Evaluating System Dynamics as a Tool for Teaching History

Margarita María Cruz Barrientos
Mphil. in System Dynamics / University of Bergen, Norway
Corso Vittorio Emanuele 474, 90100 Palermo-Italy
+39 3345786165
macusita.82@gmail.com

Abstract

Key words: conventional education, weaknesses, behavior over time, structure, understanding, identification, transfer, learning, performance, revolutions**.**

History has been traditionally taught as a presentation of isolated facts which fill students' minds with a great volume of dates, actors and events. However, students are seldom encouraged to relate what they learn with changes over time. Thus, they present a lack of understanding of what history has to do with them. Therefore, there is a strong need of placing the understanding of this subject into a structured pattern that leads students to the understanding not only of the past, but also of the present and the future. In this paper, SD is evaluated as a tool to enhance students' understanding of history. Experiments with high school students were run, in which SD and the conventional method of teaching history were tested with groups of students with either no previous experimental experience with the teaching method or without previous experimental experiences on them. The results present the SD teaching method as a better tool to teach history with students who used SD before. however, the conventional method reveals to be persistent in student's minds when they do not posses any previous experience with SD. Important outcomes are documented for future replications of the experiment.

1. Problems with Conventional Methods of Teaching History

The goal of learning history is subject to two main interpretations by educators: *enhancing collective memory* or *the disciplinary approach*. Enhancing collective memory is the conventional goal of history education, which proposes history instruction provide learners with a base of historical knowledge that is deemed important by authority figures who guide educational policies. Under this goal, students' primary task is the chronological memorization of events, and their associated names, dates and locations. The second approach is based on teaching skills for understanding history in the way historians do, which indeed, is not the conventional or widely used method of teaching this field (Wiley & Ash, 2005).

The usual history book combines text of historical narratives with maps, charts, timelines, pictures, diagrams and paintings to convey the historical happenings. Furthermore, history teachers often supplement reading assignments and lectures with films or documentaries based about how history occurred. However, despite the use of this media, students are not provided yet with the proper understanding of the past in light of comprehending the present and interpreting possible futures (Potash 1995).

It has been widely discussed among critic educators that school education is not preparing students to cope with modern life. It is argued that the main failure is revealed in the form of corporate executives who misjudge the complexities of growth and competition, government leaders who are at a loss to understand economic and political change, and publics that support inappropriate responses to public concerns (Forrester, 1992). Furthermore, since students are overloaded with facts without having a framework of reference for making those facts relevant for the complexities of life, it is strongly discussed that there is an important part of the learning process, which is discarded. Thus, there is strong agreement in the fact that traditional education, because of its fragmentary nature, becomes less relevant as society becomes more complex, crowded, and tightly interconnected (Forrester, 1992). In addition, there has been extensive accordance in the fact that Conventional teaching methodology seems to make many students lose motivation about some fields that are difficult to relate to the present. Furthermore some students after being taught with the Conventional method fail to recall previous knowledge and thus, they make mistakes due to ignoring lessons of the past (Forrester, 1992), (Burton, 2006)[1], (History News Network, 2006). This is the case of social sciences, especially history, a subject that is widely taught as a mere presentation of dates, events, and actors, which are hardly related to causes, consequences and effects that have to do with other fields of study and with real life. Hence, history is widely taught without analyzing the context that gave life to the happenings and without relating similar behaviors on a shorter time scale that a student can experience in a week or a year (History News Network, 2006). As a result, students find difficult to relate to changes over time in the present and future. They can hardly understand the reason why they learn something that has happened already and is not going to change by its study (Forrester, 1992), (History News Network, 2006).

Beyond the previous thoughts and impressions of those critic educators lay in students' performance as the determinant factor about whether or not students, certainly, understand what it is presented to them about history, and about whether or not such critics are truth.

Results of this Conventional schema of teaching history speak for themselves. Embedded into this frame, students –especially those in high school- dislike those subjects that are barely related to real life, that seem to be impractical. When they list their favorite subjects, history always comes in last. They consider it the most irrelevant of 21 school subjects; "boring" is the description most often applied (Loewen, 1996). Up to 20% of the student population at Fisher School in United States have classified history as the least favorite subject they are ever taught (Jianting, 2006). On the other hand, it has been found that history has not only been classified by students as the most boring subject but also as one of the hardest and most difficult to understand and recall in mind after the examination periods, just 10% of the population at Fisher School find

[1] Orville Vernon Burton is Professor of History and Sociology at the University of Illinois, Urbana-Champaign (UIUC). He is also a Senior Research Scientist at the National Center for Supercomputing Applications where he heads the initiative for Humanities and Social Science projects. Further information at: http://historymatters.gmu.edu/d/6164

it easy to learn (Jianting, 2006), (Burton, 2006). Furthermore, Pinette[2] has stated that because the typical high school history curriculum requires students to do little more than memorize names, dates, and major historical events, students often begin college with the belief that history is "Truth" rather than interpretation, which may highly intervene in their performance by limiting the learning sense of such field (Pinette, 2002).

Such performance confirms, then, what educators tend to highlight with the intention of claiming for improvements in history education. Therefore, this overview cannot be nullified and deserves full attention, especially when history is seen as a discipline whose learning and understanding takes students into the commitment with social reality (Burton, 2006). Furthermore, history’s understanding becomes crucial because mistakes made in the past recur often because there is not sufficient knowledge and understanding of the lessons of the past (Burton, 2006). Hence, it is clear that history education requires much more than the presentation of snapshots of what happened in the past. There is an evident weakness of the Conventional method in giving the students the tools to interpret more effectively the world around them. The method, then, call for its enhancement in order to fulfill the dynamic understanding demanded from students when they face real life complexities.

Innovative teachers and schools have made several efforts to improve the learning process of history and to motivate students to learn more. Some professors have encouraged their students to think about other ways of problem solving that could have been used in the past to make historical happenings to occur differently. This methodology has taken students into a deep analysis of the historical facts and made them relate a piece of history to some other historical patterns (Burton, 2006). More recently, informatics technologies have been introduced by some institutions as the main tool to improve student’s learning about history. Digital History is a website developed to support the teaching of American history in schools and colleges, which intends to be innovative by presenting interactive learning modules where history is taught as a part of the society, and visual effects are used to emphasize the important issues in history (Mintz, 2003). However, there has been no assessment of the performance of students using such methodologies.

The concept of multimedia learning, defined as acquiring knowledge in a domain through interacting with an educational environment that presents information using multiple sources, has also been applied to history in an attempt to improve the learning process. Two main reasons support the use of the multimedia learning theory in history: (1) Multiple-source environments attempt to make history learning more like the activities of real historians, and (2) Graphics or archives are often used to make the context of the time more engaging, vivid, or personally relevant for the learner. The approach is based on a constructivist point of view, which proposes that learning that is done as a form of inquiry leads to better understanding of the subject matter than learning that is transmitted through lecture or memorization. In a history classroom with this approach, rather than being simply told to believe a single story or learn what is in the textbook, students are presented with information from a variety of sources and perspectives, and taught the standards of historical inquiry, investigation and debate.

[2] Denise Pinette Domizi is the designer of the instruction “Constructing History: How historians see the light” whose purpose is to help students examine their beliefs about history and to assess a conceptual change for them to understand better this field (Pinette, 2002).

One of the most representative studies of the multimedia learning in history is the work of Kathrin Spoehr, in which a corpus of high school hypermedia instructions (called ACCESS –American Culture in Context: Enrichment for Secondary Schools) was created to enhance student understanding by supplementing text book materials and class instruction. Several assessments of learning outcomes show that this use of multimedia was beneficial to students with ACCESS: Those students who were taught with ACCESS outperformed those who were taught with the non-ACCESS class instruction. However, those improvements cannot be entirely attributable to the impact of multimedia learning on history because students in the ACCESS classroom profited from the role played by the teachers who were accompanying the process in the ACCESS classrooms, whose advance knowledge on the field could help students intensify and maximize the construction of the hypermedia corpus. Thus, the only conclusion that can be drawn from such experiment is that when multimedia learning is used in problem-based inquiry tasks, with teachers who are involved in the construction of a corpus and who think about the best ways to integrate the corpus into ongoing instruction, then improvements in historical understanding can be obtained. Furthermore, other studies on multimedia history instruction report pre/post test gains in learning. However, those gains cannot address whether the multimedia environment is better than learning from a text book or a lecture (Wiley & Ash, 2005). Thus, advantages of the multimedia learning method over the Conventional teaching method of history are not strongly conclusive so far.

However, even when some students and teachers find these methods as innovative and useful, they have not been widely recognized as a Conventional method to teach history (Forrester, 1992). In addition, despite these and other intentions to improve history instruction, students who are taught by these methodologies are not involved in a real thinking environment in which they can find out how history happened, what causes change over time, how lessons of history could be interpreted to the present, and how consequences of the historical happenings affect other matters along the years. Thus, it is important to mention what Donovan and Bransford (2005) have proposed for making history a useful and enjoyable discipline for students.

After several meetings with teachers, Donovan and Bransford concluded that history cannot be learned unless it is presented to students in a general framework in which some concepts inherent to it must be considered. Such concepts include: *time* taken for the historical event to be developed, *change* in the state of affairs, *causes* that make the change occur, *empathy* that leads to understand people's ideas in the former times, *evidence* of the historical situations which is important to be interpreted for the understanding and learning of the discipline, and most important, what they called *accounts* which looks for accumulating knowledge of history through the learning life of students. Once students have been through such a learning process of history, they find this subject as a discipline that really contributes to the understanding of the past, the present and the future. Some short experiments were run with students of high school and after being analyzed qualitatively, the author concluded that students do well when history is conceived as a dynamic and interrelated discipline (Donovan & Bransford, 2005). Given this perspective, history stops being perceived by students as a mere recall of events and becomes a dynamic discipline that, even when it is based on former happenings, its implications go through time. Thus, if there is an approach to teach history, which can meet the tenets proposed by Donovan and Bransford, it would be really helpful in broadening students' understanding and motivation.

Rather than looking for media that eases the communication of the history, it is required an approach that allows making real the learning methodology proposed by Donovan and Bransford, which lets students understand why and how history occurred and changed over time, and how those changes affected different fields and situations that happened in the past. Furthermore, once such methodology is brought to life, historical patterns must be able to be related to other situations that behave similarly, and thus, students may be committed themselves to the society and all its complexities. Such approach is described with this paper.

2. Using System Dynamics to teach History

2.1 Usefulness of SD as a teaching tool in social sciences and History: looking over the evidence

From the definition of its founder[3], System Dynamics (SD) is a perspective that combines theory, methods and philosophy for analyzing the behavior of systems. In such perspective, the world is understood as a whole, rather than as the result of isolated pieces. Thus, SD shows how thing really change through time (Forrester 1998). Complementary, SD is defined by John D. Sterman as *"a perspective and set of conceptual tools that enable us to understand the structure and dynamics of complex systems. It is also a rigorous modeling method that enables us to build formal computer simulations of complex systems and use them to design more effective policies and organizations (Sterman 2000).*

Thus, SD seems to be a promising methodology to meet the proposal done by Donovan and Bransford (2005) given its nature of expressing relationships between different elements, changes over time, and causes and effects related to past, present and future happenings. SD provides as well the possibility of understanding history within a certain context.

In the field of education, SD has been found as a useful tool to facilitate students' performance[4]. The dynamic approach based on S&F and CLD provides a foundation that is transferable from field to field, a basic that places the knowledge into a structured pattern, which augments the learning and ability to transfer knowledge to other happenings. There are several schools doing excellent work in this respect. Pioneering schools are scattered mainly over the United States, extend into the Scandinavian countries and Germany (Forrester, 1992; Skillings, 1997; Road Maps, 1992). Before 1992, no network has existed for interchanging information regarding SD among those involved in pre-college education. Thus, the Creative Learning Exchange has been founded in order to distribute SD materials among all those who are interested in it (Creative Learning Exchange, 1992).

Road Maps is a series of self-study guides that use modeling exercises and selected literature to provide a resource for learning about the principles and practices of SD; cases studies are related to social sciences, biology, management (Road Maps, 1992).

---

[3] Jay Wright Forrester

[4] Besides education, SD has been even more widely used to approach management, urban, and development problems, performing an important role due to its contribution for a better understanding and improved policy making of these issues (Forrester 1992)

Roberts concluded that SD, as an organizing framework, can be a very useful tool to teach and improve performance of fifth and sixth grade students, through the implementation and evaluation of a curriculum to teach children about SD and its application in many fields of study. Though the sample was small, the results seem positive enough to warrant further experimentation with this strategy for teaching and understanding problems (Roberts, 1978).

The Feedback Method is an SD approach to teach macroeconomics to college and high school students. Assessment in this regard has revealed that students not only preferred the explanations given by the SD approach, but also they perform better when they are taught with it (Wheat, 2007).

In the field of History, Greg Reid has explored the causes of the American Civil War in light of systemic thinking and SD tools. Instruction made of five one hour lessons given to students, aimed to supply the students with an understanding of why the Civil War took place, and that the war was not an isolated event that came out of the blue, but a conflict that began brewing during the foundation of the country. Students were encouraged to help the teacher complete causal loop and S&F diagrams that illustrated the gradual increase in the main reasons of the civil war (Reid, 1996). Another case study is the Easter Island population problem, which has been taught using S&F diagrams by Diana Fisher as part of her curriculum. A population model has been utilized to explain the reasons of the depletions of the population of this isolated island. Concepts, such as the carrying capacity of the main resources, are introduced for explaining the behavior over time of the population. Learning by doing is the methodology followed by students. As they are taught about what happened in the Island, they also model by themselves what they are taught. Policy making also makes part of the case study, thus, students are encouraged to answer what would happen with the population if more coconut palms were planted to support the nutrition of the island (Fisher, 1992). Unfortunately, there is not available information about the way Fisher's and Reid's lessons were assessed and how much they enhanced students' understanding regarding populations' problems.

Doubtless, the most significant contribution to the use of SD in the history curriculum has been done by Jeffrey Potash and John Heinbokel, who have adopted great interest in using SD in the social sciences' classroom. A conventional curriculum based on SD for the teaching of social sciences and mainly human history became their main focus of attention after they have realized that repeated patterns of behavior are present along history, in which populations have shaped the course of human history because of constraints given by the interaction with resources and human attitudes. Furthermore, they believe that understanding how and why history has unfolded replete with recurring patterns, will be determinant for the student's motivation and to enable them to learn from the past when thinking about the present and future (Potash 1995; Potash, 2005; Potash & Heinbokel, 2006).

