Exploring the Issues in Simulating a Semi-Structured Learning Environment: the SimGrad Doctoral Program Design

David Edgar K. Lelei and Gordon McCalla

ARIES Laboratory, Department of Computer Science, University of Saskatchewan
davidedgar.lelei@usask.ca and mccalla@cs.usask.ca

Abstract. The help seeking and social integration needs of learners in a semi-structured learning environment require specific support. The design and use of educational technology has the potential to meet these needs. One difficulty in the development of such support systems is in their validation because of the length of time required for adequate testing. This paper explores the use of a simulated learning environment and simulated learners as a way of studying design validation issues of such support systems. The semi-structured learning environment we are investigating is a graduate school, with a focus on the doctoral program. We present a description of the steps we have taken in developing a simulation of a doctoral program. In the process, we illustrate some of the challenges in the design and development of simulated learning environments. Lastly, the expected contributions and our research plans going forward are described.

Keywords: Simulated learners, Simulated learning environment, Agent-based simulation, Help seeking, Doctoral learners, Multi-agent system.

1 Introduction

Artificial Intelligence in Education (AIED) is one of the research fields whose focus is the use of technology to support learners of all ages and across all domains[1]. Although, one shortcoming of AIED research is the limited research attention that very dynamic and semi-structured domains, such as a graduate school, have received. There is little research that investigates how technology can be used to help connect learners (help seeker and potential help givers) in the graduate school domain. Consequently, there is a gap in our understanding of how such technology may mitigate graduate learners' attrition rates and time-to-degree. We have suggested the use of reciprocal recommender technology to assist in the identification of a suitable helper [1]. However, the nature of graduate school means that validation of any education system designed to be used in a semi-structured environment would take a long time (measured in years). This paper aims to address this challenge by exploring the use of

[1] http://iaied.org/about/

simulated learning environment and simulated learners as a potential way of validating educational technologies designed to support doctoral learners.

In this paper, we first describe the nature and the metrics used by interested stakeholders to measure the success or lack therefore of a doctoral program. Following this, we briefly discuss the uses of simulation as it relates to learning environment. We then introduce the research questions we are interested in answering using simulation. We go on to describe the architectural design of our simulation model. Further, we show how data about the 'real world' target domain is used to inform the parameters and initial conditions for the simulation model. This provides the model with a degree of fidelity. Throughout this model development process, we illustrate some of the challenges in the design and development of simulated learning environments. We conclude the paper with a discussion of the expected contributions and our research plans going forward.

2 Understanding Doctoral Program

Graduate school is a very dynamic and complex social learning environment. A doctoral program in particular is a dynamic, semi-structured, and complex learning environment. Most doctoral programs have some structure in the sense that there are three distinct stages that doctoral learners must go through: admission stage, coursework stage, and dissertation stage. While coursework stage is fairly structured, the dissertation stage is not. Further, the dissertation stage have various milestones that include: comprehensive exam, thesis proposal, research, writing, and dissertation defense. As time passes, learners move from one stage to the next and their academic and social goals change. There is need for self-directed learning and individual doctoral learners are responsible for their own learning pace and choice of what to learn especially in the dissertation stage.

The dynamic nature of the program ensures that there is constant change; there are new learners joining the program, other learners leaving the program either through graduation or deciding to drop out, and still other learners proceeding from one stage to the next. There are two key aspects that influences learners to decide whether to persist or drop out of a learning institution: academic and social integration [2], [3] which are impacted by learner's initial characteristics and experiences during their duration in the program. The various stages of the doctoral program (e.g., coursework) and learning resources can be seen as factors that directly influence the academic integration of a doctoral learner. Peers and instructors/supervisors can be viewed as supporting the social aspects of the doctoral program and hence, directly impact the social integration of doctoral learners. As time passes, doctoral learners continually interact with both the academic and social facets of the doctoral program. As a result, there is constant change in learners' commitment to their academic goal and the social sides of the learning institution

