Spring 2010

Determining the Utility of Technology for Education

Brian Spencer Dillon
Old Dominion University

Follow this and additional works at: https://digitalcommons.odu.edu/msve_etds

Part of the Computational Engineering Commons, Computer Engineering Commons, Educational Methods Commons, and the Educational Technology Commons

Recommended Citation
Dillon, Brian S.. "Determining the Utility of Technology for Education" (2010). Master of Science (MS), Thesis, Computational Modeling & Simulation Engineering, Old Dominion University, DOI: 10.25777/ks4w-kb84
https://digitalcommons.odu.edu/msve_etds/103

This Thesis is brought to you for free and open access by the Computational Modeling & Simulation Engineering at ODU Digital Commons. It has been accepted for inclusion in Computational Modeling & Simulation Engineering Theses & Dissertations by an authorized administrator of ODU Digital Commons. For more information, please contact digitalcommons@odu.edu.
DETERMINING THE UTILITY OF TECHNOLOGY FOR EDUCATION

by

Brian Spencer Dillon

B.S. May 2007, Virginia Commonwealth University, USA

A Thesis Submitted to the Faculty of Old Dominion University in Partial Fulfillment of the Requirements for the Degree of

MASTER OF SCIENCE

MODELING AND SIMULATION

OLD DOMINION UNIVERSITY

May 2010

Approved by:

Yuzhong Shen (Director)

Ginger Watson (Member)

Jiang Li (Member)
ABSTRACT

DETERMINING THE UTILITY OF TECHNOLOGY FOR EDUCATION

Brian Spencer Dillon
Old Dominion University, 2010
Director: Dr. Yuzhong Shen

There is a substantial debate concerning the place of technology in education. This debate has been going on over every technology introduced through at least three millennia although the modern version takes place in journal articles. Technology, in the general sense, is already in the educational arena in ways which have become so commonplace that we hardly question them today. More modern technologies, however, meet perpetual resistance in achieving their potential. This paper calls into question the most recent trend in research and suggests a new research paradigm which focuses not only on student achievement, but on a variety of other issues which are known to affect the learning experience. As an example and first step in using this new research paradigm, Modeling and Simulation (M&S) was experimentally examined.

The experiment was planned and executed with the cooperation and consent of the Richmond County School System of Warsaw, Virginia. The experimental subjects were minors, specifically second grade students. The group was divided randomly into control and experimental subjects. Through careful preparation, the experiment was integrated with normal classroom procedures in order to pose the least possible risk to these vulnerable test subjects. Testing instrumentation was
used before and after the experiments to measure student knowledge. The scores were compared in order to gauge the learning which had taken place during the experiment.

Three distinct instructional methods were used to introduce M&S into the classroom. A separate software simulation was produced for each of these methods. The simulations focused on simplicity and non-numerical analysis. This decision was in keeping with the knowledge and anticipated skill set of the test subjects. Instead of numerical data, each of the simulations made use of simple graphical animation in order to visually depict quantities and directions. The simulations were verified by qualified educators and validated mathematically to ensure that the simulations produced accurate results. In one case the validation was carried on by a US Navy civilian expert in electromagnetic radiation.

An analysis of the experimental results showed no comparable difference between the experimental and control subjects in any of the experimental instructional methods. Under the new research paradigm this finding justifies the use of M&S in the classroom. At the same time anecdotal evidence was searched for any indications of potential educational benefits. This evidence was delineated and several potential benefits were selected for future research. Under the new research paradigm these benefits will need to be quantified in future research in order to validate the contribution of M&S to the classroom.
Copyright, 2010, by Brian Spencer Dillon, All Rights Reserved.
ACKNOWLEDGMENTS

I wish to express my many thanks to Dr. Yuzhong Shen who has acted collaboratively and with extremely supportive measures in all cases. Though in a position of total authority, he has undertaken the role of mentor instead and has made a great difference. Also I wish to thank all of the committee members who have agreed to assist, inform, critique and finally approve this thesis.

I also wish to express my great thanks to the teachers and staff of the Richmond County School System. Their support was instrumental in preparing for this research and their participation in the experiment ensured the safety and wellbeing of the subjects.

Finally, may I extend a hearty thanks to my wife and family who have been patient during the preparation for this thesis, the experiments, the writing and have been hugely supportive of my conclusions.
TABLE OF CONTENTS

Section Page

LIST OF FIGURES ix
LIST OF TABLES xi

Chapter Page

1.0 INTRODUCTION ............................................................... 1

1.1 Definitions ................................................................. 1
1.2 The Case for Modeling and Simulation in Education ................. 4
1.3 Motivation and Expected Outcomes .................................. 6
1.4 Thesis Outline ............................................................. 7

2.0 QUALIFICATION OF TECHNOLOGY FOR EDUCATION .......... 8

2.1 Current Research in Educational Technology ....................... 8
2.2 The Argument over Technology in Education ....................... 9
2.3 An Examination of the Argument ..................................... 11
2.3.1 The Non-Traditional Argument ................................... 11
2.3.2 The Unnatural Argument ........................................... 12
2.4 Reframing the Argument ................................................ 14

3.0 METHOD ........................................................................ 17

3.1 Experimental Design ..................................................... 17
3.1.1 Sample Selection ....................................................... 17
3.1.2 Treatments ............................................................... 19
3.1.3 Statistical Methods ................................................... 19

3.2 Items of Interest ........................................................... 20
3.2.1 Experimental Instructional Methods ............................... 20
3.2.1.1 Simulation-Based Instruction ........................................... 20
3.2.1.2 Simulation-Based Laboratory ........................................... 21
3.2.1.3 Discussion Through Simulation ......................................... 22
3.2.2 Participating Teachers ..................................................... 22
3.2.3 Ethical Considerations of Human Subjects Research ................ 23
3.2.4 Simulation Characterization .............................................. 23
  3.2.4.1 Magnetism Simulation ............................................... 24
  3.2.4.2 Water Cycle Simulation ............................................. 26
  3.2.4.3 Symmetry Simulation ............................................... 27
3.2.5 Simulation Verification and Validation .................................. 29
3.2.6 Metrics and Instrumentation ........................................... 31
4.0 EXPERIMENTAL RESULTS .................................................. 36
4.1 Symmetry in Simulation-Based Instruction Method ..................... 36
  4.1.1 Pre-Experimental Knowledge Base .................................... 36
  4.1.2 Execution of Experimental Instruction ............................... 38
  4.1.3 Post-Experimental Results ............................................ 41
4.2 Water Cycle in Simulation-Based Laboratory Method .................. 42
  4.2.1 Pre-Experimental Knowledge Base .................................... 42
  4.2.2 Execution of Experimental Instruction ............................... 43
  4.2.3 Post-Experimental Results ............................................ 45
4.3 Magnetism in Discussion Through Simulation Method .................. 46
  4.3.1 Pre-Experimental Knowledge Base .................................... 47
  4.3.2 Execution of Experimental Instruction ............................... 49
  4.3.3 Post-Experimental Results ............................................ 50
### LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>National Average Cost Per Student</td>
<td>5</td>
</tr>
<tr>
<td>1.2</td>
<td>Average SAT Test Scores Per State by Revenue Per Student</td>
<td>6</td>
</tr>
<tr>
<td>3.1</td>
<td>Screen Capture from Magnetism</td>
<td>25</td>
</tr>
<tr>
<td>3.2</td>
<td>Effects of Positive and Negative Sources on various Indicators</td>
<td>25</td>
</tr>
<tr>
<td>3.3</td>
<td>Screen Capture from Water Cycle</td>
<td>26</td>
</tr>
<tr>
<td>3.4</td>
<td>Sprites Indicating Moisture Position and Quantity</td>
<td>27</td>
</tr>
<tr>
<td>3.5</td>
<td>Screen Capture from Symmetry</td>
<td>28</td>
</tr>
<tr>
<td>3.6</td>
<td>Example of Bilateral Symmetry and Line of Symmetry</td>
<td>29</td>
</tr>
<tr>
<td>3.7</td>
<td>Transitions of Moisture through State Diagram</td>
<td>30</td>
</tr>
<tr>
<td>3.8</td>
<td>Alteration of Image Processing Questions for Test Distinction</td>
<td>33</td>
</tr>
<tr>
<td>3.9</td>
<td>Sample Questions from Magnetism Test</td>
<td>34</td>
</tr>
<tr>
<td>4.1</td>
<td>Pre-Test Interquartile Range (IQR) for Experiment 1</td>
<td>37</td>
</tr>
<tr>
<td>4.2</td>
<td>Sample of Symmetry Pre-Test Completion Question</td>
<td>37</td>
</tr>
<tr>
<td>4.3</td>
<td>Subject Response to Initial Symmetry Tasking</td>
<td>38</td>
</tr>
<tr>
<td>4.4</td>
<td>Subject Response to Secondary Symmetry Tasking</td>
<td>39</td>
</tr>
<tr>
<td>4.5</td>
<td>Subject Response During Post-Instructional Playtime</td>
<td>39</td>
</tr>
<tr>
<td>4.6</td>
<td>Subject Work Showing Purpose and Apparent Randomness</td>
<td>40</td>
</tr>
<tr>
<td>4.7</td>
<td>Histogram of Learning Delta for Experiment 1</td>
<td>41</td>
</tr>
<tr>
<td>4.8</td>
<td>Comparison of Pre- and Post-Test Scores for Experiment 1</td>
<td>42</td>
</tr>
<tr>
<td>4.9</td>
<td>Pre-Test Interquartile Range (IQR) for Experiment 2</td>
<td>43</td>
</tr>
<tr>
<td>4.10</td>
<td>Reproduction of Instructional Illustration of States of Matter</td>
<td>44</td>
</tr>
<tr>
<td>4.11</td>
<td>Assessment Illustration of Water Cycle</td>
<td>44</td>
</tr>
<tr>
<td>4.12</td>
<td>Histogram of Learning Delta for Experiment 2</td>
<td>46</td>
</tr>
<tr>
<td>4.13</td>
<td>Comparison of Pre- and Post-Test Scores for Experiment 2</td>
<td>46</td>
</tr>
<tr>
<td>4.14</td>
<td>Pre-Test Interquartile Range (IQR) for Experiment 3</td>
<td>47</td>
</tr>
<tr>
<td>4.15</td>
<td>Sample of Magnetism Pre-Test Force Question</td>
<td>48</td>
</tr>
<tr>
<td>4.16</td>
<td>Reproduction of Instructional Illustration of Magnet</td>
<td>49</td>
</tr>
<tr>
<td>4.17</td>
<td>Comparison of Simulation and Subject Representation</td>
<td>50</td>
</tr>
<tr>
<td>4.18</td>
<td>Histogram of Learning Delta for Experiment 3</td>
<td>51</td>
</tr>
</tbody>
</table>
4.19 Comparison of Pre- and Post-Test Scores for Experiment 3 ...............52
4.20 Sample of Magnetism Post-Test Force Question........................................52
5.1 Contradictory Interface Controls ..........................................................55
5.2 Subject Work Showing Marked Preference for Stamps .......................59
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1 Comparison of Richmond County and US Population by Race</td>
<td>18</td>
</tr>
<tr>
<td>3.2 Sample Questions Reciprocated for Pre- and Post-Tests</td>
<td>32</td>
</tr>
<tr>
<td>3.3 A Variety of Question Formats</td>
<td>34</td>
</tr>
<tr>
<td>5.1 Elementary Education Degrees in Top Virginia Schools</td>
<td>66</td>
</tr>
</tbody>
</table>
CHAPTER 1

INTRODUCTION

Technology which has found its way into the classroom does so more because of social acceptance than any particular measure of utility. There are a myriad of individuals and disciplines with a stake in this decision. This document explores a new measure of technological utility specifically for the classroom and puts forth the data from an experimental example of that metric.

