# Gender Differences in Learning Physical Science Concepts: Does Computer Animation Help Equalize Them?

by

Laura Lee Jacek

A DISSERTATION

submitted to

Oregon State University

in partial fulfillment of
the requirements for
the degree of

Doctor of Philosophy

Presented May 8, 1997
Commencement June 1997

UMI Number: 9734840



UMI
300 North Zeeb Road
Ann Arbor, MI 48103

Doctor of Philosophy dissertation of Laura Lee Jacek presented on May 8, 1997.

APPROVED:

Major Professor, representing Geography

Head of the Department of Geosciences

Dean of Graduate School

I understand that my dissertation will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my thesis to any reader upon request.

Laura Lee Jacek, Author

AN ABSTRACT OF THE DISSERTATION OF

Laura Lee Jacek for the degree of Doctor of Philosophy in Geography presented on May 8, 1997. Title: Gender Differences in Learning Physical Science Concepts: Does Computer Animation Help Equalize Them?

Abstract approved:

A. Jon Kimerling

This dissertation details an experiment designed to identify gender differences in learning using three experimental treatments: animation, static graphics, and verbal instruction alone. Three learning presentations were used in testing of 332 university students. Statistical analysis was performed using ANOVA, binomial tests for differences of proportion, and descriptive statistics. Results showed that animation significantly improved women's long-term learning over static graphics ($p = 0.067$), but didn't significantly improve men's long-term learning over static graphics. In all cases, women's scores improved with animation over both other forms of instruction for long-term testing, indicating that future research should not abandon the study of animation as a tool that may promote gender equity in science. Short-term test differences were smaller, and not statistically significant. Variation present in short-term scores was related more to presentation topic than treatment.

This research also details characteristics of each of the three presentations, to identify variables (e.g. level of abstraction in presentation) affecting score differences within treatments. Differences between men's and women's scores were non-standard between presentations, but these

differences were not statistically significant (long-term $p = 0.2961$, short-term $p = 0.2893$). In future research, experiments might be better designed to test these presentational variables in isolation, possibly yielding more distinctive differences between presentational scores. Differences in confidence interval overlaps between presentations suggested that treatment superiority may be somewhat dependent on the design or topic of the learning presentation. Confidence intervals greatly overlap in all situations. This undercut, to some degree, the surety of conclusions indicating superiority of one treatment type over the others. However, confidence intervals for animation were smaller, overlapped nearly completely for men and women (there was less overlap between the genders for the other two treatments), and centered around slightly higher means, lending further support to the conclusion that animation helped equalize men's and women's learning. The most important conclusion identified in this research is that gender is an important variable experimental populations testing animation as a learning device. Averages indicated that both men and women prefer to work with animation over either static graphics or verbal instruction alone.

ACKNOWLEDGEMENTS

Special thanks to all of those individuals who put time into looking over drafts of this manuscript, and made valuable comments and suggestions, most especially Dr. A. Jon Kimerling, Dr. David Hibbs, Dr. Charles Rosenfeld, Dr. Julia Allen Jones, and Dr. W. L. Schroeder. My thanks also to Vetle Torvik, who tirelessly worked with me to assure that the statistical analysis in this dissertation was both accurate and understandable. His willingness to put his time and energy into this project has been a true gift.

Along with those who have directly contributed to my research, there are those names that are typically not heard, and only rarely receive the notice that they so richly deserve. These are the friends and family that stand behind each graduate, selflessly giving of their time and energy. In my life, I would like to especially thank my family, Joe Wisecaver, Heidi Fassnacht, Kathleen Avina, Marganne Allen, and Darlene McCabe for the times they have listened, hugged, encouraged, sympathized, and simply been there when I needed them, even at the expense of themselves. To each of you I would like to say, I love you very much my friends; thank you for the greatest gift that you could have given me, the gift of your friendship. My greatest thanks goes to my Lord and Savior Jesus Christ, without whose help, none of this would have ever been possible. Truly this dissertation is the work of far more than simply its author.

## TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

## LIST OF FIGURES

## LIST OF FIGURES (Continued)

## LIST OF FIGURES (Continued)

Page

## LIST OF TABLES

# Gender Differences in Learning Physical Science Concepts: Does Computer Animation Help Equalize Them?

## INTRODUCTION AND JUSTIFICATION

Do men and women benefit to different degrees from the use of animation in classroom computer based instruction (CBI) while learning physical science concepts? Computer animation (graphics with a moving or changing component) has experienced steadily increasing usage in instructional programs and education over the past two decades. This is true, even though consistent empirical support for animation as a valuable tool in education is lacking (Rieber 1989a). There are several theoretical discussions of animation present in the literature, but few empirical studies testing the efficacy of animation are available.

As the computer industry turns out new products right and left, research struggles to keep up with the surging tide of educational technology. There aren't enough studies being done to determine with any degree of certainty the general usefulness of animation in CBI. It would be of value in and of itself, to be able to conclude whether gender is an important variable in studies examining student performance using animation in CBI. Gender may not significantly influence experimental results in animation. If it does, however, the elimination of gender as a possibly confounding variable, currently seldom accounted for in research designs, could serve to elucidate future experimental results and further strengthen research conclusions badly needed in the physical sciences educational arena.

Theory backing the use of animation in education rests mainly on Paivio's (1971) dual-coding theory, a theory well accepted and frequently cited in the literature of several disciplines. Paivio's theory postulates that pictures and words are coded independently of one another. Thus, the redundant coding of a picture and words will result in greater recall.

> Dual-coding is more likely to occur when the content is highly imageable...Animation, like any graphic, should be expected to aid the recall of verbal information when it serves to precisely illustrate a highly imageable fact, concept, or principle. (Rieber 1989a, p. 375)

Results are mixed with respect to animation in education, but for the most part, in adult learning, animation does not appear to facilitate subject visualization (called imaging in Rieber's paper) as measured through improved information recall (Rieber 1989a). Children, however, appear to benefit conditionally from animation, and animation may be a valuable tool for anyone who has a low capacity for visualization (imaging) (Blake 1977, Rieber 1989a).

One group of individuals who appear to have a lowered capacity for visualization are women. Studies of women and women's learning aptitudes, have historically identified several areas where women's learning strengths appear to differ from men's. While evidence is sketchy and debate high for many of these differences, those studies citing gender differences in certain visual-spatial abilities seem to be accurate and well-supported (McGee 1979, Linn and Petersen 1985). Following this train of logic, if animation facilitates visualization in lowered visual-spatial ability groups, and women have lowered visual-spatial ability, then women will benefit from the use of animation in instructional presentation more than will men.

In my Master's research (Edgeman 1994), I compared the effectiveness of teaching successional vegetation dynamics using animated and static graphics. Learning two different sets of information, results showed that subjects were at worst, not harmed by informational presentation in animated format, and at best, were aided by computer animation. Further, all students using the animated format scored above the traditional 50% cut-off line for exam failure, while all students using a static graphic format scored below the 50% cut-off (Jacek and Brewer in review 1997).

An interesting sidelight to my research showed women benefitting consistently from animation, while men did not. Using one of the two animated presentations, both men and women benefitted. On the second animated presentation, however, women's scores benefitted, while men's suffered. These differences were neither significant, nor part of the thesis' critical analysis due to the small sample size, but they do raise interesting questions. Tentative results then, seem to point towards conditional usefulness of animation as a learning tool. This is consistent with findings of previous researchers. One of the possibilities that could account for the difference between the presentations is the visualizability of the information being presented. One set of information might have been more easily imaged without graphics, the other clearly benefitted from presentation in a graphic format (Jacek and Brewer in review 1997). Does animation aid in visualization? If so, the differential benefits gained by men and women make sense. Women, a group with lowered visual-spatial abilities, would clearly benefit more from tools supporting those aptitudes.

Since the first American Association of University Women Educational Foundation national study (conducted in 1885) sought to debunk the commonly held opinion "...that higher education was harmful to women's health" (AAUW 1992, p. v), much advancement has been made in the study of women's learning. As educational research has progressed, it has done more and more to examine and identify how traditional methods of instruction are influencing women's educations. Answers are being sought detailing women's unique needs in the classroom, and how women's representation within experimental populations may affect quantitative results in academic studies. This is certainly a valuable endeavor considering that women make up just over 50% of the student body in the United States, and yet are so often underrepresented in academic studies focusing on developing or identifying effective educational tools (AAUW 1992).

The implications that the research presented in this dissertation holds for the academic community in the field of CBI are high. First and foremost, women's education and ways of learning have been visible in the literature for years. One of the quandaries ever present in the literature is women's low persistence at the study of physical science and engineering based on physical science. If animation can be conclusively identified as a benefit to women's learning in the physical sciences, new avenues are opened for improving instructional design to better connect with a woman's unique learning strategies, perhaps facilitating an increase in persistence of women in science and engineering.

Secondarily, if differences are identified in benefits men and women gain from animated instruction, academic research focusing on animation that has not accounted for gender effects in their experimental populations will become suspect. Experimental literature on animation in CBI is scanty, and what there is often presents conflicting results, possibly resulting in the current lack of consistent support for animation in CBI. Separating out studies that have very uneven gender ratios in their experimental populations could aid in elucidating experimental results.

Before 1970, research showed that standard graphics such as those in textbooks neither enhanced nor detracted from information acquisition. Only through time and additional research did graphics become accepted as valuable learning aids under certain conditions. It is possible that animation is undergoing this same research progression; we are still searching for the recipe that will show us how animation will best enhance learning (Rieber 1989a). This research represents one more step in this process.

In this dissertation, I will test several hypotheses relating to gender differences in learning using computer animation in CBI:

1) Women benefit more than men do from the use of computer animation in CBI, when learning physical science concepts.

2) Type of instructional medium (static vs. animated) has no difference in effect between short-term and long-term memory.

3) Differences between men's and women's learning will be non-standard among presentations, owing to differences in individual presentations.

4) Use of animation in CBI enhances female students' confidence in their ability to do science, and does not affect men's confidence.

5) None of the results from these experiments will indicate that animation hinders women's learning.

6) Gender is an important source of variation in experimental populations looking at teaching through the use of animation.

In the following chapters, I will outline literature that supports the theoretical backing supporting this research, describe my research methodology, present the experimental results, and outline conclusions. Within this framework, I hope that the reader will come to better understand the theoretical foundations for the use (as of yet, unsupported) of animation in education, and come to a better understanding of the conditions under which animation is an effective learning tool. Through the introduction of pertinent literature and a discussion of the theory behind animation and learning, I will attempt to introduce other directions that may be used in future research to further elucidate results, and refine research design in the field of computer based instruction in physical science education.

In the methods chapter, I have outlined the steps I went through to execute my experiments, and have included laser prints of portions of my presentations. Microsoft Bookshelf's® plate tectonics presentation is not included, as permission could not be obtained for reprints. From the information given, this experiment should be replicable. But the objective of science is to further, not simply replicate, hence, chapter four. A little unusual for a dissertation format, a separate chapter on the presentations themselves is included directly after the methodology. This chapter discusses a theoretical division of the individual components of educational presentations utilizing graphics, and relates the literature discussing each component and its importance to the learning process. In relation to my research, this chapter lays the groundwork for some of the possible explanations for differences in student scores between presentations. It may also be looked at as suggesting future research directions. Following chapter four, the dissertation will conclude with a discussion of results, and conclusions.

# LITERATURE REVIEW

Information presented to a student in graphic form should be more easily learned and retained by the student, because it can be dual-coded (Paivio 1971, Paivio and Csapo 1973). The dual-coding theory, well supported and accepted in educational academic literature, states that imaged and verbal information are stored differently in our memory, and therefore, information redundantly encoded will be more easily retained. "Dual-coding is more likely to occur when content is highly imageable" (Rieber 1989a, p. 375). Animation is one tool that aids in learner imaging (visualization) (Friedhoff and Benzon 1989, MacEachren and Ganter 1990, MacEachren *et al.* 1992). Groups or individuals that have low capacities for imaging gain the most benefit from animation (Blake 1977, Rieber 1989a). Imaging can also be called visualization or visual perception. "Visual perception is characterized by spatial properties" (Paivio 1971, p. 33). "...it may require less spatial ability to perceive demonstrated spatial relationships in a dynamic presentation than in a static presentation" (Blake 1977, p. 978). Women have slightly lower spatial skills than do men (Linn and Petersen 1985). Women gain greater benefits from the use of motion in video than men, when teaching spatial concepts such as the movement of chess pieces (Blake 1977). This research will seek to determine whether women gain greater benefits from the use of animation in CBI than men when teaching physical science concepts, and whether any significant gender effects exist such that gender could be considered an important variable in experimental populations testing animation as a learning tool.

This literature review will be arranged in several sections, each addressing the theoretical construct behind the final question "Do men and women benefit to different degrees from the use of animation in classroom CBI while learning physical science concepts?", and subsequently discussing related assumptions. The first portion will define animation, detail Paivio's theory of dual-coding, and will explore animation as a tool in visualization (or imaging). The second section will explore differences between men's and women's learning styles, with special emphasis on visual-spatial skills, and will then look into women's lack of persistence in science, possible causes, and how the use of animation, if effective, could enhance women's science persistence. The third section will examine research done in animation, taking special note of experimental population composition. Assumptions inherent in this research design will be revisited in section four, and study limitations will be addressed.

## What is Animation?

Everyone at one time or another has seen a picture. There are pictures taken by cameras, there are impressionistic paintings, and there are drawings representing objects, concepts, or ideas. When these 'pictures' do not move, they are called static graphics. When they do, they are called dynamic. Dynamic graphics have some sort of moving, changing, or moveable component. In the case of animation, this component is not interactive, the screen (computer monitor or television screen) simply changes on its own. The screen will show a scene or graphics changing frame by frame through time. There is a parallel to this in language, both in words and sentences. Nouns or passive voice sentences represent static verbal states, while verbs of active voice sentences represent dynamic verbal states (Paivio 1971).

Animation has been visible in many disciplines in the last three decades. In geography, cartographic animation has been its main expression. On the whole however, not much work has been done with cartographic animation (Campbell and Egbert 1990), as is the case with animation on the whole (Rieber 1990, Park and Gittelman 1992). Work that has been done in geography is based on human and urban geographic issues (Tobler 1987, MacEachren and DiBiase 1991). As research in animation progresses, more emphasis is being placed on informational presentation and education than has been the case in the past. This trend is clear in the burgeoning supply of computer based education journals available in the past five to ten years such as The Journal of Computer-Based Education, and The Journal of Educational Multimedia and Hypermedia. Given that there is little empirical support for the use of animation as a valuable tool in education, it is clear that use of animation needs to be tested more vigorously, and some solid conclusions reached. In addition to simple empirical testing though, theoretical support for the medium in terms of learning theory needs to be further explored and tested. This is one of the objectives of this research.

## Animation in the Literature

This dissertation is, in part, extending my Master's research (Edgeman 1994). As previously stated, my results showed that subjects were at worst, not harmed by informational presentation in animated format, and at best, were aided by computer animation. Even so, other researcher's results have not been as encouraging, generally finding either no differences (King 1975, Collins *et al.* 1978, Caraballo 1985a, Caraballo 1985b, Rieber and Hannafin 1988, Rieber *et al.* 1989), or solely a difference in response times to information recall (Rieber *et al* 1990, Slocum *et al.* 1990, Koussoulakou and Kraak 1992).

Studies identified supporting animation as as valuable learning aid over static graphics number eight (Blake 1977, Rieber 1991, Zavotka 1987, Rieber 1989b, Baek and Layne 1988, Rieber 1990, Park and Gittelman 1992, and Jacek and Brewer in review 1997). Studies finding no significant differences between animated and static graphics number six (above). Of these, King (1975), and Collins *et al.* (1978), have been questioned as to their reliability (Rieber 1990). If we're counting, this leaves four studies finding no significant differences between student's learning using animated and static graphics, eight studies that find animation to be a superior educational tool to static graphics, and three studies that identify response times as the only difference between instructional types. Eleven out of fifteen studies find that animation has some positive effect on learning, whether in improved information recall, or decreased response times on objective examinations. Clearly animation research should not be abandoned. A valid question remains, however. Why doesn't animation *consistently* result in improved recall on quantitative examinations following instruction? There is a firm foundation of theoretical literature suggesting why animation *should* benefit student learning in this way, but the experimental literature does not conclusively and consistently support this conclusion, at least not in the literature identified thus far.

Let's back up for a second from animated visuals, and take a look at static visuals. Initially, academic research did not identify static visual aids as valuable aids to learning (Rieber 1989a). Through time and subsequent research, however, static visuals became accepted as positively benefitting verbal and written instruction under certain circumstances (Rieber 1989a).

Neither animated nor static visuals represent a consistent improvement over lack of visual aids in several papers in the animation literature (King 1975, Caraballo 1985a, Caraballo 1985b, Rieber and Hannafin 1988). This is surprising. There may be little support for animated visuals being valuable learning aids, but there is ample support for static visuals being valuable learning aids in previous literature. Is it possible that researchers have ignored the past literature that put forth guidelines outlining circumstances under which static visuals are valuable learning aids? If presentational design in animated instruction keeps with the theoretical support for visuals in instruction, will results be more clear? If instruction does not merit the use of visuals, then experimental results will show no presentational differences in student learning between animated visuals, static visuals, and text only. This is the case in most of the literature cited thus far that does not identify animation as showing improvement over static graphics. Only by assuring that some type of visual is necessary to the instruction being tested, will a clear analysis be able to determine the conditions under which animated visuals are superior to static visuals. The fact that studies showing no significant differences in learning or response times between instructional mediums also failed to identify improvement when any graphics supplemented the text or verbal instruction, points to the conclusion that these studies might have been teaching subject matter that did not need graphics to facilitate instruction. If this is the case, dual-coding would not occur, and any additional benefits that might be realized using animation would not become apparent in study results.

For this reason, this study will use a control group where students are taught using verbal instruction only. Any presentation that does not show an improvement in recall over verbal instruction after the inclusion of graphics (either static or animated), will be eliminated from the final analysis. If a subject matter does not lend itself to graphical display, the additional variable of animation will have no effect on information recall. Of the studies mentioned above, several have extremely unequal gender ratios, and many others do not report the gender in their experimental populations. Table 1 relates a list of all experimentally based animation literature discussed in this section, and gives each a code designating whether it has a gender biased sample, or does not identify static visuals as a consistent improvement over verbal instruction.

Theoretical Support for Animation as a Visualization Tool: Dual-Coding

Support for the use of visuals in educational materials comes mainly from Paivio's dual-coding theory. Paivio (1971) developed the dual-coding theory to debunk the then prevalent behavioristic paradigm in psychology that sought to devalue internal imaging. In Paivio's work, imagery and image were used to mean "...nonverbal memory representations of concrete objects and events, or nonverbal modes of thought (e.g., imagination) in which such representations are actively generated and manipulated by the individual" (Paivio 1971, p. 12). This is also an accurate definition for visualization. Paivio reasoned that verbal and visual information is coded independently, or stored in different boxes as it were. The implications for this theory are that ideas or objects coded both verbally and visually should reinforce one another, leading to greater memory retention. Also, a picture would be more likely to be dual-coded than would words, thus leading to a theoretical superiority of highly visualizable pictures over words (Paivio 1971, Paivio and Csapo 1973). The behaviorist paradigm, on the other hand, states

**Table 1**

**A Review of Experimental Studies Addressing Animation**

Studies finding no significant difference between animated and static graphics

| | | | |
|---|---|---|---|
| V N | Caraballo 1985a | R V | King 1975 |
| G V | Caraballo 1985b | V | Rieber and Hannafin 1988 |
| R N | Collins *et al.* 1978 | G | Rieber *et al.* 1989 |

Studies finding a significant difference between animated and static graphics

| | | | |
|---|---|---|---|
| N | Baek and Layne 1988 | N | Rieber 1989b |
| | Blake 1977 | | Rieber 1990 |
| | Jacek and Brewer in review 1997 | N | Rieber 1991 |
| G* | Park and Gittelman 1992 | N | Zavotka 1987 |

Studies finding a difference in response times only

| | |
|---|---|
| G | Rieber *et al.* 1990 |
| | Slocum *et al.* 1990 |
| N | Koussoulakou and Kraak 1992 |

Codes

R = Questioned in the literature as to academic reliability

G = Experimental population heavily weighted towards one gender

V = No significant difference identified for use of graphics over verbal instruction

N = Gender not reported for experimental population

* This experimental population heavily weighted by gender, women 71%, men 29% of experimental population.

that words are most certainly superior to pictures in that they are more concrete.

The level of information abstraction does have an effect on the coding system. Abstract information is more likely to be coded in verbal form, while concrete information is more likely to be coded in imaged form. Concrete in this sense would be the representation most closely resembling reality, such as a photograph. Stick figures for example, would be more abstract (Paivio 1971). Translated into a verbal example, 'freedom' is more abstract, whereas 'book' is more concrete. The key to the usefulness of images in instruction rests in their visualizability. Pictures show no superiority over words when subjects image to words; when words are highly imageable, there is no positive gain in learning or recall from added images (Paivio and Csapo 1973, Rieber 1989a). In other words, if whatever I'm lecturing on readily brings a picture to the student's mind, he or she doesn't need supplementary pictures. If, however, the student finds it difficult to grasp what I'm saying, and can't seem to form an image in his or her mind to help, supplementary pictures will help.

> The problem in the case of both of these postulated processes is to clarify their function, that is, to determine the conditions under which mental images and mental words are aroused and to identify the nature of their effects on overt behavior. *Both are theoretical constructs and whether or not it is useful to postulate either, or both, depends on the adequacy of the defining operations and the research procedures used to test the properties that have been theoretically attributed to them.* (Paivio 1971, p. 6)

In light of Paivio's statement, one of the goals of this research will be to present a recipe, whereby future researchers can repeat this experiment with like graphics presentations. The key to the additive usefulness of pictures in learning is the visualizability of the information (Paivio 1971, Paivio and

Csapo 1973). Is the subject matter easily visualized through the use of words only, or not? Being aware of the need to construct animations that aid in concepts difficult to visualize through the use of words only, will be an important foundation stone in the instructional design of this research. By assuring that the instructional content used in the experimentation is more easily understood through the use of images (in this case animation or static graphics), the experiment will be more focused on testing the amount of aid given by animated and static instructional presentation in student visualization.

## Visualization and Imaging

In this research, I draw support from five disciplines: geosciences, women's studies, psychology, education, and computer technology. Within each of these disciplines, different jargon applies to like concepts. In order to facilitate reading and understanding of this section, and the remainder of the paper, I will standardize vocabulary. Various words commonly found in literature include imaging, spatial perception, spatial visualization, visualization, visual-spatial skill, graphicacy, spatial skill, etc... In the psychology literature especially, vocabulary is prolific and ill-defined (Linn and Petersen 1985). In this paper I will henceforth use the word visualization (most common in geography and computer technology) to describe a spatial skill allowing coding of an image within a pattern, graphic, or presentation. Care will be taken to assure that when citing literature from sources with differing vocabulary, the original vocabulary is parenthetically included. Visualization will not be substituted except in cases where like definitions allow.

Visualization is a concept that is at once very simple and very difficult to understand. A word not unfamiliar to laypeople, visualization has also gained meaning in the jargon of several different disciplines. In common usage, visualization can be noun or verb. As a noun, visualization simply refers to the scene represented when an individual engages in the act of picturing or imagining something. As a verb, visualization ('visualize') refers to the act of picturing or imagining an idea, or concept. If a figure is 'seen' in the popcorn shapes and shadows of fluffy cumulus clouds, the cloud gazer is visualizing that shape. Visualization can occur by picturing something in the mind, or by sketching something to assist the artist in capturing that conceptualized image. For example, if the dimensions and angular measurements of a polygon are known, drawing approximations of those dimensions on paper or in the mind may help one to visualize or 'see' the polygon.

When dealing with complex concepts, an individual may have difficulty understanding mere numbers or words. When given the written definition of a tetrahedron, the reader may have a vague notion of what the shape would look like, but if the reader pauses and either draws the figure or thinks out the form, he or she will have a much clearer picture of what a tetrahedron is.

Consider for a moment though, that all individuals have different capacities for visualizing things, just as all individuals have different capacities for learning from verbal instruction. MacEachren *et al.* (1992) assert that visualization is often not automatic. They argue that some type of 'catalyst' may be required to aid in the act of visualization. The search for that 'catalyst' has become the focus of a major movement. Static graphics have long been recognized as visualization catalysts, but more recently, computer animation has received more frequent mention in the list of possible catalysts. Are both just as valuable? Or are these two types of image presentation simply valuable for different things?

Animation vs. Static Graphics as a Visualization Tool

Under the dual-coding theory, readily imaged or visualized graphics will be superior to no graphics, but why might animated graphics be superior to static graphics to aid in user visualization? And under what conditions might this occur? "The value of animation is that it presents information in a natural way. This is a prosaic point, but we are used to seeing things move" (Dorling 1992, p. 216). Movement is a valuable attention getting device, and is a valuable part of educational presentation (Gagné 1985). Movement in animation can also serve to direct attention, and to emphasize information (Hannafin and Peck 1988). Animation can facilitate understanding of a complex static graphic by organizing the information, directing and redirecting attention to important information, and dividing information into different 'frames', thus simplifying presentation and allowing for simultaneous representation of time and space (Monmonier 1990). In fact, animation seems particularly apt when dealing with subject matter that changes over time, such as vegetation succession, movement of objects such as sand dunes, or the rise and fall of a stream throughout a precipitation event. This division or reorganization of information into temporal or spatial 'frames' introduces a type of hierarchy. Learners tend to memorize information in 'chunks' of one kind or another, although there is no single form used most often (Eastman 1985). When an animation is constructed so that information is organized in 'chunks' (a type of hierarchy), students exhibit radically improved recall (Rieber 1990).

> Map animation is particularly attractive to earth system scientists who typically study large spatio-temporal data sets. In addition to the "visual variables" of static maps, animated maps are composed of three basic design elements or "dynamic variables"-- scene duration, rate of change between scenes, and scene order. (DiBiase *et al.* 1992, p. 201)

So animated graphics contain the imaging benefits of static graphics, along with attention gaining devices and additional design variables such as a capacity for increased informational content, plus increased presentational structure and organization.

Visualization as a Spatial Ability

What is a spatial ability? This question has received much debate over the past decades. Tests for and statements about spatial ability have been in the psychology literature for years, however different definitions of what 'spatial ability' consists of run rampant, are ill-defined, and are non-standard among experimental research designs (Newcombe 1982, Caplan *et al.* 1985, Linn and Petersen 1985). Categories overlap, test things other than spatial ability, and seek to justify theories of biological sex differences that are poorly supported at best (Caplan *et al.* 1985). What does this mean in relation to this research? Unfortunately, it makes it difficult to definitively prove that any constructed animation presentation is inherently spatial in nature. Most spatial skills listed in the various articles, however, have one thing in common; they involve mental manipulation or analysis of some information that is not readily expressed or tested in a verbal form. The information is necessarily dependent upon graphical (imaginal) display. Spatial-visualization involves mental manipulation or analysis of some information that is not readily expressed or tested in a verbal form. Some researchers have relegated all math and science (earth science in particular) into the realm of the spatial, simply because of its dependency on cerebral function and graphical display. Graphs and charts are inherently spatial in nature as well (Lunneborg and Lunneborg 1984). They require mental interpretation -- a searching for pattern if you will. Graphs and charts are more abstract forms of images (Alesandrini 1984).

MacEachren and Ganter (1990) argue for a pattern identification approach to cartographic visualization.

> "The most important role of cartographic visualization...is in prompting mental visualization of spatial patterns and relationships with schematic bits and pieces of information" (MacEachren and Ganter 1990, pg. 66).

Visualization is a function of human cognition, and the human cognitive system is one in which the recognition of pattern, and differentiation of trend from noise is integral (MacEachren and Ganter 1990). In other words, an effective visualization tool would make patterns more apparent and accessible to its user, enhancing the user's capacity for visualization. The words 'spatial', and 'visualization' commonly exist together to describe categories of spatial skills in the literature (e. g. MacCoby and Jacklin 1974, McGee 1979, Hyde 1981, Newcombe 1982, Lunneborg and Lunneborg 1984, Linn and Petersen 1985, Halpern 1986). In order to visualize information, there must be a spatial component. Pictures, images, graphs, and charts are all spatial in nature; some are more abstract, others more concrete. "Spatial aptitude is one area that may be relevant to a viewer's ability to learn from visual media" (Blake 1977, p. 977), and spatial instruction may benefit from the use of motion (Dwyer 1969, Spankenberg 1973, Blake 1977, Rieber 1989a). Whether we are imaging information that is either still or moving, or we are identifying patterns in graphs, we are engaging in spatial cognition.

## Gender Differences in Learning

It is commonly stated that men are better at math and science than are women, and that women have superior verbal skills. In standardized testing, however, there are currently no gender differences identified between men's and women's verbal scores. There have been differences identified in math and science. The gap between men's and women's math scores is narrowing, yet this is not so in science. It has been hypothesized that the narrowing gap in men's and women's mathematics performance is due to the increased attention being given to mathematics and gender equity in both literature and practice. One possible explanation for the lack of progress in science, is that the bulk of research available exploring learning differences between the sexes does not focus on science (AAUW 1992).

One hypothesis for women's lesser performance and lower persistence in science rests on women's spatial abilities. Men are commonly thought to have superior spatial ability to that of women, and science is a field of study largely dependent upon spatial reasoning and imaging. In previous sections, the inconsistent and prolific definitions for categories of spatial ability have been discussed. Results of individual research differ depending on what *exactly* is being tested in a given experiment. In meta-analyses of the spatial 'gender difference' literature, disagreement persists on the scale, scope, magnitude, and types of differences in spatial ability. While differences may be smaller than many researchers would have us believe (Hyde 1981, Caplan *et al.* 1985, Sharps *et al.* 1994), they do exist; and when differences are identified, they fairly consistently favor men (Tapley and Bryden 1977, Newcombe 1982, Linn and Petersen 1985; for an alternative opinion see Caplan *et al.* 1985).

Animation does not change the fact that a given subject matter is spatial in nature. And if the subject is still spatial, how could women possibly benefit more from animation than from static graphics? After all, if both are presenting information that is spatial in nature, and in this research I am arguing that it is women's lowered spatial abilities that make the difference in how they respond to animated instruction, what possible difference could animation make?

> Perhaps one of the difficulties women have is in visualizing spatial relations, or in manipulating them internally, rather than a difficulty in perceiving spatial relations when adequate sensory information is present. (Tapley and Bryden 1977, p. 122)

The idea that women's and men's capacities for visualization might be equalized 'when adequate sensory information is present' is a major foundation stone of this research. It is postulated that animation increases the amount of information available to the brain, not by increasing informational content, or by making that information less spatial in nature, but by making the information more easily incorporated into an individual's learning structure. Imagine someone telling you that you have to go to a local market for groceries. With verbal instruction, you get Joe, the gas station attendant, giving you directions. We've all been there. With static graphics you get a rough sketch map to guide you. The hypothesis is that animation picks you up and drives you there - all you have to do is remember the route.

## Response Times and Cognitive Load

One of the interesting points to note about animated instruction is that researchers often relate a difference in response times to information recall. Students taught using an animated format take less time to answer questions than do students taught using a static format (Rieber *et al* 1990, Slocum *et al.* 1990, Koussoulakou and Kraak 1992, Park and Gittelman 1992). This difference in response times is often dismissed as inconsequential, but taking a second look, a faster response time suggests that students having used the animated presentation were certain of their answers more quickly. Lowered response times in relation to the use of animated instruction could be an indicator of reduced cognitive load in students. Lessening cognitive load enhances learning; very high cognitive load can derail learning (Sweller *et al.* 1990). Imaging things takes time. If subjects are asked a question, and the imaging or visualization has been done for them (as in animation), their processing time is lowered. If, however, they must do this processing themselves, their time is raised. The literature indicates that women may have more difficulty forming these mental images without help. Female presence in experimental populations may be one of the reasons that response time after use of an animated presentation decreases. As a matter of fact, if you eliminate all studies from the animation literature that either have very unequal gender ratios, or do not report their gender ratios, most studies remaining find that animation either significantly enhances learning, or results in decreased response times.

It has also been found, in spatial task studies outside of animation research, that men's reaction times are typically faster than women's (Tapley and Bryden 1977). Again, I'm making the argument that reduced response times mean reduced cognitive load, and that faster response times indicate that learners had an easier time arriving at an answer. This may or may not be a valid assumption. If it is, the fact that men's response times are typically quicker than women's would seem to lend credence to the hypothesis that reduced cognitive load would benefit women differentially over men in experimentation involving spatial tasks. On the other hand, Barnsley and Rabinovich (1970) found that men's reaction times were faster even when tested on non-spatial information. It may simply be that women are inherently more careful learners than are men. Reaction time will not be tested in this experiment; this may be a useful area for future research. Accuracy is clearly more important than speed in an academic setting, and in "...any task in which response time is a dependent variable and there is a relatively high error rate, one must be concerned with the possibility of a speed-accuracy trade-off" (Tapley and Bryden 1977, p. 127).

## Concrete vs. Abstract Learning

Men and women *do* learn differently, even though exactly how and to what extent is not always agreed upon. David A. Kolb (1976, 1984) did research in which he developed four learning 'types' in something called a Learning Style Inventory. Each of these types had different characteristics. For example, Type 1 learners prefer to work with ideas based from or related to personal experience. They are good communicators, and prefer to understand why a given set of information is important to learn, either in people's lives or in their own, before they are willing to devote the time and energy to learning. Type 2 learners think very linearly and work well with detail. They are likely to adapt information to whatever best fits their world-view, and they tend not to see things 'outside of the outline'. Type 3 learners are doers. They prefer the hands-on, down-to-earth, common sense approach to every situation. They are strong in efficiency and know how to integrate theory and practice. Type 4 learners are intuitive, analytical (in a non-linear fashion), gain great enthusiasm from new ideas, learn well by trial and error, and are very adaptable. Learning Types 1 and 4 tend towards a more concrete learning style, whereas Types 2 and 3 tend towards more abstract learning.

> Current data suggest that, on the average, men and women score differently on the Learning Style Inventory. Women tend to score higher on the Concrete Experience orientation, while men tend toward Abstract Conceptualization. (Kolb 1976, p. 24)

This puts women into the categories of Type 1 and Type 4 learners most often. Estimates are that 60% of women and 40% of men are Type 1 and Type 4 learners (Kolb 1984). The composition of my experimental populations flew far afield from this generalization, however, possibly owing to the fact that my students had all voluntarily registered for a science class. It makes sense that learner Types 2 and 3 (abstract conceptualization) would be more interested in registering for such a class, thus accounting for the abnormal population (Kolb's figures applied to the general population, my population was restricted to those students who had voluntarily registered for a science class, a possible bias in the study).

Extrapolating from Kolb's learning type preferences, one might hypothesize that Type 1 and Type 4 learners (concrete experience) will receive the greatest benefits from animation. Animation can be more concrete than standard, non-moving graphics because animation adds the component of motion to a learning presentation, allowing the learning presentation to more realistically (concretely) mimic the reality of a situation (e. g. plate tectonics). Type 1 learners are very concerned with integrating their learning into their lives, their experience. Type 4 learners thrive on the integration of application and experience as well (Kolb 1984). Women not only perform better when their learning is concrete and clearly related to their life experience, but also persist within science more consistently when their style of learning is supported (Roychoudhury *et al.* 1995).

For the last 20 years, the accepted paradigm for spatial skill has been mental image rotation (MIR) (Shepard and Metzler 1971). MIR is far more consistent than any other type of spatial test study in identifying male superiority. Education meshed with concrete experience influences even these results, however. Sharps *et al.* (1994) performed two experiments designed to test gender differences in spatial learning using MIR. They found that expunging words and allusions to 'spatial' testing could reduce differences between men's and women's scoring, suggesting a socialization effect. They also found that no gender differences were identified when familiar figures were used in the testing. These authors attributed this equalization of scores to the fact that these figures were 'easier'. Is this necessarily the case, or were the authors unknowingly using testing material more in tune with a woman's way of learning?

Women's emphasis on, and preference for concreteness is interesting in light of the dual-coding theory. Both symbolic modes (verbal and visual) are likely to be useful when situations are based on more concrete information, however, verbal will be more often used for relatively abstract information (Paivio 1971). This is reaching slightly, but it's still interesting that women prefer a Concrete Experience style in their learning. This learning style would clearly be benefitted by use of both verbal and visual learning aids, whereas abstract information is more often coded verbally. Women prefer a learning style where imaging is more necessary, while men find greater congruence than women with traditional methods of instruction, which are by nature more abstract and less imagery dependent (Philbin *et al.* 1995).

While there is a dearth of research available dealing with women's learning in terms of imaging, there is slightly more literature available dealing with children's imaging. As a child develops, his or her vocabulary evolves into a greater and greater level of abstraction (Paivio 1971). What implications does this have for children being more dependent on imaging aids than adults? Children's vocabulary tends to be removed from the abstract; women's learning styles tend to be removed from the abstract. Children benefit from animation (Rieber 1989b). What about women? Also, children tend to image in static images, while adults are more capable of transformational imaging (Paivio 1971). This statement has distinctive implications for the value and usefulness of animated (transformational) learning aids in adult women's education. Do women have lowered capacities for transformational imaging as well? The paradigm for testing spatial skill in the literature for the last couple of decades has been one of mental image rotation (Shepard and Metzler 1971). Under this paradigm, women consistently test as having lower spatial skill than men. This would certainly suggest that transformational imaging is more difficult for women.

Women and Science Persistence

Interestingly enough, females' grades in school are often as high or higher than males'. This is often attributed more to socialization than to greater mental facilities. On standardized tests, however, females fall short, and earth science is the subject where the widest gap occurs (AAUW 1992, Weinburgh 1995). Women also enter fewer careers in math and science than do men, and have far less confidence in their ability to do well in science than do men (AAUW 1992). In fact, one of the only, and certainly the most consistent difference between men and women in science is women's far higher levels of 'science anxiety' (Catsambis 1995). Men also report higher satisfaction with traditional methods of instruction than do women (Philbin *et al.* 1995). It follows that any strategy found to reduce science anxiety and to better 'fit' the educational style of women could encourage women to persist in science.

