586, 2005.
- [122] G. I. McCalla and J. Champaign, “Simulated Learners,” *IEEE Intelligent Systems*, vol. 28, no. 4, pp. 67–71, 2013.
- [123] V. Devedzic and A. Harrer, “Architectural Patterns in Pedagogical Agents,” in the *Proceedings of the 6th International Conference on Intelligent Tutoring Systems*, 2002, pp. 81–90.
- [124] W. L. Johnson and J. C. Lester, “Face-to-Face Interaction with Pedagogical Agents, Twenty Years Later,” *International Journal of Artificial Intelligence in Education*, vol. 26, no. 1, pp. 25–36, 2016.
- [125] C. Y. Chou, T. W. Chan, and C. J. Lin, “Redefining the Learning Companion: The Past, Present, and Future of Educational Agents,” *Computers and Education*, vol. 40, no. 3, pp. 255–269, 2003.
- [126] A. Silvervarg, C. Kirkegaard, J. Nirme, M. Haake, and A. Gulz, “Steps Towards a Challenging Teachable Agent,” in *Proceedings of the 14 International Conference on Intelligent Virtual Agents - IVA 2014*, 2014, vol. LNAI 8637, pp. 410–419.
- [127] N. Matsuda, E. Yarzebinski, V. Keiser, R. Raizada, G. J. Stylianides, W. W. Cohen, and K. R. Koedinger, “Learning by Teaching SimStudent – An Initial Classroom Baseline Study Comparing with Cognitive Tutor,” in the *Proceedings of the 15th International Conference on Artificial Intelligence in Education*, 2011, pp. 213–221.
- [128] A. C. Graesser, “Conversations with AutoTutor Help Students Learn,” *International Journal of Artificial Intelligence in Education*, vol. 26, no. 1, pp. 124–132, 2016.
- [129] A. Vizcaíno, “A Simulated Student Can Improve Collaborative Learning,” *International Journal of Artificial Intelligence in Education*, vol. 15, pp. 3–40, 2005.
- [130] J. Vassileva, G. I. McCalla, and J. E. Greer, “From Small Seeds Grow Fruitful Trees: How the PHelpS Peer Help System Stimulated a Diverse and Innovative Research Agenda over 15 Years,” *International Journal of Artificial Intelligence in Education*, vol. 26, no. 1, pp. 431–447, 2016.
- [131] A. Gulz, M. Haake, A. Silvervarg, B. Sjöden, and G. Veletsianos, “Building a Social Conversational Pedagogical Agent: Design Challenges and Methodological Approaches,” in *Conversational Agents and Natural Language Interaction: Techniques and Effective Practices*, D. Perez-Marin, Ed. 2011, pp. 128–155.
- [132] W. L. Johnson, J. W. Rickel, and J. C. Lester, “Animated Pedagogical Agents: Face-to-Face Interaction in Interactive Learning Environments,” *International Journal of Artificial Intelligence in Education*, vol. 11, no. 1, pp. 47–78, 2000.
- [133] S. Ur and K. VanLehn, “Steps: A Simulated, Tutorable Physics Student,” *Journal of Artificial Intelligence in Education*, vol. 6, no. 4, pp. 405–437, 1995.

- [134] N. Matsuda, W. W. Cohen, and K. R. Koedinger, “Teaching the Teacher: Tutoring SimStudent Leads to More Effective Cognitive Tutor Authoring,” *International Journal of Artificial Intelligence in Education*, vol. 25, no. 1, pp. 1–34, 2015.
- [135] C. J. MacLellan, N. Matsuda, and K. R. Koedinger, “Toward a Reflective SimStudent: Using Experience to Avoid Generalization Errors,” in the Proceedings of the *1st Workshop on Simulated Learners*, held in conjunction with the *16th International Conference on Artificial Intelligence in Education*, 2013, vol. 1009, pp. 51–60.
- [136] N. Matsuda, E. Yarzebinski, V. Keiser, R. Raizada, G. J. Stylianides, and K. R. Koedinger, “Studying the Effect of a Competitive Game Show in a Learning by Teaching Environment,” *International Journal of Artificial Intelligence in Education*, vol. 23, no. 80, pp. 1–21, 2013.
- [137] J. Champaign, “Peer-Based Intelligent Tutoring Systems: A Corpus-Oriented Approach,” in the *Proceedings of the Canadian Conference on Artificial Intelligence*, 2010, pp. 393–396.
- [138] S. Frost and G. I. McCalla, “Exploring Through Simulation the Effects of Peer Impact on Learning,” in the Proceedings of the *1st Workshop on Simulated Learners*, held in conjunction with the *16th International Conference on Artificial Intelligence in Education*, 2013, pp. 21–30.
- [139] T.-W. Chan and B. Arthur, “Studying with the Prince: The Computer as a Learning Companion,” in the *Proceedings of the 1st International Conference on Intelligent Tutoring Systems*, 1988, pp. 194–200.
- [140] S. Palthepu, J. E. Greer, and G. I. McCalla, “Learning by Teaching,” in the *Proceedings of the 1st International Conference on the Learning Sciences*, 1991, pp. 357–363.
- [141] C. Epp and A. Makos, “Using Simulated Learners and Simulated Learning Environments within a Special Education Context,” in the Proceedings of the *1st Workshop on Simulated Learners*, held in conjunction with the *16th International Conference on Artificial Intelligence in Education*, 2013, pp. 1–10.
- [142] A. Gulz, M. Haake, A. Silvervarg, B. Sjöden, and G. Veletsianos, “Building a Social Conversational Pedagogical Agent,” in *Conversational Agents and Natural Language Interaction*, IGI Global, 2011, pp. 128–155.
- [143] C. Conati, L. Fratamico, S. Kardan, and I. Roll, “Comparing Representations for Learner Models in Interactive Simulations,” in the *Proceedings of the 17th International Conference on Artificial Intelligence in Education*, 2015, pp. 74–83.
- [144] J. M. Beaubien, “The Use of Simulation for Training Teamwork Skills in Health Care: How Low Can You Go?,” *Quality and Safety in Health Care*, vol. 13, no. suppl_1, pp. i51–i56, Oct. 2004.
- [145] G. Augenbroe, “Trends in Building Simulation,” *Building and Environment*, vol. 37, no. 8, pp. 891–902, 2002.
- [146] P. Blikstein, D. Abrahamson, and U. Wilensky, “NetLogo : Where We Are , Where We’re Going,” in the *Proceedings of the Annual Meeting of Interaction Design and Children*, 2005, vol. 23.
- [147] F. Qiu and X. Hu, “Modeling Group Structures in Pedestrian Crowd Simulation,”