1.1 Definitions

This thesis is intended for members of the educational field, as well as those interested in educational technology. A large portion of it concentrates on terminology most familiar to those in the field of modeling and simulation. This thesis will work from the definitions of system, model and simulation as proposed by Gene Bellinger [1]. They are repeated in their entirety below.

A system exists and operates in time and space.
A model is a simplified representation of a system at some particular point in time or space intended to promote understanding of the real system.
A simulation is the manipulation of a model in such a way that it operates on time or space to compress it, thus enabling one to perceive the interactions that would not otherwise be apparent because of their separation in time or space.

With such a general definition, it might be appropriate to question the purpose of a model or simulation. Modeling and simulation (M&S) does not have a unique, but rather a multi-faceted purpose that is met or at least addressed by several other disciplines. What makes M&S unique is the fact that it attempts to address all of these areas at once. Some of the application areas are:
• Evaluation and improvement of manufacturing processes
• Representation of and prediction for phenomena from natural sciences
• Parameterization and alteration of human behavior
• Depicting and demonstrating combinations of military units in action

This short selection of applications does much to explain the very general definition. M&S is the science of depicting a real-world system in such a way that it can be viewed from a variety of perspectives and scales in order to enhance the user's understanding and predict the behavior of the original system.

The important and operative word for the purposes of this thesis is understanding. Models can be and are built to model trivial situations that can be experimented upon in the real world. Generally, however, models are built around systems that are completely inaccessible. This characteristic of inaccessibility may be for reasons of exorbitant cost, extreme danger, infrequent opportunity, spatial/temporal distance or prior dedication of the system. In any of these cases the model of the system becomes an invaluable substitute for the actual system.

By manipulating the parameterized characteristics of the model in a simulation, the real world system can be represented in whichever time, space or orientation is convenient to improve the investigator's understanding of the real world system.

In order for that understanding to accurately reflect how the real world system would act under the parameters given, it is necessary for the model to be validated. Validation involves adequate proof that the model does indeed reflect the behavior of the real world system. Given this adequate proof, the investigator can confidently rely upon the simulation output and represent that output as characteristic of the real world system. Validation, however, can be carried out in
different ways and to various degrees. For instance, the mathematical models of quantum phenomena, some of the best validated models, have predicted phenomena that were subsequently confirmed by experimentation. On the other hand, the climate model predictions which Al Gore touted in his famous presentation are still—excuse the pun—hotly debated, primarily because the models have only been validated over a limited data set and have widely varying predictions.

It should be noted that validation in the field of M&S is not equivalent or related to the concept of validity evidence in education. Validity evidence refers to the proof that a test, interview or other method of selection among a population is indeed predictive of the selectees' eventual success, understanding or ability within that domain. As such it should never be applied to models or simulations. However, in order to diminish the ambiguity which the similarity in the terms may create, it is necessary to categorically distinguish between the two. In this document the terms valid, validate and validity in reference to a simulation will refer only to the definition under M&S and not to the common understanding of validity evidence.

Finally, the level of detail may vary highly from one simulation to another. As Mr. Bellinger quipped, "One cannot develop every model in the context of the entire universe, unless of course your name is Carl Sagan" [1]. In order to make the production and validation of a model cost effective, the investigator may choose to model only those details that are pertinent to the investigation or limit the number and frequency of measurements. These concepts, known as fidelity
and resolution respectively, need not detract from the validity of the model. A simple block diagram of the three branches of the federal government is a valid model, albeit one with a low fidelity. Any model may be validated, but it cannot produce output that has predictive validity outside of that resolution and fidelity.

1.2 The Case for Modeling and Simulation in Education

The driving concepts behind M&S are reducing the cost of experimentation by producing a model of the system valid within the resolution and fidelity required for the experiment, and then simulating the system under different parameters. Usually, this makes it more cost effective than taking experimental measurements of the real system. Industry and commercial enterprises make use of M&S to perform extensive testing and experimentation at a fraction of the cost of real world experimentation. The US Department of Defense has made M&S a key part of its acquisition process because "[M&S] will improve efficiency and effectiveness for program-level systems engineering" [2].

In each of these cases, M&S has been a powerful tool; however, it is in education that the cost savings might be most valuable. The cost of education has been rising alarmingly fast, especially in light of the frustratingly elusive corresponding rise in student achievement. The US Department of Education statistics for primary and secondary education show that the average cost per student nationally has risen steadily since 1919. Adjusted for inflation, these figures show that relative to the cost per student as little as a generation ago, today's student costs 125% more per school year [3].
These statistics would not be unpalatable if results followed. However, the numbers do not bear out the assumption that more money leads to improved achievement scores. The Department of Education keeps statistics on Revenue Per Student [4] and SAT scores [5] for each state. Comparing these statistics for 2005-06 as in Figure 1.2 shows no correlation at all. In fact, even where the disparity in revenue is greatest, comparing New Hampshire and Alaska, the results are seemingly reversed. Alaska outspends New Hampshire by 244%, but New Hampshire has more students who take the test and better scores in every category. In the future, educators can expect that costs will continue to rise, but these statistics make it clear that simply throwing more money at the problem will not increase student academic achievement.

Figure 1.1: National Average Cost Per Student (Source: [3])
Each school district in the country is looking for a magic bullet that will improve the outcome for their students. Michelle Rhee, the Chancellor of the Washington, DC Public Schools, has come under considerable scrutiny because of her recent actions to this effect. She has instituted a policy of firing teachers who do not meet her criteria for student achievement and, much to the chagrin and distress of the teacher's union in the District of Columbia, this decision has become big news nationally [6]. With the extraordinary budget restrictions which are common to all school systems, a plan—any plan—that seems to improve student achievement is likely to be implemented.

1.3 Motivation and Expected Outcomes

The educational field is a ready market for M&S technology. Indeed, the purpose of M&S, as defined above, is to promote understanding of real world systems—the very definition of education. However, an apparent harmony between the fields is not factually sufficient to recommend its use. The purpose of this research is to determine, experimentally and in a statistically significant way, whether M&S is compatible and useful in an educational setting.
1.4 Thesis Outline

The remainder of this thesis is organized as follows.

Chapter 2 examines the faults in the current research paradigm for determining educational utility, their origin, a scheme for correcting those faults and a formalized metric that incorporates the corrections.

Chapter 3 explains an example of the experimental setup and methodology required to meet the criteria for the first portion of that metric. This includes a look at the population, characterization of the simulations involved and the instrumentation used for collecting data.

Chapter 4 discusses the experimental results obtained from three distinct experiments examining various instructional methodologies utilizing M&S technology in the classroom.

Chapter 5 examines the anecdotal evidence derived during the experimental procedure and discusses their implications for the use of this technology.

Chapter 6 summarizes the conclusions and explains the reasoning behind several suggested future research topics related to the second portion of the metric.
CHAPTER 2

QUALIFICATION OF TECHNOLOGY FOR EDUCATION

In order to understand the scope of any measure of educational technology, it is necessary to explore the various viewpoints which must be satisfied by that measurement. This section looks at the views of researchers and educators followed by a new proposed metric which attempts to address the objectives of both groups.

2.1 Current Research in Educational Technology

Ongoing research in educational technology focuses on the instructional affordance of each feature. A 2002 article in the Journal of Educational Psychology compared the instructional outcomes from images, animations and animated characters [7]. A 2005 article in Educational Technology and Society compared the instructional outcomes from two educational game designs [8]. These articles and others are looking for positive outcomes from educational technology based on subject test scores. The ultimate prize under this research paradigm would be a technology which vastly improves test scores—a magic bullet. Unfortunately, this paradigm requires technology to show a statistical advantage in student achievement over traditional instruction methods and may be overlooking other measures of suitability.

Setting aside the quest for that magic bullet, the question of suitability should be framed in terms of the inherent educational value of any technology. Education is a more holistic process than this unique metric of student
achievement would indicate. Indeed, education comprises a wealth of other features such as the emotional and mental health of the students, building a desire for life-long learning and addressing the needs of the teachers and parents as well as those of the student. Inherent educational value might be defined in any number of areas of which student achievement is only one. In the current research paradigm student achievement has become, to the exclusion of all others, the defining metric in determining the utility of technology in the classroom.

2.2 The Argument over Technology in Education

In the course of the experimental procedure explained later in the thesis, an interview was conducted with various educators. Mrs. Susan Farmar, who has been teaching for 25 years, indicated some of the typical bias in this area. On several occasions she opined that any research would merely confirm the absolute necessity of the human teacher in the classroom. Speaking of the disparity between herself as a traditional, face-to-face teacher, and a colleague who in many cases replaced human interaction with technology she said, "I really despair for education. In five years [teachers like me] will retire and many of those kids will be missing a lot of that human interaction." Given this argument, it was appropriate to introduce the following quote with which Mrs. Farmar readily agreed.
"The fact is that this invention will produce forgetfulness in the souls of those who have learned it. They will not need to exercise their memories, being able to rely on [this technology], calling things to mind no longer from within themselves by their own unaided powers, but under the stimulus of external marks that are alien to themselves.... And as for wisdom, you're equipping your pupils with only a semblance of it, not with truth. Thanks to you and your invention, your pupils will be [instructed by technology] without benefit of a teacher's instruction; in consequence, they'll entertain the delusion that they have wide knowledge, while they are, in fact, for the most part incapable of real judgment"
[9].

In all fairness to Mrs. Farmar, this slightly altered quote was intentionally deceiving, but it shows how little the argument has changed. The quote could have been from a like-minded, modern colleague about any modern technology. In fact, it was spoken by Plato more than two millennia ago and, ironically, referred to the invention of the written word that has preserved his argument since that time.