Time-to-degree, completion rates, and attrition rates are important factors influencing the perception and experience of graduate education by interested stakeholders [4], [5]. Research on doctoral attrition and time-to-completion indicates that on average, the attrition rate is between 30% and 60% [5]–[8]. Long times to completion and a high attrition rate are costly in terms of money to the funding institution and the learning institution; and in terms of time and effort to the graduate student(s) and supervisor(s) [8]. Lack of both academic and social integration (isolation) have been shown to affect graduate learners decision to persist [2], [3], [9]. Learners facing academic and social integration challenges should be enabled to engage in a community of peers to foster interaction and hence, encourage peer help and personalized collaboration [10]. Understanding the nature of learner-institution interactions that foster doctoral learners' persistence to degree is important to both the learning institution and its learners. We use simulation to achieve this feat.

Simulation is an established third way of exploring research questions in addition to qualitative and quantitative methods [11], [12]. VanLehn [13] has identified three main uses of simulation in learning environments: 1) to provide an environment for human teachers to practise their teaching approaches; 2) to provide an environment for testing different pedagogical instructional design efforts; 3) to provide simulated learners who can act as companions for human learners. Our research is mainly focused on the first and the second uses – to enable deep insight into the complex interaction of the factors affecting doctoral learners' attrition and time-to-degree leading to a better design of an educational system. Therefore, our research questions are formulated around investigations of how various factors influence time-to-degree, completion rates, and dropout rates of doctoral students. We are interested in answering the following research questions:

1. How does the number of classes (as a platform for social integration with peers – potential helpers) offered by a program(s) or taken by a learner, influence learners' time-to-degree and their propensity to persist or drop out?
2. How does the average class size (as basis of learners' social integration) attended by learners, impact learners' time-to-degree and their inclination to persist or drop out? What is the optimum class size?
3. How does the overall population size of the learners (a few learners vs many learners) influence learners' time-to-degree and their likelihood to persist or drop out?
4. Does timely help affects doctoral learners' time-to-degree and their decision to persist or drop out? If so, how?
5. How does the level of reciprocation influence the formation of a 'helpful community' of learners and adaptive help seeking behavior of the learners?

Use of simulation enables us to explore the aforementioned issues in a fine-grained controlled environment. For example, it would be almost impossible in the 'real world' setting to examine the impact of different number of course to take or class size to attend. Two cohorts of learners will have different attributes. Simulation allows us to tweak the number of courses or class size without touching the other characteristics of learners. Hence, we are able to see the real impact of one variable at a time. Before any exploration and insight can be gained on these issues, there is need to design and implement the simulation model.

3 Building an Initial Prototype of *SimGrad*

In this section we demonstrate the steps we have taken in the development of our initial prototype of our simulated doctoral learning environment: *SimGrad*. We show how a designer of an educational technology can develop a model of their target learning environment and inform its initial condition with available ‘real world’ data.

3.1 *SimGrad* Design

We need to design a simulation model by addressing two key challenges. First, we need to consider issues related to the modeling of the learning environment: how do we design conceptual and computational models of a doctoral program and what stakeholders should be included in these models? The second concern is about modeling of simulated learners: what doctoral learners’ features affect persistence and time-to-degree, what factors do we model, and can we inform these features with available ‘real world’ data?