- Simulation Modelling Practice and Theory*, vol. 18, no. 2, pp. 190–205, 2010.
- [148] J. Champaign, “Peer-Based Intelligent Tutoring Systems: A Corpus-Oriented Approach,” Unpublished Doctoral Dissertation. University of Waterloo, 2012.
- [149] B. Hollocks, “Assessing Simulation Learning in Higher Education,” in the *Proceedings of the 37th Conference on Winter Simulation*, 2005, pp. 2297–2303.
- [150] E. M. Anderson and A. L. Shannon, “Toward a Conceptualization of Mentoring,” *Journal of Teacher Education*, vol. 39, no. 1, pp. 38–42, 1988.
- [151] T. Gatfield, “An Investigation into PhD Supervisory Management Styles: Development of a Dynamic Conceptual Model and its Managerial Implications,” *Journal of Higher Education Policy and Management*, vol. 27, no. 3, pp. 311–325, 2005.
- [152] A. Oluwabunmi, P. Fred, and V. Julita, “Dynamics of Trust in Group Peer Mentorship,” in the *Proceedings of the 1st International Workshop on Intelligent Mentoring Systems*, held in conjunction with the *13th International Conference on Intelligent Tutoring Systems*, 2016.
- [153] J. P. Preston, M. J. Ogenchuk, and J. K. Nsiah, “Peer Mentorship and Transformational Learning: PhD Student Experiences,” *Canadian Journal of Higher Education*, vol. 44, no. 1, pp. 52–68, 2014.
- [154] T. D. Allen, J. E. A. A. Russell, and S. B. Maetzke, “Formal Peer Mentoring: Factors Related to Proteges’ Satisfaction and Willingness to Mentor Others,” *Group & Organization Management*, vol. 22, no. 4, pp. 488–507, Dec. 1997.
- [155] F. D. Minor, “Building Effective Peer Mentor Programs,” in *From Learning Communities and Student Affairs: Partnering for Powerful Learning*, 2007, pp. 467–482.
- [156] B. Hodges, “Factors that can Influence Mentorship Relationships,” *Paediatric*, vol. 21, no. 6, pp. 32–35, 2009.
- [157] C. B. Kandiko and I. M. Kinchin, “What Is a PhD? Process Versus Product in PhD Supervision,” in the *Proceedings of the 4th International Conference on Concept Mapping*, 2010, no. 2002, pp. 451–459.
- [158] A. McCallin and S. Nayar, “Postgraduate research supervision: A critical review of current practice,” *Teaching in Higher Education*, vol. 17, no. 1, pp. 63–74, 2012.
- [159] Russell L. Huizing, “Mentoring Together: A Literature Review of Group Mentoring,” in the *Proceedings of the Northeastern Association Of Business, Economics, and Technology Proceedings*, 2010, pp. 58–71.
- [160] P. D. Lonnie D. Inzer, C. B. Crawford, “A Review of Formal and Informal Mentoring: Processes, Problems, and Design,” *Journal of Leadership Education*, vol. 4, no. 1, pp. 31–50, 2005.
- [161] B. R. Ragins and J. L. Cotton, “Mentor Functions and Outcomes: A Comparison of Men and Women in Formal and Informal Mentoring Relationships,” *Journal of Applied Psychology*, vol. 84, no. 4, pp. 529–550, 1999.
- [162] O. Adewoyin and J. Vassileva, “Can Online Peer-Review Systems Support Group Mentorship?,” in the *Proceedings of the 16th International Conference on Artificial Intelligence in Education*, 2013, pp. 737–741.