Their first effort, entitled *Plagues and People*, commenced with a relatively simple model of disease dissemination that drew upon historical epidemics to develop the historical contexts within which these operated, and to better understand the current AIDS epidemic in terms of cultural as well as biological factors. This model has been tested with students in the form of a curricular practice, in which students were taught gradually how to build and simulate models based on S&F diagrams. This methodology led the students' understanding of the case studies to be related to other facts over time, and certainly, to the context in which they were developed (Potash *et al,* 1996). Later on, *Population Dynamics and the Human Experience*, is doubtless their most exciting and ambitious project of all. They identify myriad structures which influence how, where, and why human populations have grown over the course of history; and then look at the role of human population growth as it has affected economic, political, and social systems, past and present (Potash, 2005). Correspondingly, Heinbokel and Potash with the Center for Interdisciplinary Excellence in SD (previously the Waters Center for System Dynamics) have designed Demo Dozen, a collection of 12 interactive lessons[5] about American history, bacteria reproduction, finances, which are accompanied with dynamic models (CIESD, 2001; Waters Center for System Dynamics, 2001). One of the most famous case studies taught with Demo Dozen is the Irish Potato Famine. Beyond the conventional causes found for this happening, they explain using S&F diagrams and time series graphs, how and why this happening is related to past and future happenings such as the great depletion of the Irish population even several years after the famine. Thus, a stock representing the Irish population at any time is increased by the birth rate and the death rate in Ireland. The most interesting point of their approach is the consideration of the available resources, and the attitudes and expectations of the people towards believing about getting welfare in Ireland. These factors become the most determinant factors for population loss in Ireland in the period 1846-1900, even after the famine (Potash & Heinbokel, 2006).

After a while of exploring in the history classroom, Potash and Heinbokel state that passing from the learning of discrete events regarding the social sciences to the understanding of patterns as part of a whole, is a process which is really enhanced by using generic SD concepts and tools. Furthermore, they strongly believe that *"When students develop the capacity for and the interest in understanding the powerful role of dynamic feedbacks between populations, resources, and attitudes in the past and, more critically, when they can see the relevance of that learning when applied to their own world, we will have truly made progress in our efforts to bring social studies to the level of creating systems citizens"* (Potash, 2005; Potash & Heinbokel, 2006).

Nevertheless, assessment is absent in all the practices employed by those who have intended to approach history by using SD, even in the most important contribution made by Potash and Heinbokel. Despite all of them highlight the fact that students benefit from using SD as a teaching tool, none of them has evaluated how much benefit those students get. Controlled experiments, in which it is possible to measure how much enhancement and improvement the students profit from an SD approach, is still missing and is calling for answers. In the intent of recognizing SD as a generic tool to teach different fields, scientific assessment is widely required.

2.2 SD for teaching revolutions

2.2.1 Revolutions as diseases

Revolutions are among the most often repeated happenings in history. They describe more or less similar patterns of behavior, which come from similar structures or

[5] Such collection of interactive lessons is called Demo Dozen, which is available at http://www.ciesd.org/influence/demo_dozen.shtml

relationships between elements. Thus, revolutions result an interesting case to approach history.

Crane Brinton (1965) analyzes and compares the development over time of different revolutionary processes such as the American Revolution, the French Revolution, the English Revolution and the Russian Revolution too. He finds out that all of them share common patterns that made them unfold in the way they did. He describes such patterns through the analogy of the development of a fever over time, in which three stages are present most of the time: the symptoms, the fever itself, and the breakdown. The symptoms represent the dissatisfaction of a current state versus a desired state; the fever itself represents the development of the revolution over time; and the breakdown represents the loss of morale by those who support the revolution, and therefore this stage describes the revolution's end (Brinton, 1965). Among all stages, the fever itself exhibits the core point that determines the course of revolution; therefore, the process of understanding how a revolution gains and looses power is essential for the students to manage and interpret similar situations in the past, present, and the future.

Specifically, as Brinton (1965) states, the *fever itself* goes through different stages. Likewise, once the revolution has started, it experiences different phases in which its popularity and support of the people varies over time.

The full symptoms disclose themselves and the fever of the revolution has then begun. After a hard beginning because of constructing reliability on the revolution, the fever of the revolution faces a period of great popularity. The revolution then works up, not regularly but with advances and retreats, to a crisis, frequently accompanied by delirium, the rule of most violent revolutions: reigns of terror. Specifically, the revolution gains great support, in which several chaotic happenings occur for a while until the first and main event occurred: the breakdown of the often oppressor government is reached (Brinton 1965). From the revolutions analyzed by Brinton (1965), in the specific case of the French Revolution, million of conversations between the French population spread the fear that the king and his party were about to dismiss the revolutionary assembly and rule by armed force. The revolution started to gain even more power. Paris and other French towns, therefore, rose in its might and with a sure instinct did impressive revolutionary facts such as the seizing on the Bastille. The revolutionaries stirred up France in a hundred ways: they sent orators to street corners and cafes, they distributed radical news-sheets and pamphlets, they sent agents to spread discontent among the royal troops, and they even subsidized prostitutes to get at the soldiers more effectively. People joined the Revolution because their neighbors already did it. A strong motivation to diffuse the fever of the revolution is revealed in events like these presented in France.

Once the revolutionaries have gained what they wanted, they wish to stay in power. New ways of government arise and a revolution is still on against those who do not support it. Until the moment, a process of strengthening the revolution through more support is essential to keep the power. In the meantime of these events, the incipient government lacks ability to meet the interests of all those who had hopes in the revolution, and this is the cause for stronger and radical supporters of the revolution to take over the power and start reigning. In this moment, radical and lunatic events start happening such as the terror reigns, in which many people are killed at a search of virtue of the revolution (Brinton, 1965). The revolutions analyzed by Brinton (1965) showed to be very successful in the first stage. They became actual revolutions instead of mere discussions and desires, especially after revolutionaries have beaten, or won over, the armed forces of the revolution.

Crane Brinton (1965) also states that in social systems, as in the human organism, a kind of natural healing force tends almost automatically to balance one kind of change with another and restorative change. Thus, social systems such as revolutions seek themselves for the balance, in order to gain equilibrium after a period of crisis and disorder. Once the revolution has gained power, its natural trend is to recover a state of equilibrium, in which the population feels satisfied with it. Tranquility is claimed by the population and so the revolution itself starts losing power. The equilibrium is then restored and the revolution is over. In all four revolutions analyzed by Brinton (1965) the crisis period was followed by a convalescence, by a return to a fundamental desire of calm (Brinton, 1965).

Thus, the fever itself of a revolution should be taught using an approach, such as SD, that describes the dynamics of revolutions rather than mere discrete events that happened.

2.2.2 Useful SD concepts for teaching revolutions

The description of the fever itself as stated by Brinton (1965) presents two main processes that reveal dynamics and diffusion of the passion of revolution. On one hand, the gain of strength of the revolution and all the critical happenings occurring alongside are the result of a reinforcing process, in which the desire of revolution is increasingly supported by the population. The spread or diffusion of such desire of revolution is given by the several conversations, news-pamphlets, orators, and different strategies adopted by the revolutionary parties. The diffusion process amplifies and reinforces the popularity of revolution among people. In SD terms, this process is given by a reinforcing loop describing an exponential growth. On the other hand, there is a process of seeking balance to restore calm among population. This process is then given by a balancing loop, in which a goal seeking behavior proceeded by a collapse describes the way the population lose interest in the revolution and tranquility is recovered.

Thus, in terms of SD, the fever itself of a revolution is analogous to a diffusion process given by a generic SD structure called the SI Model or Diffusion Model. This model is a simplification of diseases, which represents the spread of a disease within a population over time.

In the model, the total population of a region or community is divided into two categories: those susceptible to the disease, S, and those who are infectious, I, (for this reason the model is known as SI model). As people are infected they move from the susceptible category to the infectious category. The SI model is the simplest model of spread of epidemics and is based on the assumptions of not taking into account births, deaths, and migrations. Furthermore, it has to do with chronic infections in which once people are infected, they remain infectious indefinitely[6]. The SI model contains two

[6] Due to the simplicity of the SI model, extensions of it have been made, in which recoveries from the ill state and deaths are possible. Therefore, the stock of infectious population is decreased at the last stage. In the way revolutions are approached for the present experiment, both, the simple SI model and the

loops, the positive *Contagion loop* and the negative *Depletion Loop*. Infectious diseases spread as those who are infectious come into contact with and pass the disease to those who are susceptible, increasing the infectious population still further (the positive loop is dominating) while at the same time depleting the pool of susceptible (the negative loop). Both categories of population, Susceptible and Infectious are represented by stocks in the SD nomenclature, while the rate at which the population gets infected and migrate from susceptible to infectious is considered a flow. The infectious population exhibits s-shaped growth, in which after great difficulty of infecting new people at the beginning of the disease, new cases of infection rise exponentially the stock, which later on keeps constant as there are no new cases of infection.

Three more constant variables make part of the model: the probability of infecting a susceptible person when is contacted by infectious people, monthly contacts per infectious, and total population.

As follows the S&F diagram of the SI model is presented with its corresponding graph of behavior over time.

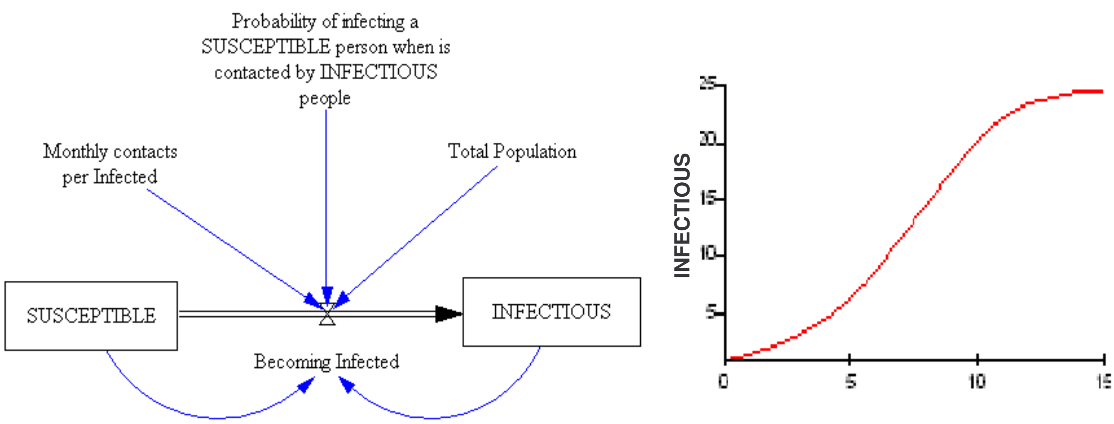


Figure 12. S&F diagram and behavior over time of diseases

From this perspective this model becomes very useful to teaching about the revolutionary process in history.

2.3 In-service Training: the outset of infecting with the SD approach

At each school in which the history experiment was run, a three hour in-service training with teachers was done in order to share the basics of SD and the core meaning of SD with which history would be explained to students in the SD groups. The teachers were taken into some theory about SD through a verbal presentation (with power point presentation as an aid), then were encouraged to play the Infection Game[7] and finally, the presentation of the Diffusion Model was used to debrief the game and transmit the understanding of situations which are spread as diseases do.

considerations about deaths are taken into account. Further information about the Diffusion Model (with all its extensions) is available in Business Dynamics, Chapter 9 (Sterman, 2000)

[7] The Infection Game is adapted from The Epidemic Game by Hill Glass at the catalina Foothills School District, Tucson, Arizona, 1993.

Main concepts of SD such as accumulation, multiple influences, feedback loops, non proportionality, and delays were taught to teachers through the explanation of the bathtub analogy. The teachers did not have previous knowledge or experience in SD; thus, in order to let them understand those new concepts, the Infection Game was played and immediately debriefed in an active plenary session. At this point, the Diffusion model was presented as the core model that leads to the understanding of how the disease was spread among all the members of the group, and how the number of people who was infected every day changed over time. The diffusion model played an important role because it led teachers to understand from which perspective the history instructional method will be approached when being taught to the students. Figure 13 illustrates the way the Infection game was debriefed using the SI as approach during the in-service session.

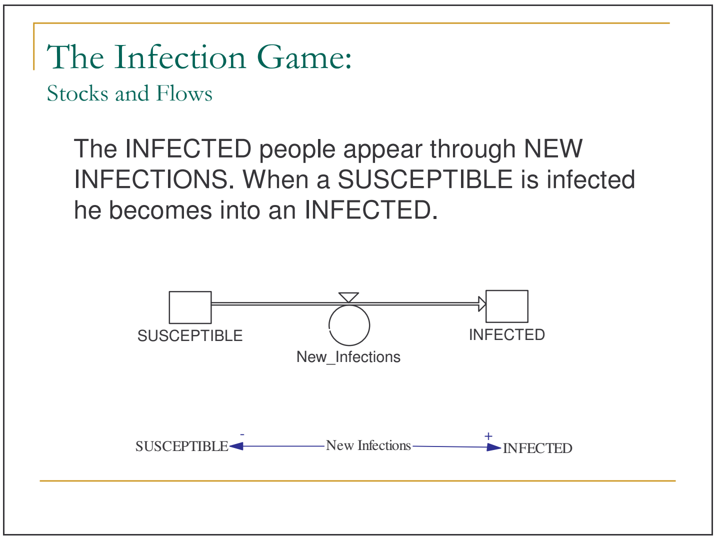


Figure 13. Slide # 8 Infection Game debriefing during the In-Service training

Interest and enthusiasm were the main reactions of teachers during and after the in-service training session. At first, their attention was focused on learning about how to improve education in K-12 and how to make the students more interested in it; but while taking them into the entire in-service session they could see the kind of generic applications that this tool can have on different fields, and thus, teachers got enthusiastic and really interested in learning how to apply such approach.

Concepts such as accumulation, flows and feedback loops were difficult for them to grasp. However, the idea of approaching different fields (such as diseases and history) from a systemic point of view was challenging and encouraging enough to make them ask further questions and to keep interest in the session. The queries were mainly about the previous applications of SD in the education field, their results in the understanding and interest of the students, and about the advantages of using this tool instead of using a conventional teaching method. Furthermore, the most eager ones were interested in knowing how to apply it in examples of physics, biology and informatics that make trouble to the students.

