University Student Desertion Analysis using Agent-Based Modeling Approach

M. C. Castellanos Rojas, L. D. Alvarado Nieto and J. E. Villamil Puentes
Departament of Engineering, Universidad Distrital, Bogotá D.C., Colombia

Keywords: Agent-Based Model, University Student Dropout, Student Retention, National Educational Ministry (MEN).

Abstract: Student dropout at universities is a worldwide phenomenon that exceeds a 40% rate of the students admitted to first semester. In Colombia, it exceeds 45%, an alerting rate that makes studies on the subject very important for Governments and universities considering that it has a huge social impact and affects the resources of the education area. Traditionally, studies are performed by statistical and mathematical methods and the Ministry of National Education acknowledges that they have been insufficient since they fail to explain the dropout behavior. Agent-based modeling and simulation (ABMS) has been considered a new way of doing science when managing social problems that are complex systems forcing an evolution simulation as a useful approach to develop this phenomenon.

1 INTRODUCTION

Students´ dropout rate at Colombian universities is of 45.8%, a major phenomenon in every sense for the country´s education (MEN and *Permanencia* Group, 2015). The society, the Government and the universities themselves are concerned about the study of dropout structure and behavior. Not only Colombia, but also countries like Mexico show desertion rates per cohort of 40%, Argentina 43%, Venezuela 52%, and Chile presents 54%. There are several researches in this field, but the Colombian Ministry of National Education -MEN- suggests that studies are insufficient considering that adjustments resulting from these studies do not decisively impact the problem (MEN & *Qualificar* Group, 2015).

Understanding this phenomenon is difficult due to multifaceted factors which do not allow establishing a single structure and behavior.

In the course of his career, a student goes through states such as: a normal performance, dropout, low performance, subject completion and graduation; Today, each state depends on one or more of the following factors: academic, attitudinal, socioeconomic, personal and institutional.

The majority of studies about these problems are based on what has been called “traditional methods” (statistical and mathematical), but they fail when explaining all kind of behaviors. This paper adopts an agent-based modeling, based partly on historical data and meaningful narrative, making a difference in comparison with those methods used in previous researches.

Based on this methodology, an initial space is composed by these agents: student, subject, teacher and institution, and their corresponding attributes. Its behavior is directed by very simple rules that they modify themselves autonomously throughout time evolution as simulation progresses.

Therefore, other ways and methods of studying this complex phenomenon are meant to be found through this research on agent-based simulation.

2 MAIN STUDIES

When discussing desertion, a distinction between some studies carried out worldwide and those that have been advanced in Colombia is made, mainly addressed from qualitative and quantitative analyses.

Worldwide studies have focused their efforts on defining approaches that can be classified into five major categories such as: psychological, sociological, economical, organizational and interactionist as exposed by Cabrera et. al (1993). In turn, these studies take into account explanatory variables, which are drawn together in personal, family or institutional.

Early researches made on the topic of dropout were focused on sociological and psychological

Castellanos Rojas, M., D. Alvarado Nieto, L. and Puentes, J.
University Student Desertion Analysis using Agent-Based Modeling Approach.
DOI: 10.5220/0006777601280135
In *Proceedings of the 3rd International Conference on Complexity, Future Information Systems and Risk (COMPLEXIS 2018)*, pages 128-135
ISBN: 978-989-758-297-4

approaches based on Durkheim´s Suicide Theory in which he raises the university as a society, where the student becomes part of it by interacting with others, demonstrating their disposition, interest, attitude, skills and expectations; Although when failure appears, the student assumes a suicide-like behavior (Spady, 1970).

2.1 Main Studies in Colombia

Since 2003, the Ministry of National Education within the policy of coverage expansion, has been developing the project " Higher Education Dropout Decrease" with the support of the Studies on Economic Development Center (known as CEDE by its initials in Spanish) of the Andes University (CEDE, 2007) through the "Research on desertion in higher education institutions in Colombia". They have built different strategies in order to follow up the desertion through the System for Prevention and Analysis of Desertion in Higher Education Institutions (or SPADIES by its initials in Spanish), which now has historical data of Colombian universities in categories such as socioeconomics, individuals, institutional and academicals variables (MEN et al., 2009).