Plato's argument then and the visceral response of many near-to-retirement teachers today are understandable and prompted by much the same motives. Plato learned and taught in an age when education was entirely cognitive. The well-educated man then had nearly as many proverbs, calculations, poems and arguments on the tip of his tongue as a well-stocked library has today. From his perspective, dependence on writing did indeed give the "delusion" of a good education. Teachers who were trained before the advent of the personal computer feel much the same about the advent of such sources as Wikipedia and Google. They recognize the potential of computers but maintain that software is replacing a valuable part of our educational heritage. As this argument has
been going on for some time, a closer examination is called for in order to adequately measure the utility of any technology.

2.3 An Examination of the Argument

Because this argument is so prevalent today as well as in the past, it has become the major objection to any new technology. As such it should be examined here in order to explain its application, if any, toward the metric. Therefore, the two principle components of the argument are discussed in more detail below.

2.3.1 The Non-Traditional Argument

Plato made several arguments against the use of technology in education. The first is what might be called the non-traditional argument. This is that as a technology replaces the traditional teacher's instruction, students will only be equipped with the pretense of a good education. A student in Plato's time could recite hours of famous and epic tales and explain his most complex arguments without recourse to a single notebook. Certainly from Plato's perspective the individual who referred to the written equivalent was doing so on the cheap. Strictly speaking he was correct that this new technology leads us to the "delusion" of a wide knowledge base.

However, few would agree that this argument has been proven correct over time. The written word has proven invaluable as a form of instruction. The greatest educational institutions take justifiable pride in their extensive libraries.
Our young are instructed by the greatest authors of centuries past. We record and publish abroad some of the greatest thoughts of our day. Scientists the world over wait breathlessly for the monthly and quarterly installments of the latest and greatest research in their fields. More importantly, we preserve this knowledge in a continual stream that will be available in the next generation. In much the same way as anthropologists tell us that spoken language extended early man's wisdom to the next generation, written language has further extended that wisdom indefinitely. One wonders how Plato would feel if he knew his oral arguments were still read today.

Therefore, we must conclude in contradiction to Plato, that a new technology which violently alters the method of instruction cannot be dismissed out of hand for this reason. Just as the carbide replaced the flintlock and the machine gun replaced the carbide, advancing technology in education will eventually replace the old. We have no imperative to move forward on every technology, but we have no right to discard them out of hand because they change the status quo. Applying this argument to the introduction of M&S into the classroom, we have no imperative to use simulations just because such are available, neither do we have the right to ignore them because they are not traditional.

2.3.2 The Unnatural Argument

Plato's second argument might be called the unnatural argument. It is that there is a "natural" method of instruction and therefore the technology that
replaces it is unnatural. In Plato's day a teacher chose his students as carefully as one might a spouse. In fact there are many stories about the famous teachers which begin "A man came to a teacher and asked to be taught." The advent of large expanses of written equivalencies would have turned this paradigm on its head. The student would be able to choose not only among the teachers, but also among the lessons. The teacher no longer existed and a cold expressionless page would serve as an instructor. These "alien" marks were controlling the reader, manipulating and altering his thoughts. From Plato's perspective, what could be more unnatural than talking to a thing?

After two thousand years we have a slightly different view of the situation. Our language has acquired phrases like "curl up with a good book" that suggest that a book may become as important as a friendly embrace. We speak of a writer's voice and tone suggesting the same familiarity we might have with the sound of their verbal communication. We even have preferences among authors, showing that it is the person behind the words and not the words themselves that we appreciate. In fact, many of the individuals who nay-say our latest technology advocate that instead we should do the "natural" thing: pick up a good book.

With due deference to Plato, we cannot use the "unnatural" quality of the technology as an argument against it. Would anyone still advocate that humans should use flint blades and eat raw flesh? This was once the "natural" thing to do, but not any longer. Two centuries ago it would have seemed unnatural to fly through the air in a metal bird, but today it is the safest way to travel. In the same way, it has become natural that the great thinkers of our day will extend their
influence through the written word. "Publish or perish" is the common phrase used as it is expected that most professors will write about their most recent work. Natural is a matter of perspective and experience. Those who perceive the world through the eye of an outdated frame of reference may insist that something is unnatural, but this is more of a commentary on their own limited perspective than the thing they critique.

2.4 Reframing the Argument

A slightly different view of the question was put forward by Dr. David Kennedy of Lingnan University, Hong Kong. He quoted a quip by Tom Holt [10], "I have never ever heard someone ask the question: 'I wonder if overhead projectors make a difference in student achievement?'" The implication is that the current research paradigm may seem slightly ridiculous when applied to a more accepted technology. Dr. Kennedy's perspective has led to his research into the use of social networking technology in education. He is continually refining his method of application in order to achieve the best results. Note that these results are not limited to student achievement, but encompass a wider range of what he calls "pedagogical advantages" [10].

Dr. Kennedy's research primarily involves reaching an understanding of pedagogical advantages and then building a framework in which these may be used to great effect. Building on his work, the method for determining the utility of any technology is as follows. First, identify the pedagogical advantages each has to offer. Second, define the learning objectives for an assignment, lesson or
entire course. Third, determine and select the technology that offers pedagogical advantages best suited to facilitating and forwarding those objectives. This method may be applied to any technology, even in the instance which Holt suggests. Consider a lesson plan discussing renaissance art. The learning objectives include the visual display of masterworks and perhaps a side-by-side comparison. Assuming that the pedagogical advantages of all technologies are known, then the projector would be an ideal match. The projector has reached such a level of acceptance that this analysis seems unnecessary but not wrongfully applied.

2.5 Formalizing a Metric

By applying the method used on the projector, already the intuitive norm for existing technologies, we can justify the use of new technologies. However, this metric is missing a key indicator: student achievement. In this new metric, student achievement is not the only measure, but it does serve a pivotal role as a limiting factor. Whatever other advantages a technology may offer, it needs to be at least as effective as traditional instructional methods in order to be accepted. This "as good as" standard becomes the hinge on which the whole case hangs. Once a technology can prove it is as good as traditional instructional methods, it can be examined for the pedagogical advantages it offers.

Identifying pedagogical advantages is a two step process. The first step involves some creativity and particular attention to anecdotal evidence of potential advantages. These evidences are most likely found during the
experimental examination of student achievement. Although anecdotal evidences may only point to corollary consequences of the use of technology, when properly exploited these advantages can be of major importance. The second step involves quantifying the expected benefit which can be derived from these advantages. Again, taking the over-simplified case of the projector, an observer might note that slides may be used repeatedly over the course of several years. After quantifying the cost savings in a typical classroom, the lifetime cost savings of the equipment can be calculated. If substantial, the lifetime cost savings could be considered as a pedagogical advantage.

Therefore, this first experiment with M&S in an educational setting will focus on anecdotal evidence in addition to statistical measurements of student achievement. Testing and statistical analysis will confirm or deny that this technology meets the "as good as" standard, while anecdotal evidence and observation may provide valuable information about the advantages, pitfalls and preferred use of simulations in the classroom.
CHAPTER 3

METHOD

Because this document combines two domains, education and M&S, it is necessary to use forms which will be familiar to both groups. Educational papers are generally formatted according to American Psychological Association (APA) standards [11]. While much of this document ignores that format, this chapter which explains the experimental set up and execution is of particular importance. Therefore, section 3.1 below describes the general information required in such a chapter. In the follow on section more information is given related to the structure of proposed simulations other items of interest.

3.1 Experimental Design

This section gives information about the method for selecting a sample, the experimental treatments which were applied and the statistical methods used on the data.

3.1.1 Sample Selection

The experiment was made possible through the cooperation of the Richmond County Elementary School in Warsaw, Virginia. The school administrators agreed to support this research by providing facilities for the experiment and allowing access to the student body. Two second grade teachers agreed to offer their students the opportunity to participate in experiments during
the 2009-2010 school year. The incoming class of 100 students was randomly distributed into five classes giving a sample size of 40 students although the actual number in each experiment varied according to classroom attendance. In order to qualify participants, those with prior knowledge in these areas were excluded on the basis that their pre-test scores are extreme outliers.

Demographic data was not taken on the participants as they are a protected population; however, Richmond County has a relatively average population. I looked at income and racial composition, two key factors known to influence student achievement. The median household income [12] is $59,575, while 9.9% of the population live below the poverty line. This is comparable to the national averages [13] of $50,303 and 13.2%. Racially the county is divided primarily among Non-Hispanic White and Black populations with a smattering of various other ethnic and racial groups [12]. This represents a slight trade off between Black and Latin populations when compared to the national averages [13]. The incoming class of second graders represents a cross section of this population.

Table 3.1: Comparison of Richmond County and US Population by Race
(Source: [12][13])

<table>
<thead>
<tr>
<th>Racial Composition</th>
<th>Richmond, Co.</th>
<th>USA</th>
</tr>
</thead>
<tbody>
<tr>
<td>White</td>
<td>65.1%</td>
<td>79.8%</td>
</tr>
<tr>
<td>Non-Hispanic</td>
<td>61.9%</td>
<td>65.6%</td>
</tr>
<tr>
<td>Black</td>
<td>33.3%</td>
<td>12.8%</td>
</tr>
<tr>
<td>Hispanic</td>
<td>3.6%</td>
<td>15.4%</td>
</tr>
<tr>
<td>American Indian</td>
<td>0.1%</td>
<td>1%</td>
</tr>
<tr>
<td>Pacific Islander</td>
<td>0.1%</td>
<td>0.2%</td>
</tr>
<tr>
<td>Asian</td>
<td>0.5%</td>
<td>4.5%</td>
</tr>
<tr>
<td>Multi-Racial</td>
<td>0.8%</td>
<td>1.7%</td>
</tr>
</tbody>
</table>

Participants were not compensated for their participation but participated on a voluntary basis as part of their normal school day. Experimental and control
groups were selected at random from the 40 participants. This process was conducted independently for each of the experimental instructional methods. The sample size, although small, was sufficient for the statistical analysis even accounting for absenteeism and non-participation. Those choosing not to participate received the same quality education as the control group (i.e. from a qualified educator), but no data were taken.

3.1.2 Treatments

Three separate experimental treatments were conducted, each examining a distinct instructional methodology involving simulations. In each treatment the experimental group was instructed using the selected methodology over a two hour period. During the same time period, the control group received adequate and similar instruction by a qualified educator. Age appropriate tests in the relevant domain were administered prior to and after each experiment. These tests came primarily from the Virginia Standards of Learning and classroom worksheets. Consequently, the tests focused on the lower levels of learning, specifically assessing fundamental knowledge and comprehension. More information about the tests is available in section 3.2.6.

3.1.3 Statistical Method

The key measures of the subjects' knowledge were taken through pre- and post-tests in each domain. The interquartile range of the pre-test scores was used to determine which, if any, of the subjects are outliers that needed to be excluded
from the experiment. The subjects were divided into control and experimental groups. The difference between the pre- and post-test was calculated along with a $t^*$ distribution of that difference. The $t^*$ distributions of the experimental and control group were compared for each instructional method in order to determine if there was a statistical difference between the groups.