This session was really important because it revealed how relevant the improvement of education and the need that teachers have to approximate to the best approach for teaching K-12 students. It was revealed a strong need to enhance the understanding of students in both, the soft and hard sciences.

Because of all stated in this section, it may be thought of SD as a better approach to help students understand history within the context in which it occurred and to interpret it in light of the present life and possible futures. Specifically, it may lead students' understanding into why and how history occurred and changed over time, and how those changes affected future situations and different fields. Furthermore, generic structures of SD, such as the SI model, that can be applied to different fields of study, allow relating historical happenings that behave similarly, and thus, students may be committed themselves to the society and all its complexities. Thus, this study aims to answer this research question:

Does SD enhance students' understanding of history as a disciplinary approach?

Exploring whether students' understanding of history is enhanced when the students either have or do not have previous knowledge of SD is highly interesting to determine the role of that teaching method in the learning process of history. Assessment of the impact of the SD teaching method on students' performance is aimed to be provided to validate the contribution of SD in students' understanding of history.

3. Experimental Design

3.1 The Research Method

The chosen method of research is a laboratory experiment in which actual history classrooms constitute the "laboratory". Different instructional treatments were applied to different groups of students in order to compare their performance.

The Teaching Method and the Experimental Experiences were the two treatments applied to the groups of study. The Teaching Method treatment concerns the method used to teach history to the students, and it is made of two levels, the Conventional method and the System Dynamics method. The Experimental Experience treatment concerns students' previous experience in experiments using the teaching method assigned to them. Those without prior experience represented the First Experimental Experience, while those who participated in the Civics' Engagement Experiment by Maria Teresa Gonzalez[8] represented the Second Experimental Experience.

These treatments definitions lead to the following experimental design, in which four different groups were taught with different teaching methods and must go through different experimental experiences.

Treatments		**Teaching Method**	
		SD Method	Conventional Method
Experimental Experiences	First Experimental Experience	SD1 group	CONTROL1 group
	Second Experimental Experience	SD2 group	CONTROL2 group

Table 1. Treatments

[8] Further information about the Civics' Engagement Experiment is available by contacting its author. Email address: maria.galvis@student.uib.no or at www.clexchange.com/ftp/newsletter/CLEx16.1.pdf

The Conventional teaching method is the traditional way of teaching history in a Colombian classroom. This typically involves reading passages about the development of a historical process based on the presentation of isolated events, reading about the biography of the main characters in history, presentation of pictures, and timelines. Additionally, graphs over time accompanied the explanation of the development over time of the historical case study. Such graphs are not often used as a tool to teach history, but certainly could be added to the Conventional curriculum. By using the time series graphs in both teaching methods leaves the SD concepts as the only difference between both teaching methods. The SD teaching method provided the students with the same information as the Conventional teaching method; however, SD tools (stocks, flows, S&F diagrams, and SI model) were used to approach the case study from a SD point of view. Thus, both system thinking and S&F diagrams in the SD method were the main difference presented in the teaching methods. The Conventional method presented isolated historical events, while the SD Method presented connections and a structure underlying the happenings of such events. Both teaching methods were presented to the students in computerized slideshow format.

In the First Experimental Experience, students studied a history case study using either SD or the Conventional method. Students in the Second Experimental Experience approached the same history case study, and in addition, a Civics[9] case study, using in both experiments the same teaching method (either SD or the Conventional method). Thus, more than one field of study was approached with the teaching method.

3.2 Students' Task

Table 2 summarizes the way the teaching method was presented to each group, according to what was said above. Differences in the presentation of the teaching method between groups are expected to make their performance discern.

3.3 Teaching Method: French Revolution

As previously mentioned in the section 2, given the analogy between the development of fevers and revolutions over time presented by Crane Brinton (Brinton, 1965), the course of a revolution can be taught by using the concept of "*infection*" of the desire of revolution within a certain population. A person who makes part of the revolution is the infectious one who can transmit the desire of revolution (*the disease*) among people. *Infecting* other people to join the revolution is not an easy task at the beginning and it takes some time until people starts getting *infected* by such purpose. However, once there are more revolution's supporters (*more infectious*), the desire of joining the revolution starts to increase and to become popular among the people (*contagion loop*). After a while, there are few non-revolutionaries to who transmit the passion for the revolution, thus the number of people supporting the revolution does not increase anymore, and this can either remain constant or collapse. In figure 14, the behavior of an infectious population and a revolutionary population are presented to reflect at some extent their analogy in their behavior as stated by Crane Brinton (1965)[10].

[9] The Civics Engagement Experiment was run a day before the students in the Second Experimental Experience went through the history experiment.

[10] In figure 14, both behaviors represent at some extent the behavior of both populations over time. However, such behaviors can differ, especially at the late phase, because of the deaths or other factors that decrease the number of people in both populations. Furthermore, it is important to be aware that not

Groups			
SD1	**CONTROL1**	**SD2**	**CONTROL2**
-Teaching Method: **SD**	-Teaching Method: **Conventional**	-Teaching Method: **SD**	-Teaching Method: **Conventional**
-Experimental Experience: **First**	-Experimental Experience: **First**	-Experimental Experience: **Second**	-Experimental Experience: **Second**
-Pre test (one week in advance)	-Pre test (one week in advance)	-Pre test (one week in advance)	-Pre test (one week in advance)
-Verbal instruction about SD principles and the Infection Game. Duration: 3 hours		-Verbal instruction about SD principles and the Infection Game. Duration: 3 hours	
-Computerized slideshow about the French Revolution, using SD as approach. Duration: 2 hours	- Computerized slideshow about the French Revolution, approached with a Conventional method. Duration: 1 hour	-Computerized slideshow about Civics, using SD as approach. Duration: 2 hours	- Computerized slideshow about Civics, approached with a Conventional method. Duration: 1 hour
		- Computerized slideshow about the French Revolution, using SD as approach. Duration: 2 hours	- Computerized slideshow about the French Revolution, approached with a Conventional method. Duration: 1 hour
-Post Test	-Post Test	-Post Test	-Post Test

Table 2. Tasks per group

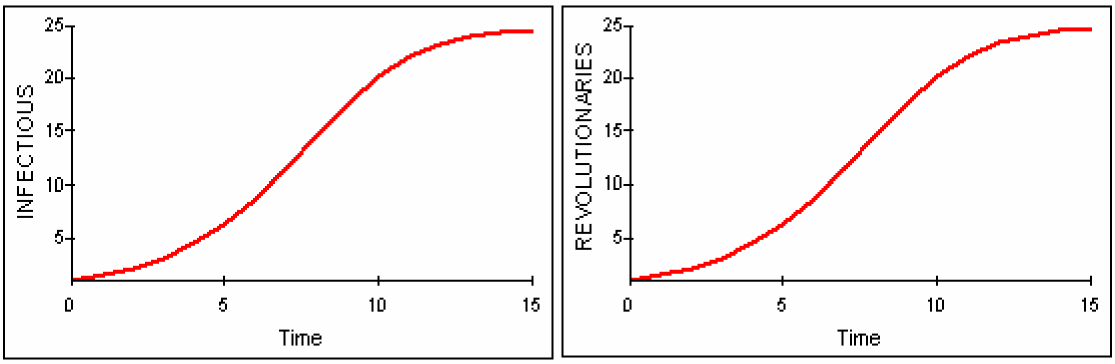


Figure 14. Behavior of revolutions as diseases.

With the purpose of clarifying this concept to the students who are learning history in this experiment, the French Revolution has been chosen as the case study. Thus, the generality of the development over time of the revolution can be transferred to the French Revolution, and the SI model can be utilized to explain students in the SD groups how the revolution changed over time in light of those persons supporting it. For those students involved in the Conventional method, the focus is on the changes that the revolution suffered over time.

The Conventional instructional method approaches the instructional goal by using readings, passages, pictures, presentation of isolated events regarding the French revolution, and presentation of graphs about the number of revolutionaries supporting the revolution. However, the SD relies on system thinking and S&F diagrams. In the SD instructional method, the case study utilized the SI model, explained in section 2.

---

every revolution behaves in the same way. Revolutions studied by Brinton (1965) are alike in this sense; however, other kind of revolutions may totally differ from the disease analogy.

3.3.1 SD approach of the French Revolution

Revolutions describe s-shaped growth in which the number of people joining a revolution changes according to the dominance of a reinforcing (strengthening) loop or of a balancing (weakening) loop. The strengthening loop is analogous to the contagion loop in the SI model and the balancing loop is analogous to the depletion loop. The groups of revolutionary and non-revolutionary people can be thought as stocks. At the same time, the people joining the revolution can be interpreted as a flow, which decreases the non-revolutionary stock and increases the revolutionary stock.

At the beginning of the French Revolution, the strengthening loop takes a while until people star believing in the usefulness of joining the revolution. So, the stock of revolutionaries starts accumulating people who make part of the revolution. New supporters of the revolution might want to involve more people in the revolution. Thus, the contagion loop starts to dominate and to convince many people to join the French Revolution. At this point the stock has started to grow exponentially and the revolution has become quite popular between the French people. However, once the revolution has gained popularity and gained lots of supporters, fewer people are susceptible to become revolutionary. Thus, the stock of revolutionary people stops to increase in that accelerating way, and begins to seek balance. The depletion loop takes dominance at this moment, and slowly all those who were not part of the revolution join it as well. The stock still accumulates revolutionary people, but the accumulation occurs slowly now because conversions from non-revolutionaries to the revolutionary party do not occur that often. When there is no more susceptible population to convince of joining the revolution, the stock stops to increase and reaches stability. The depletion loop has balanced the contagion loop and the French Revolution does not last forever.

Thus, the French revolution is suitable to be approached with the SI model mentioned in section 2. The first approximation to the French Revolution with the SI model considers the simplest SI model, without deaths of revolutionaries and returns to the non-revolutionary state are not considered. At this point the stock of revolutionary people behaves as shown in figure 14. In the last parts of the instructional method, the extension regarding the dead revolutionaries is included in the model for enhancing students' understanding of such system[11].

[11] This extension of the SI model is called SIR model, where S and I are still the Susceptible population to get infected and the Infectious population correspondingly, and R is the Recovered population from the disease. This case treats the diseases which are not chronic, such as the flu, in which the population gets recovered from the disease after some time (Sterman, 2000).

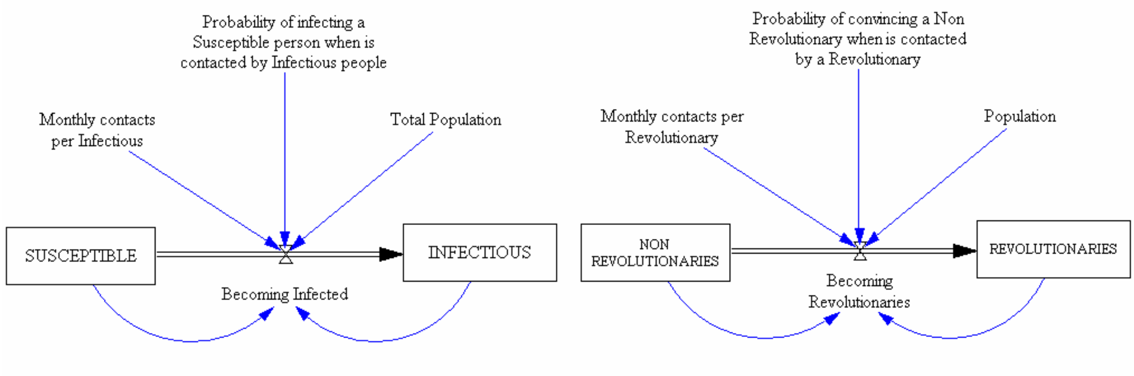


Figure 15. S&F of revolutions as diseases

Thus, the model, initially made of two stocks, one flow, two feedback loops, and three parameters was used to explain the dynamics of the French Revolution over time, considering its difficult beginning, its peak, its slow down, and its later stability. The understanding of the flow of people from being non revolutionaries to become revolutionaries is highly decisive of the dynamics associated to revolutions. This same structure gives rise to the s-shaped behavior of a disease. Enhancing the understanding and reality of the model is intended to be done by adding in the model a stock of dead revolutionaries, which die at a certain rate over time. The assumption is that certain amount of revolutionaries dies as a result of the violence of revolutions. The third stock accumulates the number of dead revolutionaries and is fed by a negative feedback loop, the deaths loop. Revolutionaries dying in the French revolution increase such stock and decrease the Revolutionaries stock. The greater the number of revolutionaries is, the greater the death rate is, and the smaller the number of remaining alive revolutionaries is (Sterman, 2000). Figure 16 shows the full S&F diagram, in which the dead revolutionaries are considered.

Because of the effect of the deaths loop, the stock of Revolutionaries after the dominance of the depletion loop is decreased. At this point, the number of revolutionaries dying per month is greater than the number of non-revolutionaries becoming revolutionaries. Thus, the characteristic s-shape growth in the simple SI model collapses when the dead revolutionaries are taken into account. Figure 17 presents a representative behavior of the stock of revolutionaries (*infectious*) with the SI model considering deaths (*recoveries*).