**Experimentation in Animation, Gender Sensitive?**

Current literature has begun to explore many of the differences between how men and women learn. Even so, there isn't a great deal of information available, and many current studies do not even identify gender as a relevant factor, or include it within their analyses (AAUW 1992). As a result, many studies examining the efficacy of CBI and animation have samples heavily weighted towards one gender or the other (e. g. Rieber *et al* 1989, Rieber *et al.* 1990, Park and Gittelman 1992, Relan and Smith 1996). Others view gender effects as so irrelevant, that differences in sample composition are not even reported (e. g. Rieber 1989b, Gooding and Forrest

1990, Rieber 1991, Koussoulakou and Kraak 1992, Larsen 1992). In fact, I have identified only two studies that specifically reported controlling for gender effects, Edgeman (1994), and Blake (1977). And Blake's study used animated video as opposed to clear-cut animation.

**Research Assumptions and Study Limitations**

The three main assumptions inherent in this experimentation are as follows: 1) learning presentations constructed for this project are testing spatial skills, 2) men have superior spatial skills to women, and 3) the presentations constructed for the experimentation will induce or facilitate subject pictorial imaging. As previously noted, there is no way to quantitatively prove or disprove these assumptions with current knowledge.

Tests that have been used for decades to test spatial skill may have a higher cognitive load in other types of abilities (Caplan *et al.* 1985). There is no empirical way to know or to test whether these learning presentations are indeed testing spatial visualization as their primary function. There is no way to prove that the presentations constructed for the experimentation will induce or facilitate subject pictorial imaging. It is not possible to ever determine concretely "...Whether an internal representation is pictorial or propositional" (Anderson 1978, p. 249). The only guidelines that can be followed to minimize the impact of this assumption is to make a qualitative assessment as to whether the concepts presented are more easily expressed with words only, and to examine the data attesting to whether or not verbal scores for any given presentation were higher than scores for either or both of the graphics treatments.

Most literature states determinedly that men's spatial skills are better than women's. Most is the key word here, however. As previously discussed, there is disagreement. Future research will hopefully elucidate the differences or lack thereof. It does not seem unreasonable, however, to assume that the bulk of the literature available on gender differences in spatial skill is accurate.

Any conclusions drawn from this research are limited in generalizability to populations of third and fourth year college students with no previous experience in any of the topics covered, and animated graphics learning presentations closely following the guidelines put forth in this research. Given that age and ethnic differences can alter results on learning and spatial exercises (Catsambis 1995), these results apply only to populations consisting of like age and ethnic compositions. Approximate ethnic composition at Oregon State University is 73.9% Caucasian, 12.9% minority. Minority groups are broken up into the following percentages (of the 12.9%): 9.4% African American, 24.3% Hispanic, 55.5% Asian/ Pacific Islander, 10.8% Native American (OSU Multicultural Center figures 1995). The ethnic composition of my experimental populations is in accordance.

# METHODS

In order to address my hypotheses, I designed three physical science learning presentations for use with CBI. Each had an animated, static, and verbal (no graphic ) instructional expression. The subject matter for these presentations was 1) acid rain and soil acidification, 2) generalized forest hydrology following a clear-cut and the installation of roads, and 3) plate tectonics. I designed presentations one and two, while presentation three was imported into PowerPoint from Microsoft Bookshelf® by permission. Each of these concepts is difficult to grasp without visual aids, and the second presentation, generalized forest hydrology following a clear-cut, is mainly graphical in nature, even in the animated treatment of the presentation. Understanding of graphs and charts has been identified as being strongly related ($p < 0.05$) to performance on spatial ability testing (Lunneborg and Lunneborg 1984). A sample frame from the animated presentations has been included for soil acidification (Figure 1) and forest hydrology (Figure 2). Reprints from the plate tectonics presentation could not be included because permission could not be obtained from Microsoft Bookshelf®. Complete animation presentations for the other two presentations may be found in Appendix E. The plate tectonics animation may be viewed by obtaining a copy of Microsoft Bookshelf® from Microsoft®. The plate tectonics animation is in the encyclopedia, under 'plate tectonics'.

The experimental group consisted of approximately 350 students from junior/ senior level environmental conservation classes. All testing and instruction was introduced as a normal part of classroom procedure. Students also completed a pre- and a post-term survey outlining their experience level with computers and science, and their confidence in doing

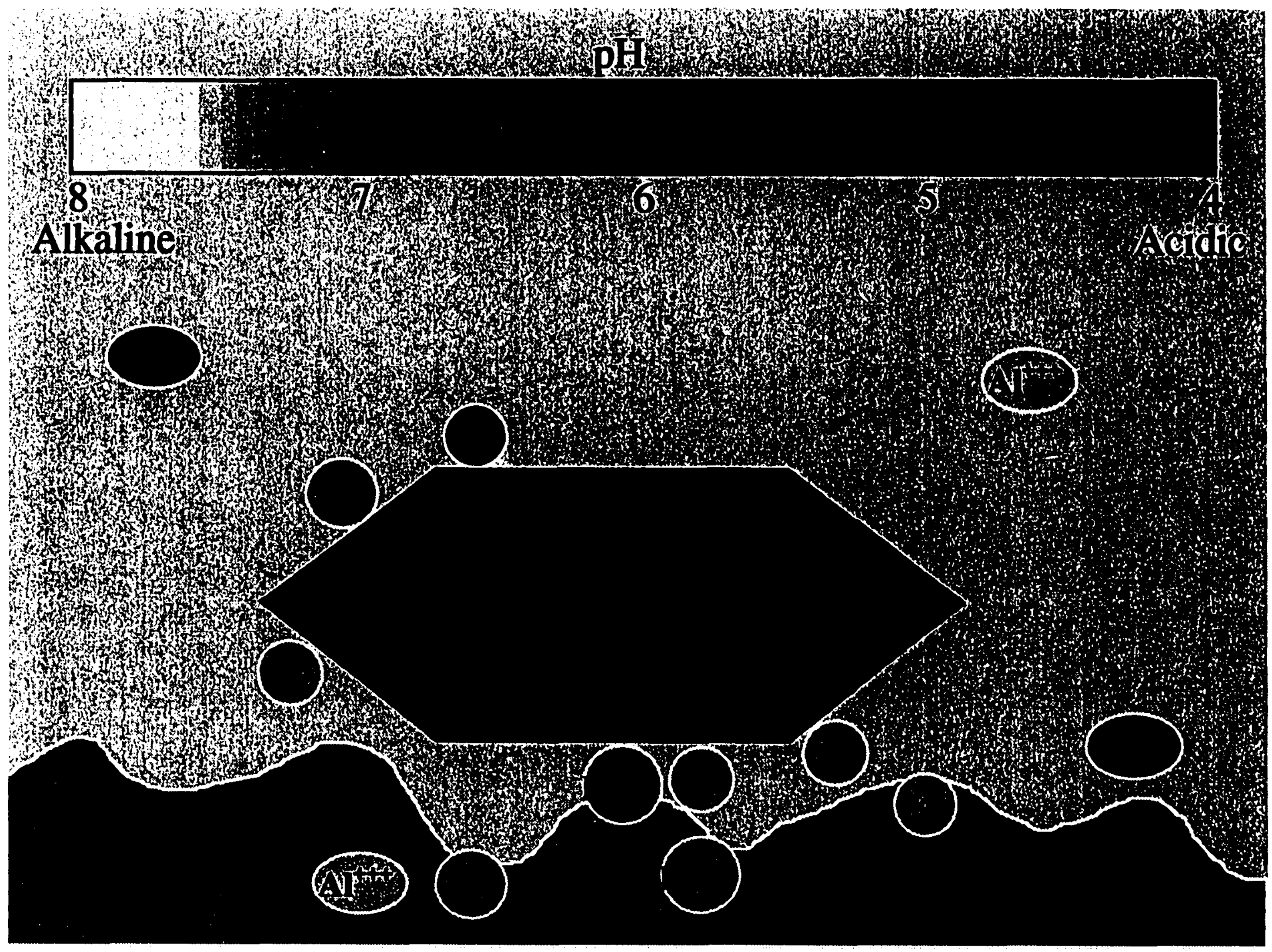



Figure 1. Sample frame from the soil acidification animation.

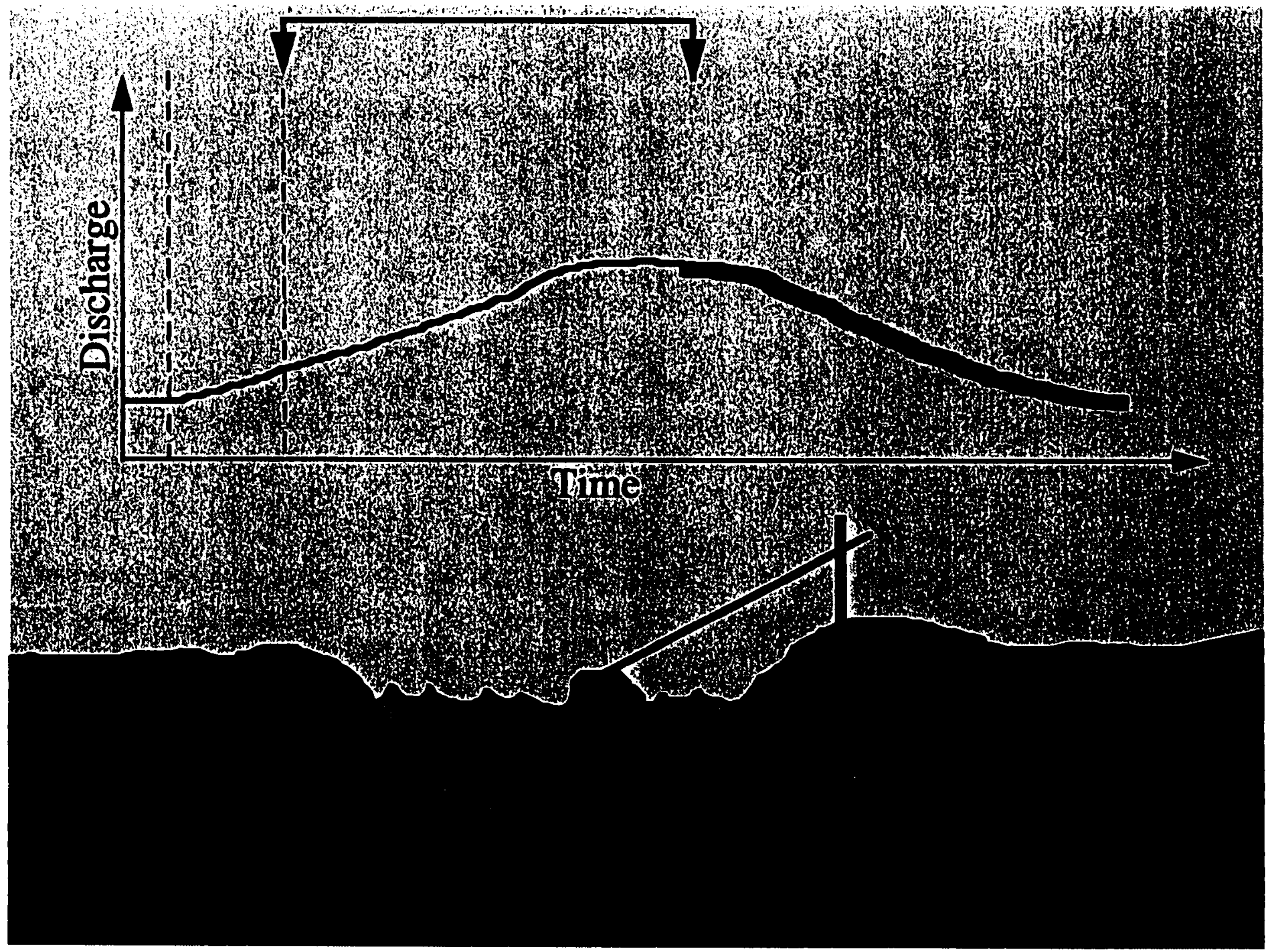



Figure 2. Sample frame from the forest hydrology animation.

well in the class. In addition to these surveys, I conducted several informal interviews per term, with an even number of males and females, in order to gauge changes in attitude, and other information that might not have been directly requested in the pre-and post-term interviews.

## Presentations

I constructed the presentations on soil acidification, and generalized forest hydrology following a clear-cut and the installation of roads using the software package PowerPoint 4. The Microsoft Bookshelf® animation covering plate tectonics was imported into Power Point in its entirety. Each presentation was incorporated into a regular class lecture. While PowerPoint is not a standard animation package, it does allow idea sequencing, 'chunking', attention re-direct, and apparent motion, all factors closely associated with animation. While another software package might have produced more intricate presentations, PowerPoint was chosen due to its ease of use, and its availability to educators. It is simple to use, probably familiar to users of CBI, and requires no special knowledge to integrate into a standard lecture.

PowerPoint is a program commonly used to present information in CBI. It is standard on the OSU campus, and it is supported in both Mac and PC environments. In this research, I used commonly available, commonly used technology that can be slightly adapted to serve an additional educational function, animation. Motion does not have to be complex, requiring advanced design and specialized software packages, to be a successful learning tool (Blake 1977). And the easier it is for educators to integrate animation into their teaching, the more prone they will be to do so.

Presentations were designed in three formats, animated, static, and verbal (no graphic). In each of the three presentations, lectures contained the same information, regardless of experimental treatment. All static graphics were the same colors, size, and resolution, as their animated counterparts. Viewing time was also standard between presentational formats. This was possible because a standard lecture was used for each presentation, whether the presentation was animated, static, or no graphic (verbal only). The only component that differed between presentational formats was the graphics being shown. There was no text included with any of the graphics presentations.

Considering the factors outlined in the literature review, how can a presentation be constructed to make the best use of an animated format? Initially, presentation content should be judged suitable for use with animation; informational content should be readily visualizable (Paivio 1971). If a subject matter does not need pictorial representation, animation will not be of use. Once it has been determined that graphics will be of benefit to the instruction, standard principles of design should apply, as well as special considerations due to the animated format of the material. Tufte (1983) most aptly described optimal presentational graphic design as "That which gives to the viewer the greatest number of ideas in the shortest time with the least ink in the smallest space" (cf. DiBiase 1990, p. 16). Presentations should be as simply constructed as possible; complexity in the form of unnecessary elaboration does not aid in memory retention (Borg and Schuller 1979).

Animated presentations were designed in 'frames'. There was a set interval of time between each frame change. The length of this interval determined how long a given 'chunk' of information will appear on the screen. It is vital that enough time is given for students to assimilate the information being presented. "The problem with animation is that if we cannot control its progress ourselves, the picture may change before we have had time to understand it" (Dorling 1992, p. 216). The time interval for each animation frame change was set by repeatedly questioning three volunteers (two males and one female, plus myself gave an equal gender ratio). When the volunteers were satisfied, times were set. While presenting the information, adequate time and examples must be given (appropriate to subject matter content and difficulty) to help the student link his or her new knowledge into long-term memory. Allowing time and examples of how the new material relates to previous knowledge is important to the retention of new information (Monmonier 1992). Women especially need and value this 'concrete experience' linkage (Kolb 1976, Roychoudhury 1995). I added a special feature into each presentation to assure that presentations could be 'paused', if need be, during a lecture. Interestingly enough, no class requested a 'pause' on any presentation except plate tectonics. This makes sense, as the plate tectonics presentation was a continuously moving set of graphics, and progressed very rapidly. Presentational speed was set by Microsoft Bookshelf®, and was not designed for classroom presentation, but rather for personal use. Frequent complaints from students accompanied the use of this animation. Students found the animation to be "Great! ...but way too fast".

A summary graphic should be, and was, included at the finish of every animated presentation, summarizing the information presented for maximum recall (Monmonier 1992). Each animated presentation had a summary graphic, including the plate tectonics presentation, which had five summary graphics (three tectonic processes were discussed in the presentation, and five graphics were necessary to adequately portray these).

The summary graphic of the animated presentations was the same graphic used for the static graphic component of the soil acidification and forest hydrology presentations in this research. The plate tectonics static graphics presentation used the animated presentation summary graphics as well, but five graphics as opposed to one were used, in order that the idea of physical change could be adequately represented. The idea behind the inclusion of a summary graphic is to give a sort of pictorial recap sentence (summary graphic) to a long explanation (animation). Both the forest hydrology and the soil acidification summary graphics were actual graphs as opposed to pictures (Figure 3 and Figure 4). As already discussed, graphs and charts are relatively abstract, and women and men have differing capacities to respond to abstract information. This factor was also looked at in the analysis. It is also important to note that differing learning types have different capacities for learning with abstract vs. concrete information. Women and men are generally found in uneven proportions in these learning types; men tend to be about 10% higher in the abstract conceptualization groups (2 and 3), while women tend to be higher in concrete experience groups (1 and 4) (Kolb 1976). Since the proportions of male and female learners in this class do not conform to the norm, learning type may be more indicative of student performance in this category. However, there turned out to be too few Type 1 and Type 4 learners to have a large enough sample for analysis of learning types. This possibly indicates an area where future research may be directed.

Certain teaching strategies applied during instruction are as essential to effective learning presentations as is sound design. Practice has been identified as a valuable part of assuring the value of an animated presentation while teaching physical science concepts to children (Rieber 1989b). Students included in this study were quizzed on material learned two

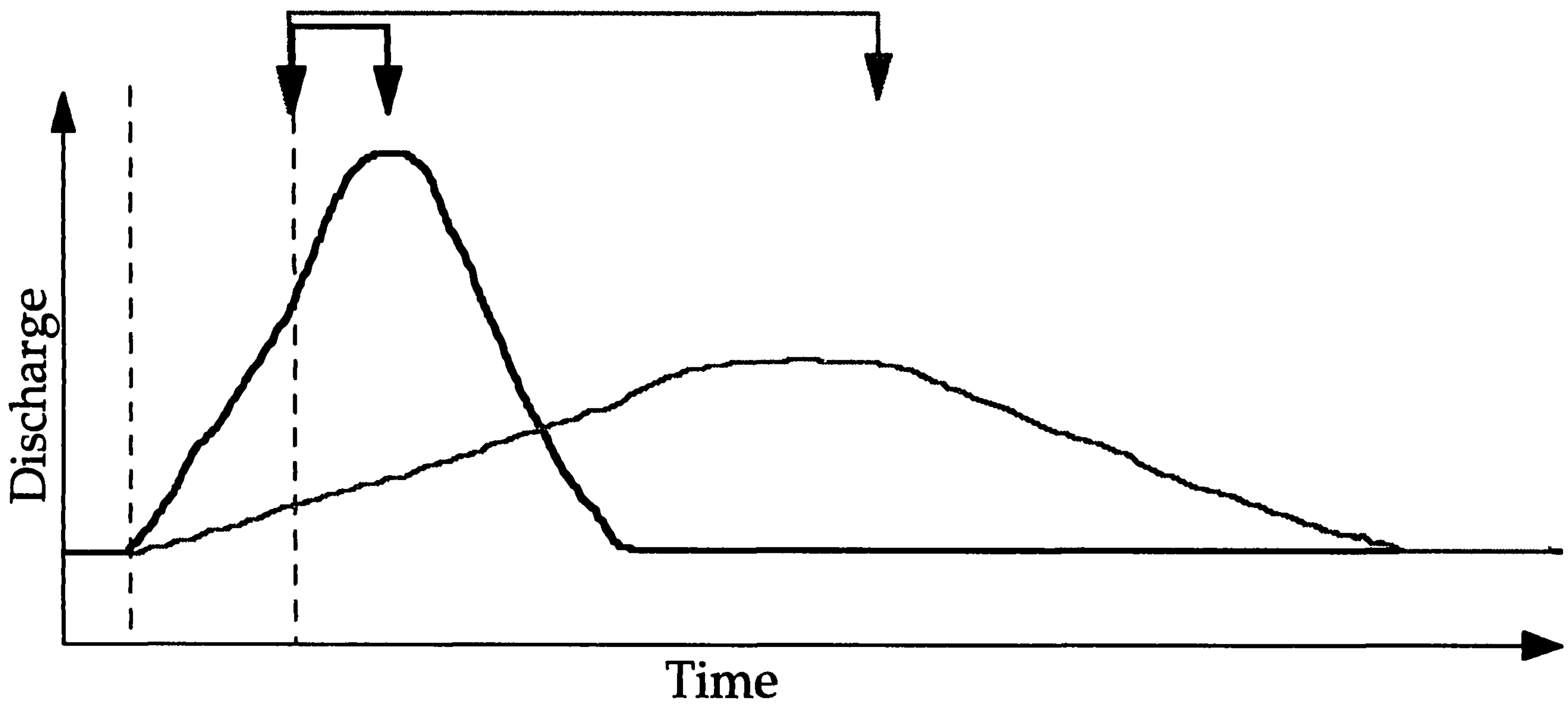



Figure 3. Summary graphic for the forest hydrology presentation.

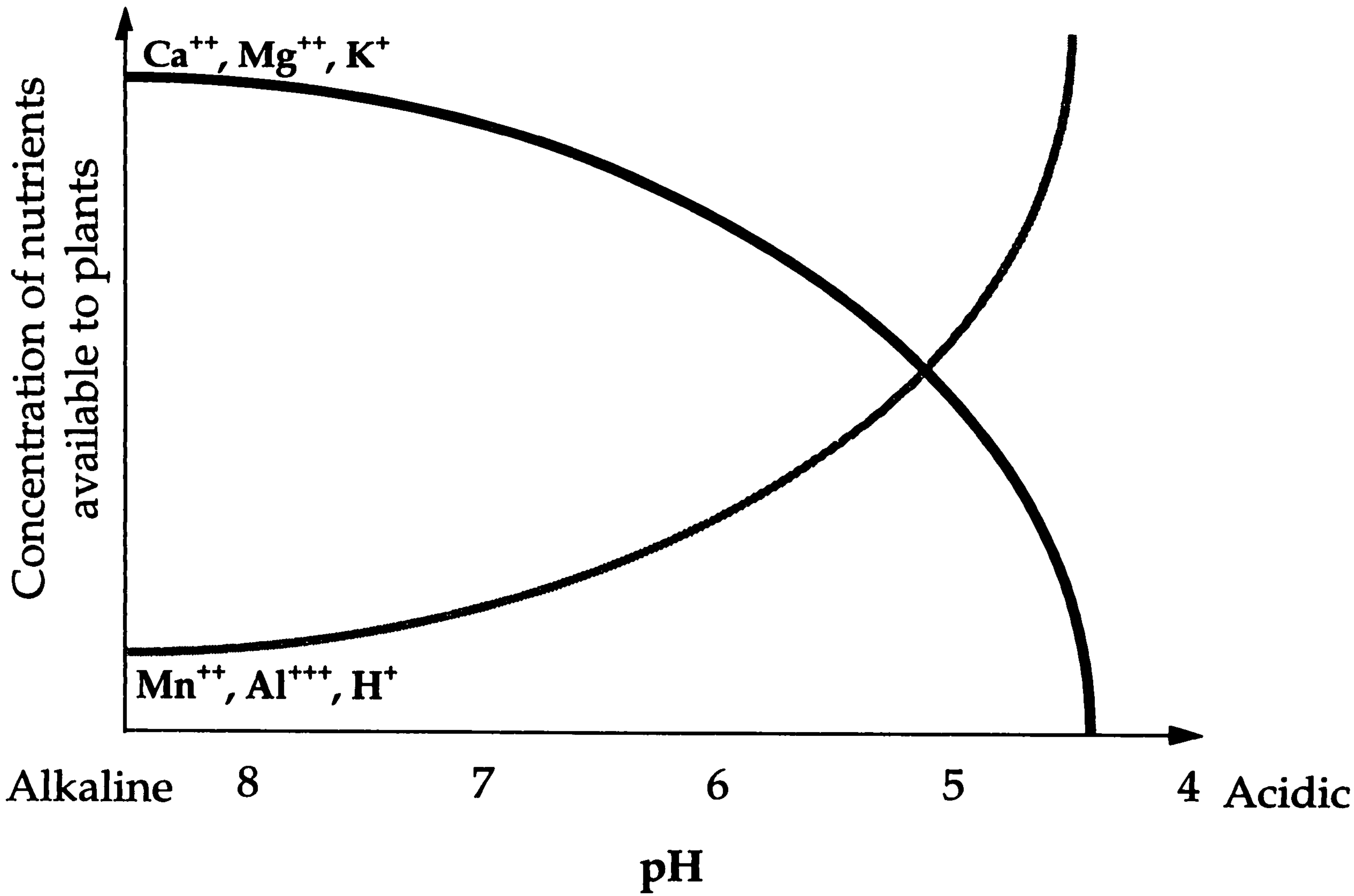



Figure 4. Summary graphic for the soil acidification presentation.

days after the original presentation, regardless of what presentation format was used as a learning device. Students not present for the short-term quiz were not included in the total (long- plus short-term) analysis for that presentation.

Instructions play a large part in women's performance in experiments testing spatial skill. Often if the word spatial, or the implication that some part of an experiment may hold material widely known (accepted) to be more difficult for women is included in instructions prior to an experiment, women will perform at a lower level than will men. If these references are eliminated, however, the differences between men's and women's performance can diminish (Sharps *et al.* 1994). Researchers are increasingly identifying gender bias in educational material, instructions, and standardized tests (AAUW 1994). The experimentation included in this research is one small part of a larger classroom experience. Special instructions did not need to be given, and were not given before experimental material was presented. In this way, it was hoped that women would not start out learning the material with the false assumption that they could not succeed.

## Subject Selection and Experimental Design

Subjects were taken from six environmental conservation classes, and all presentations were taught by the same instructor. An initial survey identified students who had had prior experience in any one of the subjects being used in testing. These students were eliminated from the analysis of that, and only that, presentation for which they had prior experience. Gilhooly *et al.* (1988) and Kulhavy *et al.* (1992) both identify experience as a factor that helps subjects to memorize material more quickly and effectively. All subjects who exhibited color-blindness were eliminated from the analysis for all presentations. Color-blindness is a gender weighted disability, and might have biased the analysis. All students failing to sign their informed consent documents were eliminated from the subject pool as well. No other students were eliminated from the subject pool.

Each of the six classes had approximately seventy-five students. It was expected that through student eliminations due to high experience levels and non-attendance, approximately one-third to one-half of each class would be eliminated from the analysis. Subject eliminations left 332 students total to be included in the final analysis. Due to the fact that different numbers of students were eliminated from the analysis of each presentation (varying numbers of students had experience with each subject matter), each presentation, in each class had a different total student count, totals ranging from 46 to 69 before elimination for previous experience, to 46 to 64 after elimination due to previous experience. (These numbers are per class, per presentation. Total number for all presentations = 332, above.) The gender ratios for the classes were not equal, however, the use of gender averages and/or ANOVAs that accounted for gender, assured that there was no gender bias in the statistical analysis.

All six classes were taught all three presentational subjects in different formats (Figure 5). The six classes were divided into two subgroups. The first three classes were arranged using a Latin Square experimental design. This design enabled me to examine how each class responded to each presentational format, across presentational contents. Each of these three classes saw one presentation in each format. For example, experimental group one, in the Latin Square subgroup, saw the soil acidification

Subgroup A: Latin Square Design

| | Class 1 | Class 2 | Class 3 |
|---|---|---|---|
| Soil Acidification | Static Graphics | Animated Graphics | Verbal Instruction Only |
| Forest Hydrology | Verbal Instruction Only | Static Graphics | Animated Graphics |
| Plate Tectonics | Animated Graphics | Verbal Instruction Only | Static Graphics |

Subgroup B: Stratified Random Design

| | Class 4 | Class 5 | Class 6 |
|---|---|---|---|
| Soil Acidification | Verbal Instruction Only | Animated Graphics | Static Graphics |
| Forest Hydrology | Verbal Instruction Only | Animated Graphics | Static Graphics |
| Plate Tectonics | Static Graphics | Animated Graphics | Verbal Instruction Only |

Figure 5. Experimental Design

presentation in the static format, the forest hydrology presentation in a verbal format, and the plate tectonics presentation in an animated format. By allowing both control (verbal instruction only) and experimental (animated and static graphics) groups to exist in the same class and the other classes, differences both within and between classes could be identified and addressed in the analysis. The choice as to which class would become which experimental group within the Latin Square was randomized. There were twelve possible Latin Square treatment assignments. One of these was randomly selected (double coin toss). I expected that class and topic would be a source of variation within that data set, therefore, the blocking in the Latin Square was a logical design to put forth, as it is the only way to compare treatments across topics and classes without confounding. The assumptions inherent in this model are that it has no interaction between the main effects, no carryover effect, and it assumes a single response.

The pairing of presentation with treatment in the second group of three classes was randomized (double coin toss) as to which class would get a given treatment, although all three treatments were allocated for a given presentation (random stratified). These three classes were still presented with each subject matter, whatever the presentational treatment was. From this point on in the dissertation, the Latin Square group of classes will be called subgroup A, while the randomized group of classes will be called subgroup B.

## Interviews and Questionnaires

There were three questionnaires distributed over the course of the quarter (OSU classes are based on a quarter system). The first questionnaire was distributed at the beginning of the term, inquiring about a student's prior experience with computers, comfort with science, and prior science background (Appendix A). Responses from these questionnaires were used to assess confidence changes in students over the course of the term, and to identify students who had prior experience with any one of the experimental subjects. Responses on the pre-term questionnaire were also used to identify possible correlations between (for example) excess reading and high scores in verbal only instructional presentations.

A post-term questionnaire was distributed at the conclusion of the class (Appendix A). Post-term questionnaire responses were used to gain a better understanding of student comfort with computer use in the classroom over the course of the term. Responses were also examined in light of attitudinal changes for women throughout the term. To supplement these two questionnaires, an informal interview process was conducted throughout the duration of each class. Six students (three female, three male) were selected at random in each of the six classes. These students were requested to reply to a number of informal questions about midway through the quarter (Appendix A). Additional, unsolicited comments were encouraged throughout the quarter, and duly noted. Responses were used to make additional inferences to quantitative results, and to determine in greater depth, student attitude toward the learning formats, and possible changes in attitude as the quarter wore on. This research adds an interesting dimension to most animation experiments currently in the literature, because the experiment ran throughout the quarter, a period of ten weeks. Due to my frequent contact with the subjects, I was able to make some qualitative determination of mental and attitudinal changes in students throughout the term.

**Testing**

In order to assess student learning, two types of testing were used. Two days following each of the three presentations, students were given a quick, short-term memory quiz on one or more aspects of the presentation. The second type of exam tested student long-term memory. The tests were incorporated into the regular class testing as a mixture of pop quizzes, midterms, and finals. Testing consisted of two formats as well. Males typically score higher on multiple choice tests than do women, while women tend to score better on short answer problems (AAUW 1992). To offset these differences, testing was carried out using both formats. Some questions were presented in short answer format, others in multiple choice. Quiz and exam questions may be found in Appendix C.

**Statistical Analysis**

As mentioned earlier, there were two experimental designs used in this research, resulting in two experimental subgroups. The Latin Square design turned out to be the best in terms of analysis, and so I will spend some time explaining this design and what it means to my ANOVA model. The Latin Square is a blocking design used to isolate sources of variation within a data set; it has two blocking factors, class and topic. The experimental unit in this design is the response for a given topic and a given class. Knowing the column (class) and the row (topic) gives a more precise estimate of the treatment effect. In this model class, topic, and treatment are designated as having fixed effects (this is due to the fact that they come from finite populations). I used the blocking design because I expected that both class and topic would have some affect on the data, but I made the assumption that the effect was not random. Independence between the main effects is assumed in the error term. This may or may not be a good assumption, but using only subgroup A, the model lacks the degrees of freedom to test for interaction effects. The Latin Square model used in the ANOVA is as follows:

$$y_{ij(k)} = \mu + \rho_i + \gamma_j + \tau_k + e_{ij}$$
$$i, j, k = 1, 2, \ldots\ldots, t = 3$$

where $y_{ij(k)}$ is the response observation in the *i*th row and the *j*th column of the Latin Square, $\mu$ = the mean, $\rho$ = topic, $\gamma$ = class, $\tau$ = treatment, and e = the error term. This model corresponds to my first ANOVA.

Blocking (Latin Square is a blocking design) is done to isolate possible sources of error in an experimental design. It's like having a pile of rocks that represents all variation from the mean in a data set. Partitioning each of the main effects takes some of the rocks (variation) and removes them from the pure error term. It's like taking some of the rocks and labeling them, putting them in separate piles, and dealing only with what's left over as 'unknown' variation or 'pure error'. The partitioned variation for each of the main effects helps reduce the pure error in the model.

In the statistical analysis for most of the dissertation, I used ANOVA tests. ANOVA tests are generally used to identify variation from the mean from any one of several sources. I did three levels of analysis using ANOVA; a top level analysis, a gender level analysis, and an analysis looking solely at student variation. In the top level ANOVA, there are nine values being compared in the ANOVA, each representing the difference between the mean men's score and the mean women's score for a given row and a given column. In the second ANOVA table, there are eighteen values instead of nine (Figure 6). My second level model (the gender level) model is as follows:

$$y_{ijkl} = \mu + \rho_i + \gamma_j + \tau_k + \alpha_l + e_{ij(k)} + (\alpha\rho)_{il} + (\alpha\gamma)_{jl} + (\alpha\tau)_{kl} + d_{ije(k)}$$

$$i, j, k = 1, 2, \ldots\ldots, t = 3$$

$$d_{ijk} = y_{ijk}1 - y_{ijk}2$$

The experimental units now each contain two values, mean men's scores and mean women's scores, not taking the difference between men and women. This allows us to analyze the data without losing information relating to the variation within men's scores and the variation within women's scores. In this second level ANOVA table, gender is crossed with three other variables, making it possible to determine gender variation by topic or by class, or by treatment. Again, however, degrees of freedom are low. The third level ANOVA allowed an analysis of all student scores and a three level ANOVA table. This third level of analysis is something of a cross-over design since some students are taking tests over different topics.

Through these ANOVAs it is possible to determine whether the difference of the experimental variables from the mean of the entire sample is due to any one factor, or a combination of factors. One might assume that regression or multiple regression would suit this problem better. However, ANOVA is generally preferred when explanatory variables are categorical - when they don't make sense in a continuous sense. Also, this analysis yields more than two explanatory variables, which makes it difficult to consider the geometry of regression surfaces (Ramsey and Schafer 1995). Results that were not appropriate for use in an ANOVA were analyzed using binomial tests for

**Top level ANOVA table:**

| Source | df | SS | MS | Expected Mean Square |
|---|---|---|---|---|
| Topics | 2 | $3\sum_i (\bar{y}_i - \bar{y})^2$ | SS/df | |
| Classes | 2 | $3\sum_j (\bar{y}_j - \bar{y})^2$ | SS/df | |
| Treatments | 2 | $3\sum_k (\bar{y}_k - \bar{y})^2$ | SS/df | $\sigma^2 + t\theta^2_t$ |
| Error | 2 | SSE | MSE | $\sigma^2$ |
| Total | 8 | $\sum_i\sum_j (\bar{y}_{ij} - \bar{y})^2$ | SS/df | |

**Gender level ANOVA table (class level is the top section and represents average over genders, not a gender based analysis; gender level is the lower portion and is designed to allow for analyzing gender differences - equivalent to first ANOVA):**

| Source | df | SS | MS | Expected Mean Square |
|---|---|---|---|---|
| Topics | 2 | $3\sum_i (\bar{y}_i - \bar{y})^2$ | SS/df | |
| Classes | 2 | $3\sum_j (\bar{y}_j - \bar{y})^2$ | SS/df | |
| Treatments | 2 | $3\sum_k (\bar{y}_k - \bar{y})^2$ | SS/df | $\sigma^2 + t\theta^2_t$ |
| Error (1) - (Trt*Class) | 2 | SSE | MSE | $\sigma^2$ |
| Total | 8 | $\sum_i\sum_j (\bar{y}_{ij} - \bar{y})^2$ | SS/df | |
| Gender Level | | | | |
| Gender | 1 | $2\sum (\bar{y}_l - \bar{y})^2$ | SS/df | |
| G*Topic | 2 | $3\sum_i (\bar{d}_i - \bar{d})^2$ | SS/df | |
| G*Class | 2 | $3\sum_j (\bar{d}_j - \bar{d})^2$ | SS/df | |
| G*Treatment | 2 | $3\sum_k (\bar{d}_k - \bar{d})^2$ | SS/df | $\sigma^2 + t\theta^2_t$ |
| Error (2) - (G*Trt*Class) | 2 | SSE | MSE | $\sigma^2$ |
| Total | 17 | $\sum_i\sum_j (\bar{d}_{ij} - \bar{d})^2$ | SS/df | |

Figure 6. ANOVA tables

*$d_{ijk} = y_{ijk}1 - y_{ijk}2$

differences of proportion. This is a test designed to compare two sets of 'dichotomized' data. In this research, binomial tests were only used twice, once to determine whether women's scores on animated vs static graphics were significantly different, and once to determine whether student use of drawings on two different presentations for animated and static graphics use was significant.

Descriptive statistics were used liberally throughout the results and discussion section of the dissertation. I expected that, because this is the first study looking at gender differences between animation and static graphic instruction, any differences found would be small, indicating trends, future research directions, or interesting patterns that might help better direct future research. I anticipated that these trends, patterns, and suggestive data might be more important than simply designating a given hypothesis 'accepted' or 'rejected'. I thus placed several graphs into the results and discussion section of the dissertation, showing averages of different treatments, topics, and genders. Please use these graphics as what they are - descriptive tools suggesting patterns and trends. These graphs are not derived from ANOVA tables, and the confidence intervals depicted on some of the graphs are not ANOVA confidence intervals (I did not have adequate degrees of freedom to identify confidence intervals for the ANOVAs). Also, it is impossible to obtain a valid standard error for an interaction. In identifying confidence intervals, I therefore chose to look at each experimental subgroup as an unrelated population unto itself. Please keep this limitation in mind, and use confidence intervals with consequent skepticism.