The District University made a specific study about low academic performance, were the following factors were considered: academic, socioeconomic, aptitude and vocational, personal and familiar, institutional, habits and study methods. In addition, two models were raised, a linear one that measures the academic performance and another of logistic regression that calculates the probability of the risk of incurring in a low yield. (Quintero, Vásquez, Torres, Estrada, & Castellanos, 2015).

Other researches worth of mention are "Discovery of dropout profiles with data mining techniques" by the University of Nariño (Timarán P, Calderón R, & Jiménez T, 2013); “Determinants of student dropouts at the University of Antioquia” (Vasquez V et.al. 2003); “Student Dropout at the National Pedagogical University” (COAE, 2006); and “A matter of survival in the National University” (Pinto, et.al. 2007). From these studies different definitions were adopted such as Dropout understood as the situation in which a student incurs when he aspires to obtain a university degree and does not succeed; and *Deserter* as the person who has been admitted to a higher education institution, but does not register in the university for 2 consecutive semesters.

Considering the above, there are two types of dropout, depending on the time and space according to Vásquez et.al. (2003) and Castaño et.al. (2004). Regarding time, as shown in Figure 1, dropout is classified as: *Premature dropout,* when a person who, having been accepted to enter the first semester in the university, does not sign up; *Early dropout*, the person who leaves its studies in the first four semesters of career; *Late dropout,* refers to the person who leaves studies in the last six semesters, that is, from the fifth semester onwards (Vásquez V. et al., 2003 and Castaño et al., 2004b); and *Ungraded late desertion* when a person abandons its studies once he´s finished subjects, but does not get a graduation.

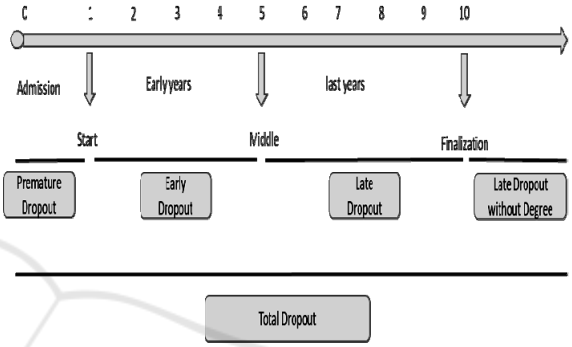

Figure 1: Classification of desertion according to time. Adapted from: (Vásquez V et al., 2003) and (Castaño et al., 2004).

Regarding environment, as shown in Figure 2, dropout is divided as follows: *Intern* or *of an academic program*: refers to the student who decides to change his academic program by another that offers the same university institution; *Institutional* as the case in which the student leaves the university; and educational *system dropout*, when the student leaves the study. (Vásquez V et al., 2003) and (Castaño et al., 2004)

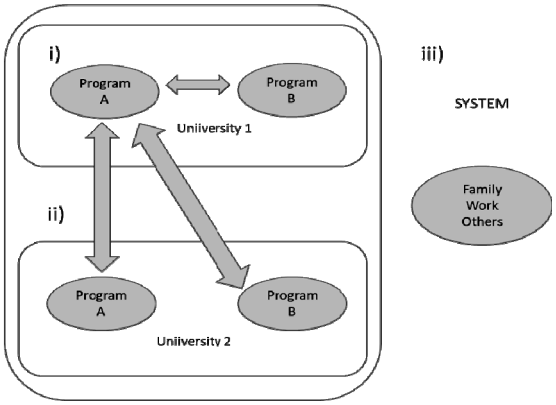

Figure 2: Classification according to the environment. Adapted from: (Vásquez V et al., 2003) and (Castaño et al., 2004a).

Relevant factors for dropout were taken into account set out by the Ministry of Nacional Education, but only variables shown in Table 1 were considered.

Table 1:Variables for models.