3.2 Items of Interest

This section gives information about other items which are not typical of an APA article, but are absolutely key to a discussion of M&S or the intended experiment. These areas include an examination of the experimental instructional methodology, the instructors who participated in the experiment and the ethical considerations.

3.2.1 Experimental Instructional Methods

The three instructional methods are described below. Each was based on a theory of learning, attempting to draw on diverse means of educating. This was done to show that simulation instructional methods would be comparable to traditional instructional methods regardless of the exact method of application.

3.2.1.1 Simulation-Based Instruction

The first method was simulation-based instructional methodology. In other words, the method required that all instruction takes place through the simulation. In this scenario the experimental group received a very brief introduction to some
of the vocabulary involved and then were allowed to play and conduct specific experiments using the simulation. This method comes directly from the discovery theory that states that "[e]mphasis on discovery, indeed, helps the child to learn the varieties of problem solving, of transforming information for better use, helps him to learn how to go about the very task of learning" [14]. This is a popular theory and fits well into the M&S structure where a student could literally hold the sun in the palm of his hand. This method makes maximum use of the unique opportunities to experience and discover many principles first hand that are afforded students by simulation technology.

3.2.1.2 Simulation-Based Laboratory

The second method was a simulation-based laboratory instructional methodology. In this scenario the experimental group had a typical classroom learning experience for the first hour and then spent the remaining time using a simulation to further investigate the concept. The use of a laboratory is well established as a method of experiencing and investigating theoretical knowledge. M&S affords a limitless supply of laboratory experiences. Situations that are inaccessible because of geographical, chronological or financial constraints will become part of the learning experience. Others which must normally be conducted under rigorous safety standards can be accomplished in the absolute safety of the classroom.
3.2.1.3 Discussion Through Simulation

The third instructional method attempted to use a simulation to prompt classroom discussion. In this scenario the experimental group was given an introductory explanation and then is allotted an hour to work with the simulation. This is followed by another period dedicated to discussion. The latter period focuses on student experiences, perspectives and understanding of the simulated concept. This method comes directly from the generative theory of learning which states that "comprehension depends directly on what students generate and do during instruction" [15]. The lecture and simulation time are used as a jumping off point to a discussion where students are able to gain a personal perspective on the material.

3.2.2 Participating Teachers

Mrs. Bonnie Slusser and Susan Farmar, both teachers at Richmond County Elementary School, agreed to participate in these experiments. These teachers have been team teaching their individual classes for some time and have 47 years of combined teaching experience. Their primary role was to specifically care for the welfare of the participants although they also provided instruction to the control groups as well as supervision of the experimental instruction. They assisted in identifying the specific Standards of Learning which would be taught during the experiments. These were areas that traditionally were particularly difficult to teach. Because of the cooperation of these trained individuals the
experiments could be fully integrated into the average school day, greatly reducing the impact on research subjects.

3.2.3 Ethical Considerations of Human Subjects Research

After selecting a population and designing the experiment in order to best protect these vulnerable subjects it was necessary to present the research to a federally mandated human subjects review board. The Institution Review Board at Old Dominion University reviewed and approved the research protocol in the spring of 2009. The review board found that in spite of the protected population, the issues raised by the research were purely educational and would otherwise have qualified as exempt research. In view of the fact that the experiment was designed in order to specifically protect the subjects, the board felt that it qualified as exempt research and passed it as such without further review.

3.2.4 Simulation Characterization

These experiments were designed to give a general measure of the utility of M&S rather than a particular kind of model or software feature. Consequently, these experiments needed to control for the features that may have been seen as potentially additive in educational value. The simulations were designed to be minimalist and simple in nature. Animations were kept to a minimum with intentionally simple graphics. There was no attempt to use multimedia. Unnecessary features were removed to keep the simulations as simple as possible for this novice learner audience.
The fidelity of an educational simulation should be considered in relation to the intended audience. As the selected sample population was entirely inexperienced with these concepts, a low or medium fidelity was important. Details are hard to absorb and more details do not improve the situation. Especially for those untrained in these domains and ill-equipped to take in a broad spectrum of inputs, excessive details would be worse than useless. Under the strain of too much input the subjects could lose track of the vital thread of the simulation and consequently experience very poor learning. Therefore, a low or medium fidelity does not indicate that the simulation is of low quality, only age appropriate [16].

3.2.4.1 Magnetism Simulation

Magnetism used static time to display the effects of magnets. After each user event the effects of magnetic force and E fields were computed at a variety of locations in a grid. A variety of indicators could be selected and these indicators were drawn statically at each location in order to visually demonstrate those effects. The resolution was moderate and the fidelity of the simulation was high. Variables were limited to the static distance between measured locations and the number, size and location of user-placed magnetic sources.
The simulation has limited user actions, namely setting a magnetic field source, altering existing sources, clearing all sources and altering indicators. At the indicator locations the prevailing magnetic force is calculated as the sum of vectors calculated by Coulomb's law simplified to $F = q/r^2$. The indicators use the prevailing force to predict the electronic or $E$ field and magnetic or $B$ force.
3.2.4.2 Water Cycle Simulation

Water Cycle used discrete time intervals and units to track the flow of water through the various states in the cycle. Animations were set to calculate current quantities and location every 0.25 seconds. The animation was also bound by the "Mountain" at left and unidirectional "wind". The resolution and fidelity were very low. Variables included winds in the range of 0 to 5 units per second and temperatures from 32 to 120 units in the water reservoir where evaporation occurs.

Figure 3.3: Screen Capture from Water Cycle
Water is simulated in one of several states which are linked together according to the state diagram shown in Figure 3.7. The various states are depicted graphically as a reservoir, evaporation, condensation into clouds of various densities, rain and runoff. These are illustrated in Figure 3.4. As indicated, the fidelity and resolution are both very low so that the reservoir is considered limitless and such values as cloud density are only shown in one of four levels. The lack of numerical data is appropriate given the age group and the purpose of the simulation. In this case an accurate quantification of water in each state is less important than the ability to visually perceive the motion of water from one state to the next.

![Figure 3.4:Sprites Indicating Moisture Position and Quantity](image)

### 3.2.4.3 Symmetry Simulation

Symmetry used static time and was driven by user events. The simulation modeled the effects of drawing in symmetry similar to the physical experience of drawing with a mirror. The resolution was very high and the fidelity was moderate. Variables included the color and location of pixels as well as the number and type of symmetric facets. For the purposes of this experiment, only
the linear symmetry setting was used. This is the only simulation with a
previously designed and therefore more extensive interface. However, apart from
the variety of drawing tools, all other features were inaccessible during the
experiment.

![Symmetry Interface](image)

**Figure 3.5:** Screen Capture from Symmetry

Using one of the drawing tools and a color selection or else by applying the
stamps it is possible to alter the pixels of the drawing canvas. Without the
bilateral symmetry the image is identical to that which might be created in any
drawing program. However, with the bilateral symmetry an invisible vertical line
of symmetry divides the canvas in two. Any alteration to the pixels on one side of
the canvas is, literally, reflected in the pixels on the other side. Thus, where the
line of symmetry is located at \( x = X_s \), two pixels are related by the equation

\[
RGB(X_s - \Delta x, y) = RGB(X_s + \Delta x, y).
\]
3.2.5 Simulation Verification and Validation

For instructional purposes, verification requires that the simulation meet a specific set of lesson objectives, while validation merely requires that the simulation be sufficiently accurate to meet those objectives. This process for educational simulations differs from simulations built for other purposes as those would be verified to meet user requirements and extensively validated to prove predictive accuracy. The magnetism simulation was validated for accuracy by a Department of Defense Naval Surface Warfare Center expert in electromagnetic phenomena. Aside from some adjustments in terminology, it was found to be accurate within the bounds of the simulation fidelity and resolution. The water cycle simulation, however, was validated using a well-known state diagram shown in Figure 3.7. The system was neither numerically nor physically accurate beyond the basic structure of the cycle. In contrasting the two processes it is
evident that the magnetism validation was much more rigorous than that of the water cycle.

Figure 3.7: Transition of Moisture through State Diagram

Verification, however, was another matter. The Virginia Standards of Learning (SOLs) for the second grade define the requirements for magnetism and the water cycle to be:

The student will investigate and understand that natural and artificial magnets have certain characteristics and attract specific types of metals. Key concepts include
a) magnetism, iron, magnetic/nonmagnetic, poles, attract/repel; and
b) important applications of magnetism including the magnetic compass.

The student will investigate and understand basic properties of solids, liquids, and gases. Key concepts include... processes involved with changes in matter from one state to another (condensation, evaporation, melting, and freezing) [17].

The magnetism simulation was designed to show the forces related to these concepts. The simulation was verified through an informal introduction in the classroom a year before the experiment and was approved without further revision from the participating teachers. It was verified to be in line with the lesson objectives as they were defined by the SOLs. The water cycle simulation
was aligned not as directly with the SOLs as with the lesson objectives designed by experienced teachers and the well-known testing criteria. In short, the former appeared to be as thoroughly verified as it was validated, while the latter was verified by less rigorous standards.

It was determined later, however, that the verification process was ill-informed in the case of the magnetism simulation. In spite of, or perhaps because of, the greater fidelity and resolution, the model failed to obtain and keep the attention of the students. The more rigorous validation and strict adherence to the legislated lesson objectives did not take into account some of the realities of teaching. Students were distracted by the amount of information. The water cycle simulation, however flawed and blocky its execution, was ideally suited for the classroom environment. This highlighted the fact that verification in education is more complex than simply meeting the state-mandated learning objectives. The importance of a complete and holistic verification process cannot be overstated in an educational setting.

As a final note in this section, the symmetry simulation was based on mathematical equations and validated by visual inspection.

### 3.2.6 Metrics and Instrumentation

As discussed above, the tests that were administered to measure student achievement were designed by Brian Dillon from the Virginia SOL tests. The school staff supplied various worksheets, test questions and related material from which the test questions were drawn. The pre-test was designed to be different
from, but substantially similar to, the post-test. Principally, this required altering the direction of questions. For instance, one test gave the term and asked for the definition while the other test gave the definition and asked for the term. Below are some sample questions from the pre- and post-test which demonstrate how this was accomplished.

**Table 3.2: Sample Questions Reciprocated for Pre- and Post-Tests.**

When liquids becomes solid we call that _____.
When a liquid freezes it turns into a _____.

When a gas condenses it turns into a _____________.
When a gas _____________ it turns into a liquid.