- [163] N. Van Labeke, J. Kirby, and T. E. Roberts, "Personalised and Adaptive Mentoring in Medical Education – the myPAL Project," in the Proceedings of the *1st International Workshop on Intelligent Mentoring Systems*, held in conjunction with the *13th International Conference on Intelligent Tutoring Systems*, 2016.
- [164] D. W. Shaffer and A. Graesser, "Using a Quantitative Model of Participation in a Community of Practice to Direct Automated Mentoring in an Ill-Formed Domain," in the *Proceedings of the 10th International Conference on Intelligent Tutoring Systems*, 2010, pp. 427–469.
- [165] G. Carrington, "Supervision as a Reciprocal Learning Process," *Educational Psychology in Practice*, vol. 20, no. 1, pp. 31–42, 2004.
- [166] F. J. Elgar, "PhD Degree Completion in Canadian Universities," Technical Report. Dalhousie University, Nova Scotia, Canada, pp. 1-31, 2003.
- [167] M. Walpole, N. W. Burton, K. Kanyi, and A. Jackenthal, "Selecting Successful Graduate Students: In-Depth Interviews With GRE Users," Technical Report. Princeton, NJ, 2002.
- [168] L. Declou, "Linking Levels to Understand Graduate Student Attrition in Canada," Technical Report. McMaster University, Hamilton, Ontario, Canada, 2014.
- [169] M. Torrance, G. V. Thomas, and E. J. Robinson, "The Writing Strategies of Graduate Research Students in the Social Sciences," *Higher Education*, vol. 27, no. 3, pp. 379–392, 1994.
- [170] C. Jungbluth, I. M. Macfarlane, P. M. Veach, and B. S. Leroy, "Why is Everyone So Anxious?: An Exploration of Stress and Anxiety in Genetic Counseling Graduate Students," *Journal of Genetic Counseling*, vol. 20, no. 3, pp. 270–86, Jun. 2011.
- [171] R. Tamburri, "The PhD is in Need of Revision," *University Affairs*, 2013. [Online]. Available: http://www.universityaffairs.ca/the-phd-is-in-need-of-revision.aspx#latest_data. [Accessed: 26-Dec-2016].
- [172] L. Charbonneau, "PhD Completion Rates and Times to Completion in Canada," *University Affairs*, 2013. [Online]. Available: <http://www.universityaffairs.ca/margin-notes/phd-completion-rates-and-times-to-completion-in-canada/>. [Accessed: 31-Jul-2014].
- [173] C. of G. Schools, "Ph.D. Completion and Attrition: Analysis of Baseline Demographic Data from the Ph.D. Completion Project Executive Summary," Technical Report. Council of Graduate Schools. 2008, pp. 1–4. Available at: <https://cgsnet.org/phd-completion-and-attrition-analysis-baseline-demographic-data-phd-completion-project-0>
- [174] B. C. Seagram, J. Gould, and S. W. Pyke, "An Investigation of Gender and Other Variables on Time to Completion of Doctoral Degrees," *Research in Higher Education*, vol. 39, no. 3, pp. 319–335, 1998.
- [175] J. Rodwell and R. Neumann, "Predictors of timely doctoral student completions by type of attendance: The utility of a pragmatic approach," *Journal of Higher Education Policy and Management*, vol. 30, no. 1, pp. 65–76, Feb. 2008.
- [176] L. McAlpine and J. Norton, "Reframing Our Approach to Doctoral Programs: An Integrative Framework for Action And Research," *Higher Education Research & Development*, vol. 25, no. 1, pp. 3–17, Feb. 2006.