Individuals	Type of Registration
	Gender
	Age
	Compensation
	Resignation
Academics	Department
	Project
	State
	Icfes score
	Average
	Kind of test
	Number of academic tests
	Approved subjects
	Failed subjects
	Academic performance
Institutional	Cohort
	Entry year
	Entry period
	Alimentary support indicator
	Type of registration
Socioeconomic	Social stratum
	Birthplace

3 COMPLEX SYSTEMS

There are many definitions of complex systems, but in general terms, systems are formed by numerous parts or components, which interact typically in a nonlinear way; these systems can arise and evolve through self-organization, so that they are neither completely regular nor completely random, allowing the development of collective behaviors (emerging) in macroscopic scales (Sayama, 2015).

These systems are characterized for: emergency, self-organization, nonlinear behaviors, sensitivity to initial conditions, they are unpredictable, do not have a centralized control, they can present sudden changes, managing of uncertainty, evolution, adjustment, among others.

3.1 Agent-based Modeling

From the point of view of complexity sciences, the study of the complex systems, given its characteristics, requires certain particular forms and methodological strategies to analyze the resulting behaviors of the interaction between different agents, which not only function in unstable environments but also can be or not homogeneous and in addition can be adaptive and even auto-organize and evolve; then, a computer simulation allows the comprehension of their structure and the study of their behavior, but also lead to possible emergent behaviors.

Simulation is a particular type of modeling (representation) and introduces the possibility of a new way of thinking about the processes social and economic, based on ideas of the appearance of the complex conduct of relatively simple activities (Simon 1996), mentioned in (Gilbert and Troitzsch, 2005).

Axelrod considers a simulation of a third route of knowledge, given that the resultant model improves the comprehension, representation or explanation of complex processes; " The simulation is the third way of doing science, in contrast with the induction and the deduction " (Axelrod, 2005).

His logic lies on the idea of a possible reproduction of structures, behaviors or global functions of a system, from the characterization of its components (agents), the environment and the local interactions agent-agent and agent-environment (Gómez, 2016).

Some patternmakers consider any type of independent component, already belong to software or a model to be an agent (Bonabeau 2001). On the other hand, other authors insist that the behavior of a component also must be adaptive in order to be considered as an agent. Casti (1997), mentioned in (Macal and North, 2014).

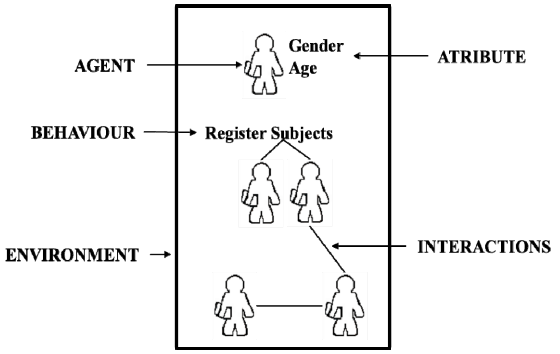

Figure 3: Elements in Agent-based Modeling. Source: authors.

4 METHODOLOGY

This method is based on approaches made by (Gilbert and Troitzsch, 2005), (Cioffi-Revilla, 2014), (Wilensky and Rand, 2015) that agree on defining a reference system posing a question of what you want to resolve, considering three stages: model design, model making and model analysis.

Model Design: precise scope, what is included and what is possible for the model, agents and types, attributes and behaviors, the environment, possible events in each interval of time, input and output, measurements and validation based on significant abstractions. This stage can be presented through flux diagrams, UML (unified modeling language) or Natural Language used in this project.

Model Making: either using a specific software or an existing package for simulation (Repast, Netlogo, Java, Python); incremental development with progressive iterations, from a basic to a final model (Wilensky and Rand, 2015).

Model Analysis: verification comes first by establishing that the simulation program works as expected; then comes validation by analyzing the results to asses if simulation is a good model in accordance to the objective.

For practical reasons, this methodology was made to be developed in two cycles, the first one covers an agent-based simulation based on the logistic regression model and the last one uses an agent-based simulation based on autonomous agents. This document addresses the first of these, using Netlogo as tool, along with a PostgreSQL database, to store the simulated data for each student per semester and with output to a file in CSV format.

5 MODEL

In this case, the model base is made from 2924 records of students belonging to five careers part of the Engineering Faculty at the District University, with information supplied by the Advisory Office of Systems, on cohorts from 2010-1 to 2015-3. Included variables were: academic, personal, socioeconomic and institutional.