When magnets push each other apart we say they repel attract
When magnets repel each other they pull together push apart

The two ends of a magnet are North/South East/West
Circle the names for the two ends of the poles North East South West

Because the tests were adapted from existing instruments, they differ greatly. Pulling from several worksheets the students were asked to highlight, circle or cross out images among a wide selection based on some property. Similar banks of images were used in each of the tests. In order to distinguish the tests, these images were reordered or new images were substituted. In the case of the water cycle tests, a large image of the cycle was used as shown in Figure 4.11. The image was identical in each test, but the questions were altered. In the pre-test the number referred to a question on that portion of the cycle. In the post-test the numbers were used to match terms related to the cycle. Using these techniques the instrumentation was able to test the same knowledge without resorting to exactly the same questions.
The tests also made use of a variety of question types. Some were multiple choice, others required matching items from two columns. In a few cases there were blanks to fill, but in each of these cases a word bank provided possible answers. In only one case was there a free answer question, which required the students to identify the energy source for the water cycle. One unusual question relating to magnetic materials came directly from the SOL tests and appears below. All these testing materials were reviewed by the participating teachers and amended to reflect their notes. At that point they were determined to be both within the ability of the subjects and capable of probative value. No further validity data, here referring to the term as used in education settings, was taken.
Table 3.3: A Variety of Question Formats

**Draw a line to match these terms**

<table>
<thead>
<tr>
<th>Frozen water</th>
<th>Water Vapor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Steam</td>
<td>Ice</td>
</tr>
<tr>
<td>Milk</td>
<td>Gas</td>
</tr>
<tr>
<td>Air</td>
<td>Liquid</td>
</tr>
<tr>
<td>Solid</td>
<td>Rock</td>
</tr>
</tbody>
</table>

_____ is a liquid.  
_____ is a solid.  
_____ is a gas.

When a solid ________ it becomes a liquid.  
When a liquid ________ it turns into a gas.

<table>
<thead>
<tr>
<th>solid</th>
<th>condenses</th>
<th>melts</th>
<th>evaporates</th>
<th>liquid</th>
<th>gas</th>
</tr>
</thead>
</table>

What gives water the energy to evaporate?  ____________________________

**Which of these piles could be cotton wool?**

![Options]

**Figure 3.9: Sample Questions from Magnetism Test**

The measurement of student achievement was obtained from the difference between the pre- and post-test scores, thus it is called the learning delta. The value may range from negative to positive values. In the case of a positive value the subject experienced an increase in knowledge presumably due to the learning experience. A negative value indicated a lower score on the post-
test than the pre-test. A zero value indicated that no change had occurred in spite of the learning experience. Test scores would have been affected by any number of unrelated variables including guessing. It is possible that chance may have given a high pre-test score in spite of a limited knowledge base. It is more likely that after a learning experience the subject may be able to make a more educated guess and substantially increase the post-test score. Although individual results may vary, variations of this type will be marginalized in the group results.
CHAPTER 4

EXPERIMENTAL RESULTS

The section on experimental results looks at each of the experimental instructional methods. In each case the pre- and post-test data is examined as well as a more detailed explanation of the instructional method. The detailed events of the instructional method are key indicators of the pedagogical advantages for this technology.

4.1 Symmetry in Simulation-Based Instruction Method

This experimental instructional method used the simulation for nearly all of the allotted instructional time. This method was an ideal means of examining the instructional affordances of M&S technology.

4.1.1 Pre-Experimental Knowledge Base

The initial knowledge baseline showed a fairly tight spread around a mean of 41.7% correct answers. This is a direct result of the test questions themselves. Because 90% of the test questions required either a yes or no answer and of those questions a significant number could be scored correct through no action at all, a blank test would receive a 32.5% score. This grouping suggests that the subjects were able to successfully guess the correct answer approximately half the time.
More telling than the random answers were the four figures which needed to be completed. In the pre-test the only figure which the children consistently completed was an airplane, shown below. This appears to be from the familiarity of the shape more than an understanding of the concept of symmetry. Therefore, disregarding the results from this figure, 87% of the subjects scored a zero on the remaining questions. In fact, this analysis also revealed that only two of the subjects scored perfectly in this area. Not surprisingly, these subjects scored extremely high in the test, but the IQR test does not justify removing those individuals from the experiment.

From the pre-test results we can draw two conclusions. First, the students were not able to do significantly better than guessing. Second, the majority of the students were unable to successfully apply the concept of symmetry. From these conclusions it would seem that, with a few exceptions, the selected population has a very limited knowledge base in this area.
4.1.2 Execution of Experimental Instruction

The experimental instruction began with an extremely brief introduction to the basic vocabulary terms: symmetry, symmetrical, line of symmetry. The subjects were then given approximately half an hour to "play" with the program in the non-symmetrical method. During that time the subjects were given no instruction but were encouraged to try the interface features. After that half hour of introduction to the interface, the subjects turned on the linear symmetry and attempted to make an "A" using symmetry. Largely these attempts were clumsy and consistent with the subjects' unfamiliarity with symmetry. Below are two samples.

![Figure 4.3: Subject Response to Initial Symmetry Tasking](image)

The next task was to attempt to make a "D". This was much more successful. Several of the subjects did an admirable job as shown in the first several samples. There were several subjects which were less successful as shown in the second set of samples. There were still some students who didn't grasp the concept completely. These individuals drew several things, including baldly non-symmetric versions as shown in the last sample. The number of these "cheaters" reduced with each additional task. After attempting "O" and "Q", as well as a patently non-symmetrical "F", the "cheaters" were entirely eliminated.
Subsequently, the subjects were again allowed additional "play" time for the remainder of the instructional period. The subjects were encouraged to create their own designs and to share these designs with their neighbors. Consequently, some of these designs, such as the skateboard below, quickly gained celebrity. Other subjects produced a variety of symmetrical figures. Most prominently among these were faces and bodies. These samples showed that the students largely understood the concept.

Figure 4.4: Subject Response to Secondary Symmetry Tasking

Figure 4.5: Subject Response during Post-Instructional Playtime
Other samples were less revealing. For instance, some continued to expand upon a basic theme, such as the original "A" task, or produced apparently random scenes. While these did not appear to show a particular grasp of the concept of symmetry, several of the individuals who produced these works also produced complex figures that did. In the table below the same subject produced the figure on the left and right in each row. While no direct correlation was drawn between these captured images and the test results, this does suggest that the apparently random play may have served to heighten the subjects' understanding.

Figure 4.6: Subject Work Showing Purpose and Apparent Randomness
4.1.3 Post-Experimental Results

The post-experimental test results were compared with the pre-test results to obtain a measure of change in the knowledge base per student or learning delta and the results are shown in Figure 4.7. A 90% confidence interval was calculated for the learning delta from sample populations of 19 experimental and 18 control subjects. The control interval was $-7.2\%$ to $+46.1\%$ with a mean of $19.4\%$. The experimental interval was $-9.9\%$ to $+49.9\%$ with a mean of $20\%$.

![Figure 4.7: Histogram of Learning Delta for Experiment 1](image)

By correlating the pre- and post-test scores we can look for patterns. In the figure below, the yellow line running through the graph highlights the "break even" point where a student shows no learning having taken place. Although there are a few subjects who individually did score below that line, the vast majority have improved. These comparisons indicate that there is no statistical advantage for either method of instruction.
4.2 Water Cycle in Simulation-Based Laboratory Method

This experimental instructional method used the simulation only for specific classroom experiments in the laboratory style. The laboratory method highlights the educational affordances which are unique, specifically the ability to manipulate and experiment on inaccessible systems.

4.2.1 Pre-Experimental Knowledge Base

The initial knowledge baseline showed a relatively wide spread around a mean of 50.1% correct answers. This test was significantly different from the symmetry test as most of the questions were multiple choice or matching a subset of terms from a list, drastically reducing the likelihood of inadvertent correct answers. The remaining questions are open ended fill-in-the-blank questions. However, given the range of responses, it would seem that the sample population has a moderate amount of knowledge in this area.
This existing knowledge base is unexpected but should not be altogether surprising. Even young subjects such as these have a daily familiarity with the states of water and weather. Ice, snow, water, hot drinks, rain, lakes, rivers and similar terms are frequently available in the average backyard and the test subjects would be familiar with many of the terms in the test. This may explain the existence of a higher-than-expected knowledge base. Regardless of how this knowledge base came to exist, its effects on the learning delta are a concern.

4.2.2 Execution of Experimental Instruction

The experimental instruction began with the introduction of several vocabulary terms related to the states of water specifically and other matter in general. The terms were displayed in a hand drawn image, reproduced below, as well as recited in a group and alone. After introducing the terminology, these states were linked to several common occurrences which would be part of the subjects' experience. The students were offered scenarios such as "Your clothes were wet and now they are dry. The water must have...EVAPORATED!" The answer was said spontaneously as a group, repeated by the instructor for clarity and then repeated in chorus again. After several of these scenarios, the subjects were able to coherently produce a chorus of correct answers when prompted.
Having explained the states of matter and, more specifically, water, the instruction passed on to the water cycle. This discussion centered on another hand drawn image substantially similar to this image from the testing materials. The students followed the path of a single "drop" of water from the liquid state, through the vapor state and back to liquid form. Each of the states and transitions was again repeated as a group. In this case a special effort was made to mention the part of the sun in supplying sufficient energy to drive the system.
Following the instructional period the subjects were given the simulations and asked to perform four experiments reproducing various climates. The first was a hot windy climate. The second was a cold windy climate. The third was a hot still climate. The fourth and last was a cold still climate. In each case the subjects were asked to comment on the frequency and duration of precipitation as well as the amount and spread of cloud cover.

4.2.3 Post-Experimental Results

The post-experimental test results were compared with the pre-test results to obtain a measure of change in the knowledge base per student or learning delta and the results are shown in Figure 4.12. A 90% confidence interval was calculated for the delta from sample populations of 17 experimental and 16 control subjects. The control interval was $-24.9\%$ to $+28.2\%$ with a mean of $1.6\%$. The experimental interval was $-33.9\%$ to $+31.4\%$ with a mean of $-1.2\%$. As with the first experiment, these result intervals were statistically indistinguishable. This suggests that the students benefited as well from the simulation laboratory as from traditional instruction techniques.
These test results shown in Figure 4.13 also indicate that 63.6% of the subjects experienced negative or no learning. This group was much larger than expected and far worse than the results seen in the first experiment. It is possible that this is a consequence of the increased knowledge base measured in the pre-test. High pre-test scores do not directly correlate with negative or no learning, but we might conclude that the increased familiarity with these concepts has made initial guesswork more successful giving the impression of a loss of knowledge.

4.3 Magnetism in Discussion through Simulation Method

This experimental instructional method used the simulation to expand on the same material which was explained during the instructional period. The
discussion method makes use of the educational affordances of M&S in order to further the students' understanding of the topic.