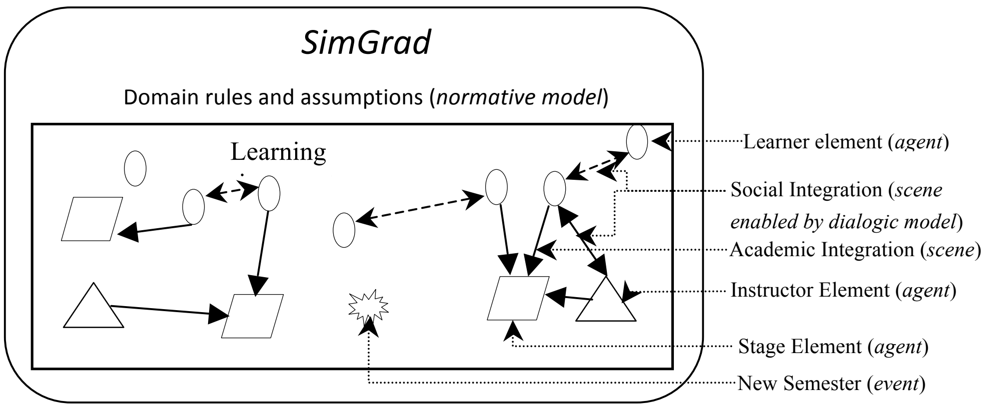

Fig. 1. SimGrad conceptual framework, its three elements, and the possible interaction between the elements

We have designed our conceptual model of the different aspects of simulated doctoral learners and doctoral learning environment based on the simulated learning environment specifications suggested by Koper et al. in [14], and features for building an electronic institution proposed by Esteva et al. [15]. We name our conceptual framework, *SimGrad*. Its core elements include: *normative model* - specifies requirements and constraints to guide agent actions and behavior; *dialogic model* – deals with interaction strategies and communication mechanism; *events* – refers to happenings in the model that trigger (re)action by agents; *scene* – description of an interaction between elements; *elements (agents)* – represent key stakeholders of the target domain that are modeled. Elements are modeled as agents. Each of the agents has attributes and behavior which are informed by our assumptions guided by our research questions and factors that influence learners’ decision to persist. Every element of interest is to be modeled within the learning environment and all possible interactions and operations within the learning setting is guided by domain rules represented by the normative model. See **Fig. 1**.

In our simulation model, we have chosen to model three types of elements: class, instructor, and learner. In this paper, in keeping with model simplicity, both the class and the instructor agents are passive while the learner agent is modeled to be active and reactive to its environment. Also, the only instructor's attributes we are interested in are related to classes (see **Table 1**). We modeled only one type of instructor agent. Another instructor type agent that can be modeled is the supervisor.

Each learner agent has the following properties: autonomy, social ability, reactivity, proactivity, and a degree of intentionality. We have also identified the following key attributes for our agent learner model: state – (busy, available), program, stages, course taken, peer interactions (pertaining challenges), academic integration, social integration, and motivation (see **Table 2**). In our model, peer interaction and state contribute to a learners' social integration, while research area, stage, course taken impact to their academic integration. Motivation combines both the social and academic integration and hence, is the main factor that determines whether an agent continues to persist or chooses to drop out of the program.

Table 1. Comparison of computed attributes of the three agent types

Attribute – data (value range)	*Agent learner*	*Agent instructor*	*Agent class*
Total number of classes take, taught, or frequency of offering within 10 years – *numeric (0-20)*	X	X	X
Grade obtained, average awarded, or average obtained by learners – *numeric (0,12)*	X	X	X
Take classes from, teach classes in, or class offered in various programs – *textual (program id)*	X	X	X
Instructors teaching a class – *array list (instructor id)*	X	-	X
What is the class size – *numeric (1-5)*	X		X
Number of classes taken or taught per year - *numeric (0,4)*	X	X	-
Which classes are taken or taught – *textual (class id)*	X	X	-

The main intentions of each agent is to persist through doctoral requirements to graduation and to do so in a timely manner. However, each of these agents reacts to the different challenges at various stages of graduate school in divergent and autonomous ways. At the coursework stage, agents have the goal of taking courses that are relevant to their field and that they will perform well. When facing a course choice challenge or any other particular challenge, we have modeled our agents to proactively associate with peers to seek help. Each peer makes individual choice on whether to or not to respond to a request for help from others. The dialogic model handles the agent to agent interaction and communication through a message passing mechanism [16].

Table 2. Attributes and parameters considered for an agent learner model for learners, their description and how each of them changes.