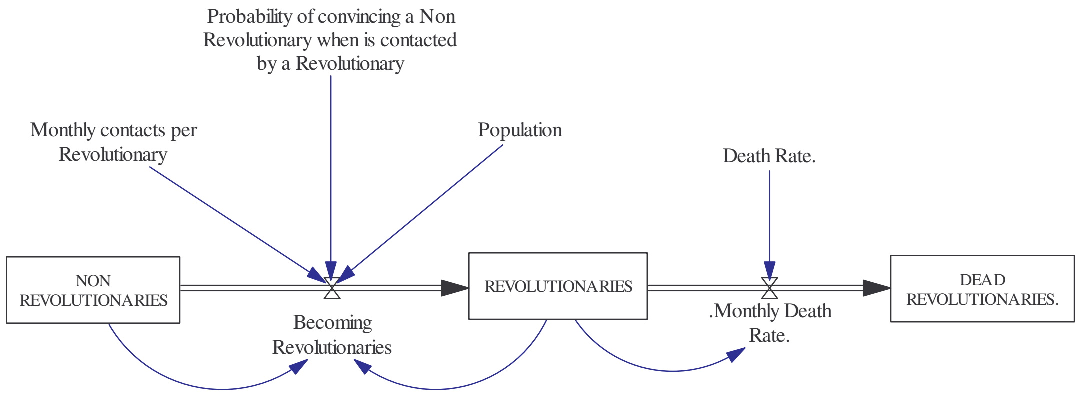


Figure 16. S&F Revolutions

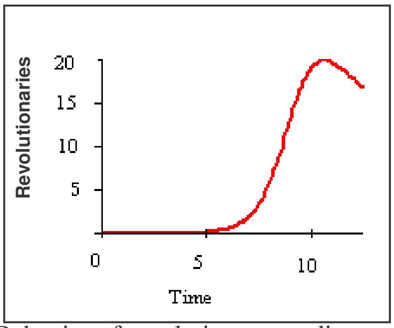


Figure 17. Behavior of revolutions according to the SIR model

3.4 Test Instrument

As history teaching has been oriented to the memorization rather than to the understanding of relationships that made history to occur in a certain way, there are no previous elaborated tests useful to assess students' performance regarding revolutions and their dynamics. Thus, the test used to measure students' performance has been self designed by the author. Rather than recalling knowledge regarding the French revolution, the test intended to measure the general understanding of revolutions, based on their dynamics (change over time and the relationships that give rise to that change) and the transferability of such understanding to other happenings. Transfer tests require that the learner applies what was learned to a novel situation. The characteristic that distinguishes someone who learns by understanding from someone who learns by rote is the ability to transfer knowledge (Mayer, 1999). Thus, the understanding of the case study was required to be transferred along all the questions of the test, due to the questions encouraged the students to apply the understanding of the relationships underlying and explaining the dynamics of revolutions.

The test was made of 14 multiple choice questions which are presented in Table 3. The questions were classified in three types according to their main measurement purpose[12].

The transferability of the structure underlying the development of the revolutionary population among the French revolution was one of the tasks that students should face most often in the test. These questions were presented in the form of analogies between the case study and other historical happenings, and in the form of questions asking for correct explanations of why certain happenings occurred during revolutions. Questions number 1, 5, 6, 7, and 8 belong to this type of question. This type of question was relevant because it tested the understanding of different relationships between the elements that caused a revolutionary population to develop in a certain way. Furthermore, these questions reveal whether a student is able to associate the understanding of the case study to another topic that may come from similar structural patterns.

A second set of questions had the purpose of measuring the understanding of the behavior of revolutions over time and the application to such knowledge to new

[12] Question number 2 was not included within the groups because its objective was not intended to measure knowledge nor learning process but to test students' ability to interpret graphs in the pre test because the instructional methods were based on several graphs.

situations. This type of question was presented in the form of analogies with other fields of study and of graphs recalling possible development over time of the populations of the revolutions. The importance of this set of questions is the possibility of measuring whether or not a student is able to associate the general behavior of revolutions with another situation that may unfold in the same way than revolutions over time. Questions number 3 and 4 belong to this group.

The last questions intended to measure the application of the understanding not only of the structure underlying the French revolution development, but also the understanding of the behavior over time of the populations involved in such revolution. These questions are the "what would have happened if…" type, referring to forecasting and thinking about hypothetical scenarios based on the understanding of the structure and the behaviors underlying the case study. Questions 9-14 belong to this type of questions.

All set of questions required different cognitive skills. Questions in the groups of behavior and structure may be easier answered by students than what the last group of questions can be. In the structure based questions the main requirement had to do with identifying constituent parts and functions of a concept. In the behavior questions it was demanded the identification and relation between parts and function of a process. While the last group of questions had to do with making an assessment of elements, relationships, values and effects. The latter process shows a more difficult process for the students to perform correctly.

Even though the aim of each question is to measure different levels of understanding of revolutions, there is still chance for the students to answer them properly without having the expected understanding required for the question.

The test was applied as pre test and post test, thus results due to the treatments' application may be compared.

3.5 Measures of Performance

Both treatments, the teaching method and the experimental experiences, represent the independent variables of this experiment in that they are controlled by the experimenter. The application of these variables determined the students' performance, in other words, the dependent variables. The dependent variables are the response variables, which are given by the performance of the students in both tests after receiving the different treatments.

Assessment of students' performance has been done by two different measures: Students who improved per group (SWI), and Effective Improvement (EI). The first measure talks about how wide the effect of the treatment was on the students and excludes the amount of improvement per student. The second measure shows how much improvement of the maximum gain[13] based on the pre test scores is reached by the students. In this measure, not only the amount of improvement is considered per student, but also such improvement is contrasted with student's learning potential. The benchmark to be reached by the students is an EI that equals the learning potential.

[13] Maximum possible gain is defined as Learning Potential = (100-Correct Answers % in the Pre test)

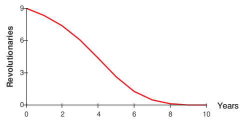
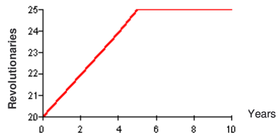
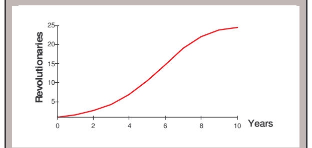
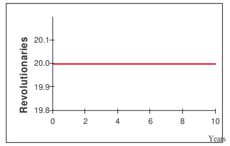
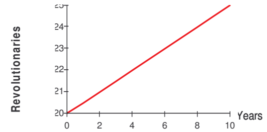
Question	a.	b.	c.	d.
1. What makes a Revolution to gain strength in a group of people?	The increasing support of a group of people	The meetings offered by Revolutionaries	The prohibition by the state of not letting anyone to become part of the Revolution	b) and c)
2. Take a look to the following graph:  Which one of the following statements describes what is shown by the graph?	The number of Non Revolucionaries does not decrease significantly in the early years	Once the Non Revolutionaries start to decrease, the depletion of the group occurs very fast until it stops decreasing and remain in a fix level	The group of Non Revolutionaries does not decrease	a) and b)
3. Which of the following graphs describes best the number of Revolutionaries supporting the Revolution along its duration?				
4. Which of the following sentences is more similar to the development of a Revolution over time?	To the speed used by a rocket to fly to the space	The interest accumulation in bank account	The spread of a recent news	a) and b)
5. Which pair of elements presented at each possible answer (a. to d.) is more similar to the relationship of the people joining the revolution and the number of people who don't support it?	The relationship between water flowing out from the drain of a bathtub and level of water in the bathtub	The relationship between the juice filling a glass and the level of juice in the glass	The relationship between the people who go to the stadium and the soccer players in the team	a) and c)
6. The relationship between "the people joining annually the Revolution" and "those who are part of the revolutionary group" is similar to:	The relationship between the central American towns being conquered annually by the Aztecs and the total of towns in Aztec Empire's care	The relationship between the peasant people who migrated to the cities after the Industrial Revolution and the people who lived in the countryside before the Industrial Revolution	The relationship between the people who annually were killed in the guillotine during the French Revolution and the number of people who supported the French Revolution	None of them

Table 3. Test Instrument

7. The relationship between “the possibility of convincing people of joining the Revolution” and “the number of people annually joining the Revolution” is similar to:	The relationship between the probability of being a country liberated by Simon Bolivar and average lifetime of Simon Bolivar	The relationship between the probability of believing in the ideas of the reason supported by the Enlightment Movement and the number of people supporting the Enlightment annually between 1700-1789	The relationship between the number of the Spanish conquistadores in America and the number of native people who died during the conquest of America	None of them
8. The relationship between “the number of Non revolutionaries in a revolution” and the “number of revolutionaries in a revolution” is similar to:	The relationship between people infected with HIV and the babies that annually are born with HIV	The relationship between people who are not aware of the World's news and People who are aware of the World's news	The relationship between people purchasing new products in the market and the quality of the new products	b) and c)
9. Which of the following strategies might be applied by a Governor in order to stop a revolution and thus, to avoid the negative effects of it:	To keep the people satisfy	To isolate the Revolutionaries from the Non Revolutionaries	To declare the Revolution as illegal	a) and b)
10. What would happen if a Governor, with the attempt of stopping a Revolution, tries to arrest all those who look suspicious of making part of the revolution?	The revolutionaries would reject those who might want to belong to the revolutionary party.	All those who do not support the Revolution yet would follow the rules of the Governor	The development of the Revolution would be slower along the time	a) and b)
11. Which of the following strategies could be applied by a Governor in order to delay the development of a Revolution?	To use advertisements in order to highlight the bad and cruel actions done by the revolutionaries	To keep quiet about the actions of the Revolution for not altering it	To kill some of the revolutionaries	b) and c)
In order to answer the following questions, please read the coming paragraph: If the Governor of a country, which faces a Revolution, decides to grant one time some rights and privileges as citizens to those who do not want to join the Revolution, how would it affect the following items?				
12. The number of revolutionaries supporting the Revolution:	Less people would be willing to join the Revolution	People would take longer to decide joining the Revolution	The Revolutionary party is not radically affected by such policy	a) and b)
13. The number of dead bodies during the Revolution	Every year, there would be less dead bodies than the previous year.	The number of dead bodies would be the same	The deaths would be less that otherwise would have been	d. a) and b)
14. The development of the Revolution along the time would be:	Slower	Faster	The same	Does not have anything to do with the policy

Table 3. Test Instrument

Measure	Definition
Students who improved (SWI)	Total Students per group with improved scores in the Post Test / Total number of students per group
Effective Improvement (EI)	-If Total Score Post Test > Total Score Pre Test: (Correct Answers% Post Test-Correct Answers % Pre Test) / (100- Correct Answers % in the Pre test) -If Total Score Post Test < Total Score Pre Test: (Correct Answers% Post Test-Correct Answers % Pre Test) / Correct Answers % Pre test

Table 4. Measures used to calculate improvement in the performance of students

These two measures of performance have been used in both the full test and the groups of questions defined in the previous section. Measuring students' improvement in the full test leads to have an overview of their general performance, while knowing how their performance was along the groups of questions reveals were such improvement is located. Strengths of SD and CONTROL groups in the understanding of revolutions can be identified through the analysis per group of questions.

The benchmark in the performance per group of questions is given by consistent improvement in the three groups. When students improved only in one group of questions, the performance is not considered as a better one. However, if they show better performance in all groups of questions their improvement turns consistent and reveals that there was less likelihood of answering the questions by guessing. Thus, knowing the importance of the questions, in which students improved, provides a complementary tool to determine whether or not SD groups had better performance than the CONTROL groups.

Even though all questions demand certain knowledge of SD, it is straightforward to all students to answer correctly questions 5-7, 1, 8, and 3 after either the CONTROL or SD instructional methods were presented, though SD groups are more skilled to answer them more correctly. However, questions 4 and 9-14 are more demanding even to the students who went through the SD treatments due to understanding of dynamics of revolutions is required.

Besides the independent variables, some uncontrollable and immeasurable ones influenced as well students' performance during the experiment. The main nuisances were the finalization of the scholar year at the same time than the experiments' application, and the cognitive training some students had during the first semester of 2006 to pass the final examination for graduating. Regarding the first one, students from all grades in the school were experiencing high load of tasks and exams to be done in order to finish the year properly. Some students had more exams than others because they did not present them on time. Therefore, to diminish the effect of this nuisance, the students for the experiment were picked from the bunch of students who did not have such great pressure and stress as those who had exams left. This helped to have on average the same level of stress and cognitive load on the students. Regarding the latter nuisance, not much was to be done to diminish its effect, due to absence of chances given by the school to mix those students, who received training for specific exams, with the rest of the students who did not receive any training of this kind.

Emotional and social differences in students had a diffusion effect (nuisance) on their performance in the experiment, which unfortunately were not susceptible to be controlled somehow by the design itself.

3.6 Hypotheses

The following hypotheses are intended to test whether SD is a better method to enhance understanding of history as a disciplinary approach. Both perspectives from which the research question is approached, students with and without previous SD knowledge, will be tested by making use of such hypotheses as well. Hence, there are two sub-sets of hypotheses, which will use the measures of performance to be tested and give answer to the research question.

Hypotheses 1: *Regarding groups with different teaching method*[14] *and the same experimental experience*[15]

SIW Null Hypothesis: There is no significant difference in the SWI of groups who were taught with different teaching method but attended the same experimental experience.

$$H_{0,\ SWI}: SWI_{\mathbf{SD1}} = SWI_{\mathbf{CONTROL1}}$$

$$H_{0,SWI}: SWI_{\mathbf{SD2}} = SWI_{\mathbf{CONTROL2}}$$

EI Null Hypothesis: There is no significant difference in the EI of groups who were taught with different teaching method but attended the same experimental experience.

$$H_{0,\ EI}: EI_{\mathbf{SD1}} = EI_{\mathbf{CONTROL1}}$$

$$H_{0,EI}: EI_{\mathbf{SD2}} = EI_{\mathbf{CONTROL2}}$$

Hypothesis 2: *Regarding groups with the same teaching method and different experimental experience*

SIW Null Hypothesis: There is no significant difference in the SWI of groups who took the same teaching method but attended different experimental experience.

$$H_{0,\ SWI}: SWI_{\mathbf{SD1}} = SWI_{\mathbf{SD2}}$$

$$H_{0,SWI}: SWI_{\mathbf{CONTROL1}} = SWI_{\mathbf{CONTROL2}}$$

EI Hypothesis: There is no significant difference in the EI of the students who took the same teaching method but attended different experimental experiences.

$$H_{0,\ EI}: EI_{\mathbf{SD1}} = EI_{\mathbf{SD2}}$$

$$H_{0,EI}: EI_{\mathbf{CONTROL1}} = EI_{\mathbf{CONTROL2}}$$

[14] Either the Conventional or SD teaching method
[15] Either the First Experimental Experience or the Second Experimental Experience

All hypotheses will be tested using a two-tailed test with a level of significance equal to 0.05.

3.7 Sample Selection Process and characteristics of the sample

The experiments were run in the Fray Rafael de la Serna School in the city of Medellin, Colombia, during the first two weeks of November, 2006. In Colombia the scholar year goes from late January until late November. Thus, the time of the year in which the experiments were done was crucial in determining the availability of the students who could be subjects under study. During the last month of school, all students are required to take final exams, and for those students, who did not perform excellent along the year, several assignments are required to be handed in as well. Therefore, students who did well the whole year were under less pressure than those who had to repeat some tasks in which they did not do well.