I have made every effort to to be clear, in the discussion of the data, when I am talking about statistical conclusions, and when I am looking at averages and trends in the data. In all phases of the analysis, subgroup B data may be used as a replicate data set. Some of the analysis uses subgroup B data statistically, other parts of the dissertation use subgroup B as a replicate only in the descriptive part of the discussion. In these sections, subgroup A data has been analyzed statistically, and descriptive statistics for both subgroup A and subgroup B have been included in graphs, in order to allow the reader to compare trends and patterns in that data for both subgroups. I'm working under the assumption that if trends and patterns in the data are comparable, even though statistical replication is not available, research conclusions are strengthened. Conversely, if statistical conclusions are not strong, and descriptive trends in the data are conflicting, statistical conclusions are not strengthened.

# PRESENTATIONAL ANALYSIS

One of the more valuable pieces of information that may come out of this research, is a determination of what components of learning presentations may result in the greatest amount of difference between genders. In order to identify these differences, we must first determine what the differences are between the presentations themselves. After I have laid out a template describing the various components of graphics-based learning presentations, I will put forth an in-depth literature review on each variable on the list, linking cognitive processes to each, such that statements of causal inference may be made in the "Results and Discussion" chapter, relating presentational differences to explanations derived from past research. Since I am not testing several presentations, each differing in only one way (thus enabling me to isolate that as a distinct variable), I will use this section to help identify possible sources of variation between presentations in a qualitative sense only. An important area for future research would be a more in depth, statistically oriented exploration of the variables identified in this chapter. Caution should always be used, however, to maintain linkages between cognition and instructional design, as stimulus-response studies deal only with surface issues and do less to advance the knowledge of learning as a whole entity. In addition to strengthening research conclusions, this chapter outlines the component parts of each presentation such that replication is possible. Part of good research lies in its replicability. In order that future research may replicate results, and build on these data, it must be made perfectly clear how to replicate the presentations themselves (in terms of component parts).

This chapter is arranged in three parts. Initially, I present a list of all quantifiable variables that can be identified in computer learning presentations (non-interactive). Subsequently, I identify which of these variables are most pertinent in an animated presentation. I then explore the literature detailing linkages between what an instructional designer wants to get across, and learner cognitive processes. Finally I explore past research identifying whether these experimental variables actually do, or do not, catalyze positive learning outcomes. At the end of this section, I include a section on gender differences in visual cognition and relate them to the presentational components. I also include a table detailing how each of the three presentations differs with regard to each of the five variables identified in the following sections (Table 4).

## The Variables

In order to identify variables within a presentation which affect learner processes, we need to look at the problem from two sides, the instructional designer's purpose, and the learner's processes. When I construct a learning presentation, I have a purpose in mind for that presentation. "With my learning presentation, I wish to convey..." Within whatever it is that the designer wishes to convey, he or she will choose elements to include within a graphics presentation that will (should as supported by the literature) facilitate that purpose. Let's say, for example, that I want my audience to notice two countries on a map that I've animated, both having the same population growth rates. I may have both countries blink to brighter colors several times to call them to the attention of my user. I may also be taking a concept that my user (learner) is already familiar with and elaborating on it in order to extend that learner's knowledge base. Either way, my goal is to add knowledge to whatever learning base the user is starting out with.

There are many graphics presentation elements that I could choose to facilitate the fulfillment of my purpose. The question is, which of these presentation elements is the most effective, and does any particular presentational element have a special relationship with gender? Is a more abstract presentation going to handicap female learners? Will a concrete format be more helpful across gender? In order to execute an experiment that can help make this determination, I must initially tear my presentations down into discrete bits, pieces that are readily quantifiable. Then I can to turn to the literature to understand how these particular presentational elements have affected learner processing in previous experiments. I don't care what my purpose is as a designer, if I use a graphics presentation element that will not facilitate a positive response in my learner's cognitive process, I accomplish nothing. "A response is said to be under the control of a stimulus when the stimulus can influence the probability of its occurrence" (Gropper 1963, p. 75).

Imagine the problem as if the graphic presentation elements are resistors to the flow of electricity. Flow of electricity cannot occur unless a current can be maintained from the designer's purpose to the learner's cognitive processes *through* the graphics presentation elements. If the graphic element that is used to aid user pattern recognition (encoding) doesn't aid in encoding, it doesn't matter how effectively I use the device. All of this assumes positive use and impact of graphics devices, which is never a good assumption. Using the blinking on and off of an object again as an example, what if everything on the screen is blinking? There is a point of diminishing returns.

Looking at Table 2, you will see that I have made a lists of designer purposes, graphics presentation elements, and learner processes (learner processes taken from Marzano *et al.* 1988) that I believe are relatively inclusive assuming an educational focus to a presentation (for definitions to each 'designer's purpose', see Appendix B). Graphics presentation elements specifically related to animation are discussed below.

In order to be meaningful within the context of my research, a graphics presentation element has to be important to both the researcher's and the learner's process, and be a factor addressed specifically through animation. There are several items on my list of graphics presentation elements that are not specifically related to animation. From the remaining elements, I have developed my list (Table 3).

The first graphics presentation element under discussion is concurrent information, which is present in almost all instructional material, either verbally or textually. Instead of presence or absence as categories, the percentage of the presentation that has concurrent information has been determined. In other words, is the subject being required to take in verbal information, visual information, or both at the same time? This category relates to the student's need to divide her or his attention between two or more pieces of information at the same time. Related to this is the amount of time that the student has to process each byte of information ('speed of presentation', subcategory of 'technical'). Therefore, I have also determined the average ratio of time per byte of information presented to the student during the presentation.

The other important areas are the Graphics and Technical categories. Within the Technical designation, 'speed of presentation' has already been addressed. The other subcategory, 'type of change or movement', can be

## Table 2

**Designer Purposes, Graphics Presentation Elements, and Learner Processes**

**Designer Purpose**

- **Additive knowledge/ Process**
  - **Elaboration**
    - Abstract - concrete
    - Generalization
    - Representation
  - **Attention**
    - Focus/ emphasis
    - Grouping/ chunking
    - Organization/ structure

**Graphic Presentation Elements**

- **Concurrent information?**
  - Presence of text
  - Use of lecture
- **Subject matter**
  - Number of topic changes
  - Bytes of information contained
- **Context**
  - Within learning presentation?
  - Stand alone?
- **Graphics**
  - Number of graphics on screen
  - Types of graphics
  - Use of color
  - Level of abstraction/ detail
- **Technical**
  - Speed of presentation
  - Motion?
  - Type of change/movement
- **Presentation**
  - Time spent on each topic
  - Length

**Learner Processes**

- **Focusing skills**
  - Defining problems
  - Setting goals
- **Information gathering skills**
  - Observing
  - Formulating Questions
- **Remembering skills**
  - Encoding
  - Recalling
- **Organizing skills**
  - Comparing
  - Classifying
  - Ordering
  - Representing
- **Analyzing skills**
  - Identifying attributes and components
  - Identifying relationships and patterns
  - Identifying main ideas
  - Identifying errors
- **Generating skills**
  - Inferring
  - Predicting
  - Elaborating
- **Integrating skills**
  - Summarizing
  - Restructuring
- **Evaluating skills**
  - Establishing criteria
  - Verifying

**Table 3**

**List of Experimental Variables**

Concurrent verbal/written information (% of presentation)

Coincident visual information (% of presentation)

Speed of the presentation (bytes per unit time)

Type of change or movement

- Additive information (parts of a graph or organizational chart being revealed one part at a time)
- Informational change (change of subject from one frame to another)
- Attentional movement (blinking)
- Representative motion (actual movement)
- Rotation (showing more than one aspect, not in order to show movement, but in order to show multiple views).

Type of graphics used

- Pictures/photos
- Pictorial representations
- Analogical representation
- Abstract symbology (color, symbol)
- Flowchart, graph, matrix

related to a categorical classification, as follows: additive information (parts of a graph or organizational chart being revealed one part at a time), informational change (change of subject from one frame to another), attentional movement (blinking), representative motion (actual movement), and rotation (showing more than one aspect, not in order to show movement, but in order to show multiple views). Each of these can be quantified as to the percentage of the presentation exhibiting this type of motion, or the information in the presentation dependent upon this type of motion.

In terms of the graphics within the presentations, fewer categories seem to be relevant. The 'number of graphics on the screen' may affect comparison and relationship identification. The analysis in this category looks at concurrent graphical information. How much of the presentation requires a division of attention for the viewer between two (or more) graphics?

'Types of graphics' affects many of the learner processes, but is difficult to quantify. Graphics types that designers use will be suited to the needs of the presentation (flowcharts, graphs, photos, etc...). There are many confounding variables that can interfere with producing meaningful results while analyzing relationships within this category. For example, is the type of graphic not suited to that presentation, or is the type of graphic simply poor overall? In order to alleviate this source of confusion, I have linked 'types of graphics' with 'level of abstraction/detail', creating a continuum consisting of five discrete categories lying along this continuum. These are listed from concrete to abstract under the heading 'types of graphics used'.

The use of symbol/color is certainly valuable to many learning processes, but this category also is difficult to analyze meaningfully in isolation from other categories. (Note: color as an aesthetic characteristic does not influence learning effectiveness, {Baek and Layne 1988}). To a great extent, however, this category relates to the previous one, and was therefore incorporated into analysis of that category (type of graphics used).

## Linkages

In essence, as the students in this research seek to learn a given set of information, they are executing specific cognitive processes to that end. Cognition is the name given to the processes used by individuals to assimilate information into their memory structures. A cognitive strategy "...can be defined as the method or process by which an individual solves a given problem or task..." (Blades and Spencer 1986, p. 4). These strategies, and the student's choice of which strategies will be employed, determine how effectively an individual is able to learn (Gilhooly *et al.* 1988, Griffin 1983, Thorndyke and Stasz 1980). Psychology literature is rife with evidence indicating that visual imaging is important to the execution of our cognitive tasks (Reed 1992). Unfortunately, however, relatively little is known about exactly what cues cause an image to be 'coded' into visual memory. The identification of these cues is of paramount importance to this section of the research. Any relationship determining the visual cues that influence cognition will give strong evidence of what factors within a learning presentation will be most likely to impact learning processes. In 1971, Paivio came up with the dual-coding theory, discussed in previous sections. In this theory, he stated that verbal and imaginal objects are 'coded' or stored in different boxes in our minds. This theory has been widely supported in the academic journals, and has led to widespread speculation as to exactly what our visual 'sketchpad' actually looks like. Do we see exact pictures in our mind's eye?

> A person can describe an image in many ways, including information about contents, vividness, clarity, color, shading, shapes, movement, foreground and background characteristics, and other spatial relationships. Furthermore a person can often tell how the image entered his awareness, its duration, associated emotions, the relationship of the image to objects in the external world, efforts to change or dispel it, and the sequential or simultaneous arrangement of a series of images. While people can describe image contents, they are usually unaware of all the underlying processes or motives which go into image formation. (Horowitz 1970, p. 3)

Keep in mind that when talking about visual stimuli in relation to cognition, we must take a bi-directional view of things (what do we put in, and what can we pull out?). What is it about a scene or visual object that people code into their memories? Do we simply see snapshots of everyday scenes, available for instant recall? No, more likely we encode the attributes of a feature that are useful and necessary to us (Reed 1992). Imagine a penny. What are the exact words stamped across the top of the penny? Most subjects tested using various drawings of penny with different words written across the top (including myself) picked the wrong penny (Reed 1992). Consider though, that we encode the essence of a penny by virtue of its size and color. These two things are most useful in differentiating a penny from other coins, but not helpful in making the previous determination. Once in the memory stores, each visual image, or visualized image, is available to be used, cognitively, to some extent. The extent to which we may use this visual memory cognitively depends heavily on what clues, cues, or stimuli we have incorporated in relation to that mental image. In a learning presentation specifically, the hope is that the most important information will be readily apparent to the student, and thus most easily picked out, linked to long-term memory, and retained.

In the following sections, I will outline each category on the list in Table 3, and detail the literature related to it.

Variables 1 through 3

Concurrent verbal/written information, coincident visual information, and speed of presentation can all be looked at in terms of cognitive load, and parallel processing. How well an individual learns has a lot to do with the relative mental twists and turns that he or she must take to incorporate the new knowledge into his or her memory structures. The relative number of these twists and turns is called cognitive load. Have you ever thought through a long string of related items to find a solution, only to get halfway done and realize that you've lost the question? If you have, then you've experienced cognitive overload. Sweller *et al.* (1990) looked at the way we write out technical material, and determined that by the time you've read it in the text and looked down at a diagram, you've lost it. Too many steps. This tells us something about the human processing load. Given too much information at once, a learner simply won't be able to put it all together and keep hold of it. Whether the information is being presented concurrently, or too quickly, may not matter. This may indicate that information presented too rapidly, whether one graphic or more are on the screen concurrently, will be a factor that is significantly difficult for female learners.

On the other hand, some researchers insist that animations need narrations/ concurrent information, whether written or verbal (Mayer and Anderson 1991). This does support Paivio's dual-coding theory; ideas will be better recalled if coded as both picture and word. This doesn't, however, tell us much about two pictures presented simultaneously. If two stimuli are presented simultaneously, does an increase in cognitive load result, cancelling out the positive gains from dual-coding, or does it reinforce dual-coding? Research indicates that information given concurrently can vastly improve learning over successively given information (Mayer and Anderson 1992, Reed 1992), an effect that holds most strongly for high spatial skill groups (Mayer and Sims 1994).

Type of Change or Movement

What does motion of any sort do for the viewer? "Research has shown that people are often predisposed to select information based on physical characteristics (e. g. color and motion) as well as information that is novel or unique" (Rieber 1991, p. 6). A valid question, however, is how do different types of motion affect the learning process? There is a substantial body of literature that supports attentional motion as valuable (Blake 1977, Dwyer 1968, Rieber 1989). Additionally, there are many different kinds of motion tested as animation in the literature. An apparent gap, however, rests in the fact that no one has tested these varying types of motion (additive information, informational change, attentional movement, representative motion, rotation) against one another. Most researchers use a single presentation to test their results, and even when more than one presentation is used, the type of motion is constant. Four of the five types of motion are represented in this study, however, often two types of motion are present in the same animated presentation. This will make it fairly difficult to determine whether the type of motion has a great effect on presentational variation.

Type of Graphics Used

The category 'types of graphics used' may benefit from an exploration of information dealing with experimental response to differential levels of abstraction. Studies indicate that learners do not gain more from progressively higher detail graphics, and in fact may experience retention problems when detail levels are too high (Borg and Schuller 1977, Moore *et al.* 1979), so some level of abstraction, or at least generalization, is positive. Each of the four learning types has a different capacity to learn from abstract information (Kolb 1976). Eastman (1985) examined the way a map user incorporates visual information into a memory structure. Eastman used the concept of 'memory chunks' to determine the best hierarchical graphic structure to aid in map memorization. He determined that individuals did memorize information in chunks of one kind or another, but Eastman could not find a single form that was used most often. Rieber (1991) also found that grouping images and text improved recall. The experimental subjects in Eastman's and Rieber's research imposed a visual hierarchy onto the mapped information with the express intention of using that hierarchy to improve their recall. Each subject did this differently, but overall it worked. Eastman's research makes a fairly strong argument that organization, and association, are valuable visual-spatial cognitive cues. Other researchers have identified orientation (Shepherd and Metzler 1971) and color (Dwyer 1971) as important in cognitive processing of information and memory retention. In order to better understand how the presentations used in this experiment fit into these categories, I have included a table that details how the presentations in this experiment fit within the variables described above (Table 4).

**Confounding Variables: Individual Schemata**

This dominant paradigm in cognitive thought treats the cognitive system as an elaborate pattern matching system (Margolis 1987). Within this paradigm, everything we encounter is a pattern, and when we incorporate a pattern into our memory structures and it becomes part of our schema ("A mental codification of experience that includes a particular organized way of perceiving cognitively and responding to a complex situation or set of stimuli" {Webster's Ninth Collegiate Dictionary}), it is available to 'cue' our next pattern (Margolis 1987). This paradigm encompasses all areas of human experience, verbal, experiential, physical, and visual, but in this research I am solely concerned with the visual. Processes involved in visual cognition are commonly called visual-spatial skills (Halpern 1986). This is interesting in relation to the research already discussed in the literature review detailing gender differences in spatial skills.

Pattern recognition makes perhaps the most sense in terms of visual cognitive processes. When faced with this paradigm, people are often drawn to categorizing visual pattern recognition features down into discrete pieces. This is not really possible, however much it might be desirable. Cognitive experimenters have tried for years to come up with a computer program that can handle pattern matching even as well as a six year old child can, and have failed (Margolis 1987). No computer can match a human being in critical problem solving either, as evidenced by the relative failure of artificial intelligence systems (AI) to date. "So it is essentially universal to concede an

**Table 4**

**Experimental Variables for each Presentation**

<u>Concurrent verbal/written information (% of presentation)</u>

100% of all presentations have concurrent verbal and graphic information.

<u>Coincident visual information (% of presentation) - animation portion only</u>

Soil Acidification - two graphics on screen at same time - 95% of presentation

Forest Hydrology - two graphics on screen at same time - 95% of presentation

Plate Tectonics - two graphics on screen at same time - 80% of presentation

<u>Speed of the presentation (bytes per unit time)</u>

Soil Acidification - 27 bytes per 3 minutes 50 seconds (1 byte/8.52 secs.)

Forest Hydrology - 27 bytes per 3 minutes 0 seconds (1 byte/6.67 secs.)

Plate Tectonics - 30 bytes per 2 minutes 15 seconds - animated (shown twice) (1 byte/4.5 secs.); 4 min. 30 seconds - static (shown once) (1 byte/9.00 secs.)

<u>Type of change or movement - animation portion only</u>

Soil Acidification - Representative motion

Forest Hydrology - Additive information, Attentional movement, Representative motion

Plate Tectonics - Representative motion, Informational change

<u>Type of graphics used</u>

Soil Acidification - Graph (animation also has pictorial representations)

Forest Hydrology - Graph (animation also has pictorial representations)

Plate Tectonics - Pictorial representations essential role to the cuing of patterns and patterned responses; but the articulation of just what is happening when a pattern is recognized is an unsolved problem" (Margolis 1987, p. 3). In this research, by chopping my presentations into quantifiable aspects, I am artificially imposing a non-inclusive categorization on external aspects of a presentation in order to try and isolate how they affect internal processes. Unfortunately, as explained above, it is not possible to isolate the internal processes themselves.

The backlash to this lack of quantifiable, identifiable 'cues' has led Pylyshyn, a major voice in cognitive theory, to argue forcefully against visual-spatial cognition and for propositional cognition in relation to visual images. *Propositional theory* is a theory in which our visual cognitive processes are based on our interpretations of images rather than generalized replicates of them (Wood 1983). In quantitative, experimental study, it is not actually possible to prove whether an internal representation is pictorial or propositional (Anderson 1978). Pylyshyn may insist that any type of image that we do 'see' in our mind's eye is simply 'functional architecture' as opposed to cognition (Margolis 1987), but I tend to accept the pattern recognition theory more easily. Imagine for a moment a bookshelf in your office. Pick out a blue book from the shelf. How did you pick this book? Were your methods propositional, or were they visual and associative? I know that I can tell you exactly where most of the books are in my bookshelves specifically because I know what size objects are in what pattern on the shelves, and I can see the colors in the pattern in my mind's eye. Without my pictorial mental image I'm lost.

There are other factors that must be considered. Schema are certainly important to individual differences in cognition, and these are unique to each of us (Margolis 1987). In fact, our response times for visual recall or pattern matching increase proportionally with the number of unusual objects or features that we are exposed to (Intons-Peterson and McDaniel 1991). All sensory input that we receive throughout our lives is linked with our schema. We remember things better if they are linked with concrete experiences in our lives (Paivio 1971); this is especially true for women (Kolb 1976, Roychoudhury *et al.* 1995). Part of this variable was isolated in this research by eliminating subjects having specialized schema giving them an advantage over other subjects in the study (experience). Altogether, however, individual schema are too detailed to identify completely. They are a red herring in any psychological analysis.

## Gender Differences in Visual Cognition

Differences in visual-spatial abilities exist, although the exact magnitude of these differences often comes into question (Tapley and Bryden 1977, Newcombe 1982, Linn and Petersen 1985). There is evidence that ethnic differences exist as well (Catsambis 1995). On tests of visual-spatial ability, women's scores are often more variable than men's. Even so, some tests reveal striking differences between men's and women's performance, while others do not. In this section, I have outlined the literature available on gender differences within each of the categories found in Table 3.

### Variables 1 through 3

We can relate the presentation of concurrent information, either verbal, written, or visual, in terms of simultaneous and successive processing. Successive processing is usually looked at as a lower level type of learning, and one more prevalent in children. It is interesting that simultaneous processing is also consistently referred to as spatial in nature (Das *et al.* 1979). There is a spatial component to any information that we process simultaneously. If the spatial aspect of simultaneous processing is an accurate assessment, it is interesting that children tend to successively process. This indicates a plausible relationship between women and successive vs. simultaneous processing as well. If simultaneous processing is spatial, and children lack advanced spatial skill, it makes sense that they would preferentially use successive processing. If women also have lowered spatial skill, a learning presentation requiring simultaneous processing might be more difficult for them. Concurrent verbal/written information, coincident visual information, and speed of presentation may all be important variables in terms of gender differences.

Type of Change or Movement

One of the most clear-cut, and widely cited, differences between genders in spatial skill is noted in object orientation (Shepherd and Metzler 1971, Tapley and Bryden 1977, Halpern 1986). Orientation is not being tested in this research, however, and no other literature has been identified as to gender differences in response to different types of motion.

Type of Graphics Used

The categories in 'types of graphics' exist on a continuum from concrete to abstract. Women tend to prefer concrete learning (Kolb 1984), and tend to perform better on tasks dealing with concrete phenomena (Roychoudhury *et* *al.* 1995, Sharps *et al.* 1994). It is expected that the placement of a given presentation on the concrete-abstract continuum will have an observable effect on differences in men's and women's scores for each of the different presentations.

Tests that measure discrimination capacity for chroma and hue tend to show no gender differences. Color-blindness, however, is more prevalent in men than in women. In order to keep from introducing gender bias into my experimental population, all color-blind individuals have been eliminated from the analysis.

## Explanations for Gender Differences

There are various theories that attempt to explain gender differences in visual cognitive skills. These include sex hormone differences, brain lateralization, biological determination, genetic dominance differences (Darwinian - hunter vs. gatherer skills), and psycho-social conditioning. Of all of these, the brain lateralization theory is taken the most seriously, but also has the most confusing evidence. The two loudest voices on the subject are both postulating opposite hemispheric specialization (Halpern 1986). Psycho-social conditioning has a large following as well, supported mainly by the lack of visual-spatial skill gender differences before the age of thirteen. Whichever theory or theories you look at, no overwhelmingly convincing evidence has been produced in support. The only sure thing that can be said is: differences do exist.

# RESULTS AND DISCUSSION

This section is arranged by topic. Initially, I review the research design in light of the results, and address how the experimental design may have affected the results. Subsequently, each of the six hypotheses is addressed in turn. Following this, qualitative data are disclosed, and data are examined that, although extraneous to the hypotheses themselves, suggest interesting ideas about the experimental population and experimental conclusions.

All statistical results were obtained through the use of ANOVAs at a significance level of 0.1, unless otherwise noted. ANOVA tests are generally used to identify variation from the mean from any one of several sources, in order to determine whether two or more samples come from different populations or not. Using an ANOVA in a blocked design (Latin Square) allows us to partition various sources of error within the data set, thus lowering the amount of variation in the data set that is unaccounted for, or the 'pure error'. The overall population means being compared, in this case, would be the quantitative scores on experimental material of the male and female students using the animated graphic, static graphic, and verbal (no graphic) presentations, for long-term and short-term testing. Through successive ANOVAs (short-term, long-term), it was possible to determine whether variation from the mean in the data set is due to any one factor, such as presentation topic, class, treatment, gender, or some combination of these. The ANOVAs for this experiment were performed on three levels of detail. At a more general level, data sets were examined for a significant difference between genders within topic and treatment, using class averages. At the next level, analysis was carried out by crossing gender with several other variables to identify more targeted variation within the data set. Finally, data were analyzed at the student level. At this detailed level, almost all of the variation in the data set was accounted for in the model. More detail on the ANOVAs used, and the exact experimental model may be found in the methods section.

## Research Design in Retrospect

This study was designed using two experimental subgroups. Subgroup A, three topics X three treatments, was carried out in a Latin Square design. In other words, each class viewed all three topics, and received all three treatments, such that no topic and no treatment was repeated in any given class. Within the three classes of the subgroup, data were produced for each topic using each treatment. This is a randomized experimental design; for more detail, see the methods section. Subgroup B (three topics X three treatments) was carried out in a stratified random experimental design for treatment placement within topics. The purpose of subgroup B was to have a set of replicates for subgroup A. It was subsequently determined, however, that the differences in experimental design made the inclusion of subgroup B into statistical analysis, sketchy at best. The limitations imposed by the use of subgroup B as a replicate follow.

First, in subgroup B, treatment data averaged across the three presentations cannot be used. Whereas in subgroup A each class receives one of the three treatments on one of three topics, subgroup B has one class that receives the same treatment on all three topics, and its other two classes receive two of the same treatments on two topics (Figure 7). Since class variation is significant within subgroup B ($p = 0.0001$), class variability could compound (or conversely obscure) any treatment variability in the data set.

Subgroup A: Latin Square Design

| | Class 1 | Class 2 | Class 3 |
|---|---|---|---|
| Soil Acidification | Static Graphics | Animated Graphics | Verbal Instruction Only |
| Forest Hydrology | Verbal Instruction Only | Static Graphics | Animated Graphics |
| Plate Tectonics | Animated Graphics | Verbal Instruction Only | Static Graphics |

Subgroup B: Stratified Random Design

| | Class 4 | Class 5 | Class 6 |
|---|---|---|---|
| Soil Acidification | Verbal Instruction Only | Animated Graphics | Static Graphics |
| Forest Hydrology | Verbal Instruction Only | Animated Graphics | Static Graphics |
| Plate Tectonics | Static Graphics | Animated Graphics | Verbal Instruction Only |

Figure 7. Differences in Sampling Design

Another source of variation within subgroup B exists in that two of the three classes in subgroup B were taught at another school, Western Oregon State College. The inclusion of classes taught at Western Oregon State College (even though instruction was done by the same instructor, on the same subjects, in the same order, and to the same class level) is a significant source of variation in subgroup B ($p = 0.0001$). The school variation would introduce a confounding variable into the analysis.

Aside from the differences in experimental designs and the variation introduced by school in subgroup B, other problems exist. In the two classes in subgroup B that were not taught at Oregon State University, fewer students chose to participate in the study, and a higher absence rate resulted in much missing data. These differences resulted in the acquisition of unbalanced data between the two subgroups. As a cumulative effect of each of these problems, except in rare cases, p values will not be reported for subgroup B, as interpretation of these values would be difficult to perform with any degree of certainty. Subgroup B will be used as a replicate for subgroup A in a limited sense. Interestingly enough, with all of the problems associated with subgroup B, its descriptive data trends parallel, very closely, those identified in subgroup A (although at a different level, and within a different range). Numbers are different between the subgroups, but patterns in the data replicate nicely. Results from subgroup A form the foundation of this analysis. Results from subgroup B should be used with coincident caution.

Student differences accounted for a large proportion of the variation identified in the data set. This should not come as a great surprise; if a high level of variation did not exist in a normal student population, we would not be as easily able to assign a wide range of grades within a classroom population. The data in this experiment were collected for short-term and long-term testing. Student variation is not significant in the short-term data for subgroup A, $p = 0.1400$. In fact, student scores were both high, and more homogeneous in the short-term data, indicating that the short-term tests were not difficult enough to produce great variation within the data set. Gender was the only variable that did show significant variation within the short-term data set (subgroup A, $p = 0.0496$). The lack of overall variation in the short-term data set indicates that long-term scores may be the best indicator of student performance. This is further supported by the fact that long-term learning is the objective of most educational endeavors. Student variation in subgroup B is significant ($p = 0.0010$). These differences are, once again, probably attributable to school differences. Experimental design in subgroup A has subsequently been identified as the more efficient of the two experimental designs for this experiment. In future research, replicates should each receive this 3 X 3 treatment successively (3 X 6, or 3 X 9, etc...), forming a Latin Rectangle experimental design. This design would allow for increased degrees of freedom, and possible presence of interaction effects could be identified. For this experiment, however, we are left to deal with the limitations introduced by the experimental design in subgroup B.

## Hypothesis One

In this hypothesis, I attempted to identify gender differences within experimental treatment groups. Hypothesis one states: women benefit more than men do from the use of computer animation in CBI, when learning physical science concepts. Since short-term scores do not produce significant differences, I looked only at long-term scores for this section of the analysis. I expected that if women benefitted more than men from animation, I would be able to look at the quantitative scores and identify an improvement in women's scores from verbal to static to animated graphics across the presentational topics. This hypothesis suggests that women's improvement should be more pronounced than men's.

I initially analyzed variation within the data set using an ANOVA. Subgroup A alone is used in this section of the analysis, as subgroup B's experimental design and limitations render it inappropriate for the questions asked in this section (discussed above). When looking at the interaction of gender and treatment in the analysis, if women benefit significantly more than men do from the use of animation, we should see a high amount of variation in the data set due to gender between treatments, and we do not. Gender effects between the experimental treatments are non-significant for long-term data ($p = 0.4270$). When we look at the graphed data of average student score for men and women across treatment types, however, differences are visible in terms of how women's scores change between experimental treatments, and how men's scores change across experimental treatments (Figure 8 and Figure 9). Note that the difference looked for here is not so much the 'difference' between men's and women's scores, but the different ways that men's and women's scores either do or do not improve from verbal to static to animated graphics. Women show little change from verbal instruction to static graphics, but show definite improvement from static graphics to animation. These patterns are comparable in both subgroup A and B. The lack of statistically significant gender differences through the use of the ANOVA tells us that men and women are not statistically different enough in their scores across all treatment types to obtain significance. However, this hypothesis asks a more basic question. Did animation help women learn? To answer this question I analyzed the data using a binomial

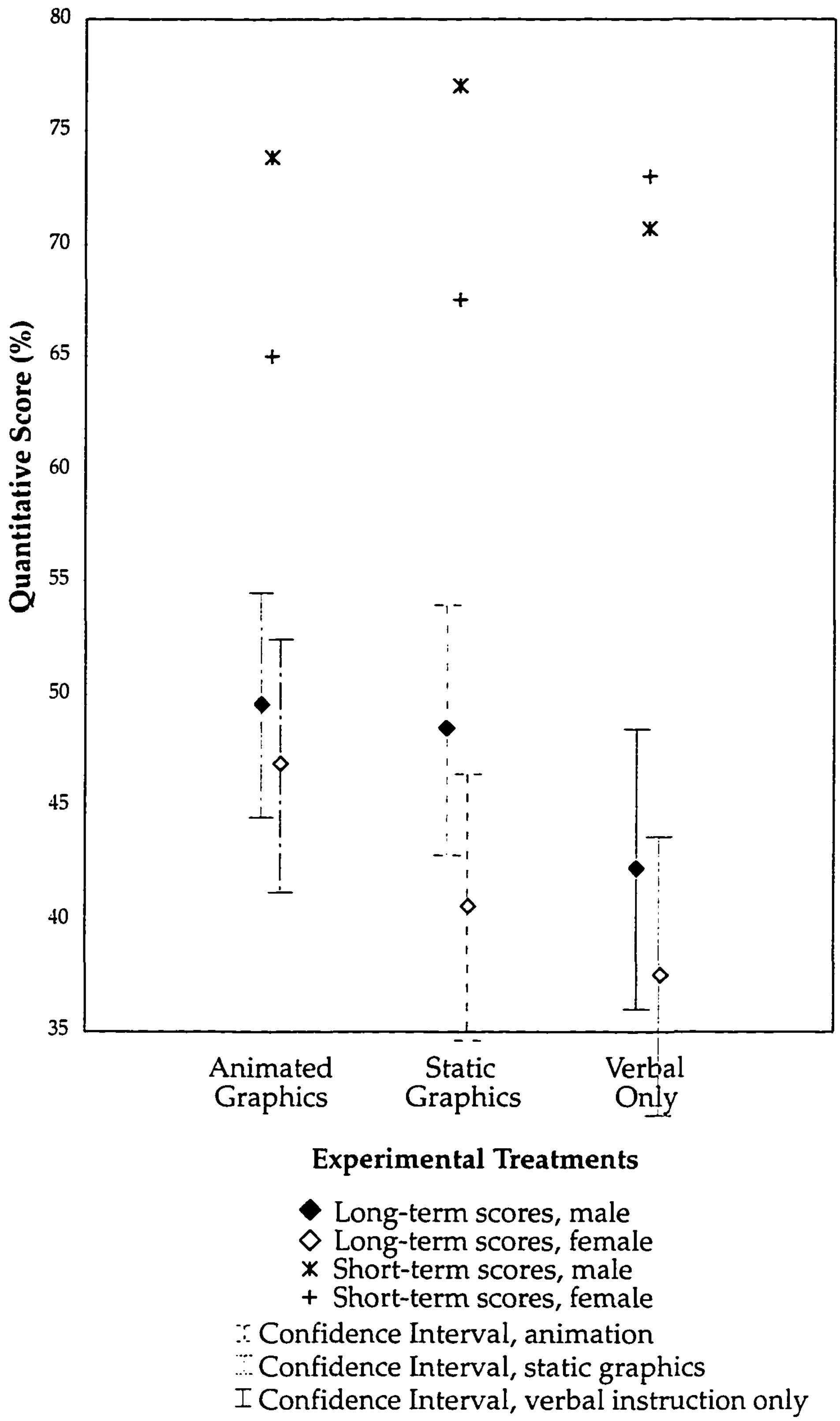



Figure 8. Changes in student performance between experimental treatments, subgroup A.

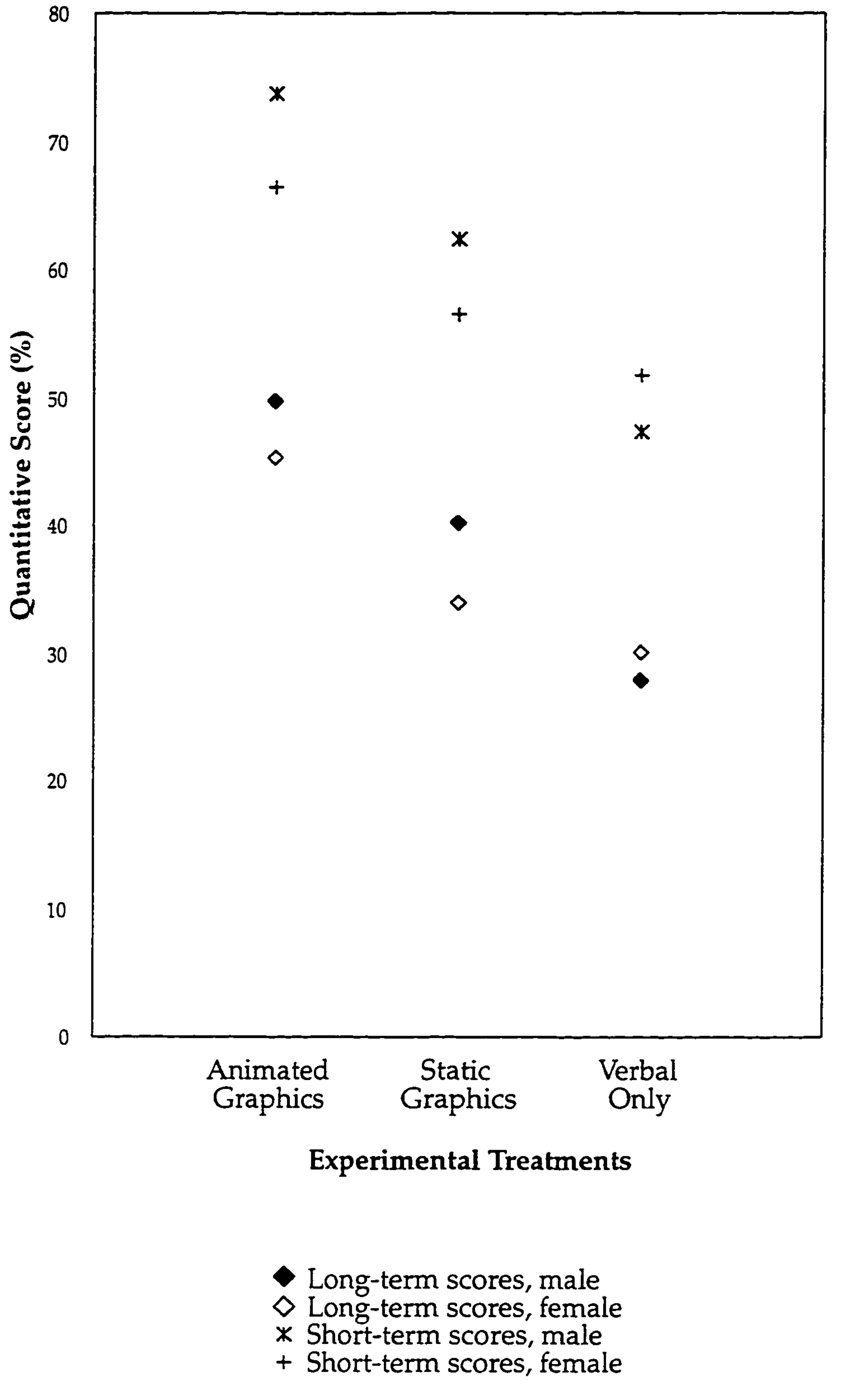



Figure 9. Changes in student performance between experimental treatments, subgroup B.

test for differences of proportion (methods section) looking only at the difference between women's scores using static vs. animated graphic instruction. Women's improvement from static to animated graphics (averaged over all three presentations) is statistically significant (binomial test for differences in proportion, $p = 0.067$). Men, on the other hand, seem to do almost the same whether using animation or static graphics, but improve quite a bit from any treatment including graphics over verbal instruction alone. This difference is not significant. The difference between animated and static graphics is the focus for this research (verbal instruction alone is simply a control group), and women clearly and statistically, benefit more than men do from the use of animation (over static graphics).