- [177] T. Mainhard, R. van der Rijst, J. van Tartwijk, and T. Wubbels, "A Model for the Supervisor-Doctoral Student Relationship," *Higher Education*, vol. 58, no. 3, pp. 359–373, 2009.
- [178] T. Heath, "A Quantitative Analysis of PhD Students' Views of Supervision," *Higher Education Research and Development*, vol. 21, no. August 2013, pp. 37–41, 2002.
- [179] L. L. Baird, "Disciplines and Doctorates: The Relationships between Program Characteristics and the Duration of Doctoral Study," *Research in Higher Education*, vol. 31, no. 4, pp. 369–385, 1990.
- [180] T. Darisi, V. J. Davidson, K. Korabik, and S. Desmarais, "Commitment to Graduate Studies and Careers in Science and Engineering: Examining Women's and Men's Experiences," *International Journal of Gender, Science and Technology*, vol. 2, no. 1, pp. 48–64, 2010.
- [181] C. M. Golde, "The Role of the Department and Discipline in Doctoral Student Attrition: Lessons from Four Departments," *The Journal of Higher Education*, vol. 76, no. 6, pp. 669–700, 2005.
- [182] L. A. Garcia, "Factors of Attrition in Cohort Doctoral Education: A Self-determination Theory Perspective," Technical Report. The University of Texas at Arlington, USA, 2014.
- [183] F. D. Ampaw and A. J. Jaeger, "Completing the Three Stages of Doctoral Education: An Event History Analysis," *Research in Higher Education*, vol. 53, no. 6, pp. 640–660, 2012.
- [184] V. Tinto, "Taking Student Success Seriously: Rethinking the First Year of College," *Ninth Annual Intersession Academic Affairs Forum*, vol. 19, no. 2, pp. 1–8, 2005.
- [185] V. Tinto, "Dropout from Higher Education: A Theoretical Synthesis of Recent Research," *Review of Educational Research*, vol. 45, no. 1, pp. 89–125, 1975.
- [186] A. Ali and F. Kohun, "Dealing with Isolation Feelings in IS Doctoral Programs," *International Journal of Doctoral Studies*, vol. 1, no. 1, pp. 21–33, 2006.
- [187] M. L. Orellana, A. Darder, A. Pérez, and J. Salinas, "Improving Doctoral Success by Matching PhD Students with Supervisors," *International Journal of Doctoral Studies*, vol. 11, pp. 87–103, 2016.
- [188] E. Lavelle and K. Bushrow, "Writing Approaches of Graduate Students," *Educational Psychology*, vol. 27, no. 6, pp. 807–822, 2007.
- [189] I. J. West, G. Gokalp, E. V. Pena, L. Fischer, and J. Gupton, "Exploring Effective Support Practices For Doctoral Students' Degree Completion," *College Student Journal*, vol. 45, no. 2, pp. 310–323, 2011.
- [190] N. Z. Abiddin and A. Ismail, "Effective Supervisory Approach in Enhancing Postgraduate Research Studies," *International Journal of Humanities and Social Science*, vol. 1, no. 2, pp. 206–217, 2011.
- [191] C. Askew, R. Dixon, R. McCormick, K. Callaghan, G. Y. Wang, and B. Shulruf, "Facilitators and Barriers to Doctoral Supervision : A Case Study in Health Sciences," *Issues in Educational Research*, vol. 26, no. 1, pp. 1–9, 2016.
- [192] F. Brouns, H. Vogten, J. Janssen, and A. Finders, "E-portfolios in Lifelong Learning," in

Proceedings of the First International Conference on Technological Ecosystem for Enhancing Multiculturalism - TEEM '13, 2013, pp. 535–540.

- [193] A. M. Law, A. M. Law, C. M. Macal, and M. J. North, “Proceedings of the 2008 Winter Simulation Conference S. J. Mason, R. R. Hill, L. Mönch, O. Rose, T. Jefferson, J. W. Fowler eds.,” pp. 39–47, 2008.
- [194] A. M. Law, “How to Build Valid and Credible Simulation Models,” in *Proceedings of the 2008 Winter Simulation Conference*, 2008, pp. 39–47.
- [195] M. Esteva, J.-A. Rodriguez-Aguilar, C. Sierra, P. Garcia, and J. L. Arcos, “On the Formal Specification of Electronic Institutions,” in the *Proceedings of the Agent Mediated Electronic Commerce Conference*, 2001, pp. 126–147.
- [196] N. Gilbert, *Agent-Based Models*. SAGE, 2008.
- [197] J. J. Odell, H. V. D. Parunak, and M. Fleischer, “The Role of Roles in Designing Effective Agent Organizations,” in *Software Engineering for Large-Scale Multi-Agent Systems*, 2003, pp. 27–38.
- [198] H. V. D. Parunak and J. J. Odell, “Representing Social Structures in UML,” in *International Workshop on Agent-Oriented Software Engineering II*, 2002, pp. 1–16.
- [199] Z. A. Pardos and M. V. Yudelson, “Towards Moment of Learning Accuracy,” in the *Proceedings of the 1st Workshop on Simulated Learners*, held in conjunction with the *16th International Conference on Artificial Intelligence in Education*, 2013, vol. 7, pp. 61–70.
- [200] Y. Wang and N. Heffernan, “A Comparison of Two Different Methods to Individualize Students and Skills,” in the *Proceedings of the 16th International Conference on Artificial Intelligence in Education*, 2013, vol. LNAI 7926, pp. 836–839.
- [201] J. T. Folsom-Kovarik, R. E. Wray, and L. Hamel, “Adaptive Assessment in an Instructor-Mediated System,” in the *Proceedings of the 16th International Conference on Artificial Intelligence in Education*, 2013, pp. 571–574.
- [202] M. Ueno, “Adaptive Testing Based on Bayesian Decision Theory,” in the *Proceedings of the 16th International Conference on Artificial Intelligence in Education*, 2013, vol. LNAI 7926, pp. 712–716.
- [203] U. Faghihi, P. Fournier-Viger, and R. Nkambou, “A Cognitive Tutoring Agent with Automatic Reasoning Capabilities,” in *Connection Science*, 2011, pp. 2010–2011.
- [204] S. Frost and G. I. McCalla, “Exploring Through Simulation an Instructional Planner for Dynamic Open-Ended Learning Environments,” in the *Proceedings of the 17th International Conference on Artificial Intelligence in Education*, 2015, pp. 578–581.
- [205] J. Champaign and R. Cohen, “A Multiagent, Ecological Approach to Content Sequencing,” in the *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems*, 2010, pp. 10–14.
- [206] H. J. C. Berendsen, D. Van Der Spoel, and R. Van Drunen, “GROMACS: A Message-Passing Parallel Molecular Dynamics Implementation,” *Computer Physics Communications*, vol. 91, no. 1, pp. 43–56, 1995.
- [207] S. Bourke, A. Holbrook, T. Lovat, and P. Farley, “Attrition, Completion and Completion Times of PhD Candidates,” in *AARE Annual Conference*, 2004, vol. 28, pp. 1–14.