According to this, the Principal of the school decided to provide the students required for the experiment, from those students who only had to present exams and few assignments left. Students made part of 8th, 9th, 10th and 11th grades[16], whose age ranged between 15 and 17 years. Thus, 30 students were assigned to each group. Table 5 describes the characteristics of the sample size per group.

Characteristics	SD1	CONTROL1	SD2	CONTROL2
Sample Size	30	30	30	30
Students in Grades	-8 Students from 9th grade -22 Students from 10th grade	-10 Students from 8th grade -20 Students from 11th grade	-5 Students from 9th grade -25 Students from 10th grade	-4 Students from 8th grade -26 Students from 11th grade
Students' Age	15-17 years old	15-17 years old	15-17 years old	15-17 years old
Female Population	0%	13%	0%	13%
Students repeating the grade	7%	3%	3%	0%
Day Time for the experiment	Morning	Afternoon	Morning	Afternoon
Grades in Social Sciences	Outstanding	Outstanding	Outstanding	Outstanding

Table 5. Categories according to the type of question

Since it was not possible to elaborate a full random process to select the students, one of the factors that helped determining certain homogeneity in the initial conditions of all groups was the grades of the students in the field of Social Sciences and History. The grades of the students per group were averaged out and then compared. All four groups had an average grade of "O-Outstanding (above average)"[17] in such subject, thus, despite all the differences between groups, they shared certain homogeneity and equal initial conditions to perform the experiment.

[16] In the school each grade had two different groups: the 11th grade had group 1 and group 2. Likewise for the 10th, 9th, and 8th grade.
[17] All four groups presented in average a grade of "Outstanding (above average)" among a qualitative scale in which the highest grade is E= excellent, followed by O= Outstanding, A= Acceptable, D= deficient, and I= insufficient.

Another important factor to be mentioned is the discipline and attitude of the students towards learning activities. Three of the four groups presented similar behavior regarding the attention paid and the attitude to perform those scholar demanding tasks. SD1, CONTROL1 and CONTROL2 groups were undisciplined and disobedient, noisy and few attempting to the scholar activities in general. Those students are used to receive ticking off by the teachers of the school, who are the same used to tell them off often. In contrast, the discipline and attitude of the SD2 group is rather positive oriented to new scholar tasks. They get easily concentrated and motivated by new things to do, thus, teachers do not give ticking off regularly. Thus, since most of the students were undisciplined, a teacher was accompanying each experimental session, in order to establish authority and order in the classroom. However, the teachers were not allowed to interfere at any stage of the experiment.

As it was stated previously, the students of the 11th grade who participated in the CONTROL groups were finishing the last year of school, in which all students must take official exams measuring their knowledge gained during the 11 grades of school life. This exam is relevant in the sense that it is the first admission criterion that Colombian universities consider for conceding places for the students and so, it is enough reason for the students to get training in such exam along the previous months to the exam in order to perform well at it. Both 11th grades groups had special training to perform well in such exam and to review the most important issues they were taught in the school. However, most of the questions of this test are the rote knowledge type, in which the knowledge is required to be repeated rather than to be understood and to be applied. Hence, these students gained some training to learn how to resolve such questions and how to recall the knowledge. Therefore, this special training is expected to help these students perform well at the lowest type of question in the test. However, this training is not expected to have influence on the performance of these students on the highest type of question, mainly because this test is the transfer knowledge type and its demands result much less susceptible of being overcome by training in answering certain types of questions rather than by understanding the subject itself.

The experimental design presented in this section shapes students' performance, thus, results are presented as follows.

4. Results

4.1 Students who improved -SWI-

The SD1 group had better performance than the CONTROL1 group based on the SWI measure of performance for the full test. As table 6 shows, 53% of the students in the SD1 group had improved scores, compared to the 43% of students in the CONTROL1 group. For those groups in the second experimental experience, the differences are even greater: 60% of the students in the SD2 group had improved scores, compared to the 33% of students in the CONTROL2 groups who improved their performance. From the groups who were taught with SD and assisted to different experimental experiences, the SD2 group had more students with improved scores.

	SD Method	Standard Method	P-values
First Experimental Experience	**SD1** SWI=**53%**	**CONTROL1** SWI=**43%**	P_{value}=0.447
Second Experimental Experience	**SD2** SWI=**60%**	**CONTROL2** SWI=**33%**	P_{value}=0.039
P-values	P_{value}=0.610	-	-

Table 6. Students who Improved –SWI-

Two-tailed t-tests have been run to test the statistically significant differences of the hypotheses concerning SWI. A significance level of 0.05 was used to test the hypotheses. For the groups in the first experimental experience, the better performance of the SD1 groups was not statistically. However, the better performance of the SD2 was statistically different. Acceptance or rejection of the SWI null hypotheses depends on whether the group had additional SD training. After additional training, the differences were large enough to justify rejection of the null hypotheses.[18]

4.2 Effective Improvement -EI-

Table 7 shows results of the full test according to the EI measure of performance. While the CONTROL1 improved more than the SD1 group in the case of the first experimental experience groups, the SD2 greatly outperformed the CONTROL2 group in the case of the second experimental experience groups. However, t-tests showed that there is no statistically significant difference at the 0.005 level in the performance of groups who were taught with different teaching methods and assisted to the same experimental experience.

In regard of those students taught with the same teaching method during different experimental experiences, clearly, the SD2 group outperformed the SD1 group. However, statistically significant difference in their performance has not been revealed by the t-test at the significance level of 0.05[19].

	SD Method	Standard Method	P-values
First Experimental Experience	**SD1** EI=**1.15%**	**CONTROL1** EI=**4.23%**	P_{value}=0.707
Second Experimental Experience	**SD2** EI=**10.96%**	**CONTROL2** EI=**-0.92%**	P_{value}=0.114
P-values	P_{value}=0.173	-	-

Table 7. Effective Improvement –EI-

[18] Since ANOVA provides considerably more flexibility in testing group differences when there are more than two groups to be compared and more than one independent variable (treatment) affecting de dependent one, an ANOVA has been applied to test the statistical difference between all groups. The results agree with the t-tests' results: There is no statistically significant difference in the SWI of all groups. The P value for the ANOVA test has been 0.182 and the F statistic is 1.65.

[19] An ANOVA has been applied to test the statistical difference in the EI of all four groups. Results show that there is no significant difference between them. The Pvalue is 0.425 and the F statistic is 0.938. Thus, this confirms what was found with the t-tests.

An often used measure of performance is the Absolute Gain (AG)[20], which indicates how much students improved. Results with the AG for the four groups of students in this experiment are in the same direction than the results with the EI[21]. However, the EI provides a more trustworthy judgment of students' performance because the number of improved questions is compared to the size of their learning potential. In addition, the AG excludes the improvement of many students because of the ceiling effects. Thus, in this assessment the EI is chosen as the measure of how much improvement the students got.

4.3 Comparisons per groups of questions

The measures, SWI and EI, are utilized to test students' performance per groups of questions as defined in section 3. The results support the main conclusions given from the EI and SWI in the full test. SD groups tend to outperform CONTROL groups, when more training with the SD method is given to students. Though, statistically significant difference is only found between groups in the second experimental experience.

Structure based questions revealed that more students in the SD1 group got –in average- more strength over the CONTROL1 group in understanding the structure underlying the diffusion process of a revolution. In contrast, CONTROL1 group demonstrated to have more students with great strength in the groups of questions of Behavior, and Policies and Forecasting. In regard to the EI, the CONTROL1 group is more outstanding in all groups of questions. Then, CONTROL1 students seemed to understand the behavior and the structures supporting dynamics of revolutions, so that, they could get insights about what kind of policies they should apply when desiring to change the course of a revolution over time. However, statistically significant difference was not found in the performance of these groups with none of the measures of performance.

On the other hand, the SD2 group had the greatest amount of students with improvement (from all four groups) in the structure and behavior types of questions, which represent, at some extent, deep understanding of dynamics of revolutions and its transferability to other historical cases. Likewise, the EI for the SD2 group is once again the highest. In addition, based on the SWI the behavior questions were statistically better answered by the SD2 group, revealing an important difference in the performance of CONTROL and SD groups in the second experimental experience.

Some differences are present among groups with the same teaching method. With both the SWI and the EI, the SD2 group exhibits outstanding performance over the SD1 group in the structure and behavior types of questions. Statistically significant difference with the EI is found between the performances of both SD groups in the behavior questions. Thus, students from the SD2 group improved in average more than what the SD1 did in this type of question. In contrast, though a lack of statistically significant difference in the performance of both groups for the questions concerning

[20] The Absolute Gain (AG) is defined as = (Correct Answers % Post test – Correct Answers % Pre Test)

[21] In the case of the groups in the first experimental experience, the CONTROL1 outperforms the SD1, while the SD2 group greatly outperforms the CONTROL2 group in the case of the second experimental experience. The AG for the SD1 group is 1.7%, for the CONTROL1 is 5%, for the SD2 is 6.2%, and for CONTROL2 is -0.92. No statistically significant difference was found with the t-tests at the level of significance of 0.05.

Policies and Forecasting, the SD1 performance was better than the performance of the SD2 group for both measures SWI and EI.

As follows the results per groups of questions are presented.

4.3.1 Structured based questions

Based on the SWI criterion, the SD groups outperform the CONTROL groups for the structure based questions. The SD1 group had 24% of students with improved scores, while CONTROL1 group had 19%. Though, statistically significant difference in the SWI of these groups is not found, the P-value of the t-test was closer to the significance level of 0.05, revealing an important difference in the performance of both groups. In the case of the groups in the second experimental experience, their performances tend to be more homogeneous. The SD2 group had 25% of the students with improved scores, while the CONTROL2 had 21%. Statistically significance is not found with the t-test applied. Likewise, the performance of both SD groups is rather similar in this group of questions, which responds to the expectations given the SD treatments to both groups.

In the case of the EI measure of performance, the groups in the first experimental experience improved in average the same in the structure based questions. Thus, significant difference is not found with the t-test applied. In the case of the groups in the second experimental experience, the SD2 improved in average more than the CONTROL2 group, though statistically significant difference is not found after applying the t-test. Similarly, for both groups taught with the same teaching method, the improvement is much higher in the case of the SD2 group, while there is not statistically significant difference in their performance.

Structure	SD Method	Standard Method	P-values
First Experimental Experience	**SD1** SWI=24% EI=6%	**CONTROL1** SWI=19% EI=7%	**SWI** P_{value}=0.078 **EI** P_{value}=0.839
Second Experimental Experience	**SD2** SWI=25% EI=15%	**CONTROL2** SWI=21% EI=7%	**SWI** P_{value}=1 **EI** P_{value}=0.266
P-values	**SWI** P_{value}=0.647 **EI** P_{value}=0.164	-	-

Table 8. Performance in Structure based questions

4.3.2 Behavior based questions

The SWI and EI reveal wider differences in the performance of the groups in the behavior based questions. In the case of the groups in the first experimental experience, CONTROL1 outperforms SD1 with both measures of performance. However, even though, the SD1 group presented a deterioration of its performance with the EI, there was not statistically significant difference in the performance of both groups based on t-tests, nor with the EI neither with the SWI. In the case of the groups in the second experimental experience, the difference in the performance between SD2 and CONTROL2 groups is wide enough to cause statistically significant difference based on the SWI measure. Thus, students in the SD2 group tended to answer more properly those questions based on the understanding of the behavior of revolutions. In the case of the EI, the difference in the performance of both groups is still wide and the SD2 reveal greater improvement than the CONTROL2. However, statistically significant difference is not found.

Both SD groups presented huge differences in the performance of the behavior based questions. SD2 group appears to have more students improving their performance (SWI) and gaining greater improvement in the post test (EI). The difference with the EI measure of performance is higher, which leads to have statistically significant difference between both groups.

Behavior	SD Method	Standard Method	P-values
First Experimental Experience	**SD1** SWI=20% EI=-3.3%	**CONTROL1** SWI=27% EI=8.3%	**SWI** P_{value}=0.305 **EI** P_{value}=0.335
Second Experimental Experience	**SD2** SWI=35% EI=23%	**CONTROL2** SWI=18% EI=10%	**SWI** P_{value}=0.035 **EI** P_{value}=0.281
P-values	**SWI** P_{value}=0.201 **EI** P_{value}=0.046	-	-

Table 9. Performance in Behavior based questions

4.3.3 Policies and Forecasting questions

Students from SD and CONTROL groups had almost the same performance with both measures of performances in this group of questions. In the case of the groups in the first experimental experience, CONTROL1 group outperformed SD1 group with both the SWI and the EI. However, there is not statistically significant difference in their performance after applying a t-test. In the case of the second experimental experience, the SD2 group had more students improving their performance than what the CONTROL2 had. However, the difference is not statistically significant. Likewise, the SD2 group presented fewer deterioration of its performance with the EI than the CONTROL2 group, though there is no statistically significant difference in their performance with the EI measure.

For both SD groups, the performance is in average the same for the SWI, though the SD1 outperformed the SD2 for the first time. While SD1 had 17% of students with improved scores, the SD2 group had 14%. However, with the EI measure, the SD2 shows a deterioration of its performance, while the SD1 group improved 2.2%. Statistically significant difference in the performance of both groups is not found with any of the measures of performance.

Policies and Forecasting	SD Method	Standard Method	P-values
First Experimental Experience	**SD1** SWI=17% EI=2.2%	**CONTROL1** SWI=22% EI=3.9%	**SWI** P_{value}=0.561 **EI** P_{value}=0.808
Second Experimental Experience	**SD2** SWI=14% EI=-2.8%	**CONTROL2** SWI=12% EI=-5.6%	**SWI** P_{value}=1 **EI** P_{value}=0.671
P-values	**SWI** P_{value}=1 **EI** P_{value}=0.406	-	-

Table 10. Performance in Policies and Forecasting questions

4.4 Controlling for Other Influences

Results given by the SWI and the EI in the full test and in the groups of questions suggest that the SD teaching method enhances the understanding of history if students have more training in that method. Using the SWI measure of performance, for example, the SD2 group outperformed more convincingly the CONTROL2 group than the SD1 outperformed the CONTROL1 group. For the EI measure, the SD1 group did not outperform its CONTROL group, while the SD2 did. Furthermore, the SD2 outscored the SD1 group on both measures of performance. While these results are suggestive, the results of the statistical significance t-tests do not provide strong confidence.