Other interesting information is available in looking at the confidence intervals. First, it is clear that confidence intervals overlap to a large extent for each of the treatment types. This undercuts, to some degree, the surety of conclusions indicating superiority of one treatment type over others. It is important to note, however, that the confidence intervals for verbal only instruction and static graphics overlap to a far greater extent than do those for verbal only instruction and animated graphics (this is the difference between a two-thirds overlap and a one-thirds overlap in terms of women's scores on the long-term testing). Also, confidence intervals around animated graphics means for both men and women were slightly smaller than those for the other two treatment types. Smaller variation around a higher mean speaks well of animation as a learning tool. Probably more important than these observations, however, is the amount of overlap in the confidence intervals for men's and women's long-term scores within each of the treatment groups. Notice that confidence intervals are between 50 and 60% different (in terms of area of overlap vs. independent distribution) for static graphic and verbal only instruction, while men's and women's confidence intervals for the animated graphics means coincide very closely. If our goal is to determine whether animation helps equalize men's and women's learning, this graphic certainly lends support to this conclusion.

Short term data trends show the same pattern as long-term data trends in subgroup B and in men's scores for subgroup A. In women's short-term scores for subgroup A, however, women's scores decrease from verbal to static to animated. Given that the opposite pattern in women's short-term learning is observed between subgroup A and B, it is possible to state that short-term scores are highly variable across treatments. Again, long-term testing is the primary goal of any educational endeavor (with the possible exception of cramming for finals, which is not recommended), the trends in the students' long-term data are seen as more important in terms of the experimental conclusions as a whole.

When the presentations are examined separately looking only at the graphs showing mean scores for each presentation (again for long-term data), there is improvement in women's scores from verbal to static to animated graphics on the soil acidification and forest hydrology presentations. On the plate tectonics presentation, however, women's improvement moves from static to verbal to animated graphics. These patterns hold for both subgroup A and B (Figure 10 and Figure 11). Women do best in an animated format, but better in a verbal format than with static graphics on the plate tectonics presentation. This is interesting in that it may indicate that under certain conditions, static graphics are inappropriate to a subject matter, and yet animation is of benefit, at least when dealing with women's learning. This is unexpected. In contrast to these thoughts, confidence intervals for each of the experimental treatments are different to a large extent on the soil acidification

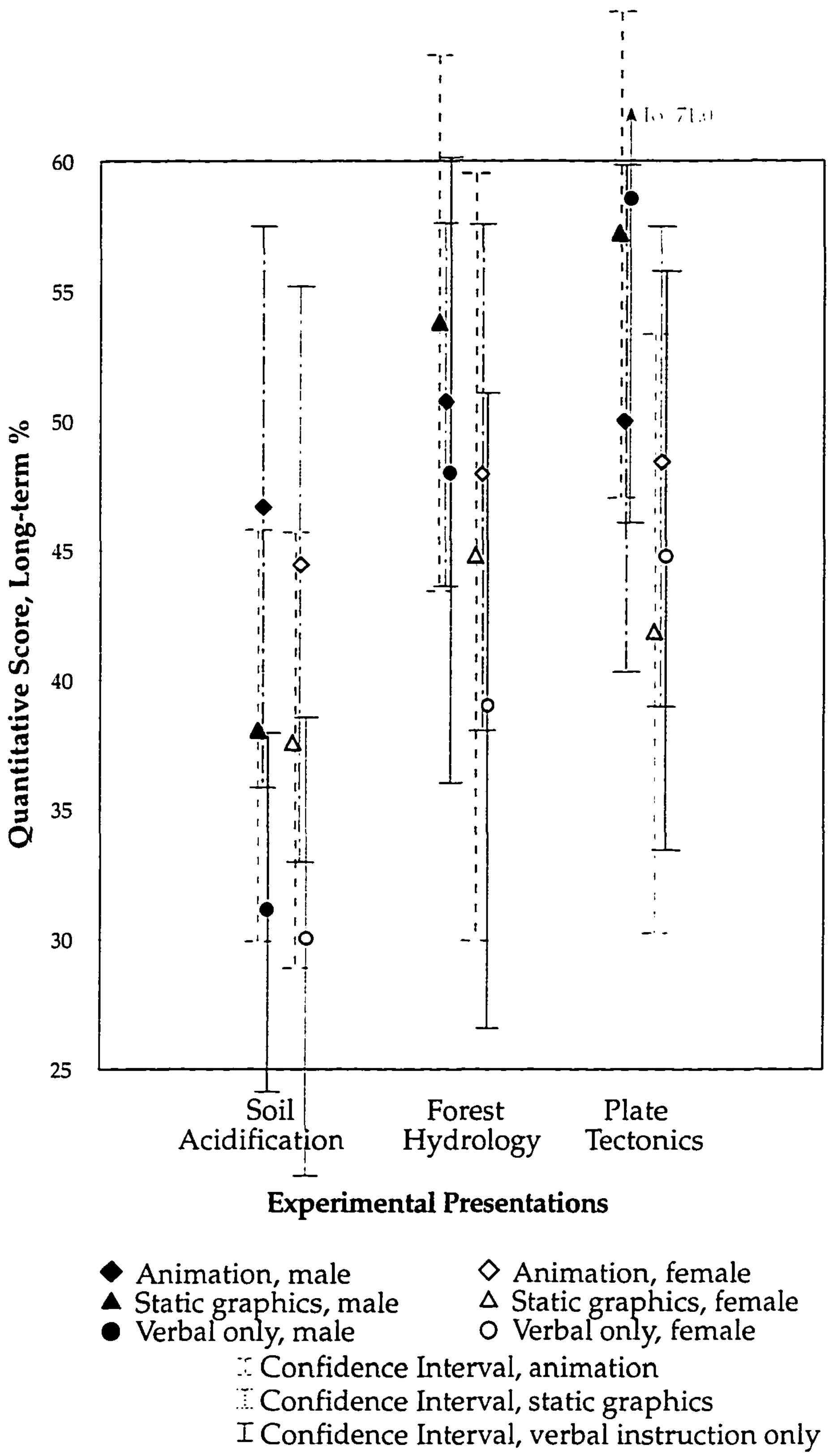



Figure 10. Changes in student performance for each experimental treatment, separated by presentational topic, subgroup A.

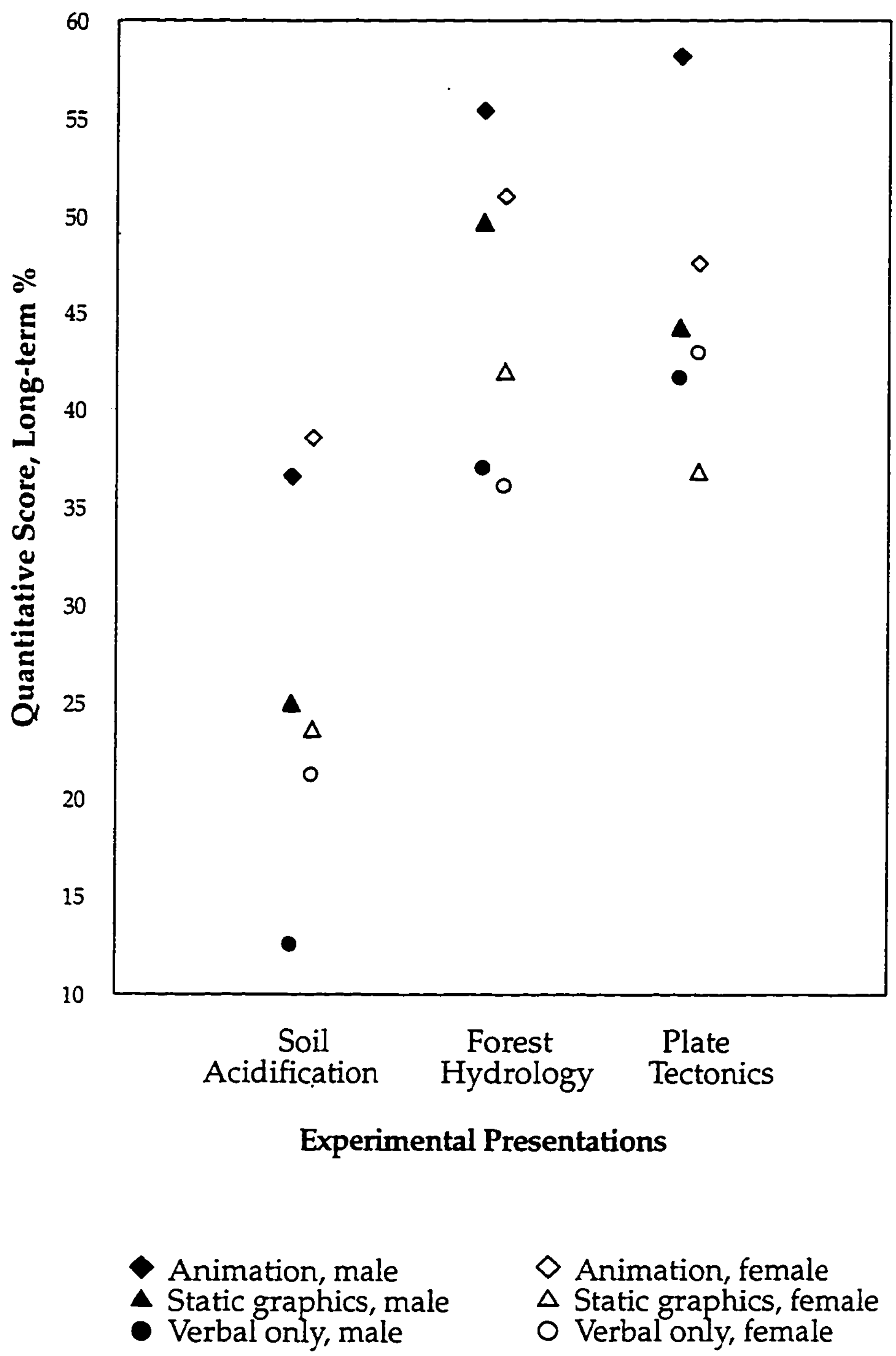



Figure 11. Changes in student performance for each experimental treatment, separated by presentational topic, subgroup B.

presentation, but coincide almost completely on both of the other presentations. This suggests that the differences between presentations in terms of which treatment type is the superior educational tool, may be dependent, to some degree, on the design or topic of the learning presentation. Lack of large, observable differences between confidence intervals, in the plate tectonics presentation especially, indicate that the descriptive data trends in this presentation might not be as indicative of treatment effects as the descriptive data trends in either of the other two presentations.

Leaving confidence intervals for a moment, and looking again at averages within each presentation, patterns in the data for women's scores hold for men's scores as well, with the exception of the plate tectonics presentation. In the plate tectonics presentation, performance increases from animated to static to verbal in subgroup A, and from verbal to static to animated in subgroup B. This odd pattern in the men's and women's data for the plate tectonics presentation is clearer in looking at raw data as opposed to simple averages. Looking at Figure 12, it is clear that the range of scores for women is approximately the same for all three treatment types in the plate tectonics presentation. (Please note that in this graph clustered data points do not represent actual values. There are several points where several data points coincide, and have been separated slightly so that data distributions are visible. For each 'cluster' of over two data points, the actual value is the mean of the cluster.) For men, the range of scores is smallest in the animated presentation, and about the same in the static and verbal presentations. Men's verbal mean score seems to be somewhat influenced by two outlier data points. Even so, 60% of the scores fall above the 50% 'failure' cut-off point; this is approximately equal to both of the other treatment types.

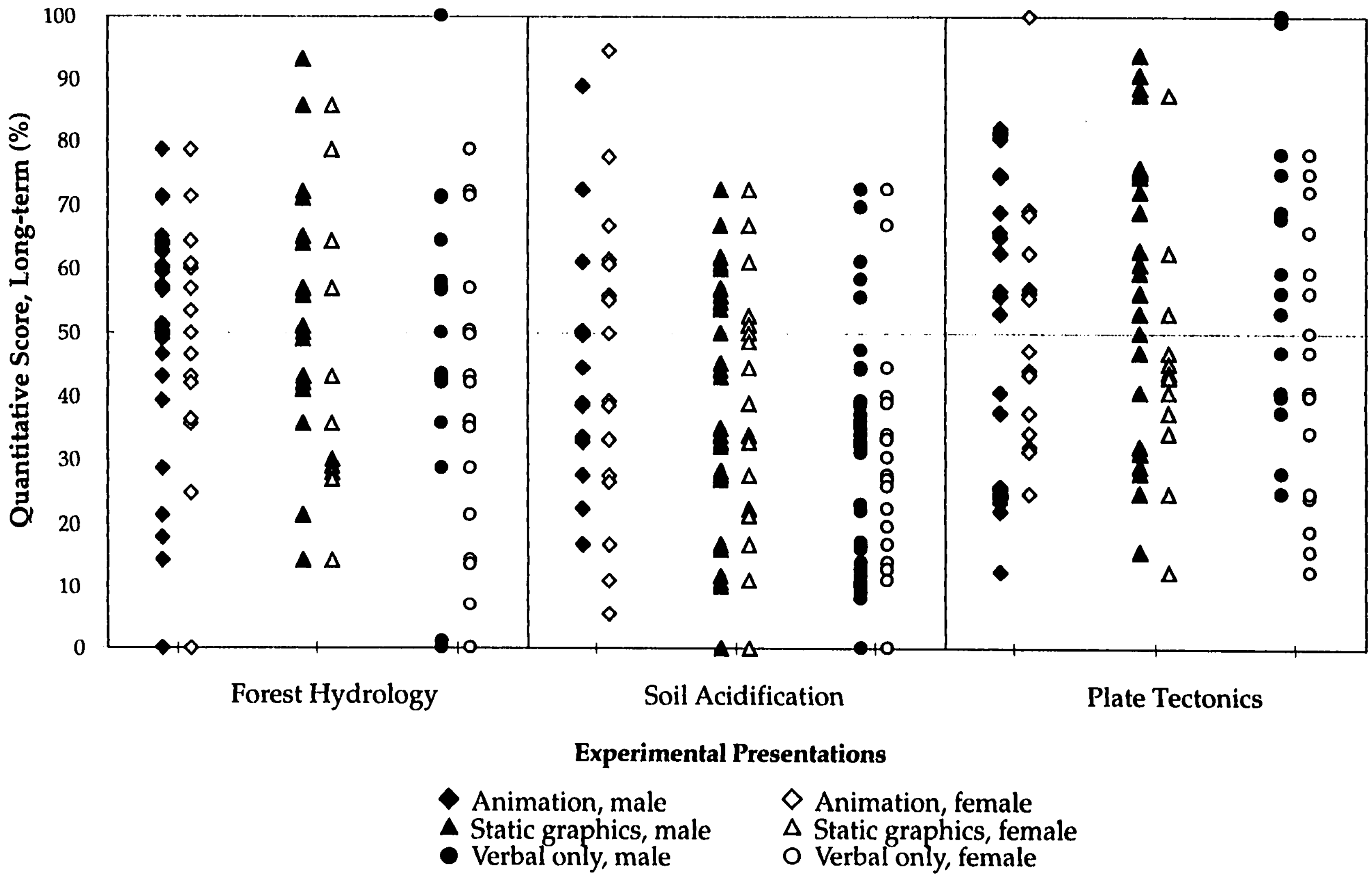



Figure 12. Raw scores for students in subgroup A by topic, treatment, and gender.

Women's scores may also have been influenced by two outlier data points, one in the animated plate tectonics and the other in the static graphics plate tectonics presentation. In terms of sheer percentage of observations over the 50% 'failure' cut-off for women, women's scores in the animated graphics data have approximately 38% of the data points over the 50% 'failure' cut-off, whereas static graphics data have only 25% of the data points over the 50% 'failure' cut-off. Verbal instruction alone is more dispersed, and has about 37.5% of the data points over the 50% 'failure' cut-off. This indicates that perhaps verbal instruction only was actually better for the plate tectonics presentation than indicated by the average scores for women. This is further supported by the confidence intervals seen in Figure 10 for the plate tectonics presentation. Although discussed earlier, it is worth repeating that confidence intervals for verbal and animated instruction on the plate tectonics presentation overlap almost completely.

Looking at the raw data for both of the other presentations, forest hydrology and soil acidification, seem to be in line with the averages recorded on Figures 9 and 10. There is one extreme outlier data point in the verbal instruction men's forest hydrology presentation, but excluding this point, if anything, would only exaggerate the pattern of the averages between treatments for that presentation.

It is important to note that in all progressions, women's scores improved with animation over all other forms of instructional treatment for long-term testing, in both subgroups, for all presentations. This is not true for men. So women may benefit more, and more consistently, than men do from the use of computer animation in physical science instruction, when we look at the data using descriptive statistics. On the basis of the both the qualitative trends in the descriptive data and the statistically significant difference between women's scores using animated vs. static graphics, this hypothesis is not rejected.

In an interesting sidelight to this hypothesis, the proportion of students who used graphics, or drawings, to explain their answers on exams differed between presentations and between treatment types. On the short answer section of the long-term testing, questions were asked about both the soil acidification and plate tectonics presentations, in which the use of drawings in an answer was appropriate. Long-term, short-answer forest hydrology questions asked students about a graphic, and thus did not require, or allow, the students to draw in support of their answers. Average use of drawings for both men and women increased from verbal to static to animated treatments on both the soil acidification and plate tectonics presentations in both subgroups. Increases in graphics use were greater for women than for men (Figure 13). Considering that graphics and the use of graphics is an inherently spatial skill or activity, the fact that women's use of graphics increased is positive, possibly indicating that women's confidence dealing with the spatial aspect of the subjects increased. It provides indirect support for the hypothesis that women benefit more than men do for the use of computer animation in teaching physical science concepts. Neither of the differences between student use of drawings for the animated and static graphics presentations is statistically significant, however (binomial test for differences of proportion).

### Hypothesis Two

Previous researchers looking at learning using animation have identified that there is a difference between short-term and long-term

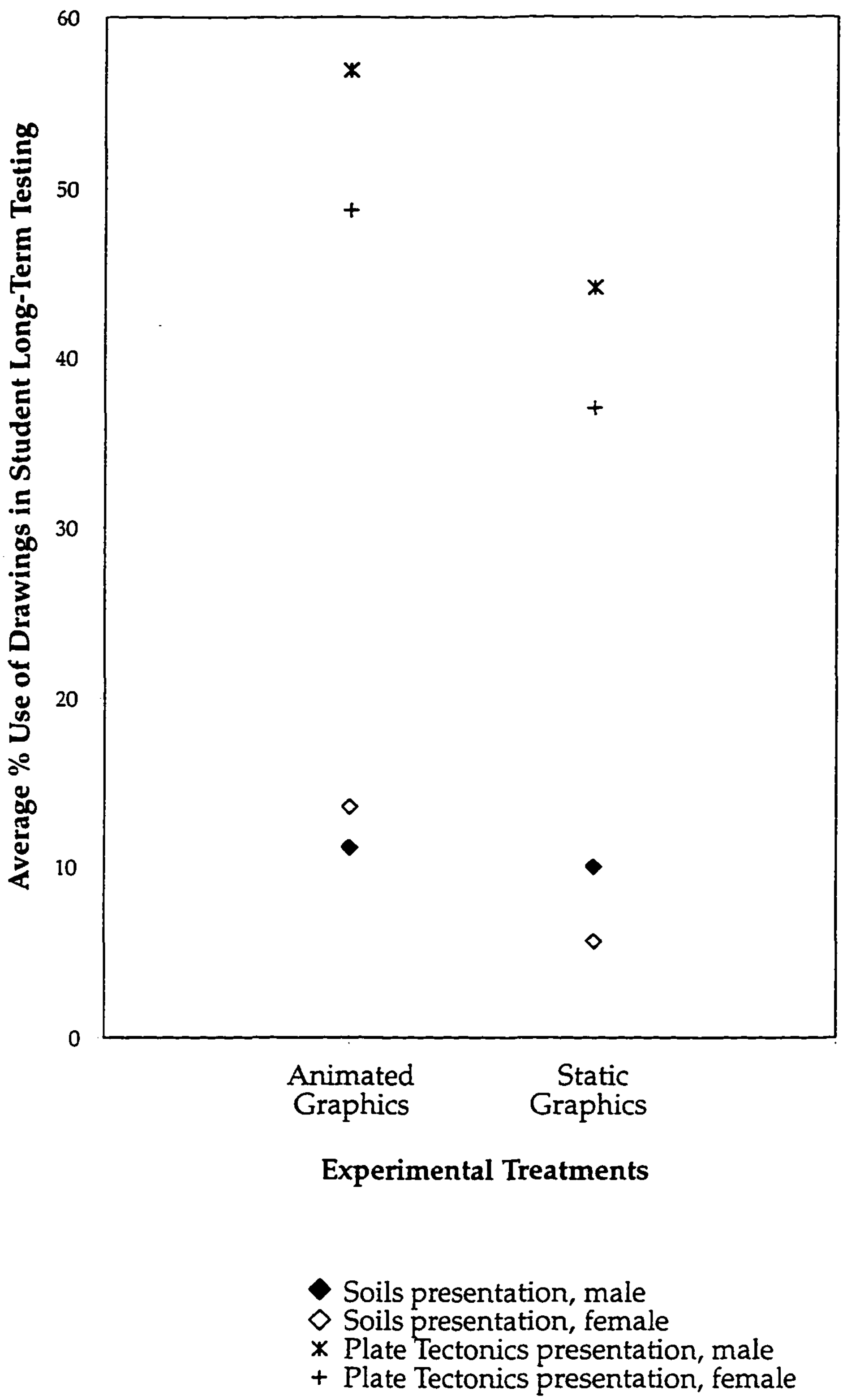



Figure 13. Student use of drawings to explain their answers during long-term testing, all classes.

learning that varies with treatment (static, animated) (Caraballo 1985a). This research does not support that conclusion. Hypothesis two states that the type of instructional medium (static vs. animated) will show no difference in effect between short-term and long-term scores. To test this hypothesis, I averaged the scores for both genders (this is not a gender based evaluation), for each of the three treatment types, animated, static, and verbal, and analyzed for variation in the differences between long- and short-term learning for experimental treatments using an ANOVA. In other words, I looked to see if long-term and short-term scores were closer together, or farther apart in one experimental treatment than in another. There is essentially no difference identified in the difference between short- and long-term learning for the three treatment types ($p = 0.8631$). The main source of variation between long- and short-term score differences was topics, not treatments. All treatments show a decrease in student scores from short- to long-term testing. This difference is standard throughout treatment types to within 3.23% (based on a comparison of average class differences), which is not surprising. Memory retention is sure to decrease over time. If genders are looked at separately, there are still no statistically significant differences between short- and long-term scores over experimental treatments. In all but one case, men's scores are above the average and women's are below. There is essentially no difference due to gender between short- and long-term scores over experimental treatments ($p = 1.0000$). Since no significant treatment effect is identified between long- and short-term scores, hypothesis two is not rejected.

Although no significant differences between long- and short-term scores across presentational treatments exist, it is interesting to note the average non-significant differences for each treatment type. The average percentage differences between long- and short-term learning for each experimental treatment are as follows: verbal instruction = 17.17%, static graphics = 18.9%, animated graphics = 12.28%. Keep in mind that short-term scores are always higher than long-term scores, thus, lowering the average difference between long- and short-term scores should be seen as advantageous. This is part of the reason that the range of animated graphics scores are slightly higher than those of the other two treatment types. Figure 14 shows a graph of total student scores (%) for each treatment type. The graph illustrates that the range of animated graphic scores tend to be slightly higher than static graphic scores, which have a slightly higher range than verbal-only scores. The difference between long- and short-term scores for each treatment type is not significant ($p = 0.1605$) using an ANOVA. Differences between the means for each of these treatment types can be found in Figures 15 and 16.

## Hypothesis Three

With hypothesis three I sought to identify variation within the data set caused by differences in the topics of the presentations. Hypothesis three states that differences between men's and women's learning will be non-standard among presentations, owing to differences in individual presentations. This research was based on the premise that animation would improve women's learning, because motion can enhance learning in low visual-spatial skill groups (Blake 1977, Rieber 1989a), of which women are members. Hypothesis three in particular relates back to this thought. It makes sense that if visual-spatial skill is our 'sticking point' in terms of equalizing men's and women's learning in the physical sciences, different

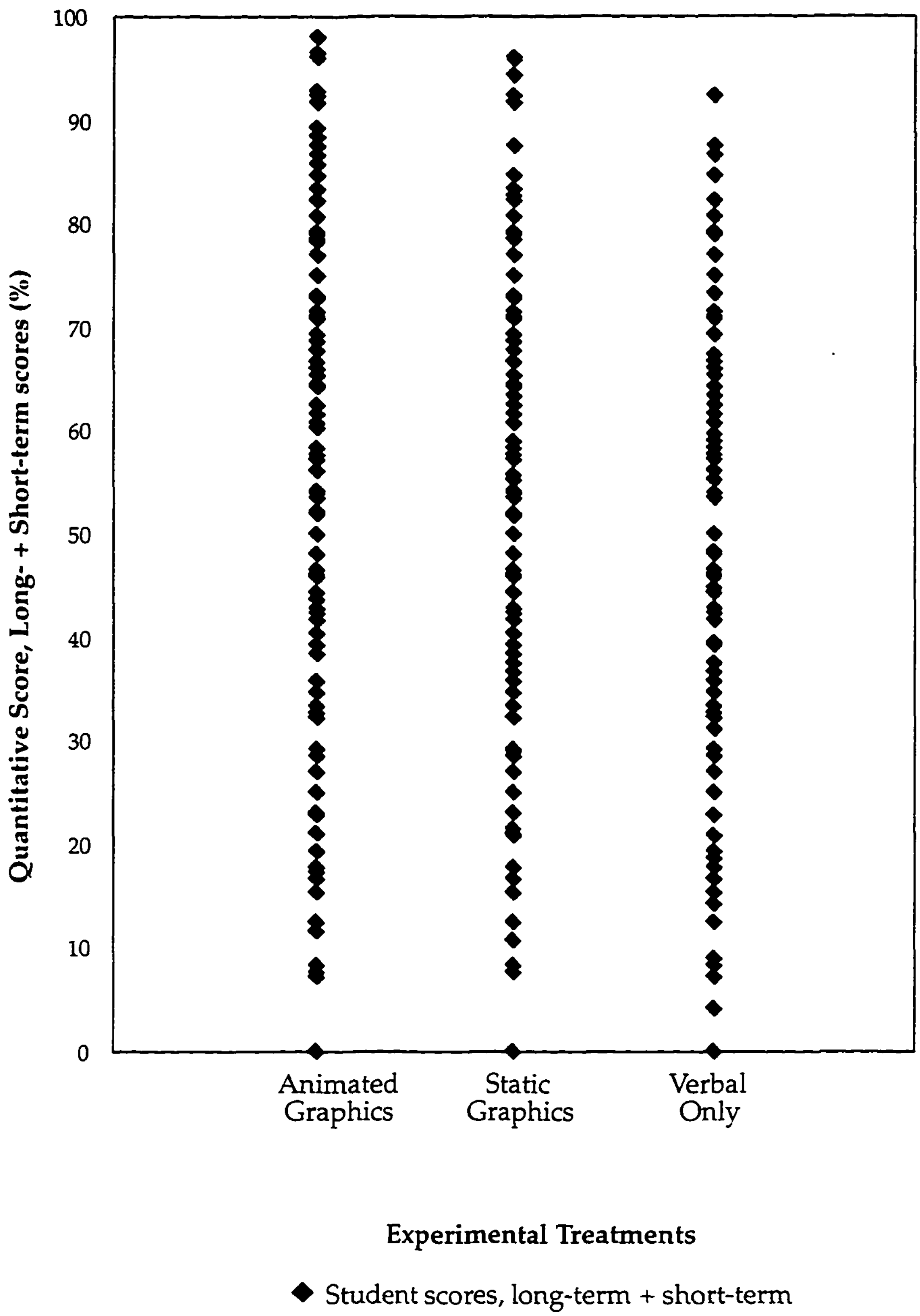



Figure 14. Scatterplot of all student scores for each treatment type, subgroup A.

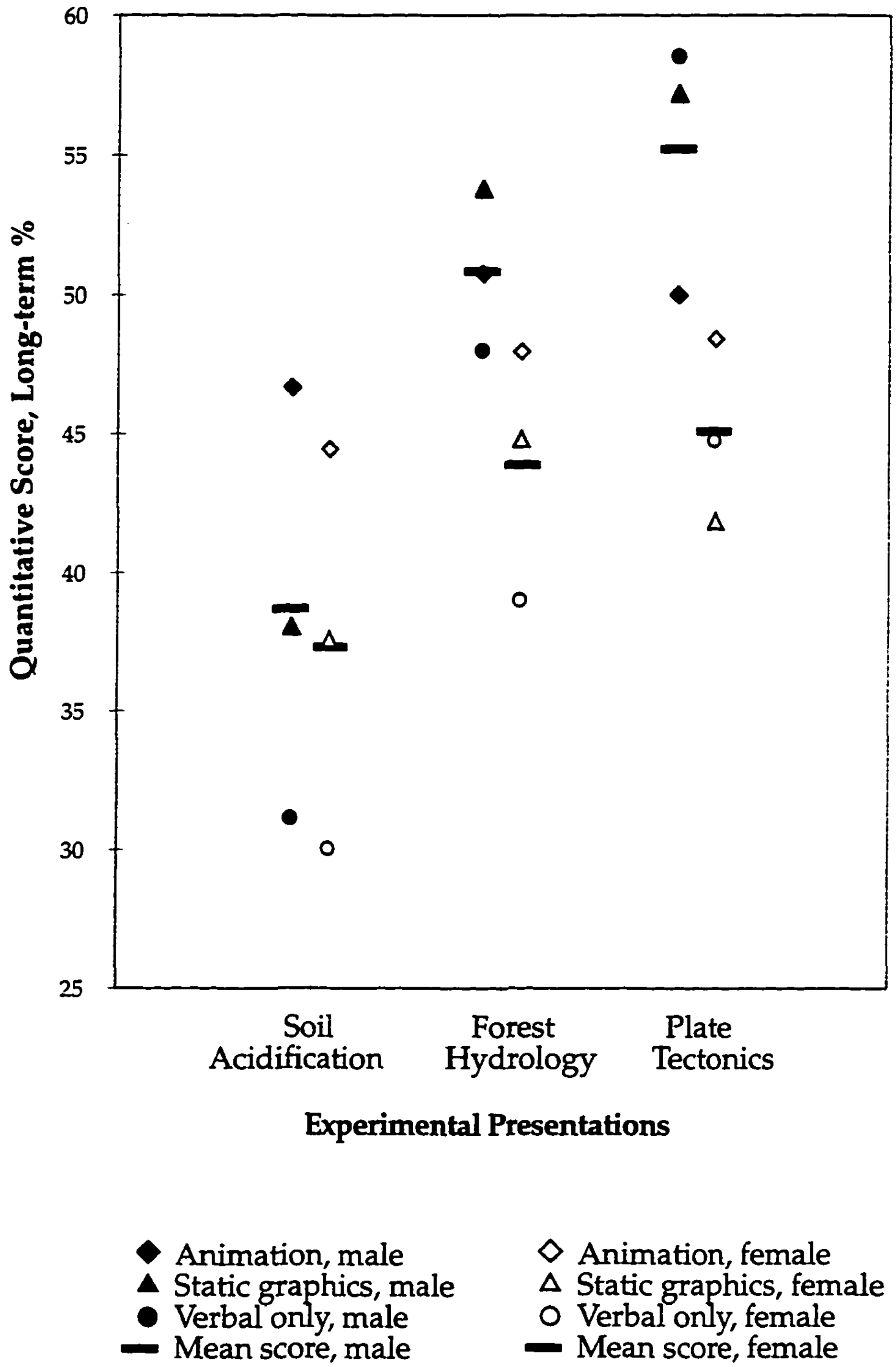



Figure 15. Mean long-term scores for men vs. women across presentations, subgroup A.

Figure 15. Mean long-term scores for men vs. women across presentations, subgroup A.

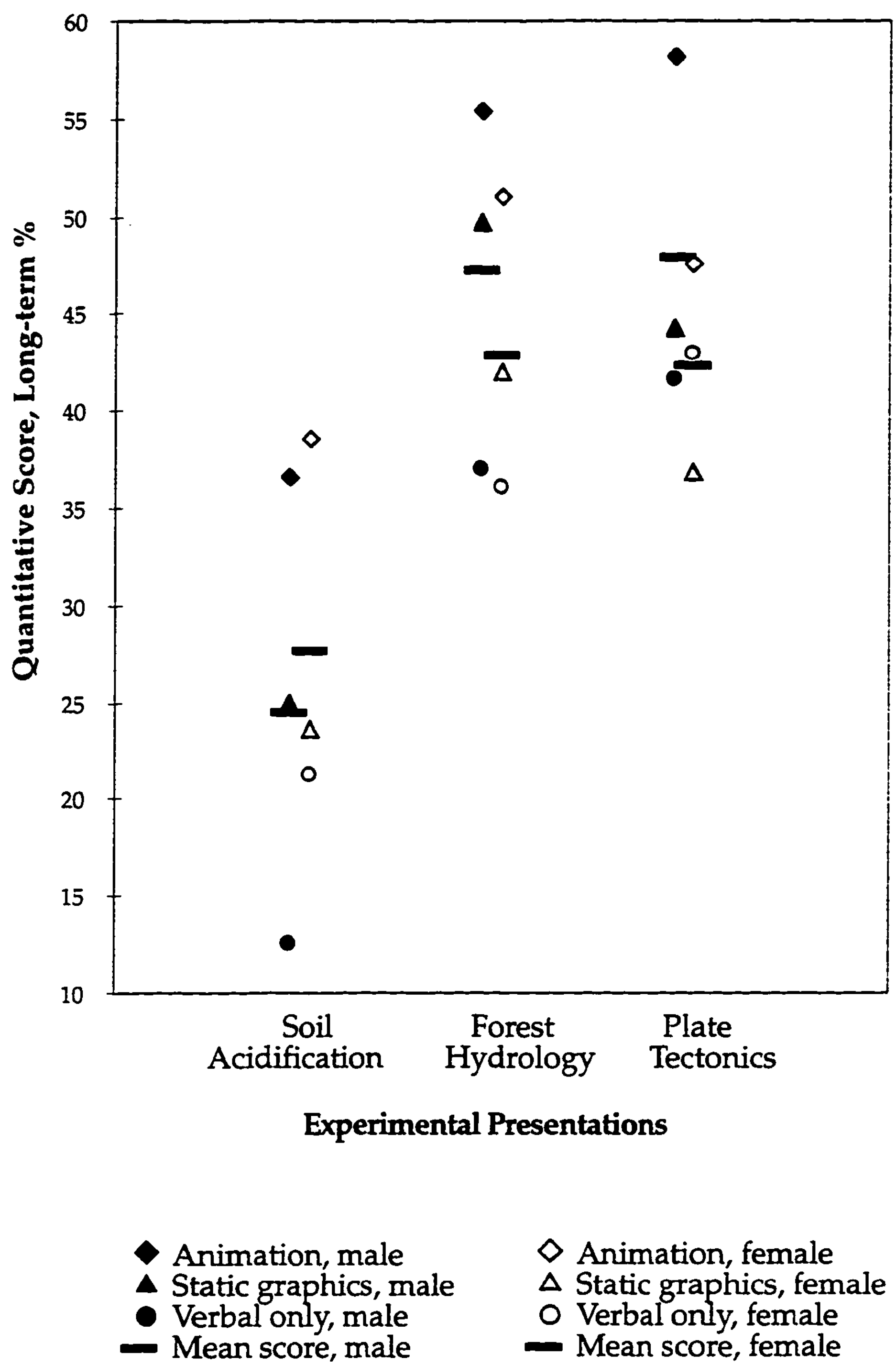



Figure 16. Mean long-term scores for men vs. women across presentations, subgroup B.

presentations may vary in the response variable being examined, and will thus exhibit varied differences between men's and women's scores. For example, if presentation Q is very spatial in nature, and presentation K is only slightly spatial in nature, it is likely that experimental subjects with higher vs. lower visual-spatial skill will perform differently on both Q and K, due to the spatial component of the presentation. This example can be repeated for a concrete or abstract nature of a presentation. Women's and men's style of learning in terms of spatial skill and abstract and concrete information has been addressed in an earlier section. In order to reject this hypothesis, we must be able to identify a consistent difference, or a lack of difference, between men's and women's scores for the three presentations. Subgroup A alone is used in the statistical analysis of this section, as subgroup B's experimental design and limitations render it inappropriate for the questions asked in this section (discussed above).