- [208] R. Conejo, E. Guzmán, and M. Trella, “The SIETTE Automatic Assessment Environment,” *International Journal of Artificial Intelligence in Education*, vol. 26, no. 1, pp. 270–292, 2016.
- [209] R. Pelánek and P. Jarušek, “Student Modeling Based on Problem Solving Times,” *International Journal of Artificial Intelligence in Education*, vol. 25, no. 4, pp. 493–519, 2015.
- [210] T. Wright and R. Cochrane, “Factors Influencing Successful Submission of PhD Theses,” *Studies in Higher Education*, vol. 25, no. 2, pp. 181–195, 2000.
- [211] D. E. K. Lelei and G. I. McCalla, “How to Use Simulation in the Design and Evaluation of Learning Environments with Self-directed Longer-Term Learners,” in the *Proceedings of the 19th International Conference on Artificial Intelligence in Education*, 2018, pp. 253–266.
- [212] C. Gershenson, “Complex Philosophy,” in *In the Proceedings of the 1st Biennial Seminar on the Philosophical, Methodological, and Epistemological Implications of Complexity Theory*, 2002, pp. 1–8.
- [213] M. A. Rau, “Do Knowledge-Component Models Need to Incorporate Representational Competencies?,” *International Journal of Artificial Intelligence in Education*, vol. 27, no. 2, pp. 298–319, 2017.
- [214] R. Carlson, V. Keiser, N. Matsuda, K. R. Koedinger, and C. P. Rosé, “Building a Conversational SimStudent,” in the *Proceedings of the 11th International Conference on Intelligent Tutoring Systems*, 2012, pp. 563–569.
- [215] D. E. K. Lelei and G. I. McCalla, “The Role of Simulation in the Development of Mentoring Technology to Support Longer-Term Learning,” in the *Proceedings of the 3rd International Workshop on Intelligent Mentoring Systems*, held in conjunction with the *19th International Conference on Artificial Intelligence in Education*, 2018.
- [216] J. Bogdoll, A. Hartmanns, and H. Hermanns, “Simulation and Statistical Model Checking for Modestly Nondeterministic Models,” in *International GI/ITG Conference on Measurement, Modelling, and Evaluation of Computing Systems and Dependability and Fault Tolerance*, 2012, pp. 249–252.
- [217] E. Koehler, E. Brown, and S. J.-P. A. Haneuse, “On the Assessment of Monte Carlo Error in Simulation-Based Statistical Analyses,” *The American Statistician*, vol. 63, no. 2, pp. 155–162, 2009.
- [218] James G. Booth and S. Sarkar, “Monte Carlo Approximation of Bootstrap Variances,” *The American Statistician*, vol. 52, no. 4, pp. 354–357, 1998.
- [219] L. T. Truong, M. Sarvi, G. Currie, and T. M. Garoni, “How Many Simulation Runs are Required to Achieve Statistically Confident Results: A Case Study of Simulation-Based Surrogate Safety Measures,” in *IEEE 18th International Conference on Intelligent Transportation Systems*, 2015, pp. 274–278.
- [220] W. Van Joolingen, “Design and Implementation of Simulation-Based Discovery Environments: The SMISLE Solution,” *Journal of Artificial Intelligence in Education*, vol. 7, no. 4, pp. 253–276, 1996.
- [221] R. B. Rosenberg-Kima and Z. A. Pardos, “Is This Model for Real? Simulating Data to