Attribute	*Value - description*	*How it changes*
Enrolment	Date (MM/YYYY) Indicate the month a year an agent enrolled in the program	Does not change
Graduation	Date (MM/YYYY) Target graduation date	Evaluated whenever an agent completes a milestone
State	Textual (busy, available) Indicates an agent availability to help others, assigned based on the smallest time unit model	Changes whenever an agent experiences a challenge
Program	Textual (program id) Identify an agent's closer community within the larger community of learners	Does not change during a simulation run
Stage	Textual (admission, coursework, dissertation, timeout, dropout)	Admission stage is like an event. Learner move to the coursework immediately after admission. They more to dissertation after completing their course load.
Courses taken	Array [course, mark, instructor id](0-6) Record courses taken by an agent and the marks obtain in each course	Every end of semester that the student took classes, this array is updated
Peer interaction	Array [learner id, challenge, result], Keep track of an agent interactions with others and the outcome of the interaction	Changes whenever two agents interact
Academic integration	Numeric (-1,1) Measures the academic satisfaction	Changes whenever an agent learner interacts with agent stage (i.e., completes a milestone or experience a challenge)
Social integration	Numeric (-1,1) Measures a learners sense of belonging to the learning environment	Changes whenever an agent learner interacts with its peers or agent instructors
Motivation	Numeric (-1,1) Measures the propensity of an agent to still want to persist. A motivation value above 0.3 indicates persistence. A value between -0.3 and 0.3 indicate help seeking needed. A value below -0.3 means the agent drops out	Whenever there is a change in the social and academic integration values. Its value is the average of the integration values.

3.2 Informing *SimGrad* behavior and evaluation functions

Having identified the important agents and their key attributes, there are two sets of important functions for each element that need to be modelled: behaviour functions and evaluation functions [17]. Behaviour functions inform the decision making of the active elements and dictates the interaction patterns between them and the other modeled elements (e.g., how many classes a given agent takes). Evaluation functions indicate whether or not various interactions between the different agents in a simulation were successful (e.g., determine what grade a given agent attains in a class it took). Informing such functions with 'real world' data allows the simulation to behave in a way consistent with reality.

Simulation model fidelity is an issues that might arise when using simulation to study a target real world phenomenon. However, the most important issue to consider is the research question to be answered. While Champaign [18] used a very low fidelity model, Matsuda et al. [19] used a model with high cognitive fidelity to reach compelling conclusion. Further yet, Erickson et al. [17] also demonstrated that is possible to use a medium fidelity model and uncover interesting results. In some situations it might not be possible to have a high fidelity model because of lack of data. A case in point is our simulation scenario. Where possible, we inform our simulation functions with data received from the U of S on their doctoral program. An investigation into the U of S data showed that we will not be able to inform every aspect of our simulation model. It would be desirable to inform every initial aspects of our simulation model with 'real world' data but, we do not have data on the dissertation stage.

We are provided information on student id, years a student is registered, year of graduation (if graduated), student's program, classes taken and marks obtained, class instructor, and students instructional responsibilities. From this dataset we are able to inform the admission and coursework stages of our model (academic integration). However, there is no information concerning the dissertation stage and the social integration aspects. While it possible to inform various behaviour and evaluation functions for our simulation model, in this paper we focus on describing the steps we took to inform two functions of our simulation: learning environment admission behaviour function, and learners' class interactions behaviour function.

As already mentioned, admission is an important part of a doctoral program that contributes to it dynamic nature. The admission process is complex and involves a lot of stakeholders and processes, but we are concerned only with determining the year to year patterns in how many students are admitted. To provide some fidelity to our simulated learning environment admission, we analyzed data provided to us by the U of S University Data Warehouse[2]. The provided dataset contained information on doctoral learners registered in the 10 years 2005-2014. In this time there were 2291 doctoral learners with a total of 52850 data points on class registration. The 2005 registration included learners who had joined the program earlier than 2005. In order to get a clean admission pattern, we only considered learners who were registered from the year 2006 onwards. This reduced the population size to 1962.