### 4.3.1 Pre-Experimental Knowledge Base

The initial knowledge baseline showed a relatively wide spread around a mean of 59.6% correct answers. This test was also primarily a set of yes/no questions that was very similar to the test used in the first experiment. However, as with the second experiment, magnets are a fairly common item and several of the subjects would have been able to successfully guess correct answers. The combination of these factors accounts for the increased spread and higher mean of the knowledge base.

![Figure 4.14: Pre-Test Interquartile Range (IQR) for Experiment 3](image)

As with the first test, however, there were a series of questions that were more revealing of the subjects' knowledge. These three questions each showed a pair of magnets with north and south poles. The subjects were asked to "[d]raw an arrow on the magnets to show the force they feel". The questions were accompanied by the sample image as shown below. At first glance, these questions netted no correct answers from 32.3% of the subjects and a majority of 64.5% got one answer correct. These results, however, belie a hidden failing in the test questions.
As can be seen above, the sample arrow points in the direction of the white-colored pole. In 54.8% of the tests the subjects followed that convention and labeled the magnets with arrows following the white pole. This led to an incorrect attraction for the first pair, a correct repel for the second pair and a tell-tale left motion for the last pair. This exact set of answers was common to each of the tests which followed the convention, indicating that those subjects were unfamiliar with the concept of like pole repulsion. In fact, although all of these subjects got exactly one answer correct, they were the only ones which correctly answered the second question. Therefore, by discounting the second question and reassessing the test results, we find that 87.1% of the subjects correctly answered no questions.

From the pre-test results we can conclude that the subjects have a limited familiarity with the use of magnets in everyday life. However, they were evidently unfamiliar with the concepts of magnetic poles and attraction. Although the test results showed that the mean score was higher than allowed for by probability, the subjects' knowledge seems to be limited to daily use of magnets and does not include these unfamiliar concepts. Consequently, the delta learning calculated from the final test may be somewhat skewed by these results, but a
closer look at the sections on magnetic poles and attraction may give better results.

4.3.2 Execution of Experimental Instruction

The experimental instruction began with an introduction to physical magnets. Each of the subjects was given one and asked to experiment for five minutes on various materials in the classroom. Afterwards, the subjects were asked to pair up and attempt to "stick" and "unstick" their magnets in pairs. The pairs were then asked to reverse one of the magnets and feel the repulsion force between the magnets. The instructor used a hand drawn image, reproduced below, to explain the existence of a north and south pole in the magnets. Having explained that the attractive and repulsive forces just experienced are a result of the different combination of poles, the instruction time concluded. In all the instructional period lasted approximately half an hour.

![South North](image)

**Figure 4.16:** Reproduction of Instructional Illustration of Magnet

For the following hour the subjects experimented with the magnetism simulation. This process began with a short period to familiarize the subjects with the interface. The subjects were then asked to simulate a pair of magnets in the four possible configurations of magnetic poles. Volunteers reproduced the lines of magnetic force by hand as shown below. The image on the left is a screen capture from the simulation showing magnetic forces. The image on the right is a
reproduction of the subjects' drawings. This drawing process was allowed to run at the pace of the subjects and consequently took the majority of the time. The simulation period ended with another brief period for free "play".

![Figure 4.17: Comparison of Simulation and Subject Representation](image)

The discussion period that followed centered on each of the subjects' experiences during the class. They were encouraged to point out materials that were attracted to the magnets and describe the forces they felt between interacting magnets. These comments were related to the corresponding drawings. The subjects were also allowed to reproduce the experiences described by their peers. In a couple of cases the subjects actively debated how the magnets interacted, demonstrating the attention which they were giving to the discussion.

### 4.3.3 Post-Experimental Results

The post-experimental test results were compared with the pre-test results in order to obtain a measure of change in the knowledge base per student or learning delta. A 90% confidence interval was calculated for the delta from sample populations of 13 experimental and 15 control subjects. The control interval was $-12.2\%$ to $+32.6\%$ with a mean of $10.2\%$. The experimental interval
was −19.8% to +32.7% with a mean of 6.5%. As with the prior experiments, these result intervals were statistically indistinguishable. This suggests that the students benefited as well from the simulation-based discussion as from traditional instruction techniques.

![Figure 4.18: Histogram of Learning Delta for Experiment 3](image)

The number of negative and zero learners from this experiment (17.9%) was far less than the findings from the second experiment (63.6%) and more comparable to those from the first (13.5%). This comparison suggests that the higher-than-expected pre-test results did not adversely affect the final results. Although most of the subjects did improve between the tests, neither of the methods shows any statistically significant advantage.
Again, the more telling force questions give a better idea of what student learning has taken place. The question read, "The arrows show the force felt by the magnets. We don’t know which pole is which. Color one side of the magnet on the right to match that force." In the post test both control and experimental subjects scored much better on the open-ended magnetic force questions. The post test reversed the question by asking the subject to color the poles to match the experienced force in the images below. One of the subjects was non-responsive, but of the remaining subjects 37.0% answered one question correctly, 44.4% answered two and 18.5% answered all three correctly. There was no marked preference for any pattern of answers. These results would indicate that substantial learning has taken place, but there was still no statistically distinguishable difference between instructional methods.
4.4 Comparison of Statistical Data

It would be helpful to compare the statistical data derived from the experiments; however, as the results were measured by three distinct sets of tests, they are in no way comparable. Instead we can make the reasonable assumption that the control subject learning is relatively comparable. Effectively, the large overlap between the control and experimental test results in each of the three instructional methods suggests that the three instructional methods are relatively comparable. While this is not conclusive, if the experiments were repeated with a different permutation of simulations and experimental instructional methods, it would be highly unlikely that they would obtain significantly different results.
CHAPTER 5
ANECDOTAL EVIDENCE

Anecdotal evidence from the three experiments abounds. They have been organized into sub-categories, specifically, criticism of the simulation, criticism of the experimental instructional methods and effects on instructional methodology. Some of the evidences and conclusions span two or more of these categories and their placement is somewhat arbitrary.

5.1 Criticism of the Simulation

Criticism of the simulation looks at specific ways that the simulation could be improved. This critique examines each of these improvements in order to optimize the effect in the classroom.

5.1.1 Professional Software Practices

In the course of the second experiment it became apparent that a capricious interface design had undesired effects on the subjects. The simulation controls for temperature and wind speeds increased as the sliders moved downward. In the case of temperature this was contrary to the hand-drawn image of a thermometer that had been used in the instructional period. Indeed, it was contrary to all terminology of "higher" and "rising" temperatures. Even in describing the required action of turning "up" the temperature this failing in the interface was noted.
This one failing was rather a minor issue. Considering the limited number of variables and controls, it should not have been more than an annoyance to the users and not a true impediment to learning and using the interface. However, the subjects reacted out of proportion and had extreme difficulties in overcoming this minor problem. Probably this extreme reaction was a consequence of the subjects' inexperience with computers in general as well as in the current lesson. This problem, which could have been resolved through the expediency of a limited user study, became a major impediment to executing the simulation-based laboratory lesson and led to time delays.

This experience suggests that an educational simulation is not a trivial piece of software and is subject to all of the same design and implementation procedures that should accompany any good software programming process. These simulations were designed to be elementary and worked with a minimal interface. The response of the subjects, however, even to a minor issue with the interface, shows that the applications themselves cannot be considered as simplistic. Professional software practices must be observed and the simulations must be well-designed and implemented. These applications are being introduced into an arena which, in its own way, is as temperamental as a nuclear reactor. The students' unfamiliarity with the topic coupled with even a minor deviation in the
software could become a major impediment to learning as well as a colossal drain of classroom time.

5.1.2 Simulation Training

Computer-based simulations are basically computer programs and users must be trained to use the software just as with any other program. Computer training generally fits into one of three categories. First, training may take place for basic computer literacy, especially in the case of the pre-personal computer generations that may display a degree of technophobia. Second, as in the case of the popular Microsoft Office applications, training may consist of a set of lessons introducing an extensive set of additional features. Third, as with proprietary business applications, training may focus on a specific set of tasks, teaching users to work with only a limited number of features.

Depending on the method of instruction, training for the use of educational software may resemble any of these three options. Where the students are completely unfamiliar with the use of computers, as is sometimes the case in early education, basic literacy training is required. Where the students will be engaged extensively in the software the teacher may be required to introduce a wide range of features. And if the teacher wishes to have the students perform a set of tasks, those tasks should be explained in detail. In short, the use of software in the classroom places the additional burden of user training upon the teacher.

This initial student software training can be looked at as a time investment and, according to one of the participating teachers, is hardly worth the trouble
when that investment will not return at least the same time savings elsewhere.

Even a minor loss will be multiplied when additional training is required for each new simulation. The most precious commodity in education is classroom time. Software training time cannot be overlooked as negligible. Therefore, the training to use ratio needs to be as low as possible.

In the case of these experiments the interfaces were intended to be overly simplistic and, therefore, very little time was allotted for interface training. Simplistic, however, did not translate into ease of use as the subjects still had frequent and often repetitive questions during two of the experiments. In the symmetry simulation, however, the users had a much better experience. Subjects were able to sit down and use the simulation in almost every case without further explanation. This, in spite of the fact that the symmetry interface was much more complex, tends to indicate another factor at work here.

In this case the simulation was based on an intuitive point-and-click design and the various features only enhanced the basic action. It was also observed that subjects learned a few features and stayed loyal to them, for instance working exclusively with the default pencil tool, stamps or the color yellow. These factors indicate that a set of easily learned central controls have a positive effect on usability. Based on these findings, a good design for an educational simulation interface, both for ease of training and usability, would be based on a basic set of controls with enhancements.

As with the Symmetry software, the central controls should be designed to be simple and intuitive. This will greatly reduce the initial investment in training
the students on the software. Additional features need to be ready-to-use extensions of those basic controls which do not require much additional explanation. Features which are only accessible through menus or function keys will require additional training just for users to find and activate them. Suppose, for example, that simulation output might require a line graph. Requiring the users to create, attach and adjust the graph will add a significant training burden in order to use that simulation. An interface which auto-attaches and adjusts the graph after a button click creation process would require less of an initial outlay in training time and provide a greater return on that initial investment.

5.1.3 Beyond Pedagogical Necessity

When the symmetry simulation was built, an informal user study was conducted with the target user audience of 6-8 year-olds. The test users indicated that “stamp” images would enhance the final product. Although these stamps added no value to the learning objectives, they were added as a user preference. Surprisingly, the experimental subjects preferred the stamp images to any other input method. While not strictly useful, this added feature may have provided much of the subjects’ motivation to experiment on their own and grasp the concept of symmetry.
The verification process for the water cycle simulation was based on teacher experience and testing objectives. During the simulation free play, however, several subjects expressed a desire to produce torrential snowfall—a feature that had never previously been considered as a pedagogical necessity. As the simulation had no provision for snowfall, nor did it account for temperature at the point of precipitation this was entirely impossible with the current version of the simulation and the subjects quickly lost interest. Although the simulation met the expected pedagogical requirements of the simulation, it failed to grab the attention of users because it lacked this one feature.