The place to begin looking for information on hypothesis three is in the graphed data showing the mean of scores for each gender over the presentations (Figures 15, 16, 17, and 18). These graphics show that men generally tend towards the higher scores for each presentational topic in long-term testing, and women tend towards the lower. This difference in mean scores is the highest in the forest hydrology and plate tectonics presentations, and the lowest in the soil acidification presentation for both short- and long-term testing. In fact, there is almost no difference between the means in the soil acidification presentation for either short- or long-term testing, and the two instances in which women's mean scores are higher than men's mean scores are found in the soil acidification presentation (subgroup B results only). If the scores are weighted, to account for gender variation across classes, the same pattern persists (weighting by total average class score by

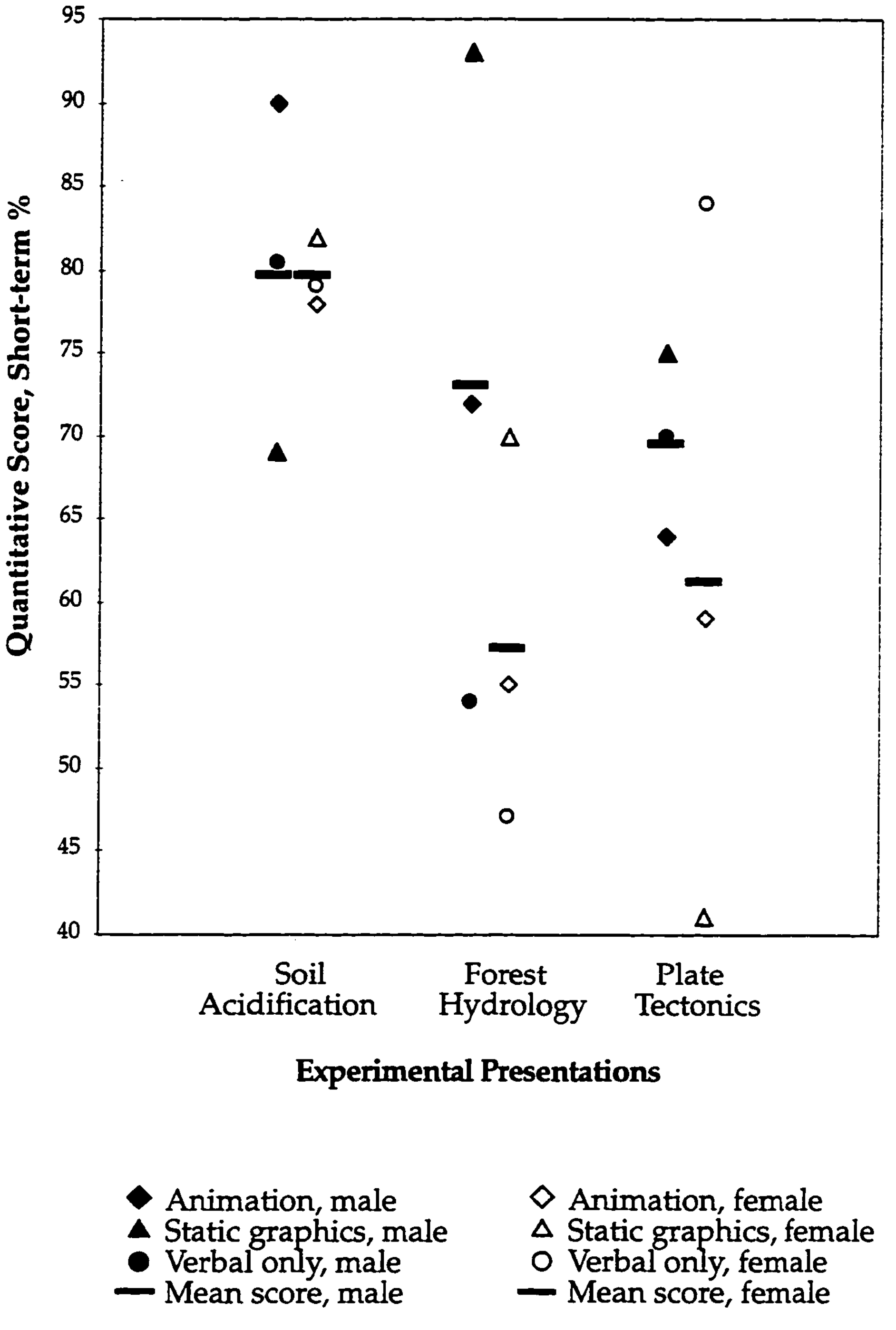



Figure 17. Mean short-term scores for men vs. women across presentations, subgroup A.

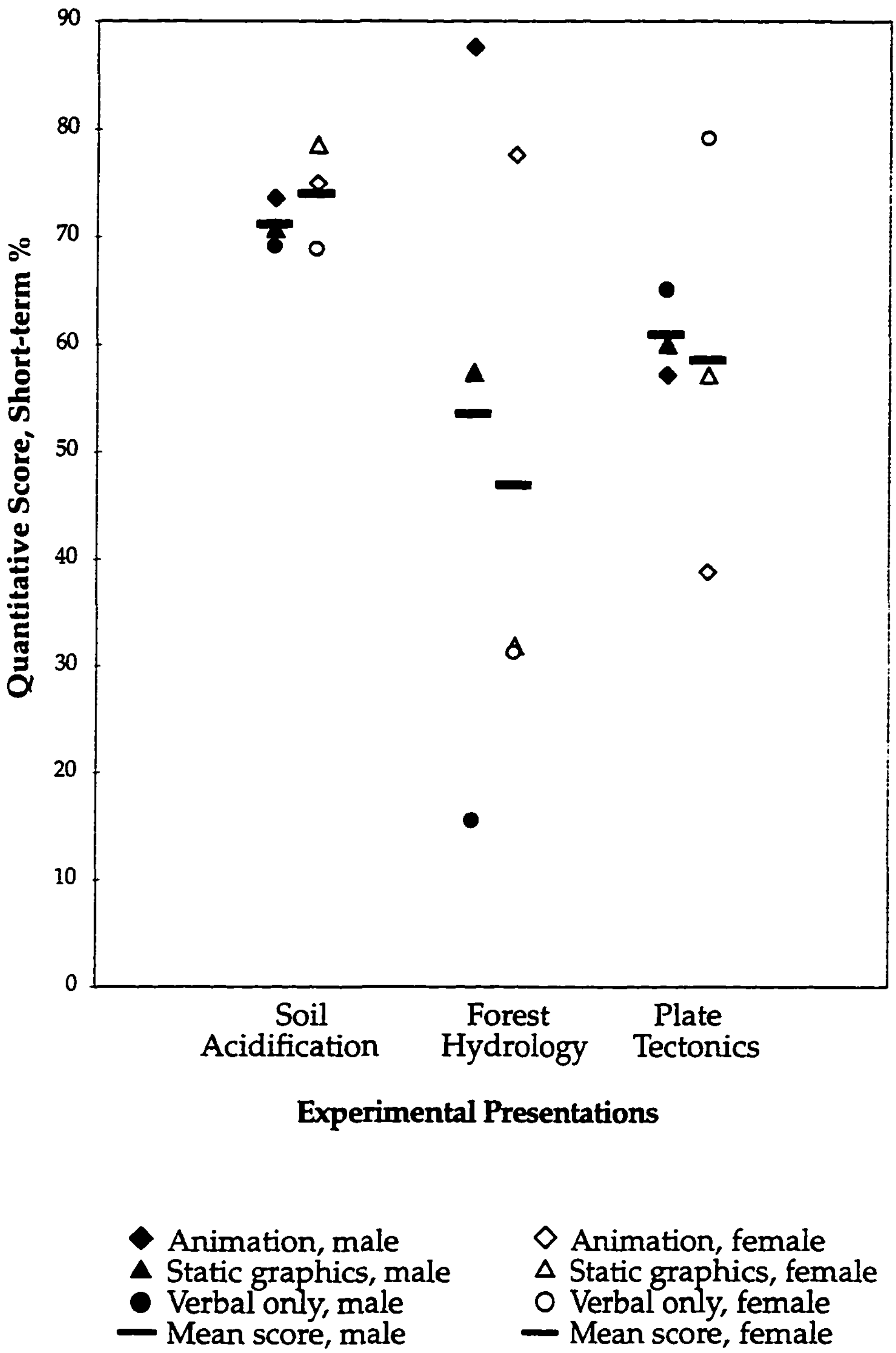



Figure 18. Mean short-term scores for men vs. women across presentations, subgroup B.

gender in order to account for the fact that overall women's scores were generally lower than men's). The differences between men's and women's scores are non-standard between presentations, but these differences are not significant (long-term $p = 0.2961$, short-term $p = 0.2893$), therefore hypothesis three is rejected. The lack of significance in looking at these differences may be due to the fact that some presentational variables were mixed within topics (e. g. both the soil acidification presentation and the plate tectonics presentation were pictorial representations, and varied in other factors, making it difficult to discern what variable was precisely responsible for a difference in scores). More detailed lists of how animated presentations differed may be found in Table 4, on page 63. In future research, experiments might be better designed to test these variables in isolation, possibly yielding more distinctive differences between presentational scores.

While gender variation between topics is the focus of this hypothesis, there is another issue that should be addressed. If genders are averaged together, there is a significant difference identified between presentational topics ($p = 0.0001$), indicating that some of the variables discussed may be more important to differences between individual learners than between genders. The variable within which the design of the presentations differed to the greatest degree was 'types of graphics used'. This category imposed an artificial classification scheme on presentational graphics from abstract to concrete. Learning types differ substantially in their preference and ability to work with concrete vs. abstract information (Kolb 1984). It was expected that since women tend to be concentrated in the learning types that prefer the concrete forms of representation rather than the abstract, this variable would produce the greatest degree of variation between genders. As discussed earlier, however, the percentage of students in class that fit into each of the learning type categories did not match Kolb's generalizations. Typically, women compose 60% of Type 1 and Type 4 (concrete experience) learners. The composition of my experimental populations flew far afield from this generalization; women filled the Type 2 and 3 learner categories to a substantial extent (classes averaged 63% female Type 2 and 3 learners) while there was a comparable dearth of female Type 1 and Type 4 learners (classes averaged 37%). This is possibly due to the fact that my students had each voluntarily registered for a science class. It makes sense that Types 2 and 3 (abstract conceptualization) would be more interested in registering for such a class, thus accounting for the abnormal population (Kolb's figures applied to the general population, my population was restricted to students who had voluntarily registered for a science class, a possible bias in the study). When experimental groups were divided into learning types, numbers were too small (often non-existent) in Types 1 and 4 categories to use in data analysis. This discussion may suggest a future research direction, however. Perhaps learning type is the variable most indicative of student preference when dealing with instructional media. Future research might direct itself towards achieving an understanding of gender variation within learning types. This understanding might prove extremely valuable in cognitively modelling women's learning.

Looking at each of the topics by either averages (Figure 8 and 9) or through raw data (Figure 12), it is clear that the plate tectonics presentation and the forest hydrology scores were consistently slightly higher than those of the soil acidification presentation. This is possibly due to a widespread lack of familiarity with soils issues. When subject eliminations are tabulated, there were roughly twice as many people eliminated on the basis of experience from the plate tectonics subject pool than the forest hydrology subject pool, and subjects were only rarely eliminated from the soil acidification pool. The differences between the three presentations are still not large for the descriptive data trends, and are not significant in the statistical analysis, thus, this hypothesis is rejected.

## Hypothesis Four

The fourth hypothesis states that use of animation in CBI enhances female students' confidence in their ability to do science, and does not affect men's confidence. This hypothesis was addressed through the use of the pre-, mid-, and post-term questionnaires (Appendix A). The pre-term questionnaires queried students about their level of confidence in their abilities to do well before the class began. The mid-term informal interview and the post-term questionnaires asked questions about changes in confidence levels, and in how each student felt about the various types of computer instruction used in class. There is a problem in terms of making a direct inference about exactly what it is that improves or decreases one's confidence. In order to approach this question from an observational standpoint, I first examine student reported confidence evaluations for men vs. women at the beginning of the classes on pre-term questionnaires. I then examine changes in men's vs. women's confidence in learning science at the end of the term as reported on post-term questionnaires, question 2. I then compare the averaged confidence improvements in classes having seen only one animation (all classes in subgroup A), to one having seen four animations (experimental group 5, three experimental, one extra). I then discuss the relationship between pre-term survey question 1 (how do I think I'm going to do in this class) with post-term survey question 1 (how do I think

I did in this class). I subsequently relate student comments on animated presentations and how each of the three treatments fits into women's vs men's preferred modes of learning as reported in student post-term surveys. Statistics are not appropriate for this section of the analysis. In order to make the above determinations, the classes in subgroup A (each having viewed one animation) must be compared with one class from subgroup B (having seen four animations), which introduces uncertainty into class differences. Also, the inferential nature of this hypothesis requires a qualitative assessment, not a statistical conclusion.

In looking at the initial confidence ratings submitted by students, men rated themselves at an average of 86.7% on the confidence scale. Males in experimental group 5 rated themselves at 86.8%, so the two are at about the same level. Women rated themselves at an average of 82.4% on the confidence scale; experimental group 5 females rated themselves at 80% exactly. Again, these are very close, experimental group 5 actually almost 3% lower than the overall average women's confidence rating. Changes in confidence, as reported by post-term surveys show that both men's and women's confidence raised slightly (Figure 19). Male confidence gains exceeded female confidence gains slightly (by 4.6%). This figure and the following percentage figures were determined by dividing the larger of the two numbers into the difference between the two numbers under discussion, yielding a percentage figure that describes how much larger the higher number is than the lower. This is meant to be a descriptive measurement rather than a statistical comparison. In a comparison of men's scores in subgroup A, and men's scores in experimental group 5, men in experimental group 5 gained 4.2% more on the confidence scale than did those in subgroup A. Women in experimental group 5, on the other hand, gained 11.9% on the

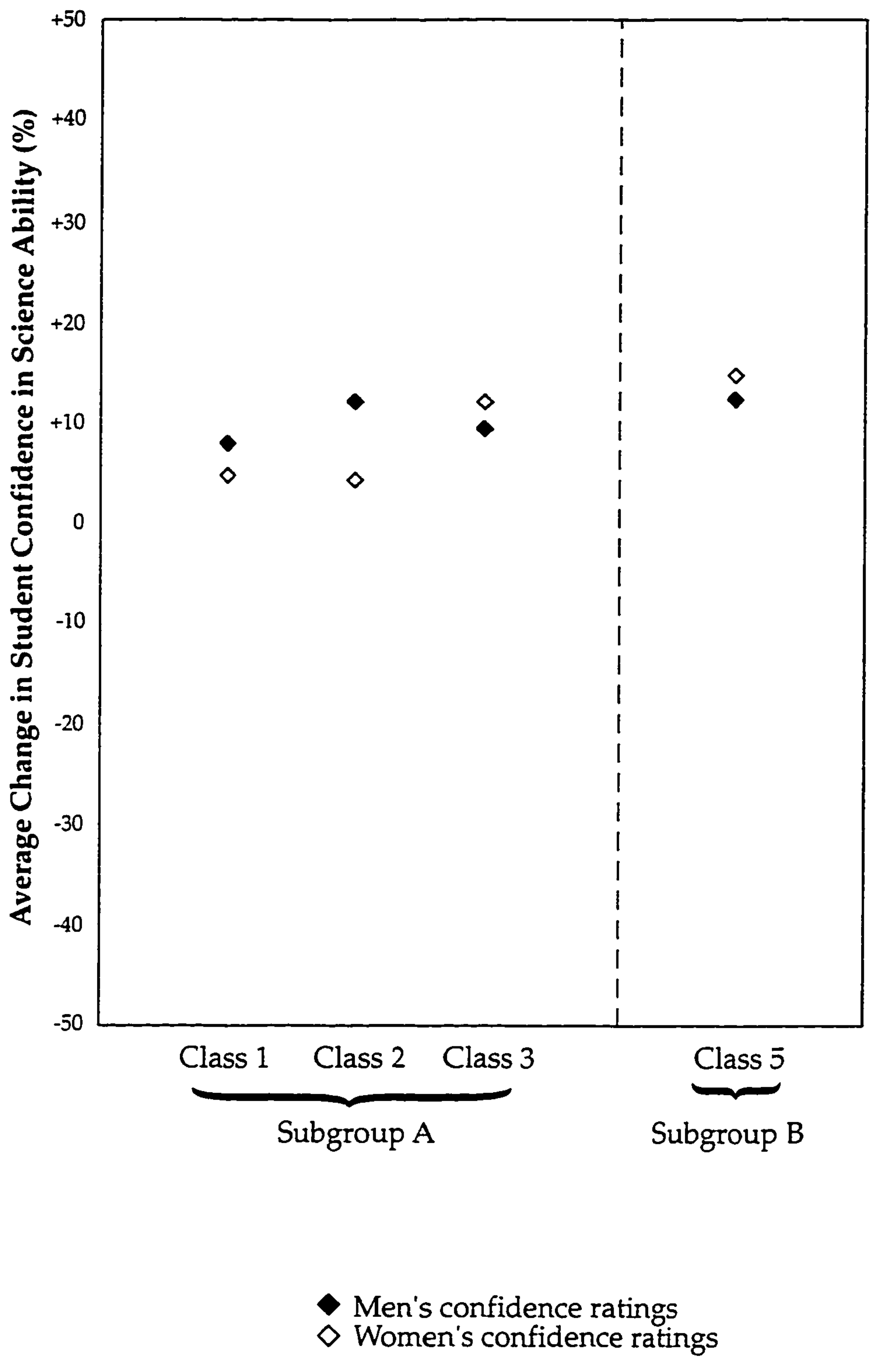



Figure 19. The difference in student confidence increases between the three classes in subgroup A (each having seen only one animation during the quarter), and class 5 (having seen four animations throughout the quarter).

confidence scale over those in subgroup A. This is quite a leap, lending qualitative support to the hypothesis that women's confidence ratings improved more than men's, when a greater amount of the class was taught with animated graphics. This is especially evident when we consider that the initial confidence levels of women in subgroup A were *lower* than those of women in experimental group 5.

The next section of data deals with the relationship between pre-term survey question 1 (how do I think I'm going to do in this class) and post-term survey question 1 (how do I think I did in this class). Figure 20 shows a scatterplot of these two values for the classes in subgroups A (all classes in subgroup A having seen only one animation per term). Figure 21 shows the same data for experimental group 5 (a class in subgroup B having seen an animation every other week during the term). If we drew a trend line on both graphs, the trend line on Figure 18 would be steeper than that of Figure 21, possibly indicating that experimental subjects in experimental group 5 felt that they had done 'better' than those students in subgroup A in retrospect. This may have been due to the fact that initial confidence scores were slightly lower in experimental group 5 than in subgroup A. Even so, these differences lend tentative support to the hypothesis. Again, these inferences should be used with caution. There is no replicate available to confirm this difference. Whether or not use of animation will prove, in the future, to statistically and conclusively benefit women's confidence in science education, present data lend support to this hypothesis.

In terms of learner preference for each treatment as an instructional device, all learners preferred animation to static graphics, and all learners preferred static graphics to verbal instruction (Figure 22). This trend was consistent for all classes. Differences between preferences for men and

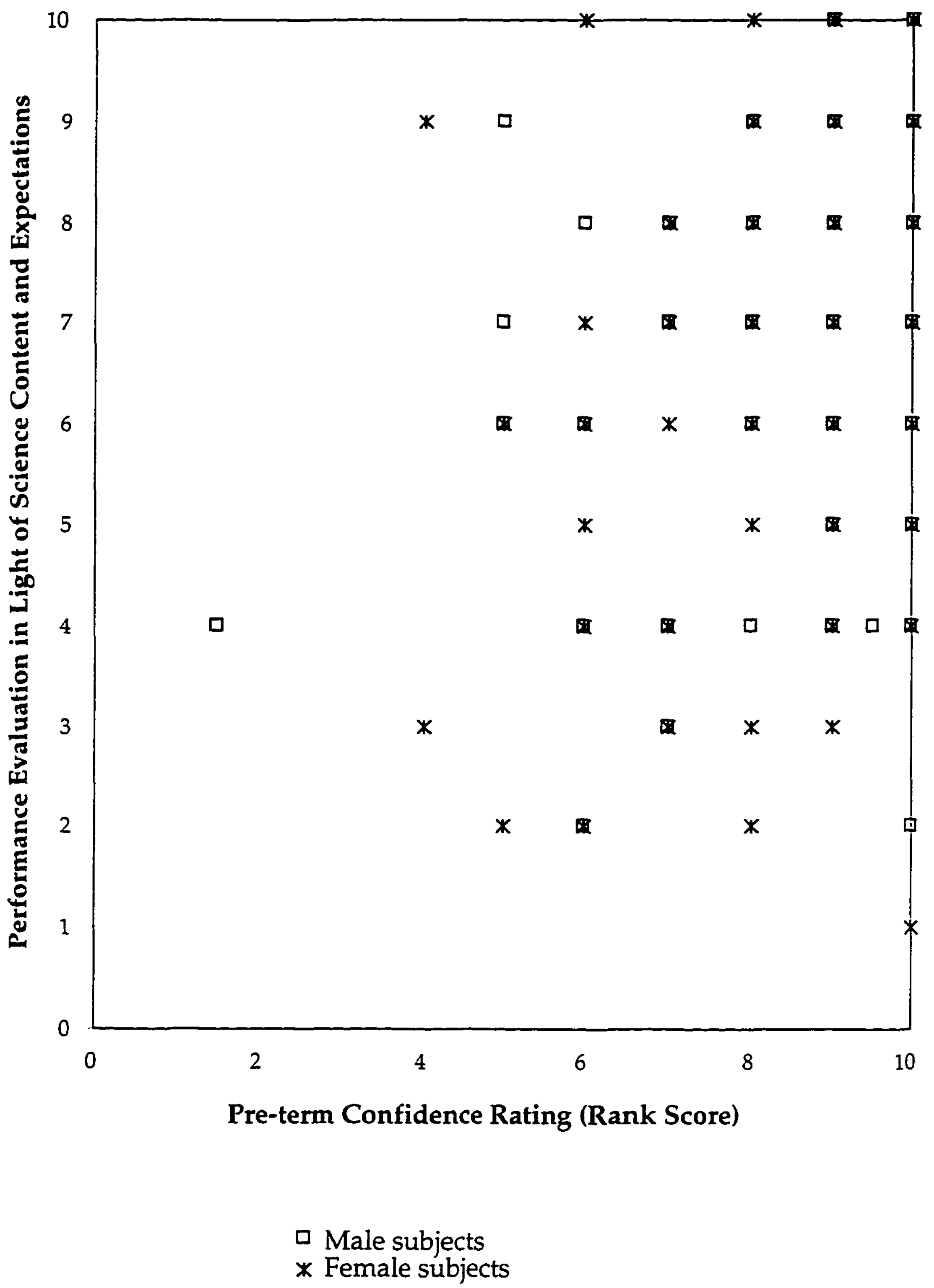



Figure 20. Correlations between pre- term confidence and post-term estimates of performance as measured by question one on the pre-term survey (X-axis) and question one on the post-term survey (Y axis). Note that since repeated measurements will be atop one another, symbols denote only presence or absence of scores for a given combination of pre-and post-term scores. Subgroup A only.

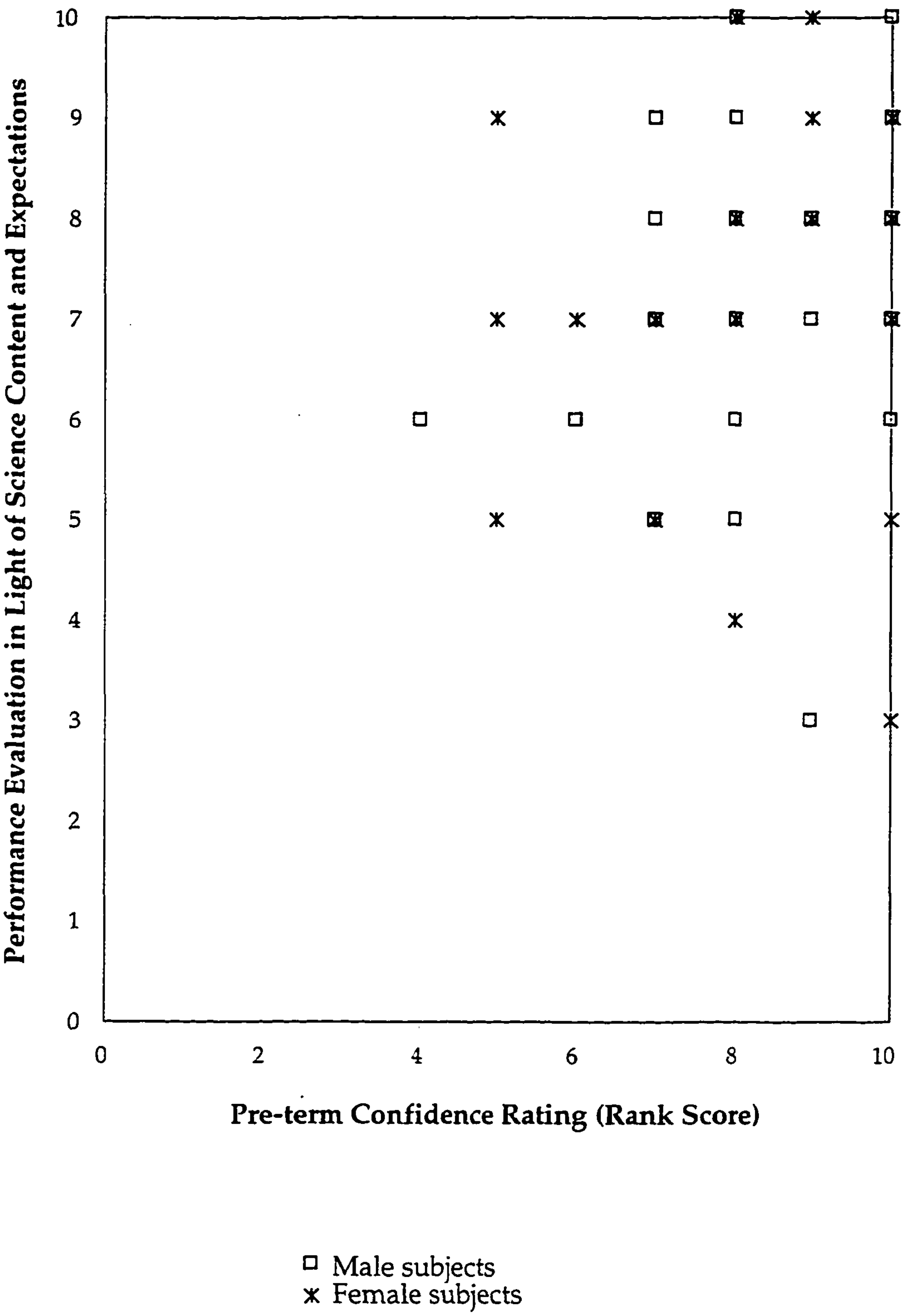



Figure 21. Correlations between pre- term confidence and post-term estimates of performance as measured by question one on the pre-term survey (X-axis) and question one on the post-term survey (Y axis). Note that since repeated measurements will be atop one another, symbols denote only presence or absence of scores for a given combination of pre-and post-term scores. Experimental group 5 only.

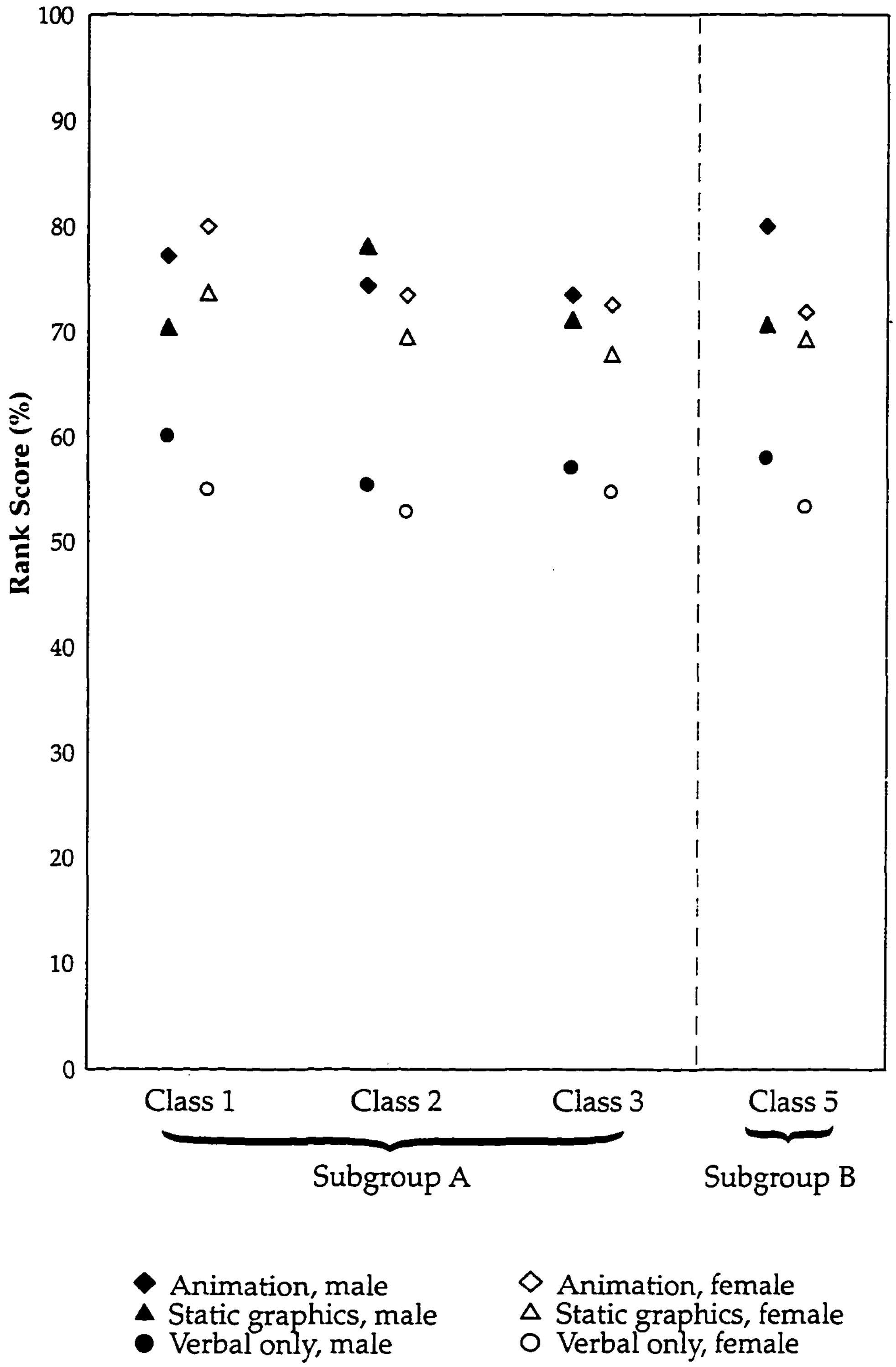



Figure 22. Men's and women's preferences for learning using each of the three treatment types, subgroup A and experimental group 5.

women increased slightly when the classes were separated into learning types. Even so, these differences were slight, and arbitrary, showing no definite trend or pattern in relation to learning type. This lack of difference in learning type categories, however, may be due to the comparatively small number of subjects in learning types 1 and 4 (in certain classes, 0). In summary, women always preferred animation to static graphics, and always preferred graphics of any kind to verbal instruction. Women's confidence did increase more than did men's in experimental group 5 (the group seeing four animations as opposed to one) over those classes having seen only one animation. These observations and figures lend strong qualitative support to hypothesis four.

## Hypothesis Five

Hypothesis five states that none of the results from these experiments will indicate that animation hinders women's learning. Statistically, there was no significant difference identified when looking at the interaction of gender and treatment. On this basis we could argue that this hypothesis should not be rejected. I also identified that women's average scores on animation were significantly higher that women's average scores on static graphics ($p = 0.067$, binomial tests for differences of proportion). Again, statistically, women do not suffer from the use of animation. But hypothesis five goes slightly farther; animated formats in any one of the physical science presentations shown must not hinder women's learning for this hypothesis to be supported. None of the presentations should show that animation test groups scored lower than static graphic or verbal instruction test groups. While statistically significant differences (or lack thereof) are nice, it would be more convincing to identify that animation does not hinder women's learning at all, whether the results are significant or not.

In examining the graphed data, it is clear that this is the case. Animation is superior for women's learning on all three presentations in long-term testing. This holds true for female students in both subgroup A and B (Figure 23 and Figure 24). This is not the case for short-term testing, however. In short-term testing there is no pattern to what treatments worked the best for any of the presentations. This is true for both subgroups. Lack of student variation in short-term learning, discussed in an earlier section, indicates that long-term testing is the more valuable measuring stick for student learning. So for long-term testing, this hypothesis is supported. In no presentation does animation hinder women's learning. This hypothesis is rejected for short-term learning, where no distinct pattern exists in treatment preference, either between presentations, or between replicates.

Before leaving this hypotheses it is important to note one more facet of this issue. If women learn better with animation in long-term testing, but dislike animated learning formats, then our point is lost. Looking back at hypothesis four, however, it is clear that women prefer to work with animated formats. In fact, out of all of the classroom data gathered for both men and women, the averages indicate that both men and women prefer to work with static graphics over verbal instruction by itself, and animated graphics over both static graphics and verbal instruction.

## Hypothesis Six

Hypothesis six is possibly the most important part of this research. It is at least the portion with the greatest implications for the academic

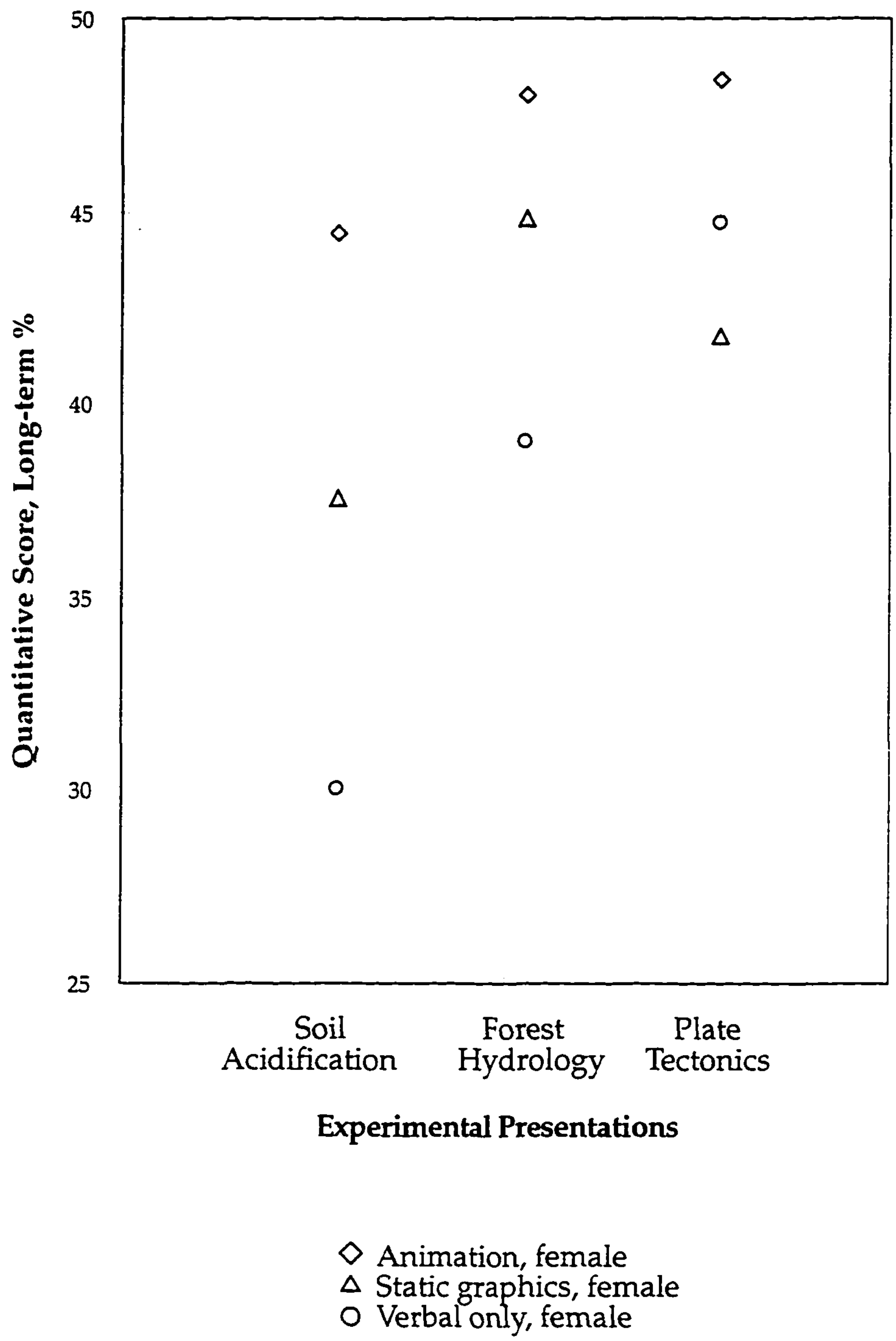



Figure 23. Women's long-term test scores across presentations, subgroup A.