Factors other than the teaching method applied to students may have influenced students' performance, and may have interfered in the perception of the absolute impact of the teaching method on the groups' performance.

The conditions and circumstances of the student selection and group assignment process suggest that that process was not random. The absence of a random process in students' assignation to each group represents an important issue affecting students' performance, mainly because guarantee of the same initial conditions for all groups was not given. Then, students may have had different capabilities before the experiment was carried out. The pre test applied to students one week in advance of the experiment was intended to measure students' initial capabilities, which were expected to be homogeneous along all groups. However, all groups showed different performance in the pre test, and with it, they revealed different initial capabilities to run the experiment[22]. Thus, an important factor to be considered in students' performance is given by the prior capabilities that students had before the experiment was carried out.

Hence, it would be expected a positive correlation between scores in the pre test and in the post test and that is seen in the data with a correlation coefficient equal to 0.44 for such relationship. With respect to the measures of performance, SWI and EI, a negative correlation with the pre test is expected because the pre test scores are subtracted to obtain both measures. Again, the data support that assumption with correlations coefficients of -0.51 between the SWI and the pre test scores, and of -0.37 between the

[22] The SD1 group answered correctly 45% of the questions in the pre test, the CONTROL1 group answered 41% of them, SD2 51%, and CONTROL2 48%.

EI and the pre test scores. It is necessary, therefore, to try to control statistically what was not controlled effectively by the group assignment process. Therefore, a multiple regression analysis was performed. This analysis includes not only the teaching method but also the students' prior capabilities as explanatory variables of the performance. Controlling other effects on students' performance different from the teaching method is the main attempt of the multiple regression analysis.

In the following equation, Y represents students' performance and is a function of the teaching method and the pre test scores (given in %). Coefficients b and c indicate the relationship between students' performance and each explanatory variable. The coefficient "a" represents the intercept or constant and the errors are also considered in the model.

Y= a +b*Teaching Method + c*Pre Test Scores + error

The inclusion of the pre test scores as an explanatory variable of students' performance leads to ensure that any given effect on the performance is not due to some effect of the prior capabilities but due to the teaching method. In other words, the linear regression helps hold the pre test scores constant, while the teaching method influences the performance according to the SD exposure received by students. Thus, the teaching method per student is defined as a number referring to the number of days of SD training received during the experiment. Thus, students in the SD1 group received a number 1 in the method variable due to one day of instruction based on SD, students in the SD2 group received a number 2 due to two days of instruction based on SD, and students in both CONTROL groups received a zero because they had no instruction based on SD[23].

All 120 students participating in this experiment are considered as a full group for the multiple regression analysis.

The performance is given by the two measures of performance previously used, the SWI and the EI. Thus, there are two multiple regressions to run:

SWI= a +b*Teaching Method + c*Pre Test Scores + error (1)

EI= a +b*Teaching Method + c*Pre Test Scores + error (2)

After applying the multiple regression analysis the values for the coefficients are given. In the case of the first (1) equation, the coefficients determine the equation as follows:

SWI= 1.111 +0.159*Teaching Method -0.016*Pre Test Scores + error (1a)

---

[23] The "number" representing the "amount" of SD training per group is certainly arbitrary. However, those numbers have been assigned in an effort to draw general conclusions about the impact of the teaching method on students' performance. This numbering scheme implies a linear relationship between the number of days of SD training and the effect of such training, while in fact the effect is probably nonlinear and increases at a decreasing rate as more and more days of training are received.

The hypotheses regarding having coefficients without impact on students' performance (coefficients equal to zero) are rejected with a Pvalue of 0.001 for the method and with a Pvalue of 0.000 for the pre test scores. As expected, the exposure to SD has a positive correlation with the SWI meaning that the more exposure to the SD teaching method, the higher SWI. Similarly, the pre test scores were expected to have a negative correlation with the SWI, because the more students did well in the pre test, the less reflected is their improvement in the SWI (measure based on the difference between post and pre tests scores).

Likewise, in the case of the second equation (2), the coefficients are certainly different from zero. Furthermore, the hypotheses regarding coefficients equal to zero are rejected with statistical difference in the case of the teaching method coefficients (Pvalue equal to 0.041), and the Pre test scores' coefficient (Pvalue equal to 0.000).

EI= 31.561 +6.208*Teaching Method -0.704*Pre Test Scores + error (1a)

Similarly to the SWI case, the correlation between the teaching method and the EI was positive as expected, indicating that the more exposure to the SD teaching method, the greater the improvement gained by students. Accordingly, there is no wonder in the negative correlation between the pre test and the EI. The higher the scores in the pre test, the less improvement the students can reveal through the EI (considering the nature of the equation, in which the nominator subtract the pre test).

Through the multiple regressions, the effect of the method on students' performance could be isolated when taking into consideration, simultaneously, the effect of the students' prior capabilities. Both regressions reveal, then, that the method of teaching history matters and that any observed effect of the teaching method on the performance is not due to the students' prior capabilities but because of the method itself. The more exposure the students have to the SD method, the higher is their EI and the more students improved their performance. Statistical difference in the impact of the explanatory variable "teaching method", for both measures SWI and the EI, evidences that certainly the SD has an important effect on students' performance[24].

These results seem to be encouraging. However, there is a potential methodological problem in the fact that the pre test scores are being subtracted in both sides of equations 1a and 2a. Given the nature of the SWI and the EI equations, the results of the pre test are subtracted from the post test scores to calculate the improvement given by students. Thus, a more trustworthy way of calculating the impact of the teaching method on students' performance must be based on the post test scores (%) as the performance variable (dependent variable), which is predicted by the teaching method and the pre test scores. The linear equation must be then:

Post Test Scores (%) = a +b*Teaching Method + c*Pre Test Scores + error (3)

This way of approaching the impact of the teaching method on students' performance provides more conservatism and confidence on the results given by this test. Any effect

[24] In addition, the assumptions regarding normality of the residuals and dispersion of the residuals according to the teaching method are fulfilled with the linear regression for both measures of performance (SWI and EI).

of the teaching method on students' performance will be more reliable given the fact that the pre test scores effect are not affecting both sides of the equation.

The results do not go far away from the conclusions provided in the two previous regressions. The coefficient for the pre test scores predicts statistically the post test scores of students. However, the level of statistical confidence resulting from this more conservative approach is at the 0.06 level, revealing a less strong effect of the teaching method on students' performance than what was revealed with equations 1a and 2a. Nonetheless, the Pvalue for rejecting the hypothesis is rather close to the significance level of 0.05, which shows results almost as strong as those mentioned before. Thus, if the sample size were larger, the power of the effect of the teaching method on students' improvement will be stronger, and the level of statistical confidence might meet that customary target level. The effect of both explanatory variables on students' performance is positive, indicating as expected that the higher the scores in the pre test, the higher the scores in the post test. Likewise, the greater the exposure to the SD teaching method, the higher the score in the post test.

Important findings result from this analysis, in which the teaching method seems to be still a strong explanation of students' performance, when keeping track of the effect of the students' prior capabilities on the their performance. Even though the implementation of the experiment may not have adequately randomized the effects of students' prior capabilities, the multiple regression helps to control that influence. Thus, it was not possible to build randomly the groups for this experiment, it can be concluded that when controlling the initial conditions regarding the prior capabilities of students, the effect of the SD teaching method on students' performance is positive and enhances students' understanding of revolutions[25].

5. Discussion

Based on the SWI for the full test, the SD groups had more students improving in the post test than the CONTROL groups. However, while the CONTROL1 group outperforms the SD1 group, the SD2 is the group with the greatest EI not only from the groups in the second experimental experience but also from all groups. Statistically significant difference between the performances of the groups has been found only between SD2 and CONTROL2 based on the SWI measure.

In the case of the groups of questions, SD1 group showed strength with the SWI in understanding the structure underlying the diffusion process of a revolution, and in using such understanding for interpreting other events that behave similarly. In contrast, the CONTROL1 group did better with the SWI in the questions based on behavior of revolutions and in the questions concerning policies and forecasting questions. In addition, the group did better based on the EI in all groups of questions. However, there is no statistical difference in the performance of both groups. The SD2 clearly outperformed the CONTROL2 group in all groups of questions and with both measures of performance (SWI and EI). Statistical difference in their performance is found with

[25] The assumptions regarding normality of the residuals and dispersion of the residuals according to the teaching method are fulfilled with the linear regression for the Post Test Score (%). This fact strengthen the reliability of the linear regression as test for measuring the impact of the teaching method on the performance of students.

the EI in the groups of questions based on behavior of revolutions. Similarly, the SD2 outperformed the SD1 group in the structure and behavior based questions with both measures of performance. In addition, questions regarding behavior had statistical difference in the EI of both SD groups.

The performance of the SD over the CONTROL groups differs depending on the measure of performance, which suggests some uncontrolled effects that may have caused the performance to differ. Thus, in an effort to control effects that were not controlled in the assignment process of the students to each group, the multiple regression analysis has been carried out in light of providing a stronger point of view to identify the role of SD in the learning process of history. In this case, the consideration of the pre test scores as an explanatory variable of the performance of the students helped isolating the effect of the teaching method on students' understanding of revolutions. Thus, the perception of the absolute effect of the method is more clearly understood. Results suggest insights in the same direction than those provided by the SWI and EI in both, the full and groups of questions. The method for teaching history matters and students' understanding of revolutions seems to be enhanced by the SD teaching method if students have more training in the method. The longer is the exposure of students to the SD teaching method, the better is their performance in the post test. In addition, this analysis also reveals the importance of considering the initial students' capabilities in light of interpreting the results. Thus, any observed effect of the teaching method on the performance of students is not given by students' prior capabilities but by the method itself. Despite absence of statistical significance in the relevance of the teaching method on students' performance, the nearness of the Pvalue builds confidence on the outcomes, especially when they strength the insights given already by the measures of performance SWI and EI in the full test and groups of questions.

Conclusions based only on either the SWI or the EI for the full test or in groups of questions do not explore totally the effectiveness of the SD teaching method on students' understanding of revolutions and their dynamics. Likewise, conclusions given from the multiple regression analysis provide more enhanced insights about the effectiveness of the teaching method when they are not analyzed in an isolated way. Thus, rather than focusing the attention on one single measure, it is more important to keep an overview of how many students improved per group, how much the group improved, on which groups of questions this improvement was concentrated, and how strong is the impact of the teaching method on students' performance. At the end, it will be possible to draw finally whether SD enhances the understanding of revolutions.

Certainly, SD2 group presented the greatest SWI. The performance of SD2 students', based on the SWI, is statistically different from its CONTROL group. In addition, the SD2 group improved more questions in the full test than what the rest of the groups did (high EI). SD2 students' improvement was focused consonantly on questions based on the understanding of the behavior of revolutions and the structure underlying such behavior. The performance of the groups in behavior questions is statistically different from the CONTROL2 group and the SD1 group also. Furthermore, the amount of SD training received by students seems to be a good explanatory variable of the students' performance. Thus, since students in the SD2 group received the largest amount of SD training in this experiment, their good performance may have been given by the effect of the method itself rather than by other effects such as students' prior capabilities.

Reliability on SD2 accurate performance, as a result of the effectiveness of the SD teaching method, is increased then by the suggestions provided by the multiple regression analysis. Thus, given the consistency of the results of all views of students' performance, it can be concluded that students in the SD2 group did not improve accidentally, and that SD may be considered as a proper tool for enhancing students' understanding of history, specifically of revolutions.

The SD1 group outperformed the number of students who improved in the CONTROL1 group, especially in the structure based questions. However, the CONTROL1 group outscored the SD1 group with the EI applied not only to the full test but also to the groups of questions. In accordance with the expected effects of the teaching method on both groups in the first experimental experience, it can be seen that performance of students in both groups seems not to be explained solely by the teaching method applied. As a first attempt to control other various effects, the students' prior capabilities have been considered as an important effect influencing students' performance and the impact of the teaching method may be analyzed more confidently. Statistical significance of the SD method is not found. However, since the Pvalue is still close to the significance level, it is possible to conclude that the amount of SD exposure received by students has certain effect on students' performance. Students without exposure to SD teaching method (CONTROL groups) are tended to have poorer performance than those who have had at least one day of SD based instruction. Nonetheless, the teaching method seems not to be a good explanatory variable for the outstanding performance of CONTROL1 group in comparison to the SD1. Thus, some other uncontrolled effects, different from the students' prior capabilities, must have influenced students' performance.

According to the treatments applied to each group of students, it was expected to find in the results of this experiment that those students, who were taught with the SD teaching method and participated in the second experimental experience, would have the best performance from all the groups in all measures of performance. Similarly but less significant, the performance of the students who were taught with SD during the first experimental experience was expected to follow the same direction than the performance of the SD2 group. None of the Conventional groups were expected to be outstanding. However, as seen before, the results differ at some extend from the expected ones, mainly in the groups in the first experimental experience. Factors that have influenced students' performance need to be discussed in light of the effectiveness of the SD method of teaching history. Factors such as the expected better performance of the SD1 group and the outstanding improvement of the CONTROL1 group over the SD1 cannot be overlooked.

There were factors, controlled within the experiment, that were expected to delimit groups' performance. The experiment design, the treatments, the instructional design were all factors that carefully defined the learning process and the understanding of the students about the revolutions over time. The research method and the experimental design were established in a way that students could show the outcome of being under certain type of treatments. Besides, content of the French revolution was provided to the students in the form of instructional methods and facilitated the chances for the students to show the skills acquired during the experiment given certain type of applied treatment. Furthermore, despite absence of a random students' assignment process to the groups, all four groups were statically homogenous and their average performance in the subject of social sciences was rather the same. At some extent, groups were under similar conditions to be guided with the treatments towards the expected results.

Students from the SD1 group were under the effect of the SD teaching method treatment and First Experimental Experience treatment. The introductory session to the SD basic concepts, the Infection game, the debriefing using the diffusion model, and the instruction about the French revolution from an SD point of view, were determinants of the students' scores. This group was taught about the concept of diffusion of a disease of a revolutionary population over time. S&F structures were the essential tools to place in a structured pattern the general understanding of revolutions and their dynamics over time. Furthermore, behavior over time was presented as a result of the interaction of the elements in the S&F diagrams, so that students were able to understand such behavior rather than merely recall it. However, even though SD1 students were taught about all mentioned above, the time for understanding this approach was few. Thus, this group was expected to have improved performance in the structure and behavior groups of questions. Some of these expectations were fulfilled. With the SWI, the SD1 group outperforms the CONTROL1 group in the structure questions.