Based on this information, quantity and percentage of students were established by the curricular project, gender, social stratum, type of inscription, birthplace, age of entry and ICFES score as shown in tables 2 to 7.

Table 2: Project were the student belongs.

Curricular project	Frequency	Percentage
Systems Engineering	579	0,20
Cadastral Engineering and Geodesy	652	0,22
Electric Engineering	489	0,17
Electronic Engineering	577	0,20
Industrial Engineering	627	0,21
Total	2.924	1,00

Table 3: Gender of the student. Faculty of Engineering.

Gender	Frequency	Percentage
F	684	0,23
M	2240	0,77
Total	2.924	1,00

Table 4: Social stratum of the student.

Social stratum	Frequency	Percentage
1	1017	0,35
2	1078	0,37
3	787	0,27
4	42	0,01
Total	2924	1,00

Table 5: Age of ingress of the student.

Age	Frequency	Percentage
Between 15 y 17	949	0,32
Between 18 y 20	1725	0,59
Between 21 y 25	214	0,07
More than 26	36	0,01
Total	2.924	1,00

Table 6: Type of inscription of the student.

Inscription	Frequency	Percentage
Normal	2.766	0,94
Minorities	119	0,04
Others	39	0,01
Total	2.924	1,00

Table 7: Birthplace of the student.

Birthplace	Frequency	Percentage
Bogotá	2.065	0,70
Surroundings	164	0,06
Outside Bogotá	434	0,15
No Data	261	0,09
Total	2.924	1,00

5.1 Model Agents

Agents were defined as: student, subject, teacher, institution, with their respective characteristics and rules of behavior. Nevertheless, phase one used a base model using student and subject or matter as

agents, but further models will be worked with all agents.

5.1.1 Student Agent

An agent student aims to travel across the curricular map, approving the study plan to obtain the degree; it has characteristics such as: name, curricular project, social stratum, Icfes mark, type inscription, birthplace, alimentary support indicator, among others; in addition, it is subject to the rules of advance, stop, loss, academic performance, not repetition, that establishes the academic regulation of the University for the student's permanency. Every agent student has an initial load of the total of credits and subjects for the first semester and changes depending on the conditions that can be: of objective aim, of test or low performance of the student.

Projects with the deserter's conditions: early deserter, late deserter, late deserter without degree, proposed in previous studies.

For the calculation of the academic performance in students, Farmer took the following model based on Quintero et al., (2015):

$$RA = 10x + 25(1- I_R) + 5I_p+10I_N + 10\frac{1}{1+n} \quad (1)$$

Where:
x is the student's accumulated average

I_R = index of repetition

$$I_R = \frac{\text{Failed subjects}}{\text{Number of taken subjects}} \quad (2)$$

I_p = index of permanency

$$I_P = \frac{\text{Number of admissions}}{\text{Number of semesters since Entry}} \quad (3)$$

I_N = index of leveling

$$I_N = \frac{\text{Number of approved subjects}}{\text{Total Subjects}} \quad (4)$$

n = number of academic tests

6 RESULTS

Simulation allows tracking a student or groups of students, distributed by gender and stratum parameters or without this classification. Since 99% of the students are stratum one, two and three, 18 simulations of systems engineering were made, six for each stratum obtaining results shown in Table 8.

For verification and analysis purposes an implementation with 200 students was made, without varying initial conditions and similar results were obtained.

As an example, actual conditions of the student A who obtained an ICFES score of 485 points over 500 as a maximum score (97%) were considered. On the first run, he fails in more than two subjects and academic performance of 61.1% is obtained at the end of the semester but improves in the following semesters and ends the career in thirteen admissions, with a final academic performance of 75.93 %.

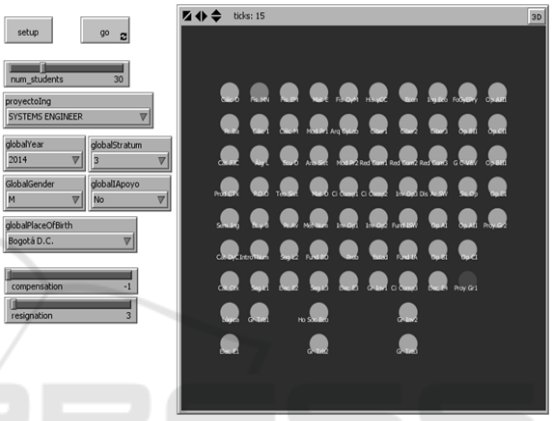

Figure 4: Functionality for male gender. Source: authors. Software Netlogo®.