[2] http://www.usask.ca/ict/services/ent-business-intelligence/university-data-warehouse.php

We were able to identify three admission periods, September, January, and May. We then obtained values for each of the admissions months for the years 2006-2014. This provided a distribution for each month that we used to generate a scatter plot of admission numbers. A sigmoidal pattern emerged. Next, we performed a non-linear curve fitting to the scatter plot so that the admission function can be represented in the form $Y = St*(c + x)$, where c is a constant, *St* is a variable dependent on the admission period, and *x* is the admission period. We then ran a regression to find values of each of these variables. This allowed us to model the admission patterns observed in the U of S dataset.

Next we derived the number of classes taken. To introduce some realism to the number classes taken behaviour, we had to further prune the data. We only considered data for students whose cohorts would have been registered for at least 3 years by the end of the year 2014 and hence, we considered class taking behaviour of 1466 U of S doctoral learners.

We obtained the number of classes each of the remaining learners we registered in and created a histogram. This histogram showed us the distribution of the number of students registered for a certain number of classes. Next, we transformed this distribution graph into a cumulative distribution function. We then took an inverse of the cumulative distribution function to achieve a quantile function. The quantile function, when run over many learners, assigns learners a class count that mimics the initial histogram. We use this quantile function to inform the number of classes a learner can take.

In this section we have described the importance of informing a simulation model with 'real world' data. We have described two functions that are informed with U of S dataset. Other examples of functions that can be informed using the U of S dataset include: class performance evaluation function, dropout behaviour function, time to degree behaviour function, and flow through behavior function (main as pertains to coursework stage). We have identified that missing data values is a major hindrance in this endeavor. There are possible ways of informing simulation attributes where there are no 'real world' data to derive from. A designer can either assign common sense values, generate and assign random values, or refer to the research literature to identify patterns that have been found by other researchers. Since we have the enrolment dates and the graduate dates for learners who graduate, we choose to derive common sense values with these two dates guiding the process and the value range.

4 Discussion, Expected Contributions, and Future Research Plans

Despite the growth in the use of simulation as a method for exploration and learning in many areas such as: engineering, nursing, medicine [20], and building design [21], research in the used of simulation within AIED is still at an early stage. There is need for more research to demonstrate that the outputs of simulation runs are desirable and informative to the AIED community. In this paper, we aim at contributing to this notion and by promoting the use of simulation in educational research and presenting an agent based simulation conceptual framework for building simulated learning environment, with a focus on the semi-structured ones. Simulated learning environment and simulated learners are important in exploring and understanding a given learning domain. Further, it helps with the generation of system validation data.

The expected contributions to AIED include: providing a conceptual framework for simulated graduate school learning environment – an architecture that enables investigations into factors affecting doctoral learners progress through their program; shedding light on learner modeling issues in dynamic learning environments; and demonstrating the importance of simulation in exploring various AIED research domains, particularly semi-structured domains.

Current research work is focused on the implementation of the simulation model and the refinement of the various behaviour and evaluation functions. Once the implementation is done, we will validate our model against the dataset we have from the U of S before proceeding to explore the impact of various environmental factors. Since we are informing the simulation with both common sense assumptions and U of S dataset, the goal is to tweak the common sense assumptions such that when the model is run we get similar results as the U of S data in terms of class performance, dropout rate, and time-to-degree. Achieving this, would give us confidence that we have captured reality in some measurable way. We can then start exploring the various impact of measures we are interested in examining. As earlier indicated, we are interested in exploring the interactions of a number of variables: number of classes taken which will impact the availability of potential peer helpers, the effect of reciprocity on help seeking and help giving, and the effect of help seeking and other factors on doctoral learners' time-to-degree and attrition rates.

Acknowledgement

We would like to thank University of Saskatchewan University Data Warehouse team for giving us access to 10 year dataset. Specifically, we would like to thank Mathew Zip for processing and explaining to the first author the nature of the dataset. We also wish to acknowledge and thank the Natural Science and Engineering Research Council of Canada (NSERC) for funding our research.