- Reveal the Proximity of a Model to Reality,” in the *Proceedings of the 17th International Conference on Artificial Intelligence in Education*, 2015, pp. 78–87.
- [222] G. Weber, “Individual Selection of Examples in an Intelligent Learning Environment,” *International Journal of Artificial Intelligence in Education*, vol. 7, no. 1, pp. 3–31, 2015.
- [223] C. Liu, “A Simulation-Based Experience in Learning Structures of Bayesian Networks to Represent How Students Learn Composite Concepts,” *International Journal of Artificial Intelligence in Education*, vol. 18, no. 3, pp. 237–285, 2008.
- [224] M. Dzikovska, N. Steinhauser, E. Farrow, J. Moore, and G. Campbell, “BEETLE II: Deep Natural Language Understanding and Automatic Feedback Generation for Intelligent Tutoring in Basic Electricity and Electronics,” *International Journal of Artificial Intelligence in Education*, vol. 24, no. 3, pp. 284–332, 2014.
- [225] M. C. Desmarais and X. Pu, “A Bayesian Student Model without Hidden Nodes and Its Comparison with Item Response Theory,” *International Journal of Artificial Intelligence in Education*, vol. 15, no. 4, pp. 291–323, 2005.
- [226] A. Yeh and M. Corp, “More Accurate Tests for the Statistical Significance of Result Differences,” in *Proceedings of the 18th International Conference on Computational Linguistics*, 2000, pp. 947–953.
- [227] M. A. R. B, N. Zimmerman, R. S. J. D. Baker, T. Titchener, and J. Cooper, “Using A Model For Learning And Memory To Simulate Learner Response In Spaced Practice,” in the *Proceedings of the 18th International Conference on Artificial Intelligence in Education*, 2017, pp. 644–649.
- [228] F. E. Ritter, M. J. Schoelles, K. S. Quigley, and L. C. Klein, “Determining the Number of Simulation Runs Treating Simulations as Theories by not Sampling their Behavior,” in *Human-in-the-loop simulations: Methods and Practice*, 2011, pp. 97–116.
- [229] R. M. Ryan and E. L. Deci, “Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions.,” *Contemporary Educational Psychology*, vol. 25, no. 1, pp. 54–67, 2000.
- [230] J. J. Cardona, “Determined to Succeed: Motivation Towards Doctoral Degree Completion,” Technical Report. The University of Michigan, 2013.
- [231] S. Ramachandran, R. Jensen, J. Ludwig, E. Domeshek, and Thomas Haines, “ITADS: A Real-World Intelligent Tutor to Train Troubleshooting Skills,” in the *Proceedings of the 19th International Conference on Artificial Intelligence in Education*, 2018, pp. 463–468.
- [232] S. Wijewickrema, X. Ma, P. Piromchai, R. Briggs, J. Bailey, G. Kennedy, and S. O’Leary, “Providing Automated Real-Time Technical Feedback for Virtual Reality Based Surgical Training: Is the Simpler the Better?,” in the *Proceedings of the 19th International Conference on Artificial Intelligence in Education*, 2018, pp. 584–598.
- [233] N. Matsuda, W. W. Cohen, J. Sewall, G. Lacerda, and K. R. Koedinger, “Predicting Students’ Performance with SimStudent: Learning Cognitive Skills from Observation,” in the *Proceedings of the 13th International Conference on Artificial Intelligence in Education*, 2007, pp. 467–476.
- [234] N. Matsuda, V. Keiser, R. Raizada, A. Tu, G. Stylianides, W. W. Cohen, and K. R. Koedinger, “Learning by Teaching SimStudent: Technical Accomplishments and an

- Initial Use with Students,” in the *Proceedings of the 10th International Conference on Intelligent Tutoring Systems*, 2010, pp. 317–326.
- [235] A. De Andrade Barbosa and E. De Barros Costa, “Simulated Learners in Peers Assessment for Introductory Programming Courses,” in the Proceedings of the *Central Europe Workshops*, held in conjunction with the *17th International Conference on Artificial Intelligence in Education*, 2015, pp. 55–64.
- [236] M. M. T. Rodrigo, A. Ong, R. Bringula, R. S. Basa, C. Dala Cruz, and N. Matsuda, “Impact of Prior Knowledge and Teaching Strategies on Learning by Teaching,” in the Proceedings of the *1st Workshop on Simulated Learners*, held in conjunction with the *16th International Conference on Artificial Intelligence in Education*, 2013, pp. 71–80.
- [237] Y. Hayashi, “Togetherness: Multiple Pedagogical Conversational Agents as Companions in Collaborative Learning,” in the *Proceedings of the 12th International Conference on Intelligent Tutoring Systems*, 2014, vol. LNCS 8474, pp. 114–123.
- [238] P. Wang, J. Rowe, W. Min, B. Mott, and J. Lester, “High-Fidelity Simulated Players for Interactive Narrative Planning Data-Driven Interactive Narrative Planning,” in the *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, 2018, pp. 3884–3890.
- [239] A. Silvervarg, M. Haake, and A. Gulz, “Educational Potentials in Visually Androgynous Pedagogical Agents,” in the *Proceedings of the 16th International Conference on Artificial Intelligence in Education*, 2013, pp. 599–602.
- [240] L. G. Pezzullo, J. B. Wiggins, M. H. Frankosky, W. Min, K. E. Boyer, B. W. Mott, E. N. Wiebe, and J. C. Lester, “‘Thanks Alisha, Keep in Touch’: Gender Effects and Engagement with Virtual Learning Companions,” in the *Proceedings of the 18th International Conference on Artificial Intelligence in Education*, 2017, pp. 299–310.
- [241] J. Self, “Student Models in Computer-Aided Instruction,” *International Journal of Man-Machine Studies*, vol. 6, no. 2, pp. 261–276, 1974.