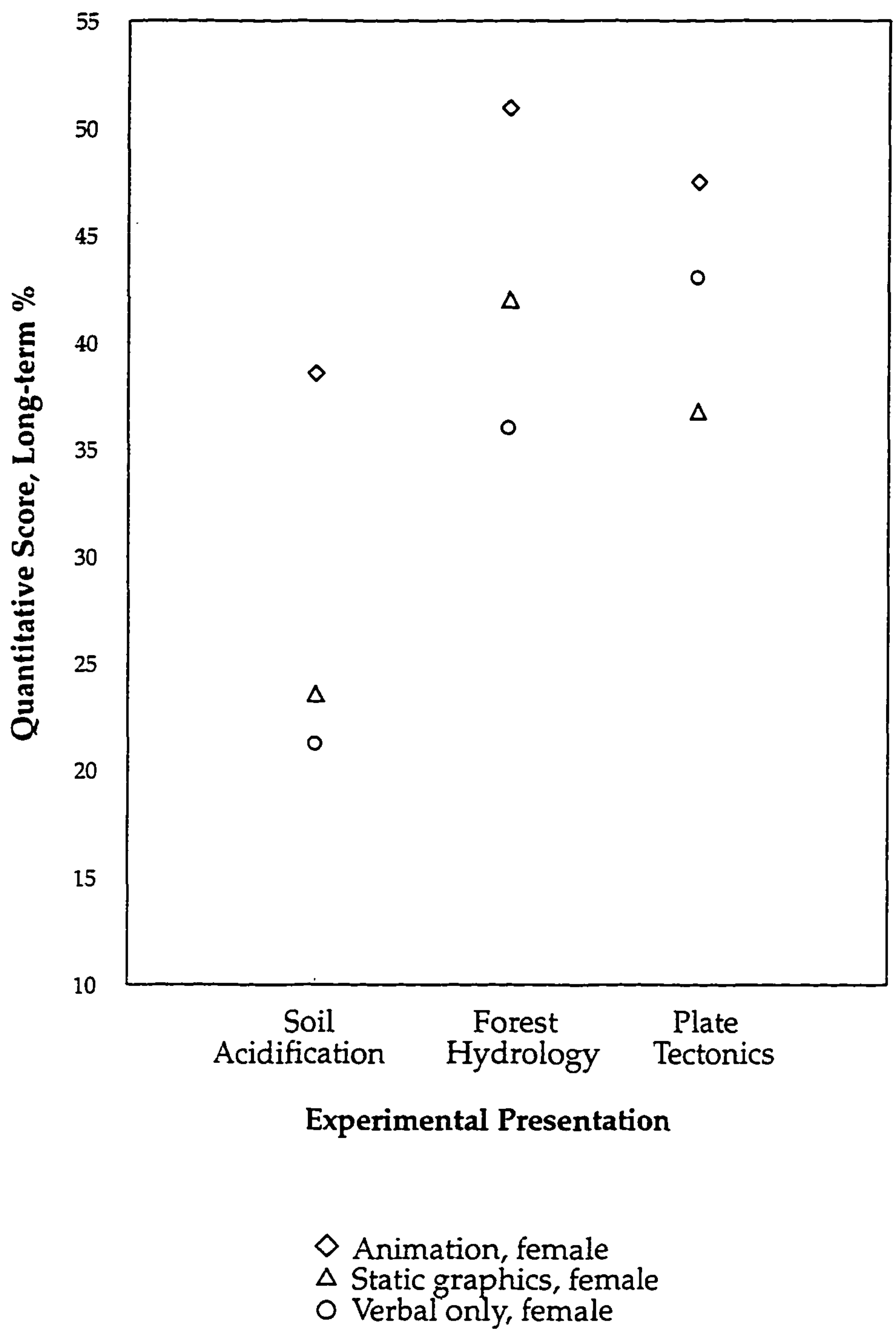



Figure 24. Women's long-term scores across presentations, subgroup B.

community at large, in dealing with teaching using animation in CBI. This hypothesis states that gender will be identified as an important source of variation in experimental populations looking at teaching through the use of animation. (Subgroup A alone is used in this section of the analysis, as subgroup B's experimental design and limitations render it inappropriate for the questions asked in this section {previously discussed}). There is significant variation due to gender in the data set when the data are analyzed at the student level: short-term data ($p = 0.0496$), long-term data ($p = 0.0157$), and total score ($p = 0.0001$). Gender is also a source of significant variation in the data set when data are analyzed at the class level. These findings give support to the idea that many of the animation experiments presently in the literature that have not controlled for gender in their experimental populations, or have large differences in the proportions of men or women in their populations, may have arrived at their conclusions in error. This hypothesis is not rejected.

**Informal Observations**

Throughout the quarter, I had ample opportunity to question students in class about their perceptions in relation to the computer technology used over the course of the experiment. I also had responses from informal interviews in the middle of each quarter, and remarks on post-term questionnaires to draw from. This section makes inferences about women's and men's confidence levels and computer technology preferences in a qualitative, observational format.

In opposition to the self-reported quantitative scores, women consistently expressed a lack of confidence in themselves, and in their work, throughout the quarter. Men did not. This may have been a socialized difference in willingness to report fears and negative emotions, rather than an actual difference in student confidence, however. Over the course of each quarter, female students were far more likely to approach me, unsolicited, and tell me how their term was going. Typically, women's statements about their confidence in their ability to do well in the course tended to increase as the quarter wore on. This increase was not substantial, however, nor was it consistent for all women. Far more pervasive, were comments heard from the women students putting down their own work, or ability to achieve. Even when the evidence, in terms of points received, strongly indicated that a female student would do well in the course, I often heard comments expressing fears that they would end the quarter with a poor grade. Men seemed to maintain the attitude that if their grades were low, I was an unfair grader, had made a mistake, or was likely to inflate the lower grade if argued with for a sufficient period of time. Women tended towards entering my office tentatively, asking what they had done 'wrong'. Neither of these differences was absolute, and again, socialization probably played big part.

In terms of the technology used in the classroom, the comments tended to be consistently favorable, from both men and women. The biggest concern expressed by both men and women, was that information was occasionally presented too quickly. This was also the most frequent comment about the plate tectonics animation. It went very quickly, and there was no way to stop the presentation in the middle. Clearly future educational animations should be designed to allow the educator to stop the presentation while playing, and information should be presented at a moderate to low speed to adhere to student preferences and comfort levels.

Comments relating to each of the three presentations, in relation to which was the 'best' were mixed, showing no distinct preference towards the treatment the presentation was shown in. This was surprising. If a student was interested in the topic content, it didn't matter whether the pictures were moving or not, the student's interest alone qualified it as the 'best'. I had my perceptions when I began this experiment, of which of the animated presentations was the most, and which the least interesting. Students did not agree with me for the most part. The presentation that I thought was the least interesting, forest hydrology, was often mentioned as the favorite, by both men and women. Related to this, I had the opportunity to question the class on the lectures that they got the most out of over the course of the quarter. Again, I was surprised by the responses. My favorite lectures were ignored completely, while many that I felt were relatively less fascinating were very popular. I mention this to indicate that individual preference is not something that is dependent on presentational format, and is not something easily understood or controlled for (or even anticipated). Future researchers who are testing more than one subject matter at a time, might want to add a section to their experiment asking for student subject preferences in a quantitative way. This might help explain some of the performance variation in the data.

One of the most frustrating and challenging things that I observed over the course of the experiment, was that many of the students decided that when an animation was being shown, they needed to put their pencils down and stop taking notes. When I said the words, "we're going to see an animated portrayal of such-and-such", the room echoed resoundingly with the sound of pencils hitting desks, and feet going up on nearby chairs. The pervasive attitude in relation to animation seemed to be that it was entertainment. This could be part of its apparent effectiveness, current literature indicates that the most effective teacher is one who is an effective entertainer.

> To be a teacher, you must give a performance of sorts, in order to communicate effectively with students...More than once we have overheard a teacher proclaim with some vehemence, "I'm paid to educate you, not entertain you." These instructors fail to realize that if they expect to educate their students, they must first attract and hold students' attention... (Tauber *et al.* 1993, p. 20)

In fact, the modern educational setting is one in which educators are being increasingly advised to think of students as the 'consumers' of the 'product' they offer. This is unfortunate in that most 'consumers' would rather not buy a poor grade or a difficult class. It would be far more productive to think of the state, or the school as the consumer of fine teaching, to be dispensed within their domain. As it is, however, education is becoming an entertainment business. Hopefully, new technology can be used to add a positive learning increase as well as serve an entertainment function. In terms of video instruction, and instruction shown using a television monitor, however, this does not appear the case. Students routinely lower the amount of invested mental effort they put into television and video lessons, because students perceive instruction through these media as 'easier' (Cennamo 1993, Salomon 1984). This perception is quite possibly what prompted my students to 'kick-back and relax' when given an animated lesson. This perception, and the coincident amount of invested mental effort involved, may be one reason why larger gains in learning are not consistently identified with student use of animation as an educational tool. Perhaps as animation and CBI become more integrated into classrooms, this perception will lose some of its influence and even larger gains will be identified for animated instruction.

## Pre- and Post-Term Questionnaire Results

In the pre-term questionnaires, students ranked their level of computer skill, indicated the time they spent reading and watching television, and related their experience in the field of science. These data were collected in order to determine whether any of these activities or skills had an observable relationship with scores on any one of the experimental treatments. Data was also collected on what type of programming students watched and what type of reading material they spent the most time with, however, the response on these inquiries left a little to be desired. When students couldn't make up their minds how to rank the types of television programming (for example), they ranked every type as a four, or a two, or another number, or even two types as threes, and two types as ones. Clearly this type of data reporting didn't leave much meaningful information to report, so these data have been ignored completely. Future researchers might want to remember that students tend to make up with creativity and imagination, answers to contingencies that researchers inadvertently leave out of instructions. In other words, give them that extra little bit of detailed direction or they will derail your data collection intentions.

Initially, I constructed a scatterplot showing science experience in relation to course grade. I expected that the more experience a student had in the sciences, the higher their grade in the class would be (on average). There is a relationship between science experience and course grade, in that students with higher levels of science experience tended towards having higher grades while students with lower grades tended to be scattered rather completely over a wide range of scores (Figure 25). This trend broke down when looking at the long-term scores on the experimental material as opposed to overall course grades in relation to science experience. The correlation of higher levels of science experience to higher scores was weak, but still present to a limited degree (Figure 26). I broke this second graph down into three, each portraying level of science experience in relation to long-term scores on one of the three experimental treatments. No definite trends are apparent (Appendix D). At the higher levels of science experience, number of women tended to be comparable to the number of men. At the lower levels of science experience, there were far more men than women (Figure 27). This may indicate that less experienced men were more likely to enroll in the class, or enroll in it earlier in their education. It may also reflect a difference in reporting between the genders. Determination of 'science experience' was done by tabulating all of the classes and periods of job experience in the field of science that an individual reported on their pre-term survey. Women may have been more meticulous and thorough in reporting all areas of their lives and educations that qualified as science experience, while men may have been more general or restrictive in their classification of 'science experience'. These are merely hypotheses however, and not provable with current information.

Level of computer skill reported by students also showed no distinct relationship to long-term scores on experimental material (Figure 28). One interesting part of this data, however, is that all students claimed at least some level of computer skill. Again, I broke this graph down into three, each portraying level of computer skill in relation to long-term scores on one of the three experimental treatments. No definite trends are apparent

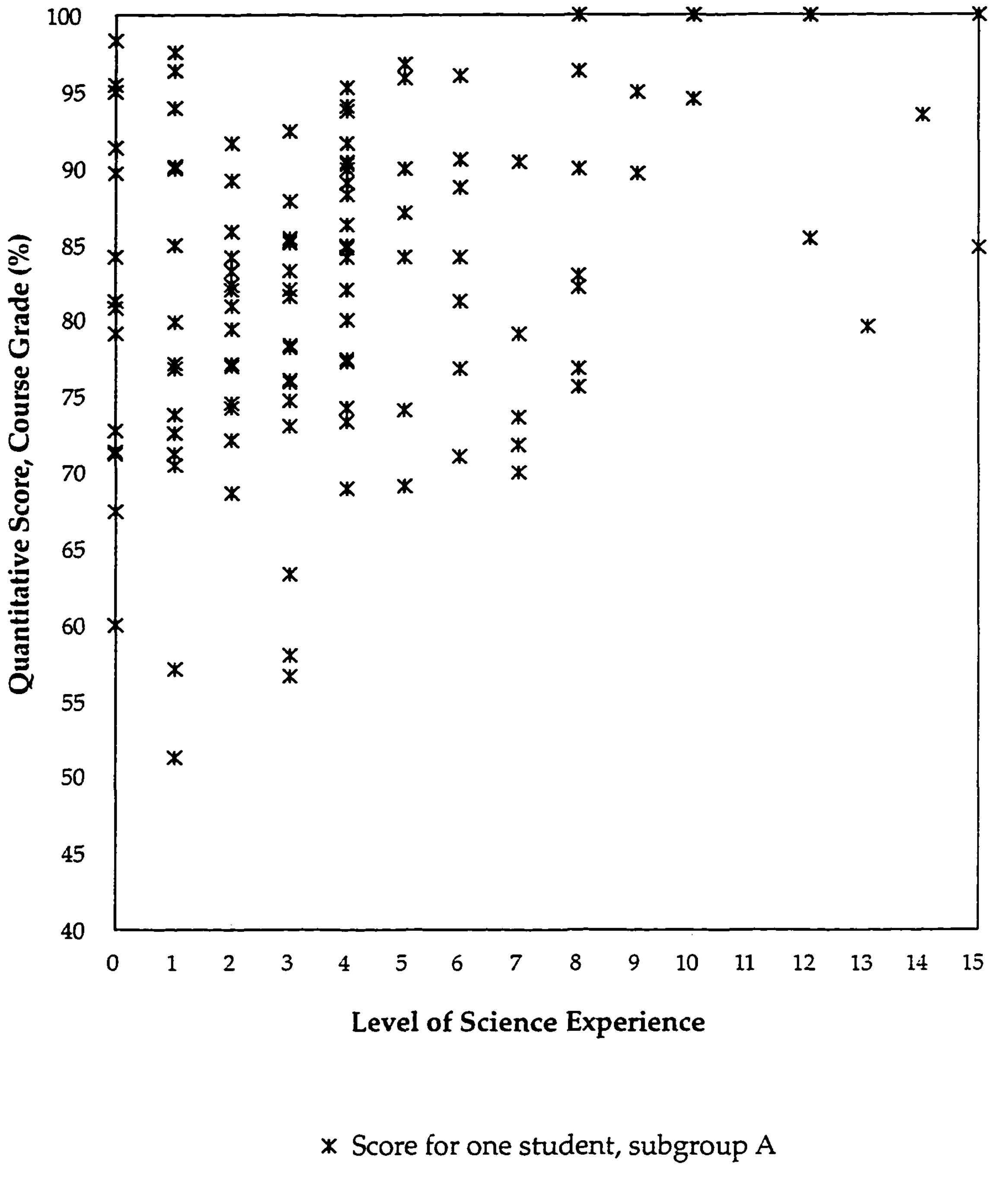



Figure 25. Relationship between self-reported level of science experience and overall course grade, subgroup A. Level of science experience determined by the number of science classes that a given student has taken, and/or job years of job experience in the field of science.

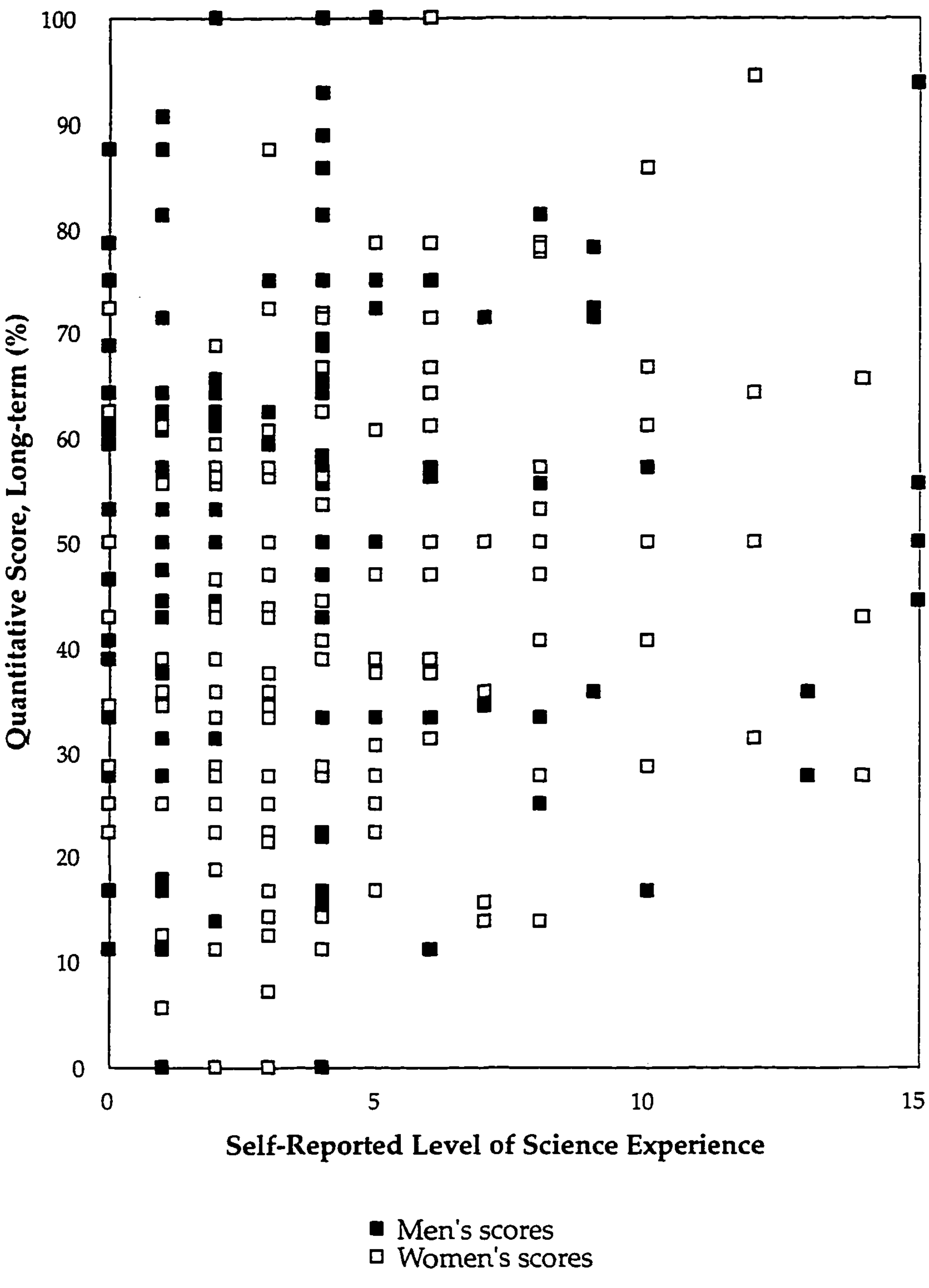



Figure 26. Relationship between self-reported level of science experience and scores on long-term testing for all three instructional treatments together, subgroup A. Level of science experience determined by the number of science classes that a given student has taken, and/or job years of job experience in the field of science.

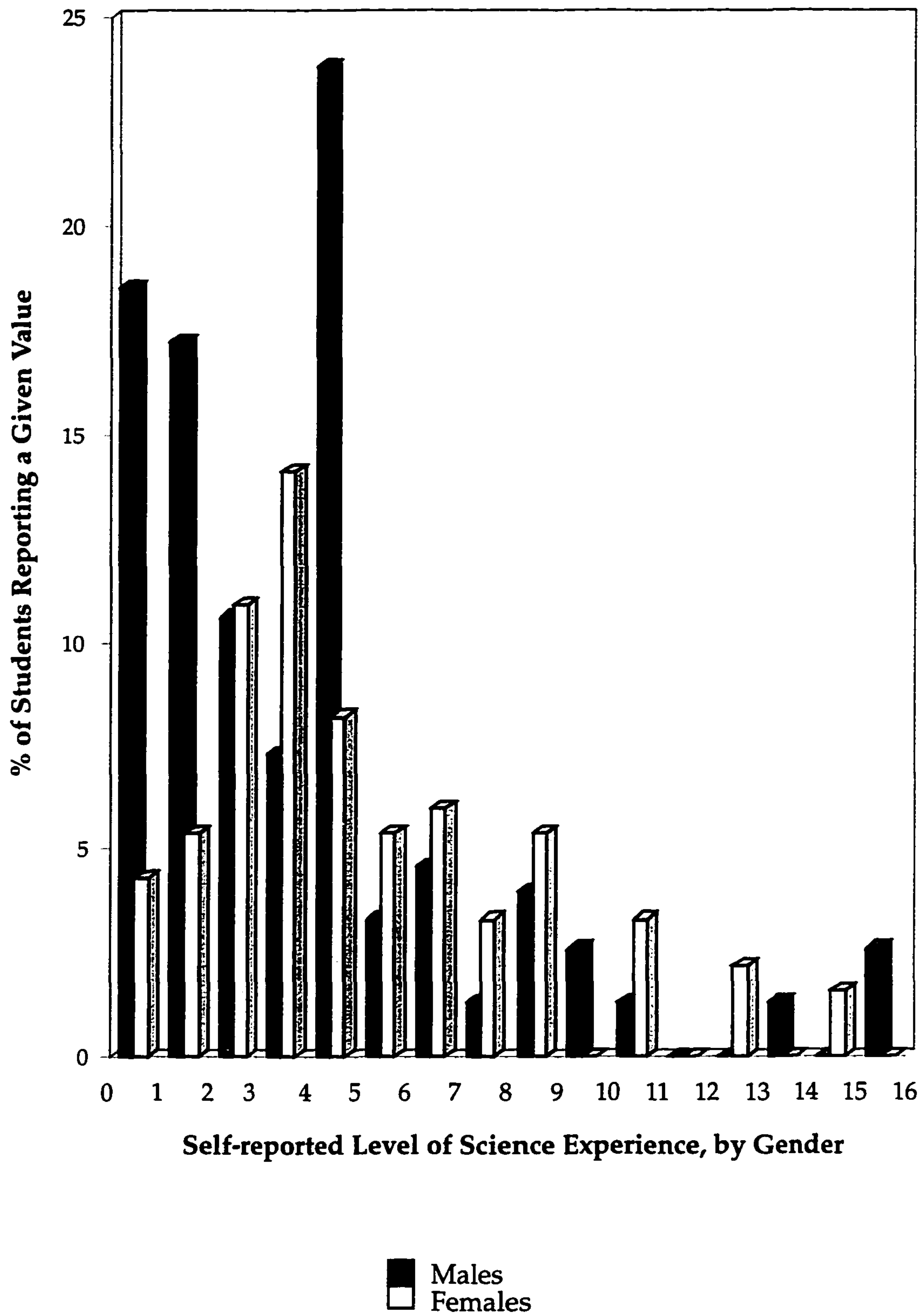



Figure 27. Self-reported level of science experience by gender, subgroup A. Level of science experience determined by the number of science classes that a given student has taken, and/or job years of job experience in the field of science.

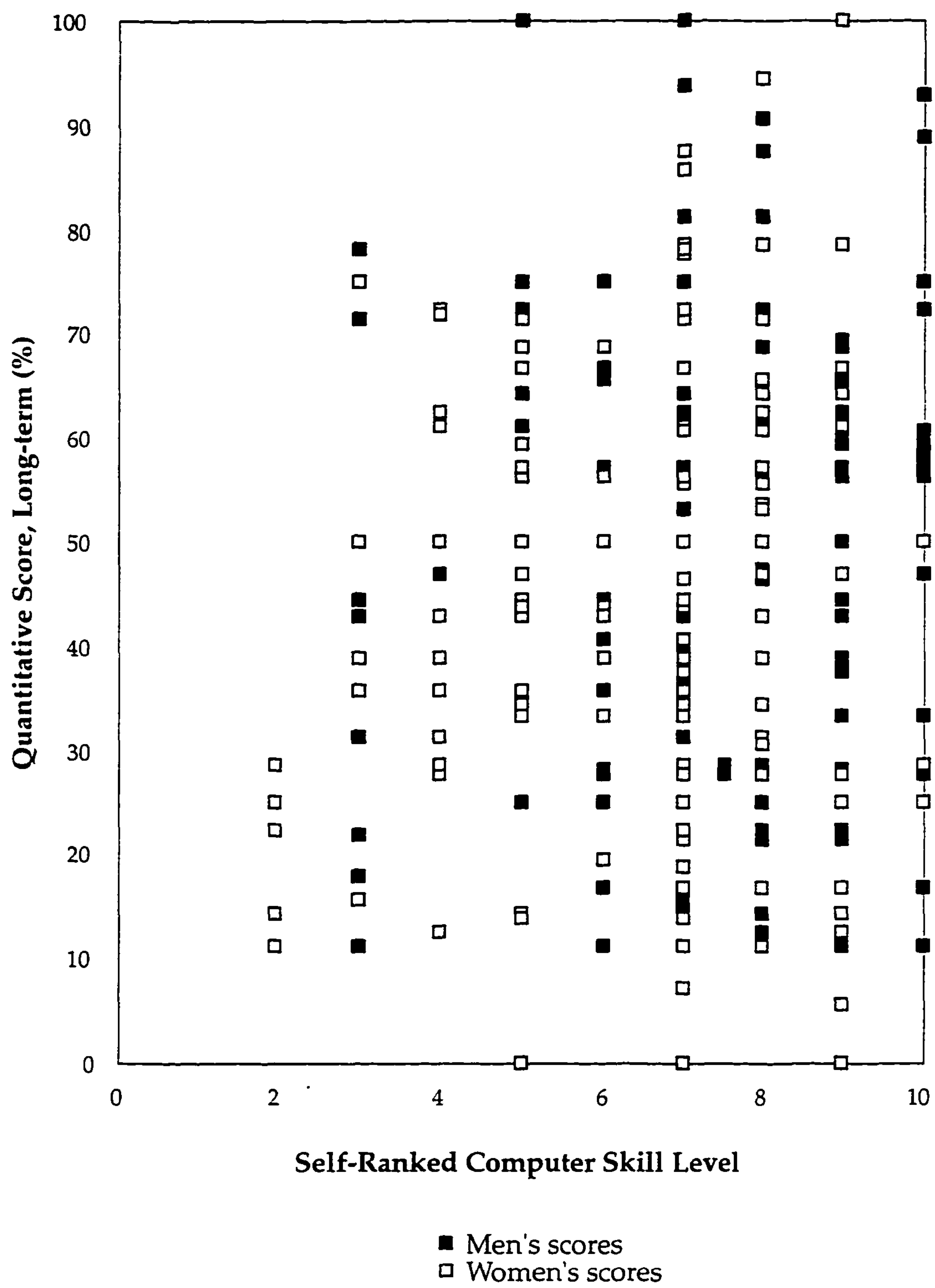



Figure 28. Relationship between self-reported computer skill level and scores on long-term testing for all three instructional treatments together, subgroup A.

(Appendix D). Figure 29 shows level of computer skill levels for men vs. women. Women tend towards lower levels of computer skill in greater numbers than men, and tend to report a moderate to high level of computer skill in numbers close to those of men; at the very highest levels, however, numbers of very experienced men outweigh women at about four to one.

Since animation is a television (or computer monitor) based instructional device, I looked specifically for a relationship between high scores on computer animation presentations, and high levels of television watching. This trend is not present, and it is far more clear that the more television one watches, the more poorly one does on long-term testing for any of the three of the experimental presentations (Appendix D). This relationship is also seen when we graph all three of the treatments together and portray them against hours per day that individuals watch television (Figure 30). This is not a distinct trend, however, indicating that other factors are in operation as well. It is interesting to note that women tend to watch less television per day than do men (Figure 31). One does hesitate to believe, though, that there are actually those out there that watch in excess of six hours of television a day. Possibly these individuals either misunderstood the question to mean 'per week', or decided to be creative.

The number of hours that each gender spent reading per day showed the opposite relationship; women tended to read more each day than did men (Figure 32). One might expect that the more reading a student does, the more likely that student is to do well in class. According to Figure 33, showing the relationship between hours per day spent reading and long-term scores on the experimental material in class, this is not the case. There is no distinct relationship between reading and score, either for all treatments together (Figure 33), or for any one type of treatment (Appendix D).

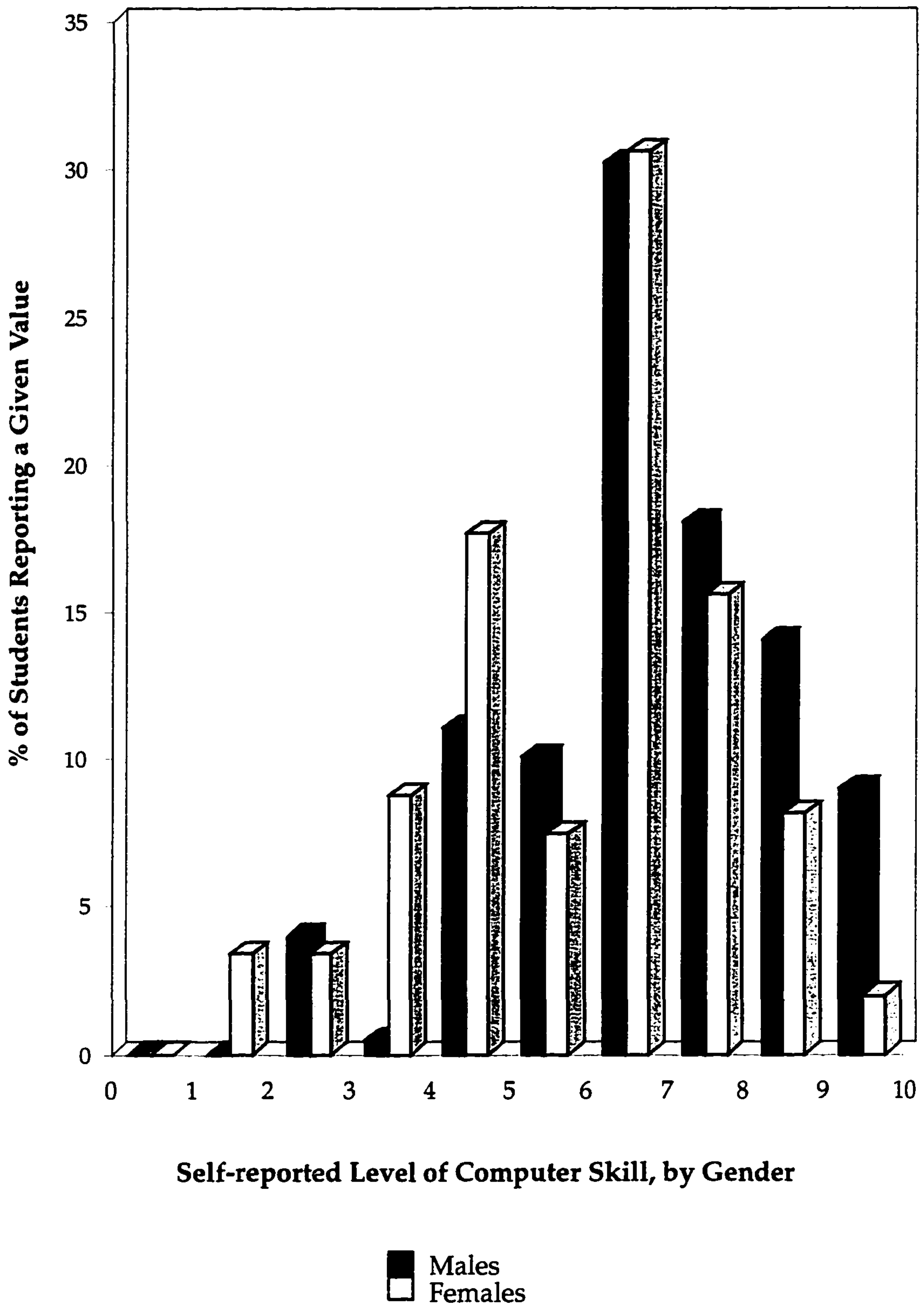



Figure 29. Self-reported level of computer skill by gender, subgroup A.

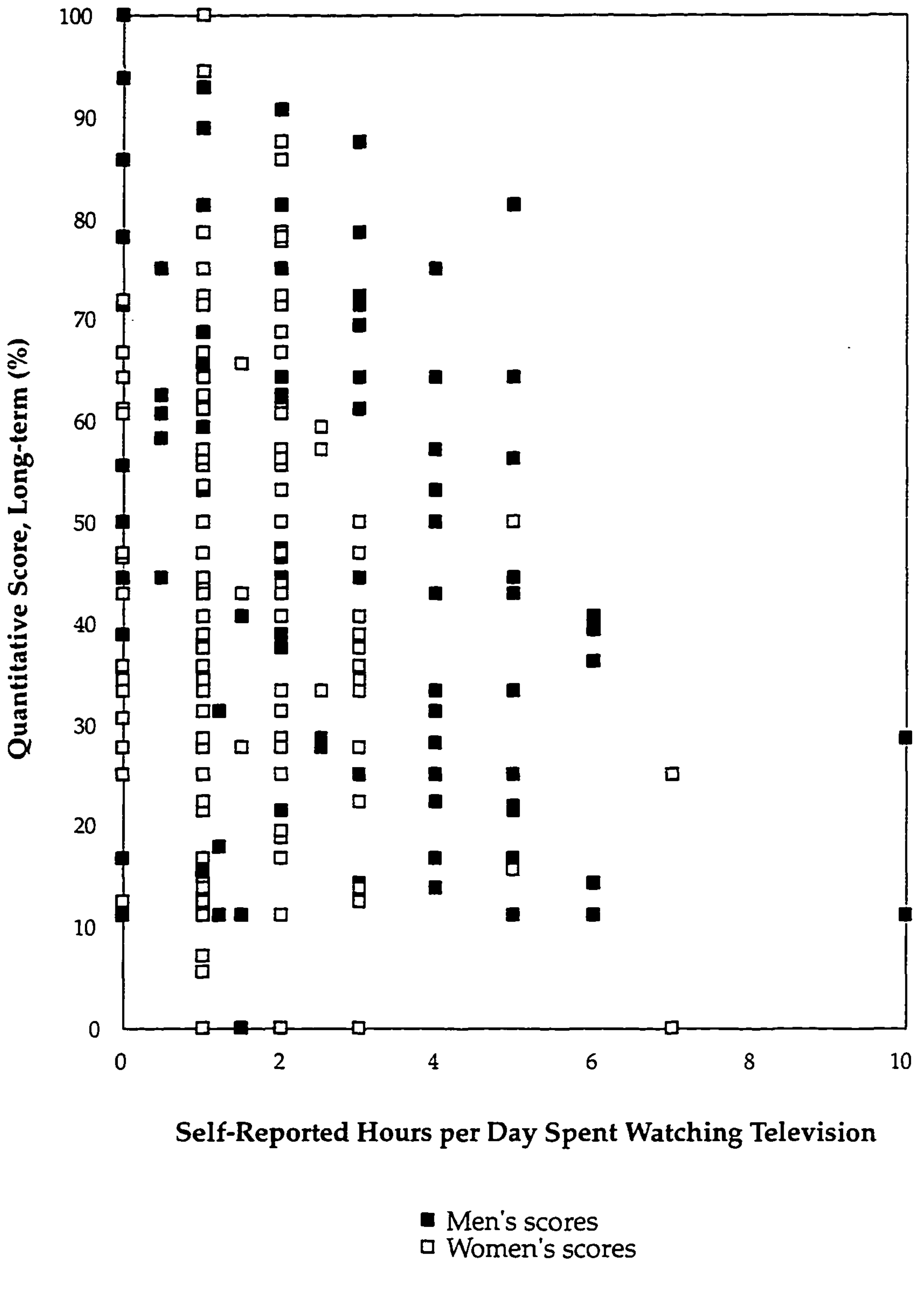



Figure 30. Relationship between self-reported hours per day spent watching television and scores on long-term testing for all three instructional treatments together, subgroup A.

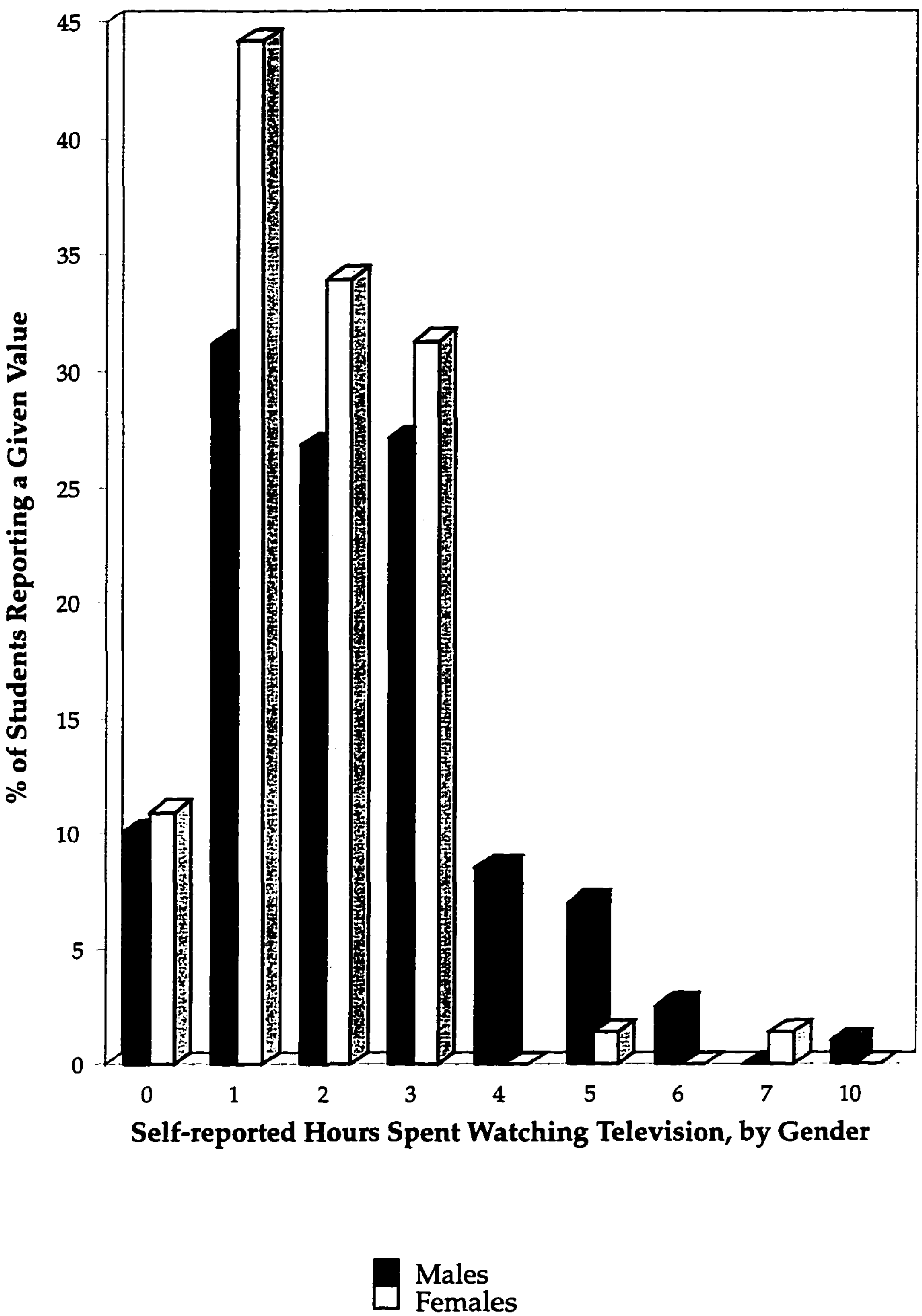



Figure 31. Self-reported hours of television watched per day, by gender; subgroup A only.