References

[1] D. E. K. Lelei, "Supporting Lifelong Learning: Recommending Personalized Sources of Assistance to Graduate Students," in *Artificial Intelligence in Education*, 2013, pp. 912–915.

[2] V. Tinto, "Taking student success seriously: Rethinking the first year of college.," *Ninth Annu. Intersession Acad. Aff. Forum,* vol. 19, no. 2, pp. 1–8, 2005.

[3] V. Tinto, "Dropout from higher education: A theoretical synthesis of recent research," *Rev. Educ. Res.*, vol. 45, no. 1, pp. 89–125, 1975.

[4] H. Groenvynck, K. Vandevelde, and R. Van Rossem, "The Ph.D. Track: Who Succeeds, Who Drops Out?," *Res. Eval.*, vol. 22, no. 4, pp. 199–209, 2013.

[5] F. J. Elgar, “Phd Degree Completion in Canadian Universities,” Nova Scotia, Canada, 2003.

[6] M. Walpole, N. W. Burton, K. Kanyi, and A. Jackenthal, “Selecting Successful Graduate Students: In-Depth Interviews With GRE Users,” Princeton, NJ, 2002.

[7] L. Declou, “Linking Levels to Understand Graduate Student Attrition in Canada,” McMaster University, Hamilton, Ontario, Canada, 2014.

[8] B. E. Lovitts, *Leaving the Ivory Tower: The Causes and Consequences of Departure from Doctoral Study*, Illustrate., vol. 32. Rowman & Littlefield Publishers, 2001, p. 307.

[9] A. Ali and F. Kohun, “Dealing with isolation feelings in IS doctoral programs,” *Int. J. Dr. Stud.*, vol. 1, no. 1, pp. 21–33, 2006.

[10] Computing Research Association, “Grand research challenges in information systems,” Washington, D.C, 2003.

[11] R. Axelrod, “Advancing the Art of Simulation in the Social Sciences,” in *Proceedings of the 18th European Meeting on Cybernetics and Systems Research*, 2006, pp. 1–13.

[12] N. Gilbert and K. G. Troitzsch, “Simulation and Social Science,” in *Simulation for the Social Scientist*, McGraw-Hill International, 2005, pp. 1–14.

[13] K. VanLehn, S. Ohlsson, and R. Nason, “Applications of Simulated Students: An Exploration,” *J. Artif. Intell. Educ.*, vol. 5, no. 2, pp. 1–42, 1994.

[14] R. Koper and B. Olivier, “Representing the Learning Design of Units of Learning,” *Educ. Technol. Soc.*, vol. 7, no. 3, pp. 97–111, 2003.

[15] M. Esteva, J.-A. Rodriquez-Aguilar, C. Sierra, P. Garcia, and J. L. Arcos, “On the Formal Specification of Electronic Institutions,” in *Agent Mediated Eectronic Commerce*, 2001, pp. 126–147.

[16] H. J. C. Berendsen, D. Van Der Spoel, and R. Van Drunen, “GROMACS: A message-passing parallel molecular dynamics implementation,” *Comput. Phys. Commun.*, vol. 91, no. 1, pp. 43–56, 1995.

[17] G. Erickson, S. Frost, S. Bateman, and G. Mccalla, “Using the Ecological Approach to Create Simulations of Learning Environments,” in *In Artificial Intelligence in Education*, 2013, pp. 411–420.

[18] J. Champaign, “Peer-Based Intelligent Tutoring Systems: A Corpus-Oriented Approach,” University of Waterloo, 2012.

[19] 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 *2012 IEEE Fourth International Conference On Digital Game And Intelligent Toy Enhanced Learning*, 2012, pp. 25–32.

[20] A. L. Baylor and Y. Kim, “Simulating Instructional Roles Through Pedagogical Agents,” *Int. J. Artif. Intell. Educ.*, vol. 15, no. 2, pp. 95–115, 2005.

[21] G. Augenbroe, “Trends in Building Simulation,” *Build. Environ.*, vol. 37, no. 8, pp. 891–902, 2002.