SD students in the second experimental experience had as great strength the fact that they already were in touch with the SD tools, mainly with the Diffusion model and S&F diagrams, in order to understand the importance of Civics. These students studied the same diffusion structure from three different fields of study: infections, civics and French revolution. Therefore, they were more skilled than the rest of the students to improve their performance in the three groups of questions. Indeed, this group has fulfilled most of the expectations in the sense of being rather outstanding not only in the full test, but also in types of questions based on the structure and behavior of revolutions. In the last group of questions, the SD2 group was not the most outstanding group; however, its performance in such group was not bad, meaning that some of the students who improved in the other groups confirmed their understanding in the Policies and Forecasting group of questions.

In contrast, students from the CONTROL groups were taught about behavior of revolutions over time and the way the French revolution was unfolded. However, they were not taught about the way to understand such behavior and the reasons why the revolutionary population in the French revolution developed in the way it did. Thus, these groups were expected to answer properly the questions of the test in which recall of behavior of revolutions was required. In the case of the CONTROL2 group, this group did not even perform well in such questions. Nonetheless, it was not expected a greater performance of CONTROL1 group in comparison to the SD1 group. Certainly, CONTROL1 group outperformed the SD1 group in behavior and policies and forecasting questions.

Thus, besides the treatment effects, some more causes (unmeasured effects) may have influenced students' performance. Specifically, other factors different from the teaching method may have, in some cases, undermined or amplify students' performance. Issues regarding implementation of the experiment (mainly duration), motivation, and cognitive loads could have undermined the possible performance of SD groups, especially the SD1 group. Furthermore, such effects may have caused the results of the CONTROL1 group to be better in categories in which it was not expected. Figure 23 shows the presence of the various effects on students' performance. It describes, as stated previously, that measured factors such as the teaching method and the students' prior capabilities have a positive influence on students' performance. The more training the students receive with SD, the more enhanced their understanding of revolutions is. Likewise, the higher the pre test scores of students, the better their performance in the post test. In addition, the figure shows that other unmeasured factors may have influenced students' performance but since those factors were not controlled by the experiment, their type of impact on the performance is not known numerically.

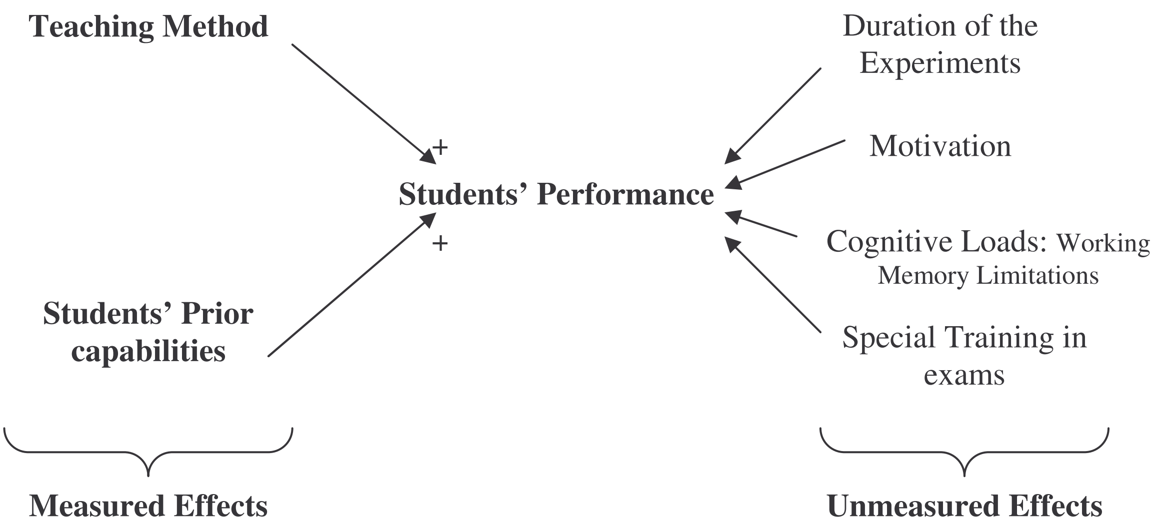


Figure 23. Influences on students' performance

Thus, as follows a description of those unmeasured effects is presented according to the way the intervened in the experiment, in an attempt to determine their role and influence on students' performance.

a. Duration of the experiment

The duration of the experiments' sessions is one of the factors that limited the SD groups' chance of maximizing the benefits from the SD teaching method. Specifically, 2 hour sessions are not enough to understand the concepts of SD. Even people who have received training in SD take time until they can understand how the process of accumulation occurs, how the feedback loops are given, and what an S&F diagram represents. Learning to think systemically and to understand changes over time implies more practice. Thus, it is very likely that SD groups did not achieve exactly what was expected from them because in few sessions they could not get the concepts. This must have certainly undermined the performance of students in the SD1 group. The evidence for this is the fact that the SD2 group showed more enhanced performance after going through 3 day sessions rather than just two. This fact leads to think that having longer time for introducing the SD teaching material to the students is essential for them to respond to the benefits of the teaching method, and with this a more enhance improvement in the understanding of revolution may be is given.

b. Motivation

A student could be motivated to learn new skills because he or she understands their potential utility or value, or because learning the skills will yield a good grade and the privileges a good grade affords. According to Self Determination Theory (SDT)[26] students can perform extrinsically motivated actions with resentment, resistance and disinterest, or alternatively, with an attitude of willingness that reflects an inner acceptance of the value or utility of the task. In the first case, one feels externally propelled into an action; in the later case, the extrinsic goal is self endorsed and thus adopted with a sense of volition (Deci & Ryan, 2000). Extrinsic motivation is aimed to be achieved in school tasks by internalizing the goal of the activity and by understanding the potential utility on such task. Likewise in the present experiment, the task itself should have been motivating enough to be self internalized by the students in order to focus their will and disposition on learning about revolutions. Besides, instructional design and the people involved in the application of the experiment aimed to motivate the students as much as possible. However, other factors undermined the motivation of some of the groups (Deci & Ryan, 2000).

The scholar year in Colombia ends at mid November, time in which these experiments were run. Certainly, students were strongly motivated to accomplish all the goals they were required to do before the academic year ended. However, it is very likely that their motivation could not be strongly oriented into an activity that demanded from them to learn something extra during such critical academic period of time. Besides this factor, the students did not receive any kind of reward for participating in the experiment. Rather, they had to follow a compulsory activity given by the Principal of the school[27]. Acceptance of the potential utility of such task was a tough process for the students, who at first tried to reject to go through more overloading activities different from the regular ones. Thus, relying only on the motivation that the instruction itself could provide to the students was not enough to ensure their best, especially because the students had many things to do alongside the experiments. This may have been essential in undermining the performance of all groups according to what was expected from those students.

The autonomy and the perceived benefits from the task are factors that promote the internalization and integration of the activity by the students (Deci & Ryan, 2000; Heckhausen, 1989). In the motivation through identification, the student has identified personally with the importance of the task and thus, accepted its regulation as his or her own. Studies have revealed that the more students are externally regulated, the less they show interest, value or effort in the tasks to be performed[28]. The more autonomous extrinsic motivation, the more associated it is with greater engagement, better performance, less dropping out, higher quality learning, and greater psychological well-

---

[26] In SDT the types of motivation are differentiated according to the reasons and goals that give rise to an action. The most basic distinction is between intrinsic motivation, which refers to doing something because it is inherently enjoyable, and extrinsic motivation, which refers to doing something because it leads to a separable outcome. Further information about Self Determination Theory (SDT) is available in (Deci & Ryan, 2000).

[27] Since the Principal chose the students who participated in the experiment, they only had to follow what he wanted them to do.

[28] The least autonomous form of motivation is the external regulation, in which some behaviors are performed to satisfy an external demand or obtain an externally imposed reward contingency. However, a more autonomous, or self determined form of motivation is regulation through identification.

being. Thus, the fact that the scholar year was finishing and students had to focus their interest on more relevant tasks at the same moment, could have caused certain degree of demotivation on them, making more difficult the process of internalizing the goal of the experiment and to visualize its potential utility in the activity (Deci & Ryan, 2000).

Furthermore, the experiment and its goals were presented to the students some minutes before the pre test were applied to them. At the beginning of this test, the students were told by a teacher about what was expected from them in the activity and what the importance of their performance was. Hence, each teacher was responsible for introducing these words to the students and for presenting the person in charge of applying the experiment. The role of the teacher on the students' motivation on the experiment (and on its consequent internalization as a students' personal interest) was rather important, and everything he said could have been a source of shedding interest or not on the students. With almost all groups, teachers kept faithful to present only what the experiment was about, however; when introducing the pre test and the experiments to the SD1 group the teacher bawl them out because they were still too energetic and noisy from the lunch break they took before. Right after, the teacher introduced the experiment and the pre test to be applied.

Cognitive Evaluation Theory (CET)[29] argues, among other tenets, that a high level of motivation cannot be achieved when a few sense of autonomy is surrounding the task to be performed. Furthermore, not only tangible rewards, but also threats, deadlines, directives and competition pressure diminish motivation. Certainly, the teacher accompanying the SD1 group took out whatsoever motivation students could have had, after giving them a tick off. Thus, it is very likely that students mixed the teacher's sermons with the sense of the experiment and thus, this single event could affect their performance, which did not entirely respond towards the expected goal. On the other hand, students from the CONTROL1 group were told about the experiment by the social sciences teacher, who indeed, told them anything else than the purpose of the experiment and what was expected from them. No berating, no ticking offs were given, and furthermore, students were already in social sciences class, which could make them relate the experiment to a subject's task instead of an extra task to be performed at the end of the year. Thus, the task itself could be more internalized and so, a high level of extrinsic motivation arose, which explains the sometimes lower performance of the SD1 group in comparison to the CONTROL1.

c. Working Memory Limitations

When dealing with novel information the working memory[30] has two severe limitations: its capacity and the duration that this information can remain in memory. All instructional methods requiring learners to deal with novel information must be processed by a structure that is minute in capacity and that retains the new information for no more than a few seconds. These limitations become successively less critical as familiarity increases (Sweller, 2005).

The French revolution was not an unfamiliar topic for the students; however, for those students whose teaching method was SD, the approach was indeed novel for them and

[29] Further information about Cognitive Evaluation Theory (CET) is available in (Deci & Ryan, 2000).
[30] The working memory is the cognitive structure in which information is consciously processed (Sweller, 2005).

provided lots of novel information about how the topic could be understood. System thinking, S&F diagrams, and understanding of behavior and structure, were just some of the novel issues students from the SD groups had to go through in order to understand the activity. The information presented was required to be recalled at any step further of the instruction. Both constraints of working memory, capacity and duration of the information, were violated. Capacity was overloaded by introducing so many SD concepts, required for the understanding of the entire instruction, in such a short period of time. Furthermore, the information was demanded to be recalled as long as the SD students went through the instruction, and so, the duration period for the information to be used was rather long. These factors may have undermined, especially, the effect of the treatments applied to the SD1 group, which were expected to enhance more students' understanding about revolutions. The more practice acquired by the SD2 group with the SD approach gave the students the chance to have one day more of getting to know the novel information. Even though instructional methods for both SD groups were the same, the experience gained by the SD2 group, due to being part of the Civics' experiment, let them understand better the novel information that were presented to them. Nevertheless, the entire learning potential that these students had was still far to be reached.

Instructional methods presented to SD groups failed in considering the working memory capacity, and added a rather high level of element interactivity[31]. Intrinsic cognitive load was inherent to the SD approach. Changing the way of approaching historical issues demands a great effort from students, especially when it is required from them to relate several elements introduced along the instruction for achieving a general understanding of the topic. Furthermore, extraneous cognitive load was presented in the need for the students to relate multiple sources of visual information, which all were essential for the understanding and were not intelligible in isolation. Thus, total cognitive load for the SD groups was rather heavy, which certainly undermined the performance of SD groups, especially of SD1 group that had less SD training.

On the other hand, students from the Conventional teaching method went through an instructional method that presented an approach of the topic that did not add complexity and novelty to the task. Extraneous cognitive load still is seen in this instruction because pictures, graphs and written texts were still present and were required to understand the topic. However, the total cognitive load of CONTROL groups was not as heavy as SD groups.

d. CONTROL groups: Special training for the State exam

Both CONTROL groups went through the final exam of the school run by the Colombian State. The exam is presented at mid September and students are trained during the first semester of the year to improve their cognitive skills for passing the exam. When the experiment was run at mid November, those students still had fresh

[31] In respect to the Cognitive Load Theory, the extraneous cognitive load is caused by inappropriate instructional designs that ignore working memory limits and fail to focus working memory resources on schema construction and automation. Intrinsic cognitive load is the cognitive load due to the natural complexity of the information that must be processed by the students. It is determined by levels of element interactivity. High element interactivity material imposes a high working memory load (Sweller, 2005).

what they got in such training. Thus, their cognition was highly increased and attentive for focusing on relevant issues taught in the instruction and to recall them in the test. This factor certainly enhanced the capacity of the CONTROL groups, in answering questions better than they were expected to do. Consequently, the performance of students in SD groups, mainly the SD1 group, can result overlooked.

Consequently, keeping in mind an integral point of view of the SWI and EI (in both the full test and the groups of questions), of the significant impact of the SD teaching method on students' performance, and of all the various unmeasured effects provides support for concluding that certainly the SD approach is a tool that enhances students' understanding and learning process in the field of history when students receive long training with the method.

Contribution to the Learning Process of History

Certainly, history is a discipline whose understanding goes beyond the accumulation of knowledge of the past. Its learning provides schemata and frameworks that when are seen as part of a whole allow individuals to identify recurrent patterns of behavior between different historical happenings. Thus, not only a single historical situation can be understood and recalled, but also such understanding becomes relevant to understand the diverse events in the society.

Teaching history with the SD method provides the students with a general framework, in which they can consider the relationships between all elements interacting in a historical happening. The methodology for teaching history proposed by Donovan and Bransford (Donovan & Bransford, 2005) is met by the SD approach. Thus, the consideration of changes of the state of affairs over time, the consideration of the time in the historical happenings, and the accumulation of knowledge are all factors that students in the SD groups revealed to grasp and to consider for the entire understanding of the case study.