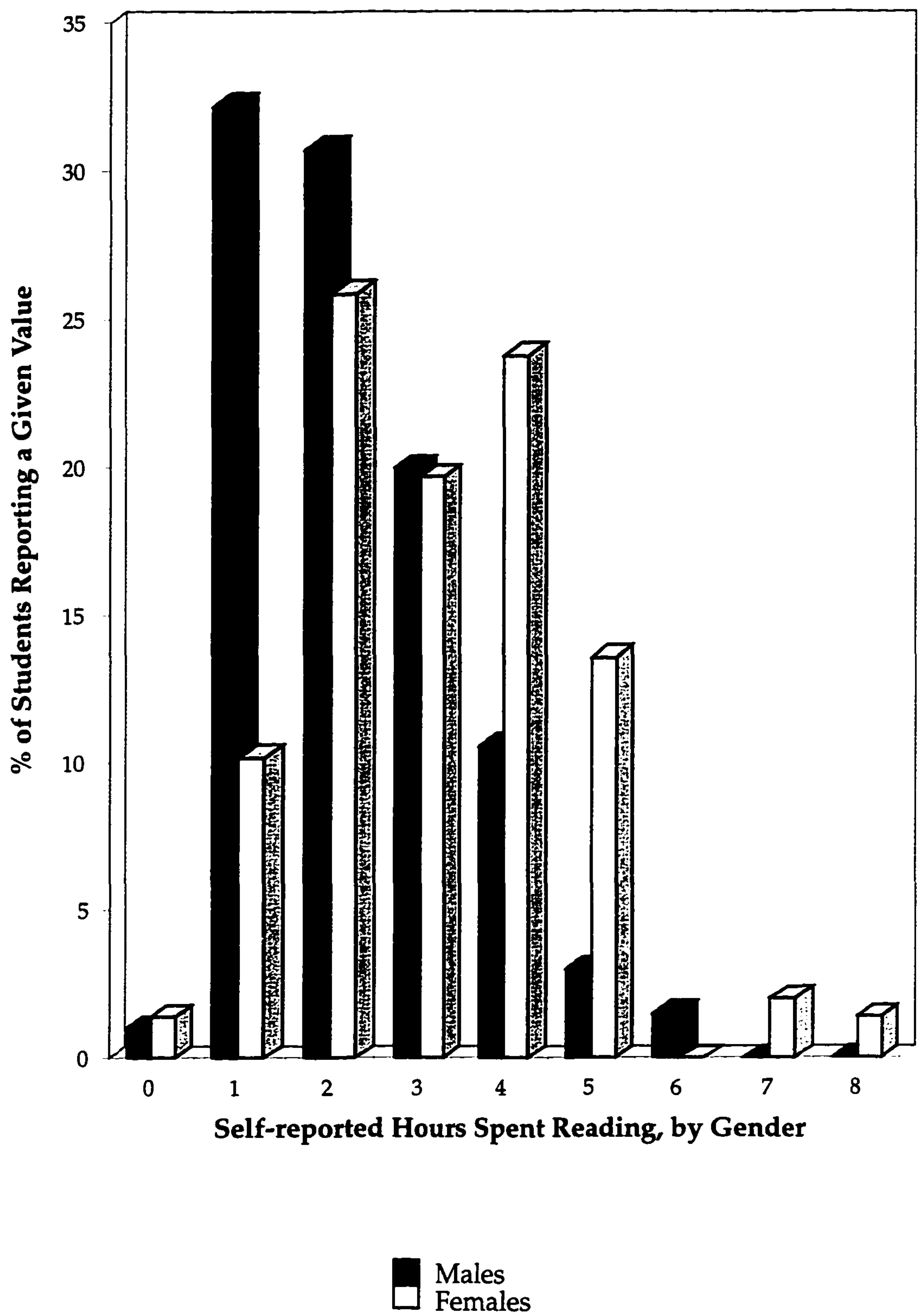



Figure 32. Self-reported hours spent reading per day, by gender, subgroup A.

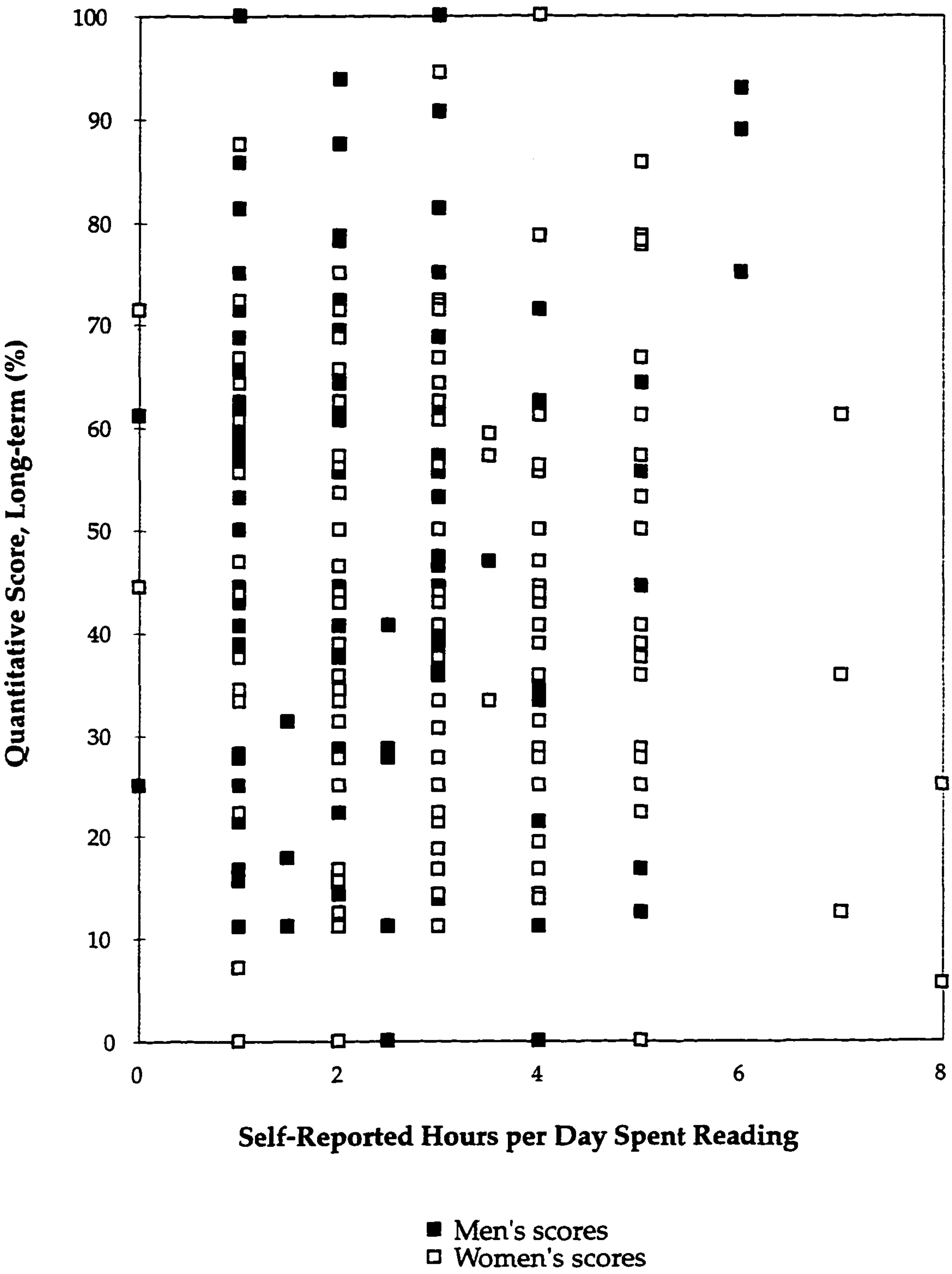



Figure 33. Relationship between self-reported hours per day spent reading and scores on long-term testing for all three instructional treatments together, subgroup A.

**Summary**

In summary, hypotheses one (women benefit more than men do from the use of computer animation in CBI, when learning physical science concepts), two (type of instructional medium (static vs. animated) has no difference in effect between short-term and long-term memory), five (none of the results from these experiments will indicate that animation hinders women's learning), and six (gender is an important source of variation in experimental populations looking at teaching through the use of animation) are not rejected. Hypothesis four (use of animation in CBI enhances female students' confidence in their ability to do science, and does not affect men's confidence) is supported through non-statistical inferential observation. Hypotheses three (differences between men's and women's learning will be non-standard among presentations, owing to differences in individual presentations) is rejected. Although not statistically significant, hypothesis three shows trends in the data that are very suggestive. In terms of this hypothesis, addressing the importance across genders of differences in topic and format of presentations, more directed research needs to be done in order to determine with greater precision whether experimental topic, within treatment varies in relation to gender. Hypotheses one and six, undoubtedly the most important parts of this research, reaffirms the need for gender equity within experimental populations when examining learning through the use of animation in CBI, and indicates a need for more directed research looking into gender and learning technology.

# CONCLUSIONS

Within the academic literature addressing animation as a viable educational tool, conflicting results exist. Through the experiments outlined in this dissertation, I have attempted to elucidate experimental conclusions present in the academic literature by isolating one possible source of variation within experimental populations, learning through the use of computer animation. My premise was that women would benefit from animation more than men. Statistically significant conclusions, as well as data trends supported animation as a valuable learning aid over static graphics for women. Results showed that animation significantly improved women's long-term learning over static graphics ($p = 0.067$), but didn't significantly improve men's long-term learning over static graphics. In all cases, women's scores improved with animation over both other forms of instruction for long-term testing, indicating that future research should not abandon the study of animation as a tool that may promote gender equity in science.

This research also detailed characteristics of each of the three presentations, to identify variables (e. g. level of abstraction in presentation) affecting score differences within treatments. Although I was certain that this hypothesis would not be rejected, statistically significant differences between scores on each of the three presentations were not identified, possibly due to the fact that the presentations in question did not differ enough in their format and design to allow subsequent variation in scores, and the hypothesis was rejected. In future research, experiments might be better designed to test these presentational variables in isolation, possibly yielding more distinctive differences between presentational scores. Confidence intervals overlapped to a great degree in all situations, however, undercutting, to some degree, the surety of conclusions indicating superiority of one treatment type over the others. However, confidence intervals for animation were smaller, showed an almost complete overlap for the genders, and centered around slightly higher means, lending further support to the conclusion that animation helps equalize men's and women's learning.

My research also sought to highlight differences in preferences for each of the three treatment types, and to identify any relationships between high scores on any one of the treatment types and several qualitative variables, including average time spent watching television, time spent reading, computer skill, etc... Averages indicated that both men and women preferred to work with animation over either static graphics or verbal instruction alone. While no direct correlation existed between any variable and an experimental treatment, it was clear that excessive television watching was not a positive in terms of attaining a high level of achievement. It was also clear that computer skill, and time spent reading, did not have a significant affect on students' scores for any of the three treatments.

My final hypothesis in this research asked whether gender was a variable that should be taken into account in experimental research designs. In this hypothesis, I sought to determine whether a researcher could perform an experiment using unequal gender ratios, and come to the same conclusion that he or she would have, had gender variation in the population been taken into consideration. In this case, my hypothesis was strongly supported, making it mandatory to take a second look at the studies available in the literature, and determine which of the studies have unequal gender ratios, or have not bothered to report gender ratios. With the information that unequal gender ratios can influence research conclusions, and may produce spurious results, studies identified in the literature that have very unequal gender ratios, or do not report gender ratios, are suspect. Eliminating these from the solid conclusions may allow us to determine whether support for animation within the literature can be identified as more conclusive, or consistent, either in support of, or against, the use of animation as a valuable learning aid in CBI.

This information does serve to elucidate conclusions in the academic literature concerning learning using computer animation. Unfortunately, it also nearly eradicates them. Out of all of the studies discussed in this research, three remain which fulfill these requirements. If we look only at these three studies, animation can be considered a valuable learning tool. In eliminating gender biased studies from the literature pool, however, we become experimentally depauperate. Now we can truly say more research needs to be done, however, a more important statement is probably, more gender equitable research needs to be done.

Advancing from the point of logical positivism, an examination of the model behind this research is in order. Currently a debate rages within the field of instructional design asking the question "Does media matter?". The battle rages fiercely on, centering the debate squarely on the question of what the question should be. On one side, researchers posit that media is only a vehicle for learning, and that it is not the vehicle that does the influencing, but it is the instructional method employed (Clark 1994). Another side argues that

> ...we must think about media not in terms of their surface features but in terms of their underlying structure and the causal mechanisms by which they might interact with cognitive and social processes. Media can be analyzed in terms of their cognitively relevant capabilities or attributes. (Kozma 1994, p. 11)

Clark (1994) argues that media may serve as a cognitive teacher. In other words, attributes of certain types of media, such as a "zoom" for example, might teach a learner unique new cognitive processes. It is Clark's voice, coincidentally, that is behind the "media is only a vehicle and will never influence learning" side of the debate. Another set of voices plead for researchers to look at instructional design from the learner's perspective, and not just through the elements of a given type of media (Jonassen *et al.* 1994).

In this research I have attempted to encompass each of the sides of this complex argument. In developing a model of women's learning, I was able to ask the question "what type of medium might benefit this model?". In designing the presentations I used, I looked at the attributes of the medium that I was using, and I linked them to learner processes, and to designer purposes. I attempted to look at the issues cognitively as opposed to behavioristically, which I believe encompasses the idea of embracing the mechanisms of learning as opposed to simply the outcomes. In this research, I embraced Reiser's (1994) theory that media may be a vehicle, but it allows us to execute the methods Clark touts, that can be used to affect learner processes. To this end, I proposed a model of women's learning. I hypothesized that some of the attributes of women's learning would be supported by the attributes identified in computer animation. I do not want to put forth a conclusion that touts animation as the wondrous learning tool for all women in science. I have used the experiments in this research to attempt to determine the degree of the cognitive match between my model of spatial information and motion in relation to women's learning processes.

A word of caution in using the results in this research; you can't model the human mind. Learning theory has come a long way in the past few decades, and behaviorism and logical positivism are finally giving way to more realistic, holistic approaches to learning and teaching. There is more to teaching than finding a technological medium that will induce a Pavlovian response in a learner. The human mind, no matter the age of the student, is not a black box, or a blank slate, just waiting for us to write on it. All students come to a class with their own knowledge, their own belief systems (some of them in direct contrast to prevailing reality), and their own ways of learning and knowing. When we, as researchers, grasp firmly to reductionistic thinking in our experimentation, we lose the richness of the human psyche. The tighter we close our fingers, the more depth is lost. When we reduce our students to populations of individuals that fit on to different lists to be analyzed, we lose so much.

> The process of learning is holistic. It cannot be understood by analyzing responses to attributes of the media that carry the messages or the methods that putatively engage learners. It is difficult, if not impossible, to isolate the effects of the affordances of media...and methods. Yet our design models (and debates) continue to focus on such issues. Our models are grounded on two essential components of reality - objectivity and causality - both integral components of Western consciousness (Jonassen 1983). Objective reality is predicated on a number of assumptions, such as commonality of perception. Objectivity enables us to observe and describe the physical world, which we accept as reality. We believe that our universe is an orderly place, but can we really depend upon objectivity for description and causality for prediction of events? (Jonassen *et al.* 1994, p. 34)

Kozma (1994) states that we must begin to 'forge' a relationship between media and learning. Basically, this involves divorcing ourselves from the behavioristic models of stimulus-response, and embracing the idea that learning is not passive, but active. There is a meeting place between the media vehicle transmitting the learning, and the active reception of that learning. At that intersection, social conditioning, cognitive skills, and perception all come into play, influencing how each individual learner will 'create knowledge' for themselves through the use of media. It is our jobs, as instructional designers and scientists, to firmly establish the links between media and cognitive processes, and then to develop our media presentations in direct relationship with those frameworks. Only by doing this will we actively succeed in 'forging the links' between media and learning.

Taking Kozma's (1994) logic one or two steps forward, I would argue that what the research conclusions in this field need most are time. Through time, we can search for the best way to embed cognitive instruction into learning. As that instruction is taken in, certain types of media will become more effective as they become more familiar. We will eventually have a generation of learners that is deeply entrenched in the technologist paradigm. When we construct experiments using these learners in our experimental populations, perhaps we will have already constructed our link between media and learning, in them.

**BIBLIOGRAPHY**

American Association of University Women, 1992. How Schools Shortchange Girls: A Study of Major Findings of Girls and Education. Washington, DC: American Association of University Women.

Alesandrini, Kathryn Lutz, 1984. "Pictorial-verbal and analytic-holistic learning strategies in science learning," Journal of Educational Psychology. vol. 73, no. 3: 63-77.

Anderson, John R., 1978. "Pictures and adult learning," Instructional Science. vol. 13: 249-277.

Baek, Young K., and Benjamin H. Layne, 1988. "Color, graphics, and animation in a computer-assisted learning tutorial lesson," Journal of Computer-Based Instruction. vol. 15, no. 4: 131-135.

Barnsley, R. H., and M. S. Rabinovich, 1970. "Handedness: proficiency versus stated preference," Perceptual and Motor Skills. vol. 30: 343-362.

Blades Mark, and Christopher Spencer, 1986. "The implications of psychological theory and methodology for cognitive cartography," Cartographica. vol. 27, no. 2: 24-46.

Blake, Tyler, 1977. "Motion in instructional media: some subject-display mode interactions," Perceptual and Motor Skills. vol. 44: 975-985.

Borg, Walter R., and Charles F. Schuller, 1979. "Detail and background in audiovisual lessons and their effect on learners," Educational Communication and Technology Journal. vol. 27, no. 1:31-38.

Campbell, Craig S., and Stephen L. Egbert, 1990. "Animated cartography: thirty years of scratching the surface," Cartographica. vol. 27, no. 2: 24-46.

Caplan, Paula J., Gael M. MacPherson, and Patricia Tobin, 1985. "Do sex-related differences in spatial abilities exist?" American Psychologist. vol. 40, no. 7: 786-799.

Caraballo, Angel Luis, 1985a. An Experimental Study to Investigate the Effects of Computer Animation on the Understanding and Retention of Selected Levels of Learning Outcomes. Unpublished doctoral dissertation, The Pennsylvania State University.

Caraballo, Jose Noel, 1985b. The Effect of Various Visual Display Modes in Computer-Based Instruction and Language Background upon Achievement of Selected Educational Objectives. Unpublished doctoral dissertation, The Pennsylvania State University.

Catsambis, Sophia, 1995. "Gender, race, ethnicity, and science education in the middle grades," Journal of Research in Science Education. vol. 32. no. 3: 243-257.

Cennamo, Katherine S., 1993. "Learning from video: factors influencing learners' preconceptions and invested mental effort," Educational Technology, Research and Development. vol. 41, no. 3: 33-45.

Clark, Richard E., 1994. "Media will never influence learning," Educational Technology, Research and Development. vol. 42, no. 2: 21-29.

Collins, Allan, Marilyn Jager Adams, and Richard W. Pew, 1978. "Effectiveness of an interactive map display in tutoring geography," Journal of Educational Psychology. vol. 70, no. 1: 1-7.

Das, J. P., John R. Kirby, and Ronald F. Jarman, 1979. Simultaneous and Successive Cognitive Processing. Academic Press: New York.

DiBiase, David, 1990. "Visualization in the earth sciences," Earth and Mineral Sciences. vol. 59, no. 2: 13-18.

DiBiase, David, Alan M. MacEachren, John B. Krygier, and Catherine Reeves, 1992. "Animation and the role of map design in scientific visualization," Cartography and Geographic Information Systems. vol. 19, no. 4: 201-214, 265-266.

Dorling, Daniel, 1992. "Stretching space and splicing time: from cartographic animation to interactive visualization," Cartography and Geographic Information Systems. vol. 19, no. 4: 215-227, 267-270.

Dwyer, Francis M., 1969. "The instructional effect of motion in varied illustrations," Journal of Psychology. vol. 73: 167-172.

Dwyer, Francis M., 1971. "Color as an instructional variable," A V Communication Review. vol. 19: 399-416.

Eastman, Ronald J., 1985. "Graphic organization and memory structures for map learning," Cartographica. vol. 22, no. 1: 1-20.

Edgeman, Laura L., 1994. Teaching Successional Vegetation Dynamics: An Evaluation of the Effectiveness of Animated and Static Graphics. Unpublished Master's Thesis: San Diego State University.

Friedhoff, Richard Mark, and William Benzon, 1989. The Second Computer Revolution: Visualization. Harry N. Abrams, Inc.: New York.

Gagné, Robert M., 1985. The Conditions of Learning. (4th edition). Holt, Rinehart, and Winston: New York, NY.

Gilhooly, Kenneth J., Michael Wood, Paul R. Kinnear, and Caroline Green, 1988. "Skill in map reading and memory for maps," Quarterly Journal of Experimental Psychology. vol. 40, no. 1: 87-107.

Gooding, K., and D. Forrest, 1990. "An examination of the difference between the interpretation of screen based and printed maps," The Cartographic Journal. vol. 27: 15-19.

Griffin, T. L. C., 1983. "Problem solving on maps- the importance of user strategies," The Cartographic Journal. vol. 20, no. 2: 101-109.

Gropper, George L., 1963. "Why is a picture worth a thousand words?" AV Communication Review. vol. 11, no. 4: 75-95.

Halpern, Diane F., 1986. Sex Differences in Cognitive Abilities. Lawrence Erlbaum Associates: Hillsdale, New Jersey.

Hannafin, Michael J., and K. Peck, 1988. The Design, Development, and Evaluation of Instructional Software. MacMillan: New York.

Horowitz, Mardi Jon, 1970. Image Formation and Cognition. Appleton-Century-Crofts: New York.

Hyde, Janet Shirley, 1981. "How large are cognitive gender differences? A meta-analysis using $w^2$ and d," American Psychologist. vol. 36, no. 8: 892-901.

Intons-Peterson, Margaret J., and Mark A. McDaniel, 1991. "Symmetries and asymmetries between imagery and perception," Chapter 2 in Imagery and Cognition. Cesare Cornoldi and Mark A. McDaniel eds., Springer-Verlag: New York.

Jacek, Laura Lee, and Cynthia Brewer, in review. "Learning vegetation succession dynamics: a comparison of animated and static graphics for visualization," Journal of Geography in Higher Education.

Jonassen, David H., John P. Campbell, and Mark E. Davidson, 1994. "Learning with media: restructuring the debate," Educational Technology, Research and Development. vol. 42, no. 2: 31-39.

King, W. A., 1975. "A comparison of three combinations of text and graphics for concept learning," Report No. NPRDC-TR-76-16. Navy Personnel Research and Development Center: San Diego, CA.

Kolb, David A., 1976. Learning Style Inventory Technical Manual. McBer and Co.: Boston, MA.

Kolb, David A., 1984. Experiential Learning: Experience as the Source of Learning and Development, Prentice-Hall: Englewood Cliffs, NJ.

Koussoulakou, A., and M. J. Kraak, 1992. "Spatio-temporal maps and cartographic communication," The Cartographic Journal. vol. 29: 101-108.

Kozma, Robert B., 1994. "Will media influence learning? Reframing the debate," Educational Technology, Research and Development. vol. 42, no. 2: 7-19.

Kulhavy, Raymond W., Doris R. Pridemore, and William A. Stock, 1992. "Cartographic experience and thinking aloud about thematic maps," Cartographica. vol. 29, no. 1: 1-9.

Larsen, Ronald E., 1992. "Relationship of learning style to the effectiveness and acceptance of interactive video instruction," Journal of Computer-Based Instruction. vol. 19, no. 1: 17-21.

Linn, Marcia C., and Anne C. Petersen, 1985. "Emergence and characterization of sex differences in spatial ability: a meta-analysis," Child Development. vol. 56: 1479-1498.

Lunneborg, Clifford E., and Patricia W. Lunneborg, 1984. "Contribution of sex-differentiated experiences to spatial and mechanical reasoning abilities," Perceptual and Motor Skills. vol. 59: 107-113.

MacCoby, Eleanor Emmons, and Carol Nagy Jacklin, 1974. The Psychology of Sex Differences. Stanford University Press: Stanford, California.

MacEachren, Alan M., Barbara P. Buttenfield, James B. Campbell, David DiBiase, and Mark Monmonier, 1992. "Visualization," in Geography's Inner Worlds. pp. 99-137. Rutgers University Press: New Jersey.

MacEachren, Alan M., and David DiBiase, 1991. "Animated maps of aggregate data: conceptual and practical problems," Cartography and Geographic Information Systems. vol. 8, no. 4: 221-229.

MacEachren, Alan M., and John H. Ganter, 1990. "A pattern identification approach to cartographic visualization," Cartographica. vol. 27, no. 2: 64-81.

Margolis, Howard, 1987. Patterns, Thinking, and Cognition: A Theory of Judgement. The University of Chicago Press: London.

Marzano, Robert J., Ronald S. Brandt, Carolyn Sue Hughes, Bean Fly Jones, Barbara Z. Resseisen, Stuart C. Rankin, and Charles Suhor, 1988. Dimensions of Thinking: A Framework for Curriculum and Instruction. Association for Supervision and Curriculum Development: Alexandria, Virginia.

Mayer, Richard E., and Richard B. Anderson, 1991. "Animations need narrations: an experimental test of a dual-coding hypothesis," Journal of Educational Psychology. vol. 83, no. 4: 484-490.

Mayer, Richard E., and Richard B. Anderson, 1992. "The instructive animation: helping students build connections between words and pictures in multimedia learning," Journal of Educational Psychology. vol. 84, no. 4: 444-452.

Mayer, Richard E., and Valerie K. Simms, 1994. "For whom is a picture worth a thousand words? Extensions of a dual-coding theory of multimedia learning," Journal of Educational Psychology. vol. 86, no. 3: 389-401.

McGee, Mark G., 1979. "Human spatial abilities: psychometric studies and environmental, genetic, hormonal, and neurological influences," Psychological Bulletin. vol. 86, no. 5: 889-919.

Moore, Martha V., Leon H. Nawrocki, and Zita M. Simutis, 1979. "The instructional effectiveness of three levels of graphics displays for computer-assisted instruction," Army Project Number 2Q763731A771, ERIC DOC ED 178057. US Army Research Institute for the Behavioral and Social Sciences: Alexandria, VA.

Monmonier, Mark, 1990. "Strategies for the visualization of geographic time-series data," Cartographica. vol. 27, no, 1: 30-45.

Monmonier, Mark, 1992. "Summary graphics for integrated visualization in dynamic cartography," Cartography and Geographic Information Systems. vol. 19, no. 1: 23-36.

Newcombe, Nora, 1982. "Sex related differences in spatial ability: problems and gaps in current approaches," Chapter 10 in Spatial Abilities: Development and Physiological Foundations. Michael Potegal ed., Academic Press: New York, NY.

Paivio, Allan, 1971. Imagery and Verbal Processes. Holt, Rinehart and Winston, Inc.: New York, NY.

Paivio, Allan, and Kalman Csapo, 1973. "Picture superiority in free recall: imagery or dual-coding?" Cognitive Psychology. vol. 5: 176-206.

Park, Ok-choon, and Stuart S. Gittelman, 1992. "Selective use of animation and feedback in computer-based instruction," Educational Technology, Research and Development. vol. 40, no. 4: 27-38.

Philbin, Marge, Elizabeth Meier, Sherri Huffman, and Patricia Boverie, 1995. "A survey of gender and learning styles," Sex Roles. vol. 32: 485-494.

Ramsey, Fred J., and Dan Schafer, 1995. The Statistical Sleuth. OSU Book Stores Inc.: Corvallis, OR.

Reed, Stephen K., 1992. Cognition. Brooks/ Cole: Pacific Grove, CA.

Reiser, Robert A., 1994. "Clark's invitation to the dance: an instructional designer's response," Educational Technology, Research and Development. vol. 42, no. 2: 45-48.

Relan, Anju, and William C. Smith, 1996. "Learning from hypermedia: A study of situated versus endemic learning strategies," Journal of Educational Multimedia and Hypermedia. vol. 5, no. 1: 3-21.

Rieber, Lloyd P., 1989a. "A review of animation research in computer-based instruction," Proceedings of the 1989 Annual Convention of the Association for Educational Communications and Technology. vol. 11: 369-390.

Rieber, Lloyd P., 1989b. "The effects of computer animated lesson presentation and cognitive practice activities on young children's learning in physical science," Proceedings of the 1989 Annual Convention of the Association for Educational Communications and Technology. vol. 11: 390-406.

Rieber, Lloyd P., 1990. "Animation in computer-based instruction," Educational Technology, Research and Development. vol. 38, no. 1: 77-86.

Rieber, Lloyd P., 1991. "Effects of visual grouping strategies of computer-animated presentations on selective attention in science," Educational Technology, Research and Development. vol. 39, no. 4: 5-15.

Rieber, Lloyd P., Mary J. Boyce, and Chahriar Assad, 1989. "The effects of computer animated lesson presentation and cognitive practice on adult learning in physical science," Proceedings of the 1989 Annual Convention of the Association for Educational Communications and Technology. vol. 11: 409-420.

Rieber, Lloyd P., Mary J. Boyce, and Chahriar Assad, 1990. "The effects of computer animation on adult learning and retrieval tasks," Journal of Computer-Based Instruction. vol. 17., no. 2: 46-52.

Rieber, Lloyd P., and M. Hannafin, 1988. "Effects of textual and animated orienting activities and practice on learning from computer-based instruction," Computers in the Schools. vol. 5: 77-89.

Roychoudhury, Anita, Debora J. Tippins, and Sharon E. Nichols, 1995. "Gender inclusive science teaching: a feminist constructivist approach," Journal of Research in Science Teaching. vol. 32, no. 9: 897-924.

Salomon, Gavriel, 1984. "Television is 'easy' and print is 'tough': the differential investment of mental effort in leaning as a function of perceptions and attributions," Journal of Educational Psychology. vol. 76, no. 4: 647-658.

Sharps, Matthew J., Jana L. Price, and John K. Williams, 1994. "Spatial cognition and gender: instructional and stimulus influences in mental image rotation performance," Psychology of Women Quarterly. vol. 18: 413-425.

Shepard, R. N., and Metzler, J., 1971. "Mental rotation of three dimensional objects," Science. vol. 171: 701-703.

Slocum, Terry A., Susan H. Robeson, and Stephen L. Egbert, 1990. "Traditional versus sequenced choropleth maps/ an experimental investigation," Cartographica. vol. 27, no. 1: 67-88.

Spankenberg, Ronald W., 1973. "The motion variable in procedural learning," AV Communication Review. vol. 21: 419-436.

Sweller, John, Paul Chandler, P. Tierney, and M. Cooper, 1990. "Cognitive load as a factor in the structuring of technical material," Journal of Experimental Psychology General. vol. 119, no. 2: 176-192.

Tapley, S. Marion, and M. P. Bryden, 1977. "An investigation of sex differences in spatial ability: mental rotation of three-dimensional objects," Canadian Journal of Psychology. vol. 31, no. 3: 122-130.

Tauber, Robert T., Cathy Sargent Mester, and Stephen C. Buckwald, 1993. "The teacher as actor: entertaining to educate," NASSP Bulletin. vol. 77, no. 551: 20-28.

Thorndyke, Perry W., and Cathleen Stasz, 1980. "Individual differences in procedure for knowledge acquisition from maps," Cognitive Psychology. vol. 12:137-175.

Tobler, Waldo R., 1987. "Experiments in migration mapping by computer," The American Cartographer. vol. 14, no. 2: 155-163.

Tufte, Edward R., 1983. The Visual Display of Quantitative Information. Graphics Press: Cheshire, CT.

Weinburgh, Molly, 1995. "Gender differences in student attitudes toward science: a meta-analysis of the literature from 1970 to 1991," Journal of Research in Science Teaching. vol. 32, no. 4: 387-398.

Wood, Gordon, 1983. Cognitive Psychology: A Skills Approach. Brooks/ Cole: Monterey, CA.

Zavotka, Susan Lee, 1987. "Three-dimensional computer animated graphics: a tool for spatial skill instruction," Educational Communication and Technology Journal, vol. 35, no. 3: 133-144.

# APPENDICES

# APPENDIX A

## Pre-term Questionnaire

Male Female (Please circle one)

How confident are you in your ability to do well in this class in light of the science content?
(not at all) 1 2 3 4 5 6 7 8 9 10 (very)

How experienced do you consider yourself in dealing with computers?
(not at all) 1 2 3 4 5 6 7 8 9 10 (very)

How experienced do you consider yourself in working on the Web?
(not at all) 1 2 3 4 5 6 7 8 9 10 (very)

On average, how many hours of television do you watch per day?
0 1 2 3 4 5 6 7 8 9 10+

Please rank the types of programming that you watch. Score the most frequently watched type of broadcast 5, and the least frequently watched type 1. If you never watch a particular type of programming, score it 0.

_____ MTV
_____ Public Broadcasting
_____ News programming
_____ Nature shows
_____ Sports

On average, how many hours per day do you spend reading?
0 1 2 3 4 5 6 7 8 9 10+

Please rank the types of reading that you do. Score the most frequently read type of material 5, and the least frequently read type 1. If you never read a particular type of material, score it 0.

_____ Academic (Textbooks, journals, etc...)
_____ Fiction (Adventure, mystery, romance, etc...)
_____ Non-fiction (biographies, historical accounts, etc...)
_____ Magazines (American Rifleman, Vogue, Sunset, etc...)
_____ Technical writing (computer manuals, hiking guides, how to books etc...)

Please list the experience you've had in science and environment (classes, jobs):

**APPENDIX A, cont.**

**Post-term Questionnaire**

Male Female (Please circle one)

In retrospect, how do you feel that you performed in this class in light of both the science content, and your own expectations?
(poorly) 1 2 3 4 5 6 7 8 9 10 (well)

In approaching classes dealing with science now, would you consider yourself more confident, less confident, or about the same as you were before this class?
(less) -50 -40 -30 -20 -10 0 +10 +20 +30 +40 +50 (more)

Looking back at the class, how well do you feel that the overall instructional style fit with your personal learning style?
(poorly) 1 2 3 4 5 6 7 8 9 10 (well)

How well do you feel that instruction using animation fit with your personal learning style?
(less) 1 2 3 4 5 6 7 8 9 10 (more)

How well do you feel that instruction using standard (still) graphics fit with your personal learning style?
(less) 1 2 3 4 5 6 7 8 9 10 (more)

How well do you feel that instruction using only lecture fit with your personal learning style?
(less) 1 2 3 4 5 6 7 8 9 10 (more)

In the space below please include any additional comments that you have on the instruction you received in this class. I am especially interested to know what you thought about the use of computer technology in the classroom.

## APPENDIX A, cont.

### Informal Interview Outline

1. What are your feelings towards science? Are you confident in your ability to do well in this class? Have these feelings changed over the course of the quarter? How?

2. How do you feel that the differences in technology used in this class influence your learning? Or do they?

3. How have your feelings and perceptions changed about this class as the quarter has worn on?

4. How have your feelings and perceptions changed about the technology used in this class as the quarter has worn on?

5. If you think back on three lessons taught in this class, forest hydrology, soil acidification, and plate tectonics, which lesson (each of which had a different presentational style) fit your way of learning the best? Please rank.

6. Additional comments?

## APPENDIX B

### Purposes for Use of Animation

Attention: Directing subject attention to a specific piece of information, or to a certain focus on the given information.

Focus/ emphasis: Examples are arrows, blinking objects, scaling into a view of an object, etc...

Grouping/ chunking: Organizational manipulation designed to relate some information to other relevant information. Put all facets of one idea together, or put an idea and an elaboration together in order to relate them in a subject's mind.

Organization/ structure: Seems most common in time series animations. Creates a hierarchical structure that allows subjects to see things in an order or format that builds upon a given knowledge base, or replicates given conditions through time.

Elaboration: Animation is simply an elaboration of given information (verbal, textual). For example, if a graph is accompanied by pictures of what is happening, this is elaborating on the initial information. This is done in order to better explain a concept or idea, and may have a lot to do with levels of detail within a presentation.

Abstract/ concrete: Graphics exist along a continuum of concrete to abstract. A graph would be very abstract, whereas a picture is more concrete. Animation can be used as an attempt to introduce concreteness into an abstract subject.

Generalization: As opposed to elaborating on an object or idea, animation can be used to take the subject in the other direction, allowing generalization from one situation to another. Found most often in teaching and learning about rules.

Representation: Take an idea or concept, and portray it in a representative or symbolic form. Usually takes the form of simplifying complex material.

# APPENDIX C

## Questions Used in Experimental Examinations

Short Term Questions

1. What is the normal pH range for soils?

a) 3 -12
b) 2 - 8
c) 6 - 10
d) 4 - 8
e) 5 - 14

2. What happens to an acidified soil's ability to bind positively charged nutrients to it?

a) increases
b) decreases
c) remains the same
d) it begins to hold only the most highly charged ions, like Mg++
e) it fluctuates wildly, allowing large scale loss of ions.

3. As the superheated areas in the asthenosphere convectively rise up under the ocean floor,

a) The plates that form the mantle are slowly pushed apart.
b) The earth's crust lifts up, and forms an oceanic ridge.
c) New sea floor is produced.
d) All of the above.
e) b and c only

4. When an oceanic plate hits a continental plate, what results?

a) The continental plate dives beneath the oceanic plate.
b) The oceanic plate dives beneath the continental plate.
c) An oceanic ridge is formed.
d) New ocean floor is produced.
e) A massive explosion occurs.