APPENDIX A – BEHAVIORAL ETHICS CERTIFICATES



UNIVERSITY OF
SASKATCHEWAN

Behavioural Research Ethics Board (Beh-REB)

Certificate of Approval

PRINCIPAL INVESTIGATOR
Gordon McCalla

DEPARTMENT
Computer Science

BEH#
13-75

INSTITUTION(S) WHERE RESEARCH WILL BE CONDUCTED
University of Saskatchewan

STUDENT RESEARCHER(S)
David Edgar Kiprof Lelei

FUNDER(S)
INTERNALLY FUNDED

TITLE
Examining Factors that Lead to Successful Help Seeking in Online Forums

ORIGINAL REVIEW DATE
05-Mar-2013

APPROVAL ON
05-Mar-2013

APPROVAL OF:
Application for Behavioural Research Ethics
Review

EXPIRY DATE
04-Mar-2014

Full Board Meeting

Delegated Review

CERTIFICATION

The University of Saskatchewan Behavioural Research Ethics Board has reviewed the above-named research project. The proposal was found to be acceptable on ethical grounds. The principal investigator has the responsibility for any other administrative or regulatory approvals that may pertain to this research project, and for ensuring that the authorized research is carried out according to the conditions outlined in the original protocol submitted for ethics review. This Certificate of Approval is valid for the above time period provided there is no change in experimental protocol or consent process or documents.

Any significant changes to your proposed method, or your consent and recruitment procedures should be reported to the Chair for Research Ethics Board consideration in advance of its implementation.

ONGOING REVIEW REQUIREMENTS

In order to receive annual renewal, a status report must be submitted to the REB Chair for Board consideration within one month of the current expiry date each year the study remains open, and upon study completion. Please refer to the following website for further instructions: http://www.usask.ca/research/ethics_review/

Beth Bilson, Chair
University of Saskatchewan
Behavioural Research Ethics Board

Please send all correspondence to:

Research Ethics Office
University of Saskatchewan
Box 5000 RPO University, 1602-110 Gymnasium Place
Saskatoon SK S7N 4J8
Telephone: (306) 966-2975 Fax: (306) 966-2069



UNIVERSITY OF SASKATCHEWAN

Behavioural Research Ethics Board (Beh-REB)

Certificate of Re-Approval

PRINCIPAL INVESTIGATOR Gordon McCalla	DEPARTMENT Computer Science	Beh # 13-75
--	--------------------------------	----------------

INSTITUTION (S) WHERE RESEARCH WILL BE CARRIED OUT

University of Saskatchewan
Saskatoon SK

STUDENT RESEARCHER(S)

David Edgar Kiprop Lelei

FUNDER(S)

INTERNALLY FUNDED

TITLE:

Examining Factors that Lead to Successful Help Seeking in Online Forums

RE-APPROVED ON

27-Feb-2014

EXPIRY DATE

26-Feb-2015

Full Board Meeting

Delegated Review

CERTIFICATION

The University of Saskatchewan Behavioural Research Ethics Board has reviewed the above-named research project. The proposal was found to be acceptable on ethical grounds. The principal investigator has the responsibility for any other administrative or regulatory approvals that may pertain to this research project, and for ensuring that the authorized research is carried out according to the conditions outlined in the original protocol submitted for ethics review. This Certificate of Approval is valid for the above time period provided there is no change in experimental protocol or consent process or documents.

Any significant changes to your proposed method, or your consent and recruitment procedures should be reported to the Chair for Research Ethics Board consideration in advance of its implementation.

ONGOING REVIEW REQUIREMENTS

In order to receive annual renewal, a status report must be submitted to the REB Chair for Board consideration within one month of the current expiry date each year the study remains open, and upon study completion. Please refer to the following website for further instructions: http://www.usask.ca/research/ethics_review/

Beth Bilson, Chair
University of Saskatchewan
Behavioural Research Ethics Board

Please send all correspondence to

Research Ethics Office
University of Saskatchewan
Box 5000 RPO University, 1607 110 Gymnasium Place
Saskatoon, SK S7N 4J8
Phone: (306) 966-2975 Fax: (306) 966-2069



Behavioural Research Ethics Board (Beh-REB)

Certificate of Re-Approval

PRINCIPAL INVESTIGATOR Gordon McCalla	DEPARTMENT Computer Science	Beh # 13-75
--	--------------------------------	----------------

INSTITUTION (S) WHERE RESEARCH WILL BE CARRIED OUT

University of Saskatchewan
Saskatoon SK

STUDENT RESEARCHER(S)

David Edgar Kiprof Lelei

FUNDER(S)

NATURAL SCIENCES & ENGINEERING RESEARCH
COUNCIL OF CANADA (NSERC)

TITLE:

Examining Factors that Lead to Successful Help Seeking in Online Forums

RE-APPROVED ON 03-Mar-2016	EXPIRY DATE 02-Mar-2017
-------------------------------	----------------------------

Full Board Meeting

Delegated Review

CERTIFICATION

The University of Saskatchewan Behavioural Research Ethics Board has reviewed the above-named research project. The proposal was found to be acceptable on ethical grounds. The principal investigator has the responsibility for any other administrative or regulatory approvals that may pertain to this research project, and for ensuring that the authorized research is carried out according to the conditions outlined in the original protocol submitted for ethics review. This Certificate of Approval is valid for the above time period provided there is no change in experimental protocol or consent process or documents.