Furthermore, Potash and Heinbokel's beliefs regarding the benefits of SD when teaching history have been, at some extent, confirmed by the present assessment of the SD tools. Thus, understanding historical patterns as part of a whole is an issue indeed given by the SD teaching method, which places such understanding in schemata that can be recalled and applied in further learning processes.

Understanding of the connections between the isolated events and variables playing an important role in the course of a revolution is determinant for altering the long term memory and to settle this understanding into the form of knowledge of history. Students who still approach the case study with the Conventional way of thinking can hardly reveal understanding of why history unfolded in certain ways. A mere recall of isolated events is the instrument for them to face history, which becomes a poor instrument when it is intended to make of history a science whose reflection and learning provides the students with tools for the understanding of society and for better approaching social problems, which reveal similar patterns than some occurred in the past, such as wars.

In addition, as Potash and Heinbokel believe, the understanding of how and why history occurred is greatly enhanced by the SD approach, and this understanding may lead students into the comprehension of the complexity of the present and the future in light of the lessons learned of the past (Potash 1995; Potash, 2005; Potash & Heinbokel, 2006). The assessment of students' performance in this experiment suggests then, that SD teaching method may be a useful approach for the students to learn from the past and the way consequences of the historical happenings affected other matters over the time in order to interpret better the present and future situations they will face.

At the same time, the present assessment has revealed more clear procedures to implement the SD teaching method as a conventional method to teach history. Longer and continuous sessions, in which different case studies are approached through the SD perspective, matter in the enhancement of students' performance. Thus, the longer the time a student has been in touch with the SD approach, and the more fields she has learned with SD, the easier to profit from such teaching method to get a deeper understanding of how certain historical happening was developed over time. Therefore, the possibility of constructing lighter instructional methods that do not load students' working memory more than necessary is an important issue for getting students' attention in learning with a different teaching method. Designing gradual increase of the cognitive demands of the SD instruction will guarantee that students do not loose any chance of improving their understanding of revolutions. Furthermore, it is required to enhance their external motivation by providing instructions and instructors that transmit autonomy and high perceived benefits to the students. Motivation seems to be an important factor influencing students' performance, which may be enhanced by the providing students with all conditions for identifying with the activity and to give worth to its utility for them.

Also, it is required to consider having much better and equal initial conditions for all groups who are tested with such experiment. Wherever possible, students must have in average the same capabilities to perform the experiment.

Thus, ensuring the control of those effects mentioned above, which undermined or amplify students' performance must be a must to do task in further experiments. Complementary, having bigger sample sizes may enhance the statistical power of the teaching method on the performance of the students. In this case, the performance of the SD groups would be highly enhanced in comparison to the performance of the CONTROL groups.

System Dynamics is still a young approach in the classroom, especially in those fields of study which have to do with social sciences, as in the case of history. Hence, further research and validation of the role of SD in the performance of the students of history is still a critical matter that deserves being at least considered as a possible conventional method to teach such subject.

6. Future Research

The present study suggests important insights regarding the usefulness of SD in the history classroom. However, it is still necessary to do more research in this field, given the few applications of SD in social sciences, and the great absence of assessment of the benefits that students really get from such approach. Therefore, in this section important aspects are mentioned in order to improve in the future the outcome of this and coming experiments for those who may be interested in the field.

Factors such as the duration of the experiments' sessions and the number of case studies approached with the SD teaching method seem to be relevant to enhance students' understanding of the dynamics and change over time of history. SD is an approach that itself provides the students with a greater understanding of the developing over time of history, but at the same time provides the students with lots of information and new reasoning that must be taught gradually over several teaching sessions. Thus, slowing down the presentation of the SD approach of history is an issue that really matters when thinking about further research in regard of these experiments.

An introductory session is still required to enroll the students into the SD thinking, to teach its basic concepts, and to play the Infection Game with its corresponding debriefing. Furthermore, the case study should be split into sessions of no more than 30 minutes, in which the content is presented step by step in the sessions. One week for one single case study would be the ideal situation in order to guarantee that students are not too cognitively loaded and that they will not loose attention because of tiredness. As mentioned in the previous section, this experiment has failed in loading the students, especially those in the SD treatments, with lots of information in very few periods of time.

Besides, teaching more than one case study, all supported on the same basic SD structure (in this case the diffusion model), let the students succeed in grasping the proper knowledge and understanding of the dynamics of history. Thus, all groups under the SD treatment should study at least two case studies. In addition, from the lessons learned in the present experiment, it is very important to assess students' understanding after every teaching session. Thus, if different case studies are intended to be taught in daily sessions within a week, assessment after each session is necessary. Every session should be proceeded by a test asking for the concepts learned during that session, and thus, the learning process of the students can be followed for measuring performance and for giving immediate feedback to the students about how they are performing in the experiment. Likewise, students' maximization of the benefits of such approach might be strengthened by running the experiment in a period of time, in which the students are not facing a too hectic moment, such as the end of the scholar year. These two considerations may be useful to lessen the cognitive load intrinsic to the SD approach itself and may increase the students' possibility to gain more profit from this teaching method.

Accordingly, a great sense of autonomy, given not only by the experiment but also by those people accompanying the learning process, is required to guarantee motivation and that students will identify with the activity's goal and will integrate it with the personal goals. In this sense, high school students have shown the need of a figure of authority to behave during teaching sessions, which are not done in the conventional schema they are used to. Thus, people accompanying the students during the teaching sessions are necessary but should limit to perform the unique role of a figure of authority and respect. Teachers must be prepared for not playing the role of the responsible of the activity, in order to avoid having different attitudes that interfere in the performance of the different groups of study.

Furthermore, a constant and immediate feedback of the performance along the sessions is highly decisive for the students' performance and their motivation. The present experiment did not provide the students with feedback about their understanding of revolutions and their performance. Thus, students could not know whether their performance was correct or whether further efforts and attention should be put in the activity. Information about their understanding of each session should encourage them either to put more efforts in the activity or to keep on doing the same well in order to finish with a high performance (Deci & Ryan, 2000).

External rewards such as monetary and academic rewards can only persuade the external motivation of students, and can easily undermine the intrinsic one. However, when the instruction contains factors such as high sense of autonomy, clear benefits from the learning experience, and immediate feedback of the performance; it is possible to increasingly persuade students' motivation with an external scholar reward such as an academic grade. Feasibility of this motivational strategy may be negotiated with the schools. A combination of all those factors will define the students to do their best in the task.

The computer based instruction is absolutely important to guarantee the provision of the same information to all students. This should be less loaded by information and should be interactive enough to present immediate feedback of the performance to the students. An Interactive Learning Environment (ILE) must be an ideal computer aid to transmit the information of history and to interact with the student along the sessions.

The test instrument is an important matter for measuring what is expected from the experiment. Thus, wherever possible, the test used to measure students' understanding regarding revolutions and their dynamics must be based on previous elaborated tests that measure the purpose of this experiment. Wherever not possible, the test should be carefully designed in cooperation not only with historians but also with system dynamicists, who can contribute with their understanding of the field and of the dynamic approach to the case study. After the design, the test should be widely tested before being applied to students in order to filter those questions that even though are intended to measure high understanding of revolutions can be easily correctly answered by guessing. Furthermore, from what has been experienced in the pilot experiments previous to this experiment, presentations in groups about different case studies analyzed with the SD perspective is a revealing instrument to test understanding of revolutions and their dynamic behavior. Sharing the personal understanding regarding revolutions with others is an important exercise to commonly build knowledge about this topic. However, designing a proper way of assessing students' understanding from the groups' presentations is a relevant issue, in which certain minimum parameters of performance might be created to measure and classify students' performance.

All the above considerations might want to be applied to CONTROL groups as well. For this purpose, all initial conditions for Conventional and SD groups must be the same in terms of cognition, knowledge, and attitude conditions. To ensure this, randomization of students' samples is essential. Each subject should be assigned randomly to the corresponding group to which she will belong. Thus, coordination with the schools should be done very well in advanced in order to control any disturbance that makes more complicated the randomization of the samples (such as the time pressure presented in the school).

Additionally, the fact that the SD2 group was the most outstanding group and that SD1 revealed high SWI, provides the insight that SD could be, indeed, a better way to teach history and to place its understanding in a structured pattern in students' minds. Thus, it is required to keep on building a conventional methodology based on SD to teach history in high school classrooms. In order to reach such goal, it is important to continue experimenting with the SD approach in the real history classroom.

References

Aires, P. and J. Sweller (2005). The split attention principle in multimedia learning. Cambridge Handbook of Multimedia Learning. R. Mayer. United States of America.

Bao, L. (2006). Theoretical comparisons of the average normalized gain calculations. American Journal of Physics Volume 74.

Bransford, J.D., A.L. Brown, and R.R. Cocking (1996). Effective Teaching: Examples in History, Mathematics and Science. How People Learn: Brain, mind, experience and school. National Research Council Comitee on Learning Research and Educational Practice. National Academy of Sciences.

Burton, O.V. "Teaching Philosophy Statement." Retrieved May, 2006, http://www.history.uiuc.edu/fac_dir/burton_dir/philos.htm

Businessballs. "Bloom's Taxonomy - Learning Domains." Retrieved May, 2006, www.businessballs.com/bloomstaxonomyoflearningdomains.htm

Brinton, C. (1965). The Anatomy of Revolutions, Prentice Hall.

Center for Interdisciplinary Excellence in System Dynamics, CIESD. (2001). "Demo Dozen & Modeling Systems Self-Taught (MSST)." Retrived May, 2006, www.ciesd.org/influence/demo_dozen.shtml

Creative Learning Exchange. (1992). Retrieved 2006, http://clexchange.org/

Deci, E.L. and R.M. Ryan (2000). Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions. Contemporary Educational Psychology Volume 25.

Donovan, S. and J.D. Bransford (2005). How Students Learn: History in the Classroom. Committee on How People Learn, a Targeted Report for Teachers. Division of Behavioral and Social Sciences and Education. The National Academic Press. Washington.

Hake, R.R (2002). "Assessment of Students Learning in Introductory courses." Retrieved April, 2007, www.physics.indiana.edu/~hake/ASLIS.Hake.060102f.pdf

Heckhausen, H. (1989). Motivation und Handeln. Springer. Berlin.

History News Network. Retrieved May, 2006, http://hnn.us/articles/849.html

Fisher, D. (1992). "Easter Island Population Model". Retrieved April, 2007, www.clexchange.org/ftp/documents/social-studies/SS1994-09EasterIslandPop.pdf

Forrester, J.W. (1992) System Dynamics and Learner Centered Learning in kindergarten through 12th Grade Education. Retrieved May, 2006, http://sysdyn.clexchange.org/sdep/Roadmaps/RM1/D-4337.pdf

____________________ (1998). Designing the Future. Universidad de Sevilla, Spain. Retrieved May, 2007, http://sysdyn.clexchange.org/sdep/papers/Designjf.pdf

Jianting, What's the favorite subject? Retrieved May, 2006, www.rjfisher.lgusd.k12.ca.us/newspaper/academics/SchoolSubjects.html

Loewen, J. W. (1995). Lies my teacher told to me: everything your American Textbook got wrong. The New Press. New York.

Marx, J. and K. Cummings (2007). Normalized Change. American Journal of Physics Volume 75

Mayer, R. (1999). Designing instruction for constructivist learning. Instructional-Design Theories and Models. A new paradigm of instructional theory. Volume II. Lawrence Erlbaum Associates. United States of America

Mintz, S. (2003). Digital History. Retrieved May, 2006, www.digitalhistory.uh.edu

Pinette, D.D (2002). "Constructing History: how historians see the light." Retrieved March, 2007, http://www.coe.uga.edu/epltt/cc_example/index.html

Potash, J. (1995). "Building Systems into the History /Social Sciences Curricula: Some preliminary thoughts." Retrieved April, 2007, http://www.clexchange.org/ftp/documents/social-studies/SS1995-10BuildSystemsIntoSS.pdf.exe

_______ (2005). Musings on Social Studies and System Dynamics: Linking "Standards" with System Citizenship. Retrieved April, 2007, http://www.clexchange.org/ftp/documents/social-studies/SS2005-11MusingsOnSSandSD.pdf

_______ and J. Heinbokel (2006). Musing on Social Studies and System Dynamics II: Thinking systemically. Retrieved April, 2007, http://www.clexchange.org/ftp/documents/social-studies/SS2006-01MusingsOnSSandSD2.pdf

_______; _____________; M. Engelken and R. Green (1996). Plagues and People, A curricular experiment. Waters Center for System Dynamics. Trinity College of Vermont.

Reid, G. (1996). Causes of the Civil War: Through viewing relationships and systems. System Dynamics Project, Catalina Foothills School. Retrieved April, 2007, http://www.clexchange.org/ftp/documents/social-studies/SS1996-11CausesTheCivilWar.pdf

Road Maps: A guide to learn System Dynamics (1992). Retrieved May, 2006, http://sysdyn.clexchange.org/road-maps/rm-toc.html

Sheskin, D. J. (2004) Handbook of parametric and nonparametric statistical procedures. Third Edition. Chapman & Hall/CRC.

Skillings, L. (1997). A connected view of history. Retrieved May, 2006, www.k12.hi.us/~lskillin/clearticle.htm

Smith, V. L. (1982). Mycroeconomic Systems as an Experimental Science. The American Economic Review, Volume 72, No.5.

_______________ (2002). Method in Experiment: Rhetoric and Reality. Experimental Economics Volume 5. Springer.

Sterman, J. D. (2000) Business Dynamics. Systems Thinking and Modeling for a complex world. McGraw-Hill.

Sweller, J. (2005. Implications of Cognitive Load Theory for Multimedia Learning. Cambridge Handbook of Multimedia Learning. R. Mayer. United States of America.

Waters Center for System Dynamics (2001). Trinity college of Vermont. Retrieved May, 2007, www.uvm.org/trinityvt.edu/waters/

Wheat, D. (2007). The Feedback Method: a System Dynamics Approach to Teaching Macroeconomics. PhD dissertation, University of Bergen (Norway).

Wiley, J. and I. K. Ash (2005). Multimedia Learning of History. Cambridge Handbook of Multimedia Learning. R. Mayer. United States of America.