Conditions: career (systems engineering), year of ingress (2014), stratum (3), birthplace (Bogota), as observed in the figure-4.

The output of simulation indicating the number of graduate students versus the number of dropout students, is also shown in Figure 5, with 21, deserters and 29 graduate students.

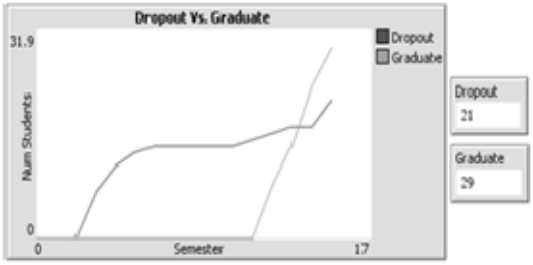

Figure 5: State of male students.

In Figure-6, there are the conditions of academic low performance of the students and the number of times that students incurred in these is observed, that is, all the times that the students failed for average, failed a subject for 3 times and the times that failed more than three subjects in the same period.

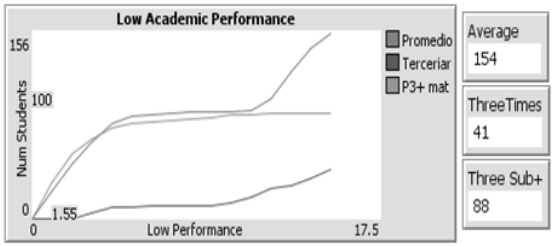

Figure 6: Academic performance in male students.

In the figure 7, there is established the degree of major resignation of a student when it has a low performance to confront its academic load.

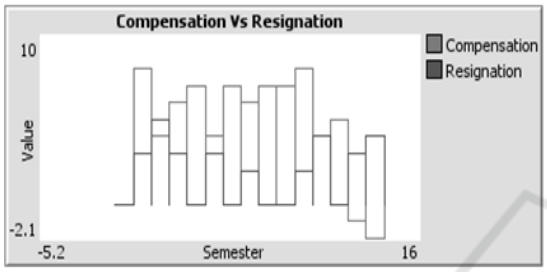

Figure 7: Compensation Vs. Resignation in male students.

Table 8 shows the final status when comparing deserter versus graduated, in order to establish the relationship between them.

Table 8: Simulation Results Deserters Vs. Graduates.

# Sim	Dropout		Graduate		Total
	Number	Percentage	Number	Percentage	
1	26	0,52	24	0,48	50
2	21	0,42	29	0,58	50
3	30	0,60	20	0,40	50
4	29	0,58	21	0,42	50
5	22	0,44	28	0,56	50
6	24	0,48	26	0,52	50
7	26	0,52	24	0,48	50
8	24	0,48	26	0,52	50
9	17	0,34	33	0,66	50
10	21	0,42	29	0,58	50
11	20	0,40	30	0,60	50
12	25	0,50	25	0,50	50
13	23	0,46	27	0,54	50
14	17	0,34	33	0,66	50
15	31	0,62	19	0,38	50
16	23	0,46	27	0,54	50
17	25	0,50	25	0,50	50
18	26	0,52	24	0,48	50
Promedio		0,48		0,52	

The model also allows the generation of a flat file, which contains the record, semester to semester of every agent-student, indicating the behavior that this one has in each of the subjects, while the whole mesh crosses curricular, together with the conditions for which it passes in every period and the indicator, if it is or not in test and for what motives. Figures 8.

7 CONCLUSIONS

According to results analysis, low-performance, desertion and loss of student quality percentages are much in line with the actual data provided by the Advisory Office of Systems of the university.

Eighteen simulations using a number of students ranging from 50 to 200, resulted with an average of desertion of 46% and graduation rate of 46% while students completing subjects without graduation reach 9%.