During the design phase even a limited user study would help to improve the simulation. Designing strictly for the classroom may meet the requirements for validation and verification, however, features beyond the pedagogically necessary should also be considered. In order to gain and retain user interest, it may be necessary to consider additional features merely for the appeal of a gimmick. If successful, the simulation will be more likely to draw the attention of the users and complete its primary function as an educational tool.
5.1.4 Using a Spiral Process

Although all of the simulations were verified by the participating teachers and validated mathematically or conceptually, it is difficult, as one of the teachers commented, to "hit the nail on the head" when it comes to meeting the teaching objectives. The unexpected failure of the magnetism simulation to anticipate some of the realities of the classroom indicates that a classic waterfall development process is bound to fail. Although the simulation served exactly as expected, the classroom experience could have usefully informed another development iteration.

Perfect design is known to be difficult to achieve in any case; however, the conflicting and unexpected requirements of the classroom exacerbates the problem. Just as experienced teachers have acquired through trial and error a store of preferred materials, the creation of simulations for the classroom should be expected to grow by experience. A spiral or iterative development process would allow for reassessment and improvement of the software, even a significant alteration of the basic design when necessary. The spiral process in particular is agile in correcting errors and backtracking to replace parts of a design which have proven to be a poor choice. At the very least the design should make allowance for suggestions and improvements when the software is first introduced in a live class.
5.2 Criticism of the Experimental Instructional Methods

Criticism of the experimental instructional methods reflect on the instructional application of M&S. These criticisms look at specific effects of M&S in the classroom, good or bad, and how these might be avoided or taken advantage of in that setting.

5.2.1 Importance of Play in Learning

Educators must be able to justify purchases for the classroom. While it is taken for granted that algebra books and pencils are a necessity, it is harder to justify the purchase of software that can be characterized as a toy or game. Similarly, teachers must account for classroom time where it is equally difficult to appraise "play" time as a worthy expenditure. However, play has tremendous value. Basing her opinion on research and the position of the Association for Childhood Education International, Dr. Joan Isenberg of George Mason University reaffirmed the value of play in learning. She said,

"We know that active brains make permanent neurological connections critical to learning; inactive brains do not make the necessary permanent neurological connections. Research on the brain demonstrates that play is a scaffold for development, a vehicle for increasing neural structures, and a means by which all children practice skills they will need in later life" [18].

The importance of play was also evident in the various experimental instructional methods. The experimental group working with symmetry was allowed a fairly open schedule and much of the time was reserved for play. This
group experienced fewer interruptions during the specific tasks and greater comprehension of the interface. They also exhibited a great deal of peer-to-peer instruction. These results were unlike those in the other experiments, where play time was severely limited. In those experiments subjects depended far more on teachers for instructions and assistance while working with the simulations. While not controlled for in any of the experiments, the additional play time in the first experiment appears to have been a great asset.

Another outcome, however, was more directly tied to play time. With the additional time and encouragement to experiment, subjects created a wide variety of imaginative images. The diversity of figures, even with the sharing and distribution of ideas throughout the group, was extraordinary. Had the subjects been limited to a narrow set of learning objectives and events, this creativity would have been lost. If the interface had been more task oriented or restrictive, the opportunity for this type of playfulness would also have been lost. The subjects benefited greatly from the opportunity to engage in inventive trial and error.

These findings, while not conclusive, suggest that the critical play time which Dr. Isenberg described should be considered while creating an educational simulation. While the ability to perform specific tasks and experiments should be included as an integral part of the simulation, a sandbox should also be available for free play. The sort of familiarity which students gain during play time should cut back on instruction time as well as software training.
5.2.2 Differences Between Instructional Methods

The three instructional methods were designed to highlight any appreciable difference between simulation technology and traditional instruction. Was there, however, any appreciable difference between the three methods? As discussed earlier, the test scores show that each method was roughly equivalent to traditional instruction methods and, assuming that traditional methods are roughly equivalent, this would not point any statistical difference between the various methods. Therefore, the anecdotal evidence will have to describe any significant difference.

The primary characteristic of the simulation-based instructional method was, surprisingly, teacher effort. As these experiments called for one simulation for each subject rather than a whole classroom or small-group simulation, the teacher needed to cycle behind each of the subjects to observe, encourage and instruct. This places a tremendous strain on the resource of a single teacher. Various means of relieving this stress might be suggested, including team teaching, engaging the students as working teams or running a single simulation for the whole class. However, this experiment included two instructors and was still an incredibly labor intensive process.

The primary characteristic of the laboratory-based instructional method was specificity. The use of a simulation as a laboratory means that the simulation need not be widely applicable as the learning objectives only require the use of the simulation within tightly constrained parameters. The simulation may have been extensive and allowed for a great deal of detail, but for this application it
need only be specific to the set of circumstances which the teacher wishes to model, such as a rain storm. The simulation needs to be well suited to this very specific set of circumstances and need not be suited to any other.

A strictly lesson-related simulation may only have a limited number of uses; however, as some of the previous evidences have suggested, the ability to expand beyond the restrictive bounds of lesson material may be valuable. Any simulation should have a wide range of potential uses; however, to make it equally accessible for specific tasks, it would benefit greatly from a set of default or user-defined parameter settings. This set of parameters would serve to assist teachers by initializing the simulation instantly and automatically. The simulation designer may want to consider these specific usage scenarios and include them in the simulation controls. The user experience would be greatly enhance by allowing the users to create a solar or lunar eclipse without additional effort. This would also lead to a decrease in training during classroom time.

The primary characteristics of the discussion through simulation instructional method were timing and interruptions. The work prior to the simulation is bound to raise several questions and comments from the students which the teacher will wish to address during the simulation phase. The simulation phase will no doubt raise additional issues which the teacher would like to cover during the classroom discussion. Some of these questions and comments will be invaluable observations. For example, a single student in the group may notice a key principle the teacher intended to point out. These
questions will come at the most inopportune time and will force the teacher to make difficult decisions.

If these questions are addressed before the teacher has laid the groundwork, the discussion may be lost on several of the students and may need to be repeated. On the other hand, delaying the question may mean losing the interest or insight of the student. As the purpose of the simulation is to increase student participation, hampering spontaneous discussion seems a poor choice. Experienced teachers are familiar with these issues and will be prepared to handle them. A proper balance between tolerating nearly continual interruptions and decreasing the amount of student participation during the discussion time will depend on teacher preference and experience.

5.2.3 Ubiquity of the Computer

In spite of programs like Labels for Education and an effort by local and state education boards, computers are not yet ubiquitous. The US Government keeps records on the estimated percentage of households with computers and internet access [19]. The latest numbers from 2003 indicate that only 61.8% of US households have a computer and 54.7% have access to the internet. Also, according to the same statistics 86% of children with access to a computer will use it. There is also an extraordinary disparity between the extremes of income. While a family income of more than $100K corresponds with 91.6% of home computer use, that number drops to a meager 41.2% for those who earn less than $25K.
These figures indicate that low income areas will continue to see many school age children without computer access for some time to come. This disparity is likely to be maintained until the home computer becomes ubiquitous or until schools begin to loan laptops just as they currently loan textbooks. Until that time teachers will not be able to use computer-based simulations in homework assignments. While teachers may, and some currently do, include computer-based enrichment activities, these cannot be 100% required or some students will be manifestly blocked from completing their assignments.

The same statistics show that the availability of in school computers has a much narrower margin from 83.8% at the lowest income levels to 90.5% at the highest. Within the classroom students have equal access to computers; therefore, the use of computer-based simulations is entirely appropriate within the school itself. Classroom assignments can be used as a substitution for computer-based homework. Teachers will still need to be selective in making the assignments, but the venue and availability of the technology will at least provide equal opportunity to all students.

5.3 Effects on Instructional Methodology

This section looks at the ways in which M&S technology would alter the typical classroom. As there appears to be a strict dichotomy between educators who prefer technology and those who do not, both views are examined.
5.3.1 Expert Availability

The background of grade school teachers is specific to the teaching and instruction of children but does not include a great deal of coursework in specific subjects. While specific curriculum varies from university to university (Table 5.1), in general an educational licensure program only requires a bare minimum of familiarity with specific academic areas, choosing instead to focus more on methods and theory of education. In theory a trained teacher is just as adept at teaching reading and writing as physical sciences and math. This was discussed with the teachers prior to and after the experiment in order to explore the value of an expert instructional simulation in the classroom. Each of the participating teachers noted that coming out of university they may have had the capability to teach any subject, but very little practical experience and, more importantly, none of the required resources. First year teachers may borrow resources from more experienced teachers, find them through a school system resource teacher or by personal initiative. It may take several years for a teacher to accumulate a full curriculum that is age appropriate and will fit the lesson objectives. Thus a software expert resource would be invaluable for new teachers, but not as useful for experienced teachers.
Table 5.1: Elementary Education Degrees in Top Virginia Schools  
(Source: [20][21][22][23][24])

<table>
<thead>
<tr>
<th>Requirements for Elementary Education from Top Virginia Schools</th>
</tr>
</thead>
</table>
| **University of Virginia** | Teacher Education - Elementary Education | General educational requirements  
3 credits each of  
Math  
Science  
Social Sciences  
6 credits literacy education |
| **Virginia Tech** | Masters in Education in Curriculum and Instruction in Elementary Education | Bachelor's degree in any field  
Option 1 or 2  
3 credits each of  
Math  
Science  
Social Sciences  
Option 3  
3 credits in Social Sciences  
9 credits in Math or Science |
| **College of William and Mary** | Master of Education - Elementary Education | 3 credits each of  
English  
History  
Social Sciences  
Math  
Science |
| **George Mason University** | Elementary Education PK-6 Initial Licensure | Several courses in methods of and integration of technology while teaching math, science and social studies |
| **Virginia Commonwealth University** | Master of Teaching - Early and Elementary Education | 12 credits in all (≥3 each)  
Biology  
Chemistry  
Physics  
3 credits in Math  
6 credits in Statistics |

Experienced and less-experienced teachers were interviewed in order to compare and contrast their opinions about an expert system. For the more experienced teachers, any additional software would at most become just another resource. It would not drastically alter the instructional method, nor the long-
established lesson plan. "I might use it for tangents and to answer questions," says Mrs. Bonnie Slusser, who has been teaching the second grade for the last two decades, "but it is not appropriate for building a curriculum." While the software might be a ready resource for finding answers and a useful enrichment exercise, the experienced teachers feel an inherent distrust of the technology and would always prefer the established repertoire of worksheets, books and classroom games. Still, they would use a software expert system if it was comprehensive and matched their educational objectives.