Any significant changes to your proposed method, or your consent and recruitment procedures should be reported to the Chair for Research Ethics Board consideration in advance of its implementation.

ONGOING REVIEW REQUIREMENTS

In order to receive annual renewal, a status report must be submitted to the REB Chair for Board consideration within one month of the current expiry date each year the study remains open, and upon study completion. Please refer to the following website for further instructions: http://www.usask.ca/research/ethics_review/



Vivian Ramsden, Chair
University of Saskatchewan
Behavioural Research Ethics Board

Please send all correspondence to:

Research Ethics Office
University of Saskatchewan
Box 5000 RPO University, 1607 – 110 Gymnasium Place
Saskatoon, SK S7N 4J8
Phone: (306) 966-2975 Fax: (306) 966-2069



Certificate of Re-Approval

PRINCIPAL INVESTIGATOR
Gordon McCalla

DEPARTMENT
Computer Science

BEH#
13-75

INSTITUTION(S) WHERE RESEARCH WILL BE CARRIED OUT
University of Saskatchewan
Saskatoon
SK

STUDENT RESEARCHER(S)
David Edgar Kiprof Lelei

FUNDER(S)
NATURAL SCIENCES & ENGINEERING RESEARCH COUNCIL OF CANADA (NSERC)

TITLE
Examining Factors that Lead to Successful Help Seeking in Online Forums

RE-APPROVED ON
09-Mar-2017

EXPIRY DATE
08-Mar-2018

Delegated Review: Full Board Meeting:

CERTIFICATION

The University of Saskatchewan Behavioural Research Ethics Board (Beh-REB) is constituted and operates in accordance with the current version of the Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans (TCPS 2 2014). The University of Saskatchewan Behavioural Research Ethics Board has reviewed the above-named research project. The proposal was found to be acceptable on ethical grounds. The principal investigator has the responsibility for any other administrative or regulatory approvals that may pertain to this research project, and for ensuring that the authorized research is carried out according to the conditions outlined in the original protocol submitted for ethics review. This Certificate of Approval is valid for the above time period provided there is no change in experimental protocol or consent process or documents.

Any significant changes to your proposed method, or your consent and recruitment procedures should be reported to the Chair for Research Ethics Board consideration in advance of its implementation.

ONGOING REVIEW REQUIREMENTS

In order to receive annual renewal, a status report must be submitted to the REB Chair for Board consideration within one month prior to the current expiry date each year the study remains open, and upon study completion. Please refer to the following website for further instructions: http://www.usask.ca/research/ethics_review/



Wivian Ramsden, Chair
University of Saskatchewan
Behavioural Research Ethics Board

Please send all correspondence to:

Research Services and Ethics Office
University of Saskatchewan
Room 223 – Thorvaldson Building
110 Science Place
Saskatoon, SK Canada S7N 5C9



Certificate of Re-Approval

PRINCIPAL INVESTIGATOR: Gordon McCalla; DEPARTMENT: Computer Science; Beh #: 13-75; INSTITUTION(S) WHERE RESEARCH WILL BE CARRIED OUT: University of Saskatchewan; STUDENT RESEARCHER(S): David Edgar Kiprof Lelei; FUNDER(S): NATURAL SCIENCES & ENGINEERING RESEARCH COUNCIL OF CANADA (NSERC); TITLE: Examining Factors that Lead to Successful Help Seeking in Online Forums; RE-APPROVED ON: 15-Mar-2018; EXPIRY DATE: 14-Mar-2019

Full Board Meeting [] Delegated Review [x]

CERTIFICATION

The University of Saskatchewan Behavioural Research Ethics Board has reviewed the above-named research project. The proposal was found to be acceptable on ethical grounds. The principal investigator has the responsibility for any other administrative or regulatory approvals that may pertain to this research project, and for ensuring that the authorized research is carried out according to the conditions outlined in the original protocol submitted for ethics review.

Any significant changes to your proposed method, or your consent and recruitment procedures should be reported to the Chair for Research Ethics Board consideration in advance of its implementation.

ONGOING REVIEW REQUIREMENTS

In order to receive annual renewal, a status report must be submitted to the REB Chair for Board consideration within one month of the current expiry date each year the study remains open, and upon study completion. Please refer to the following website for further instructions: http://www.usask.ca/research/ethics_review/

[Redacted Signature] Scott Bell, Chair University of Saskatchewan Behavioural Research Ethics Board

Please send all correspondence to: Research Ethics Office University of Saskatchewan Box 5000 RPO University, 1607 - 110 Gymnasium Place Saskatoon, SK S7N 4J8 Phone: (306) 966-2975 Fax: (306) 966-2069