Underperformance is causing the student to drop-out, with the highest percentage of failure on average, followed by the repetition of a course by more than three times.

The highest levels of desertion are detected at the ends of the career. In other words, whether during the first five semesters or once subjects are completed, but without obtaining a title, late drop without grade, the above-mentioned *Ungraded late desertion.*

Agent-based modeling and simulation allow establishing new situations of desertion and academic performance, as indicated in chapter six, particularly on the causes´ mentions. Simulation ranges fit pretty well to the actual ones.

The particular simulation of students shows that having high initial conditions, not necessarily leads to a high performance during his career and opposite conditions to not lead to desertion.

The fact of simulating student by student and students altogether, facilitates validation and verification of parameters with very similar results, strengthen the arguments of the desertion causes and will surely facilitate the academic measures for improvements.

An incoming second cycle including remaining agents to detect new variables is on its way.

Table 9: Output data file by student.

Average	Rate of Repetit	backwardness In	ReasonLowAcad	Student	Number Registra	CumulativeAver	Leveling Index	AcademicPerf	Compens	Resigna	State	CurricularProj	Gender	Stratum	ICFES	TypeRegist
2,65	0,44	0	120	5001	1	2,65	0,56	61,1	0	0	2	1	1	2	485	1
3,81	0,35	0,5	0	5001	2	3,23	0,65	65,1	8	5	1	1	1	2	485	1
3,5	0,32	0,67	0	5001	3	3,37	0,71	67,8	6	4	1	1	1	2	485	1
3,49	0,3	0,5	0	5001	4	3,43	0,7	68,8	7	3	1	1	1	2	485	1
3,86	0,24	0,4	0	5001	5	3,65	0,78	73,3	7	3	1	1	1	2	485	1
3,53	0,24	0,33	0	5001	6	3,59	0,79	72,8	8	0	1	1	1	2	485	1
3,72	0,25	0,29	0	5001	7	3,66	0,75	72,9	7	3	1	1	1	2	485	1
3,92	0,22	0,38	0	5001	8	3,79	0,78	75,2	5	4	1	1	1	2	485	1
3,47	0,24	0,22	20	5001	9	3,63	0,79	73,2	7	0	2	1	1	2	485	1
3,52	0,23	0,3	0	5001	10	3,58	0,83	71,7	8	4	1	1	1	2	485	1
4,02	0,22	0,18	0	5001	11	3,8	0,91	74,9	6	3	1	1	1	2	485	1
4,09	0,2	0,25	0	5001	12	3,95	0,99	77,7	6	2	1	1	1	2	485	1
3,56	0,2	0,23	0	5001	13	3,76	1	75,9	6	0	5	1	1	2	485	1
4,26	0,22	0	0	10	1	4,26	0,78	84,9	0	0	1	1	1	2	457	1
3,6	0,17	0,5	0	10	2	3,93	0,88	83,9	8	3	1	1	1	2	457	1
3,82	0,12	0,67	0	10	3	3,88	0,96	85,4	7	3	1	1	1	2	457	1
3,84	0,12	0,25	0	10	4	3,86	0,91	84,7	8	0	1	1	1	2	457	1
3,54	0,14	0,2	0	10	5	3,7	0,9	82,5	7	2	1	1	1	2	457	1
3,83	0,14	0,33	0	10	6	3,77	0,89	83,1	6	4	1	1	1	2	457	1
3,06	0,16	0,29	100	10	7	3,42	0,84	78,6	5	3	2	1	1	2	457	1
4,18	0,17	0,25	0	10	8	3,8	0,86	77,4	6	3	1	1	1	2	457	1
4,16	0,15	0,22	0	10	9	3,98	0,89	80	8	3	1	1	1	2	457	1
3,66	0,14	0,2	0	10	10	3,82	0,91	78,8	8	0	1	1	1	2	457	1
4,48	0,13	0,18	0	10	11	4,15	0,99	83,2	6	0	1	1	1	2	457	1
1,84	0,14	0,17	100	10	12	3	0,99	71,4	6	0	2	1	1	2	457	1
2,8	0,15	0,23	100	10	13	2,9	0,99	68,5	0	2	2	1	1	2	457	1
2,27	0,16	0,29	103	10	14	2,59	0,99	64,3	-1	3	2	1	1	2	457	1
4,47	0,16	0,33	103	10	15	3,53	1	73,3	-2	4	5	1	1	2	457	1

REFERENCES

Axelrod, R. (2005). Advancing the Art of Simulation in the Social Sciences. Gerald R. Ford School of Public Policy, 13.