For the less-experienced teacher, whose training has given them a more favorable opinion of technology in the classroom, this may not be the case. Mrs. Rachel Ferguson who has been teaching for less than five years strongly favors the use of technology. "While I don't depend on technology entirely, if I lost a computer it would hamper much of my plans for the day." For her the primary question is one of reliability. "Even if I used an off-site server-run program, it wouldn't do me any good if the school's internet connection failed." She agreed that with dependable technology a comprehensive system of simulations and lessons would be enormously useful, and she would not feel apprehensive assigning significant class work from a well-built and trustworthy application.

Both experienced and inexperienced teachers, however, agree that any expert system should meet certain requirements. Teachers can already search for worksheets, assessment materials and games online and through well-known publications. In order for an expert system to be useful, it would have to seriously reduce the search time. A system with a poor classification or search methodology
would be next to useless. Ideally, such a system could be ranked according to
topic and keywords as well as age, level or grade. If successful, teachers would be
able to access all the required material within the time normally allotted for
planning.

Another important aspect both groups agreed on was currency. The recent
changes to the Virginia SOLs have highlighted the fact that all resources are
eventually outdated. All three of these teachers will be revising their materials to
reflect the addition of the Southern Ocean, the change in an Indian tribal name
and the loss of a planet. In order for an expert system to be useful, it would have
to be edited or revised each year in order to make it as accurate as possible.
Preferably, this is work which teachers will not need to undertake themselves, but
will be included in the software upgrades and patches. This would also add
perceived value to the expert system simply by reducing the teacher workload.

5.3.2 Continuity of Technology

In discussing the three experimental instructional methods, both
participating teachers were in consensus that better results can be obtained
through a sequence of classes rather than a single instructional event. In their
opinion incorporating simulations into the daily set of teaching, assessment and
review periods over several days would be an effective instructional paradigm.
During a series of instructional periods the simulation could take the place of play
time, games, review, laboratory and any number of other uses as needed. This
continuity would give students increased proficiency with the software and lead to many opportunities for questions, answers and additional investigation.

If this inclusive method was practiced during an entire school year, it would affect the classroom in two key ways. First, the initial time invested in software training would have a better return. Especially if the simulation interface were to develop around a central control structure as described above, learning the software once would be sufficient for the whole year. Second, the software could be used to review sections of material throughout the year. Because the software would be tied directly into the lesson during class time, a review of the software would be as effective as any other form of review.

It is uncertain how far these effects could be extended. For instance, if a class of students were taught the use of a particular software suite, could the same students benefit from the continued use of that software over several years? Could the software be used by the students from home to extend the school year and diminish the loss over summer breaks? Such software would have to be extensive, spanning multiple subjects and grades. Even so, eventually the law of diminished returns must apply. This may bear further study, but as the experiment would require the concentrated commitment of several students and families for a lengthy period of time, it may be impractical for the present.
CHAPTER 6
FUTURE WORK AND CONCLUSIONS

6.1 Conclusions

This experiment was intended to complete the first requirement of the newly proposed research paradigm in order to measure the effectiveness of M&S in the classroom. Each of the experimental instructional methods proved to be as good as traditional instructional methods and no statistical advantage was observed. The experimental outcomes suggest that student achievement wouldn't be harmed if some teaching time were conducted through simulations. Simulations could be used as a substitute for traditional learning activities in specific and appropriate circumstances.

In order for M&S technology to be accepted in the classroom, the case must be made that they are efficient and economic. The economic determination will and must be based on the realities of school board budgetary decisions. Research, however, can provide descriptions, applications and explanations of the place of these technologies in the classroom, bringing evidence for the efficiency side of the case. Once this case is made to the school superintendents who are experienced in making these tough economic decisions, it is they who may make the decision in the matter of economy. In effect this is the free market system working in real time to weigh the cost and benefit. The providers of such technologies as exist will present there products and prices to the superintendents who will be equipped by research to examine both efficiency and economy.
In order to provide decision makers with the best possible information concerning efficiency, researchers can no longer be concerned with a search for the magic bullet of technology which will solve all educational dilemmas. Future research into this arena should focus on the affordance, the aptness, the expedience and the simplicity of certain technologies in the realities of the classroom. Researchers may not be able to identify a cure-all, but they can be instrumental in distinguishing the best means of conserving educational resources. Decreasing teaching time, reducing classroom sizes, increasing parental involvement and boosting student participation are just as valuable as an appreciable increase in test scores. Positively affecting the entire learning process should be the goal.

Student achievement, while not directly tied to a specific attribute or form of technology, must be maintained at the same level as traditional instructional methods. However, the potential for technology to supplement, enhance and extend the effectiveness of traditional methods has not yet been fully realized. Researchers need to consider the anecdotal evidence associated with new educational methods not just as a side effect, but as a primary consideration. The experimental evidence shows that M&S technology meets the "as good as" standard and where these anecdotal evidences point to a potential to improve the learning experience, those evidences should be examined with as much direct attention as the statistical results obtained from achievement test scores.
6.2 Future Work

In proposing this radical deviation from current research trends, there is a responsibility to take the first step. Therefore, in proposing a future course of research to expand upon these results, it is necessary to apply this new direction. If there had been any suggestion that one of the experimental instructional methods provided an advantage, a larger population experiment might have been able to show that advantage with statistical conclusiveness. However, these experimental results make it unnecessary to attempt such an experiment as it seems likely to reach the same result even with a much larger population. Instead, looking to the anecdotal evidence, it is necessary to examine M&S technology for its effects on the instructional environment. The following are four topics for future research which follow this suggested course of action.

6.2.1 Reduction in Assessment and Review

Simulation in the classroom can provide a secondary source of assessment and review which does not directly require teacher supervision or effort. As suggested in these findings, if the simulation were taught and used during a lesson, it might then be used by students on an individual basis to perform certain tasks. These tasks essentially provide additional review of the subject and may be interpreted as a form of assessment. Teachers can mandate that the students attempt these exercises as an offline tool. They may also mandate that these exercises be attempted as homework or classroom assignments.
Researchers could expand upon these results by creating a set of educational simulations, consistent with the criteria described above, and measure the time saved by substituting simulations for review and assessment. A final assessment will help to establish that the learning experiences are still comparable, but simulation tasks might be substituted in varying degrees for most or all of the typical assessments and review. If this research is able to show a savings of classroom time during a week or month long unit study, that time savings can be extrapolated into a yearly savings. Even a savings of one week of classroom time would be substantial as that time may be reapplied by the teacher to enhance, extend or enrich the learning experience.

6.2.2 Harnessing Play Time

As noted earlier, play time is known to be an essential part of the learning experience. Indeed, it is the framework upon which learning takes place. Introducing playtime simulations into the classroom may provide an alternate method of launching a new concept upon the students. Students might be more interested in a concept which they have already seen in simulated application. How, for instance, might class participation improve if one or more of the students has already measured masses on the moon and Jupiter before the teacher explains the difference between mass and weight? While this example is patently more advanced than the typical grade school science class, the potential for such a simulation is identical for any other application.
Research into this area should focus on the increase in knowledge base prior to instruction time as well as any corollary effects observed during instruction. This may include a general increase in participation or participation of a previously non-responsive group. Harnessing play time may also translate into a reduction in instructional time, so this research should focus on setting and obtaining instructional goals so that the relative efficiency can be compared. Again, a qualified teacher can capitalize on savings in instructional time or an increase in group participation in order to enhance the learning experience.

6.2.3 Providing Additional One-on-One Time

Educators are aware of the fact that class size is inversely proportional to student achievement. According to the California Department of Education [25], the goal is to have a maximum of 20.44 students per classroom. A four year study funded by the Tennessee General Assembly examined the student/teacher achievement ratio (STAR). That study "consistently and significantly demonstrated the advantage of small classes (13 to 17 students) over regular sized classes (22 to 25 students) and regular sized classes with a teaching assistant" [26]. In other words, a single teacher with a less-divided attention was more valuable to students than an additional adult assisting in the student supervision.

If a simulation could be employed in the classroom, one effect might be to decrease the amount of group time and allow teachers an increased portion of time to dedicate to one-on-one instruction. In effect, this type of experiment would replace the human assistant with a simulation related to the lesson.
Researchers would be able to determine if such a replacement has an effect on the STAR. Considering the possibility that a teacher with 25 students might manage to replace one hour a day of group instruction, in the course of a full month that teacher would have been able to meet each of the students in the classroom with nearly a full hour of individual instruction. Potentially, this research could point to another way of resolving the known problem of reducing class sizes.

6.2.4 Continuity of Technology

Could the learning experience be extended or enhanced by the substantial and continual use of a single technology through the school years? This fascinating theory was introduced to the participating teachers in order to gauge their response. In general, the theory was thought palatable, but no direct or anecdotal evidence indicates its veracity. Researchers should design an experiment to look at prolonged use of technology in the classroom, specifically focusing on related interfaces, tasking and assignments. This research should track the facility with which students use and relate to the system, assessment scores and the effect on classroom discussions as well as any unexpected results. This research may open the door to several other issues related to prolonged usage.
REFERENCES


seniors and percentage of graduates taking SAT, by state or jurisdiction:

Selected years, 1987-88 through 2007-08. [Online]. Available:


VITA

Brian Spencer Dillon

Department of Electrical
& Computer Engineering
Old Dominion University
231 Kaufman Hall
Norfolk, VA 23529
(757) 683 - 3741
bdi11003@odu.edu

Biography
Brian Spencer Dillon was born in Washington state in early 1981. He attended various elementary schools, but was home schooled through the majority of his primary education. He attended high school where he graduated with honors in the class of 1999 from Rappahannock High School in Warsaw, Virginia. He served two years as a full time missionary in Honduras and Belize, learning Spanish and acquiring a love for teaching. He married Stacey Stephens in 2002 and together they have three children. They currently home school their two oldest children in the same county where he attended high school and he continues to volunteer in the local school district.

Education
B.S., Computer Science, 2007 from Virginia Commonwealth University
M.S., Modeling and Simulation, 2010 from Old Dominion University

Professional Experience
Teaching Assistant to Dr. Lyle Lichty of the Cornell College 2003
• Produced online documentation to accompany middle school summer science program.
• Produced two computer simulations on complex harmonic motion and wave propagation.
Webmaster for Dr. Richard Wills of Virginia Commonwealth University 2005
• Created a web site which would be the central forum for a class-wide debate.
Computer Programmer at Naval Surface Warfare Center Dahlgren 2007
• Produced software on various projects for the US Marine Corps and Navy.
• Produced documentation (requirements, design, technical reports, training, online help and patent applications).
• Completed Acquisition Workforce Level III SPRDE certification.
• Applied for a patent on the Heap-Sorted Optimization Algorithm (pending).