Cabrera, A. F., Nora, A., & Castaneda, M. B. (1993). College Persistence: Structural Equations Modeling Test of an Integrated Model of Student Retention. The Journal of Higher Education, 64(2).

Castaño, E., Gallón, S., Gómez, K., & Vásquez, J. (2004a). Deserción estudiantil universitaria una aplicación de modelos de duración. Lecturas de Economía, 60, 26.

Castaño, E., Gallón, S., Gómez, K., & Vásquez, J. (2004b). Deserción estudiantil universitaria: una aplicación de modelos de duración. *Universidad de Antioquia, Lecturas de economía*, 39-65.

CEDE. (2007). Investigación sobre Deserción enlas instituciones de Educación Superior en Colombia.

Cioffi-Revilla (2014). Introduction to Computational Social Science, London, Springer

Gilbert, N., & Troitzsch, K. G. (2005). Simulation for the social s*cientist*. Shoppenhangers Road. Maidenhead. Berkshire. England: McGraw-Hill Education.

Gómez, N. (2016). Simulación basada en agentes. Una metodología para el estudio de sistemas complejos. *Centro de Innovación, Escuela de Administración*, 24.

Macal, C., & North, M. (2014). Introductory tutorial: Agent based modeling and simulation.pdf>. *Proceedings of the 2014 Winter Simulation Conference*

A. Tolk, S. Y. Diallo, I. O. Ryzhov, L. Yilmaz, S. Buckley, and J. A. Miller, eds.

MEN, & Permanencia Group, G. (2015). Educación Superior en Cifras *Boletin, 6*. Retrieved from http://cms.colombiaaprende.edu.co/static/cache/binaries/articles-350451_recurso_6.pdf?binary_rand=2733 website:

MEN, & Qualificar Group. (2015). *Estrategias para la Permanencia en Educación Superior Experiencias Significativas*. Retrieved from Bogotá, Colombia: http://www.colombiaaprende.edu.co/html/micrositios/1752/articles-350844_pdf.pdf

Montes, I., Almonacid, P., Sebastián, G., Zuluaga, F., & Tamayo, E. (2010). Análisis De La Deserción Estudiantil en los Programa de Pregrado de La Universidad EAFIT. *Estudios en Economía y Empresa. Departamento de Economía. Escuela de Administración*, 118.

Pinto, M., Durán, D., Pérez, R., Reverón, C., & Rodríguez, A. (2007). *Cuestión de supervivencia Universidad Nacional* (Vol. 1). Bogotá: Universidad Nacional de Colombia.

Porto, A., & Di Gresia, L. (2004). Rendimiento de estudiantes universitarios y sus determinantes. *La Revista de Economía y Estadística, Cuarta época, 42*,

Quintero, R., Vásquez, D., Torres, N., Estrada, J., & Castellanos, M. (2015). *Bajo Rendimiento Académico En La Universidad Distrital Francisco José De Caldas* (U. D. F. J. d. Caldas Ed.). Bogotá: Universidad Distrital Francisco José de Caldas.

Sayama, H. (2015). *Introduction to the Modeling and Analysis of Complex Systems*. State University of New York at Geneseo. Geneseo, NY 14454: Open SUNY Textbooks, Milne Library.

Spady, W. G. (1970). Dropouts from Higher Education An Interdisciplinary Review and Synthesis.

Vásquez V, J., Castaño, E., Gallón, S., & Gómez, K. (2003). Determinantes de la Deserción estudiantil en la Universidad de Antioqia. *Centro de Investigaciones Económicas Universidad de Antioquia*.

Wilensky, U. & Rand, W. 2015. An Introduction to Agent Based Modeling Modeling Natural Social and Engineered Complex Systems with NetLogo, United States of America, Massachusetts Institute of Technology