5. Draw a hydrograph showing a given storm event on a watershed both before and after installation of roads and a 90% clear-cut. Label all components.

Long-term Test Questions

1. When soil decreases in pH, which mineral elements become soluble, and are available for plant uptake?

a) Ca and Mg
b) K and Mn
c) Mn and Al
d) H and Mn
e) K and Al

2. Which of the following is true about acidification of soil.
    a) K increases in concentration to toxic levels.
    b) H ions are carried up to the surface of the soil, and are lost to the atmosphere.
    c) Mg and Ca are washed out of the soil, leaching down beyond the reach of plants.
    d) Mg and Mn take up the empty positions on soil particles, and become available to plants in toxic amounts.
    e) Ca, due to it's low charge, is easily displaced on the soil particles by Mg.

3. What is the optimum soil pH for plant growth?
    a) 3.5 - 4.5
    b) 4.5 - 5.5
    c) 5.5 - 6.5
    d) 6.5 - 7.5
    e) 7.5 - 8.5

4. Why is this the optimum soil pH?
    a) Most nutrients needed by plants are available in that range.
    b) Toxic elements do not become soluble and available for uptake in that range.
    c) The soil is most able to bind positively charged ions to it at that pH.
    d) The plant roots are not burned by excessive acidity at this level.
    e) Mn is insoluble at lower pHs, and plant nutrition suffers.

5. Which of the following has the greatest effect on the shape of curves in a hydrograph?
    a) the absence of trees
    b) the installation of the roads
    c) the size of the channel
    d) the length of time from logging to reforestation
    e) the amount of slash left on the ground after logging

6. How does the amount of water discharged from the stream in a clear-cut vs. a forested watershed differ?
    a) There is no difference in discharge amounts
    b) The forested watershed discharges more water
    c) The clear-cut watershed discharges more water
    d) The forested watershed discharges more water earlier in the storm than does the clear-cut watershed.

7. Think about the three main layers that make up the earth and pick the option that lists these from the inside out. (graphic given)
   a) mantle, lithosphere, asthenosphere
   b) lithosphere, asthenosphere, core
   c) mantle, core, asthenosphere
   d) mantle, asthenosphere, lithosphere
   e) core, asthenosphere, mantle

8. What is the relationship of plate tectonics to volcanism?
   a) Most volcanoes of the world lie along zones of active plate tectonics.
   b) There is no relationship.
   c) Plate collisions are explosive, with much volcanic matter spewed into the air.
   d) Volcanoes often result when superheated material rises up from the mantle.
   e) Volcanoes do not exist except in areas where one plate is subducting under another.

9. About how much do plates in the earth's crust move every year?
   a) a few millimeters
   b) a few centimeters
   c) a few meters
   d) a few decimeters
   e) a few kilometers

10. Explain this graph (included with this question is 3, unlabeled).

11. Explain how and why acid rain can have a negative effect on soil productivity. Be specific.

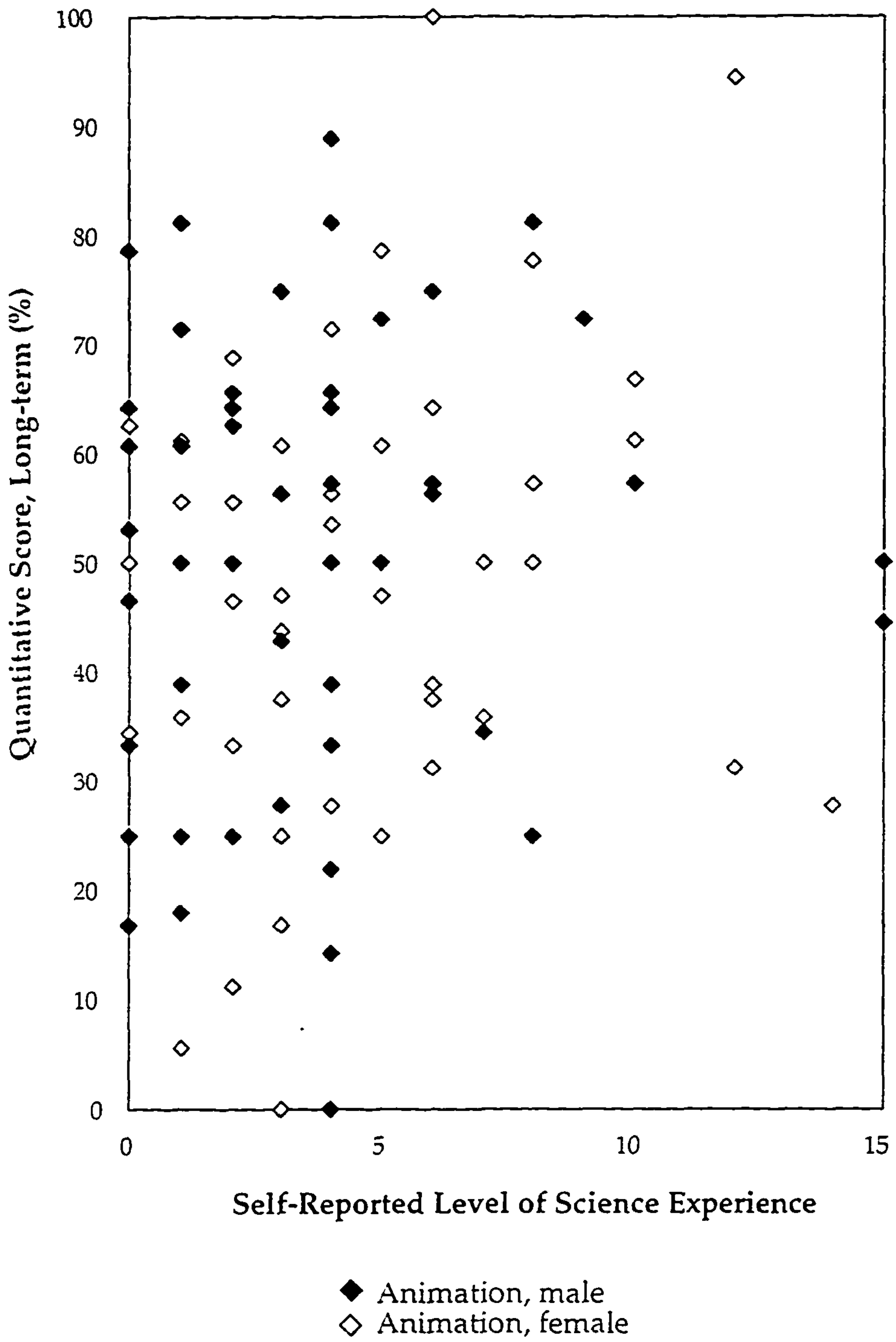
12. Explain the process of subduction in relation to plate tectonics.

13. How are oceanic ridges formed?

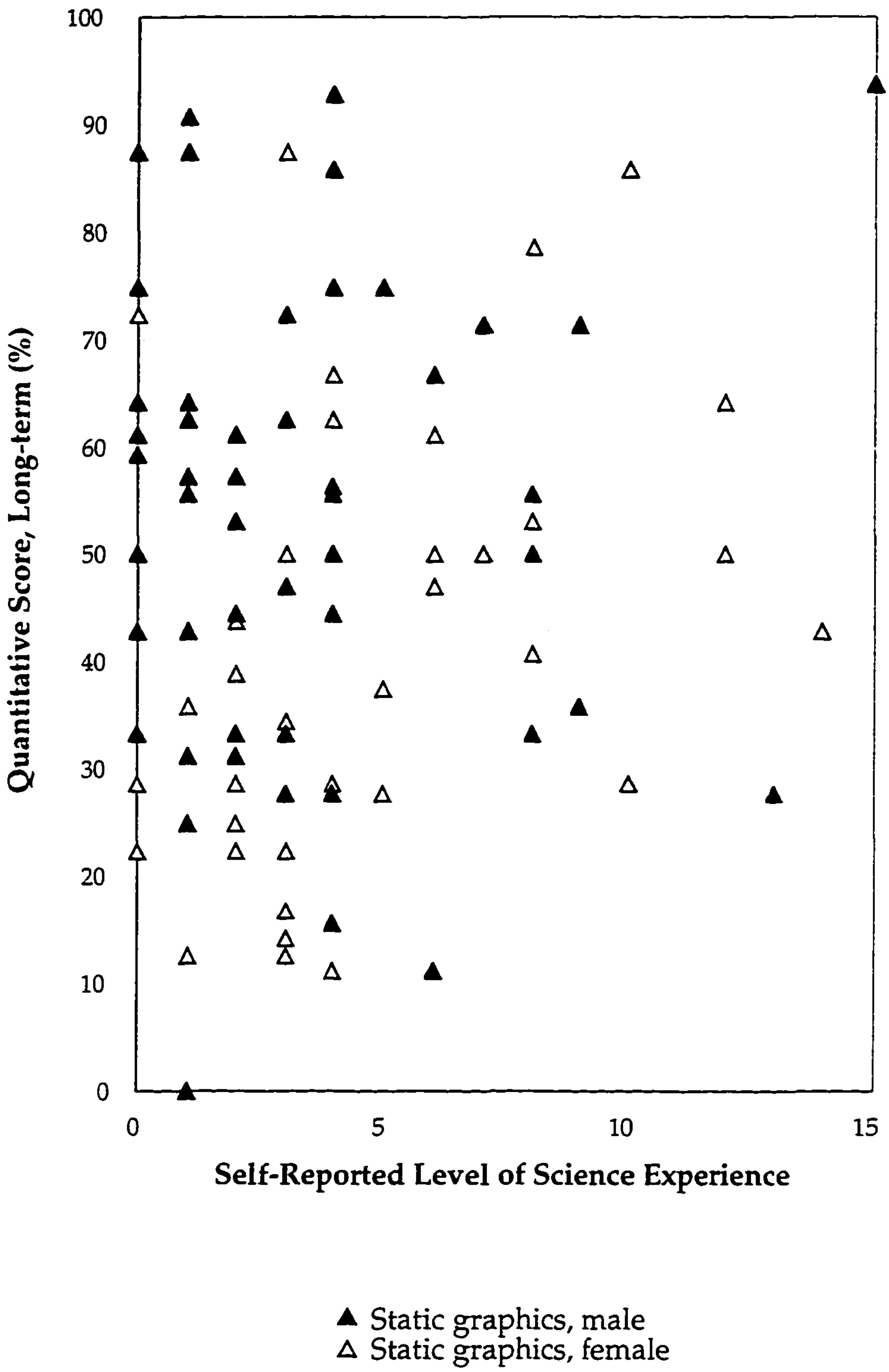
# APPENDIX D

## Pre-Term Questionnaire Correlations

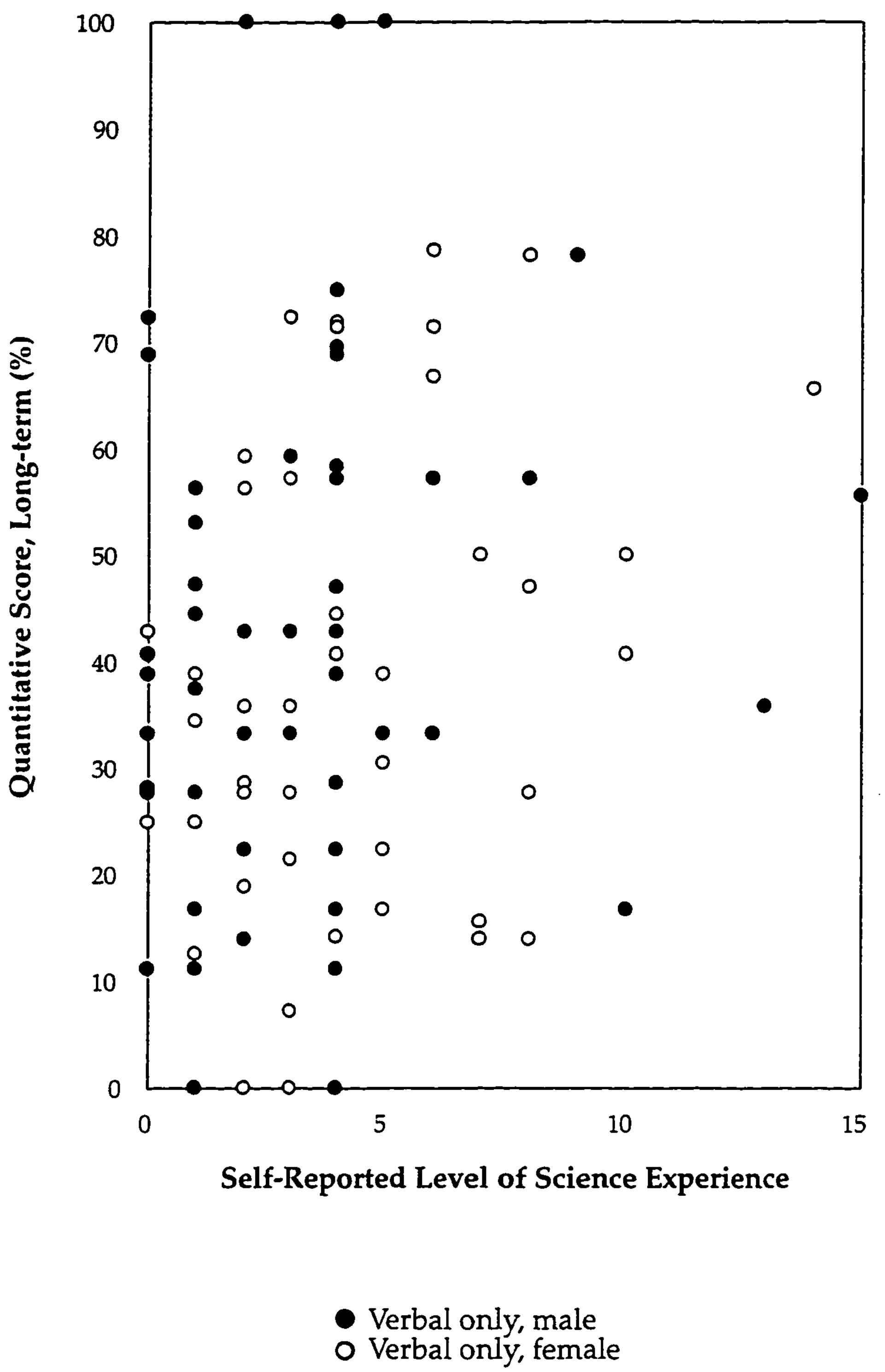
The graphs following this page each portray information gathered through the use of pre-term survey results. Each graph represents a set of responses, set against how each subject performed on one of the three experimental treatments.



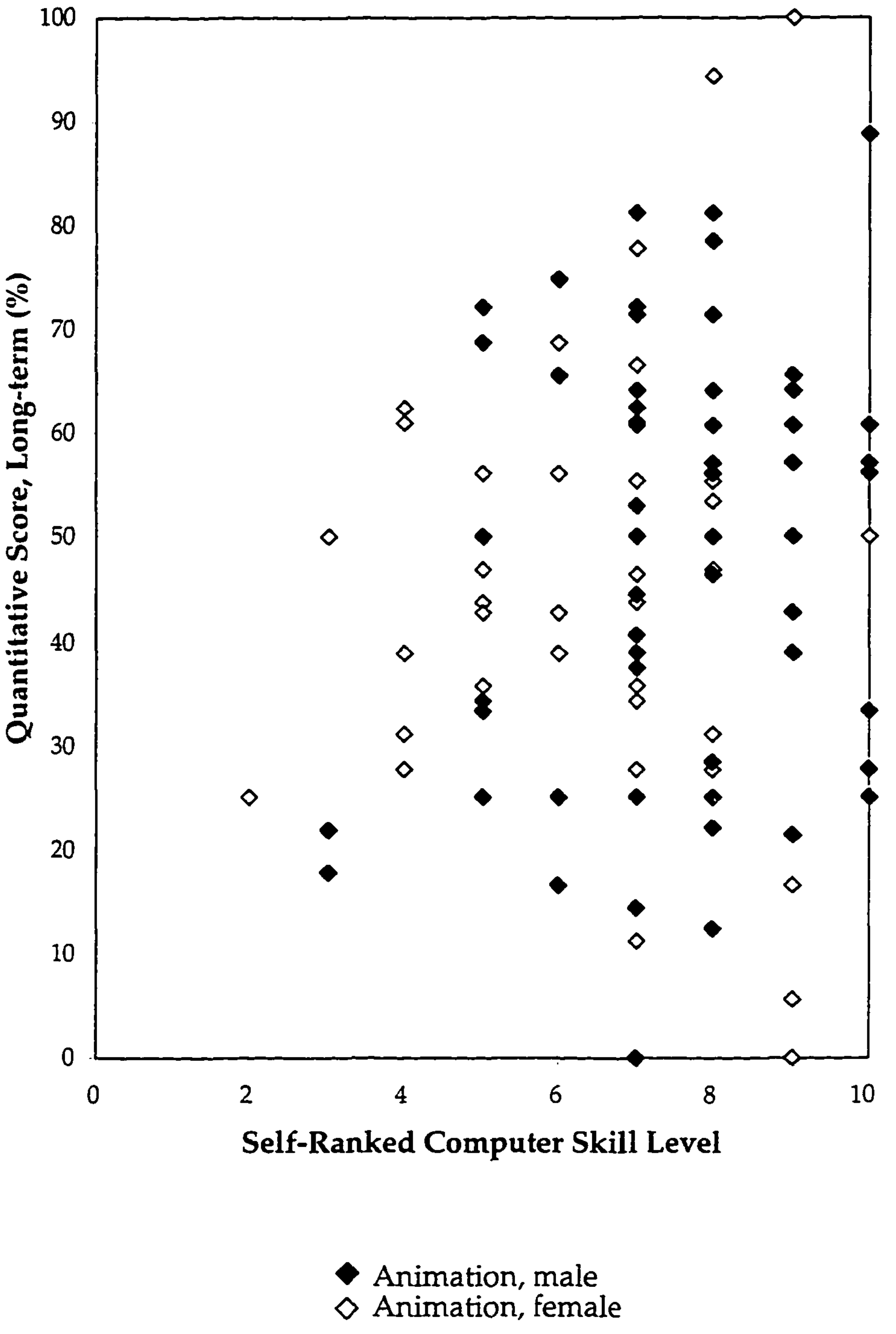

Relationship between self-reported level of science experience and scores on long-term testing in animation presentations, subgroup A. Level of science experience determined by the number of science classes that a given student has taken, and/or job years of job experience in the field of science.



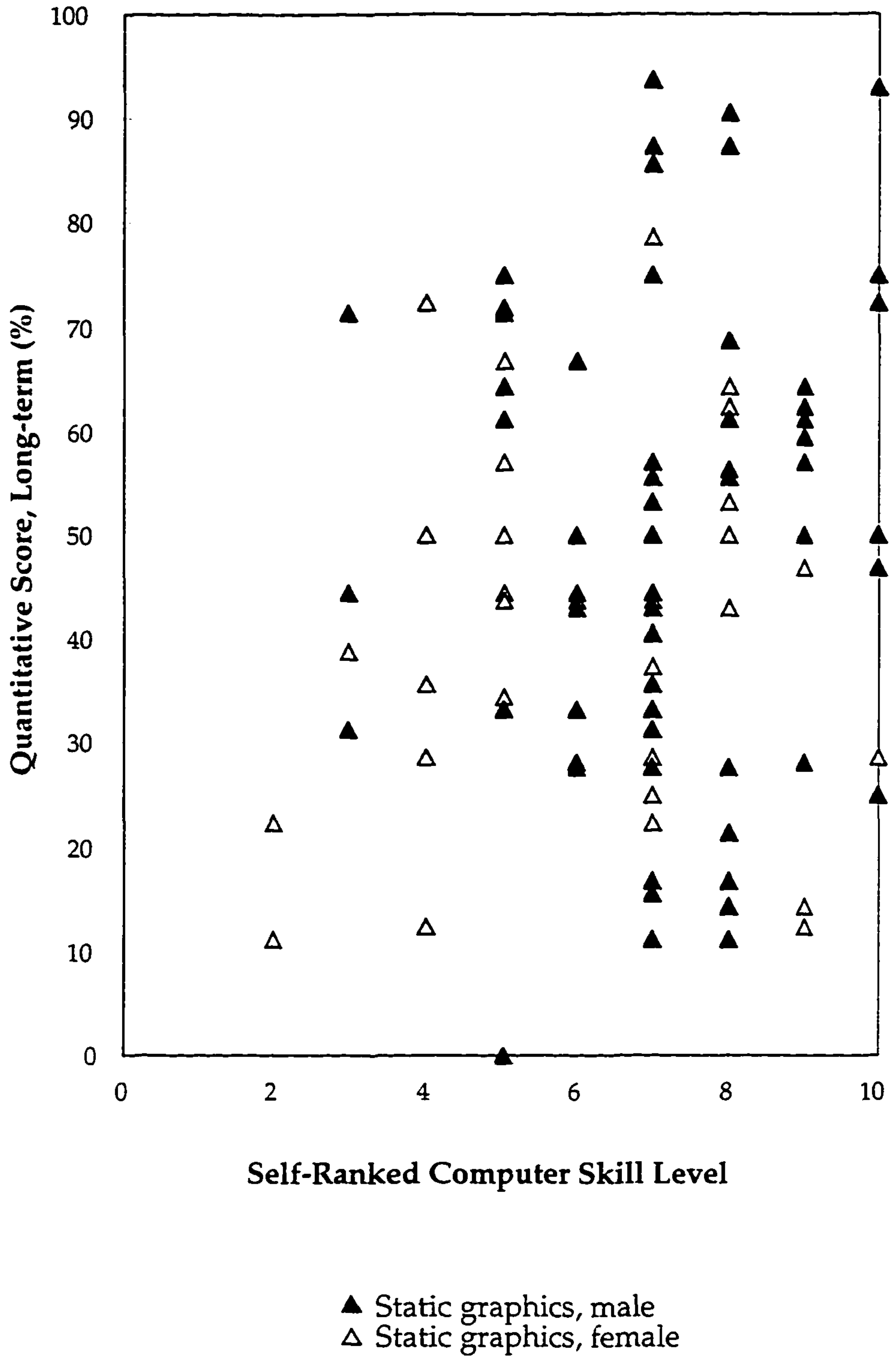

Relationship between self-reported level of science experience and scores on long-term testing in static graphic presentations, subgroup A. Level of science experience determined by the number of science classes that a given student has taken, and/or job years of job experience in the field of science.



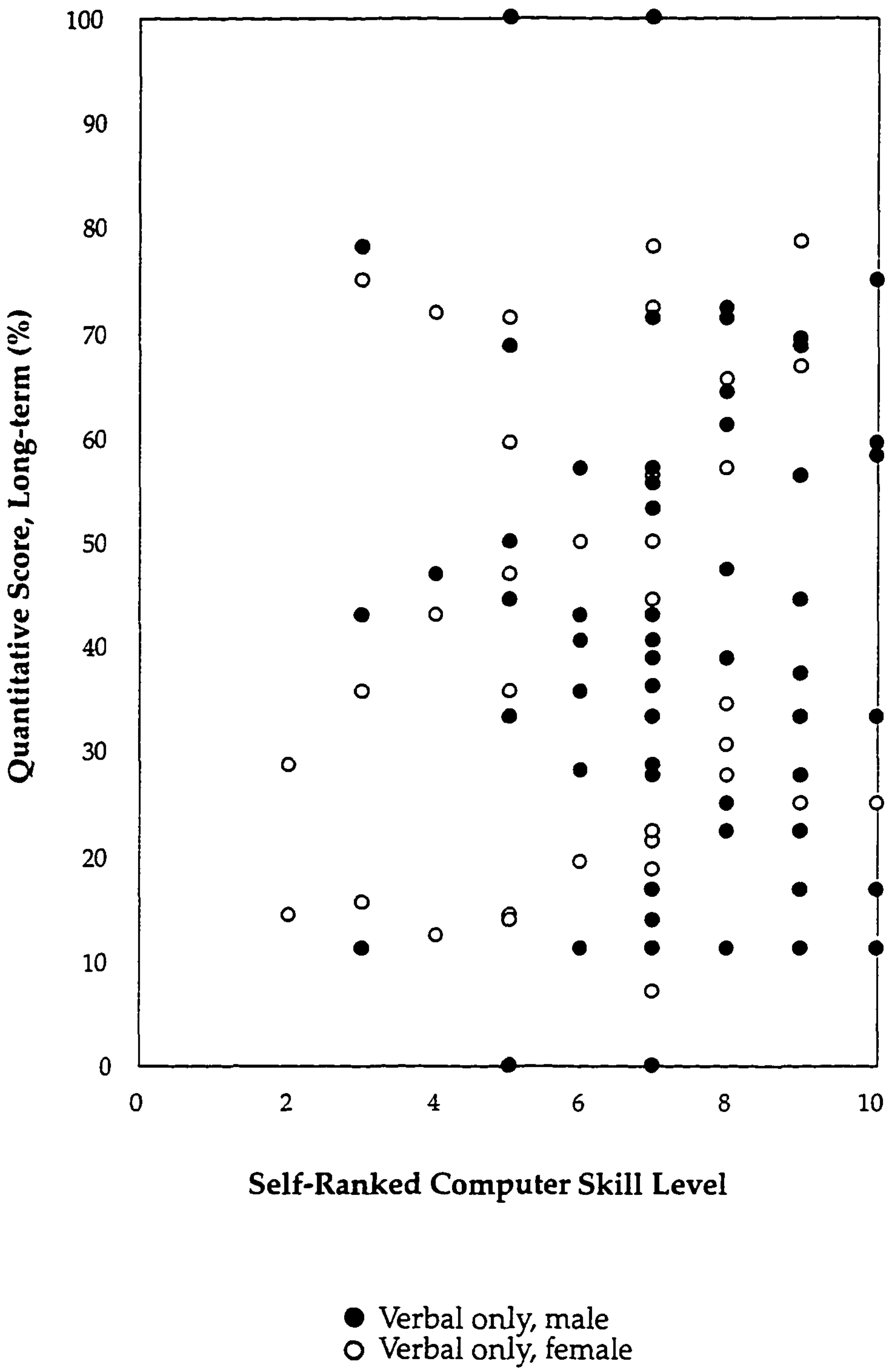

Relationship between self-reported level of science experience and scores on long-term testing in verbal instruction only presentations, subgroup A. Level of science experience determined by the number of science classes that a given student has taken, and/or job years of job experience in the field of science.



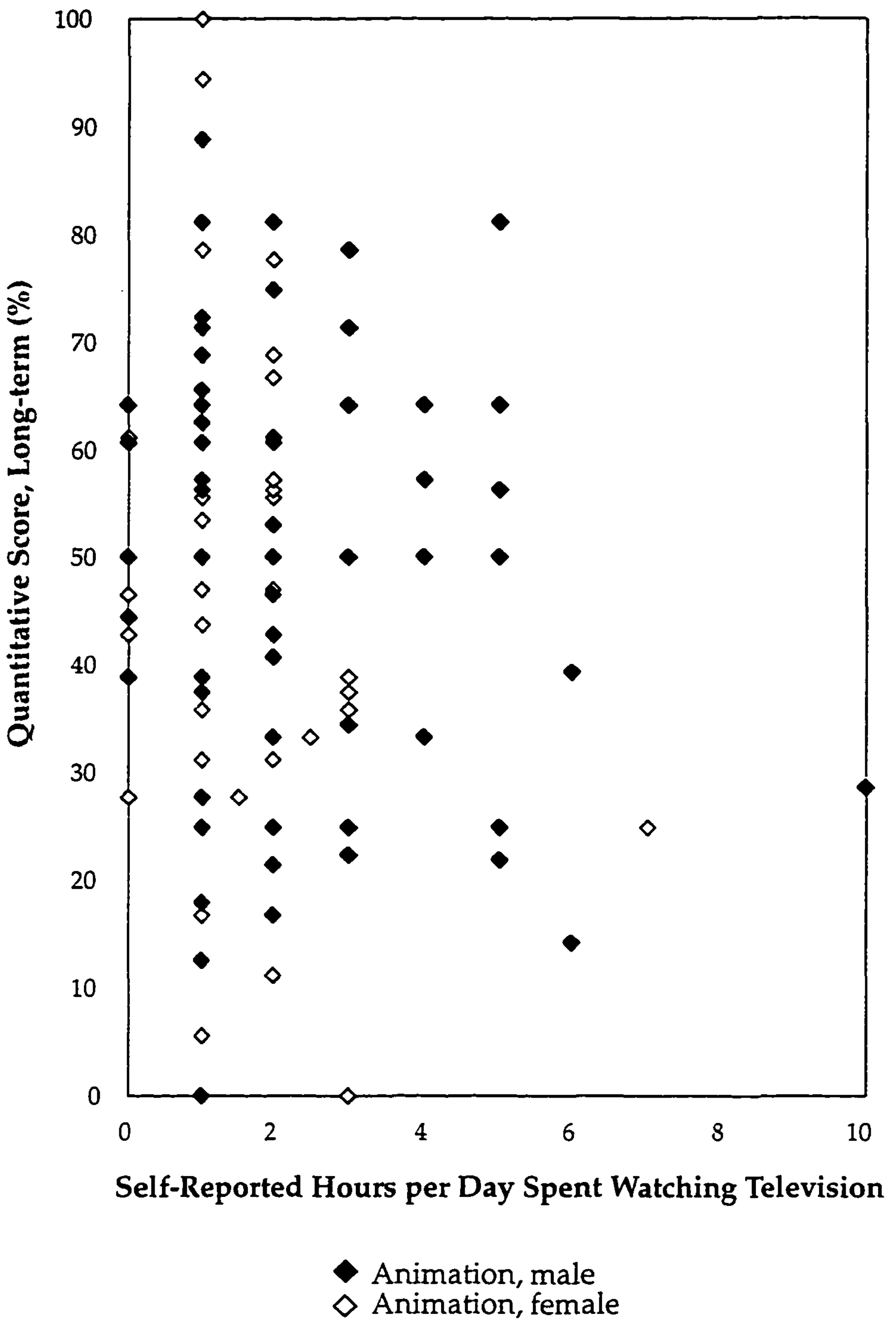

Relationship between self-reported computer skill level and scores on long-term testing in animation presentations, subgroup A.



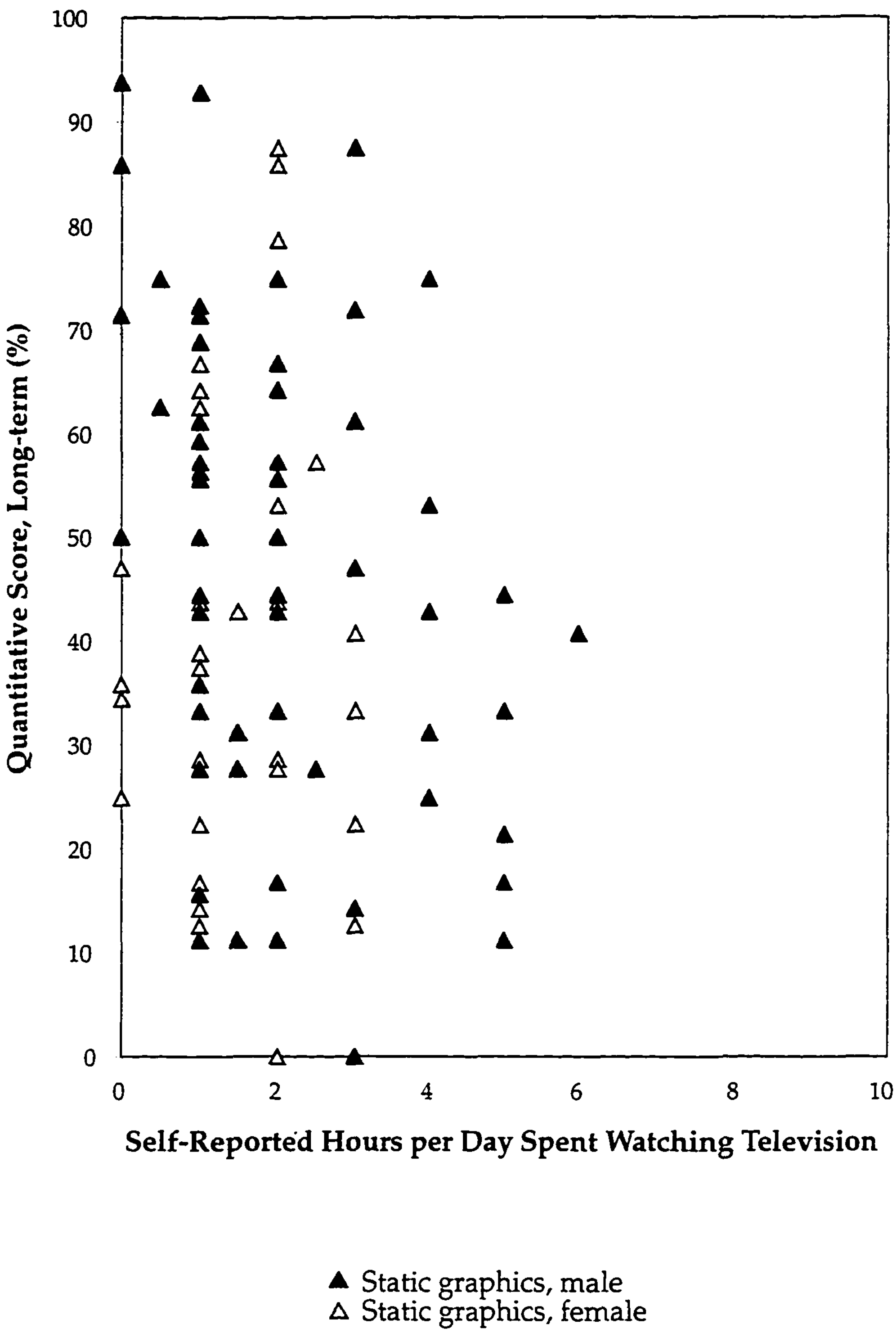

Relationship between self-reported computer skill level and scores on long-term testing in static graphics presentations, subgroup A.



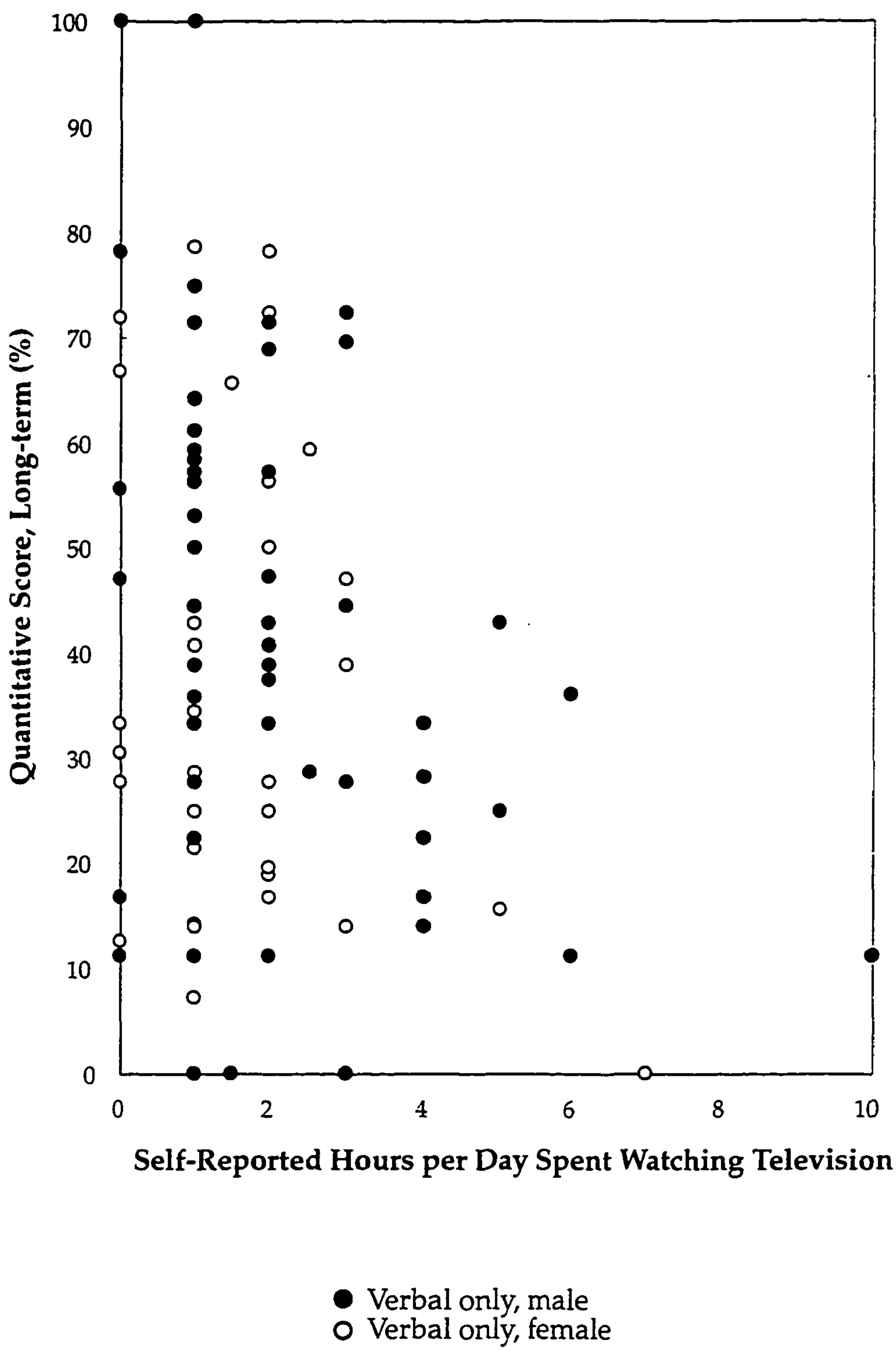

Relationship between self-reported computer skill level and scores on long-term testing in verbal instruction only presentations, subgroup A.



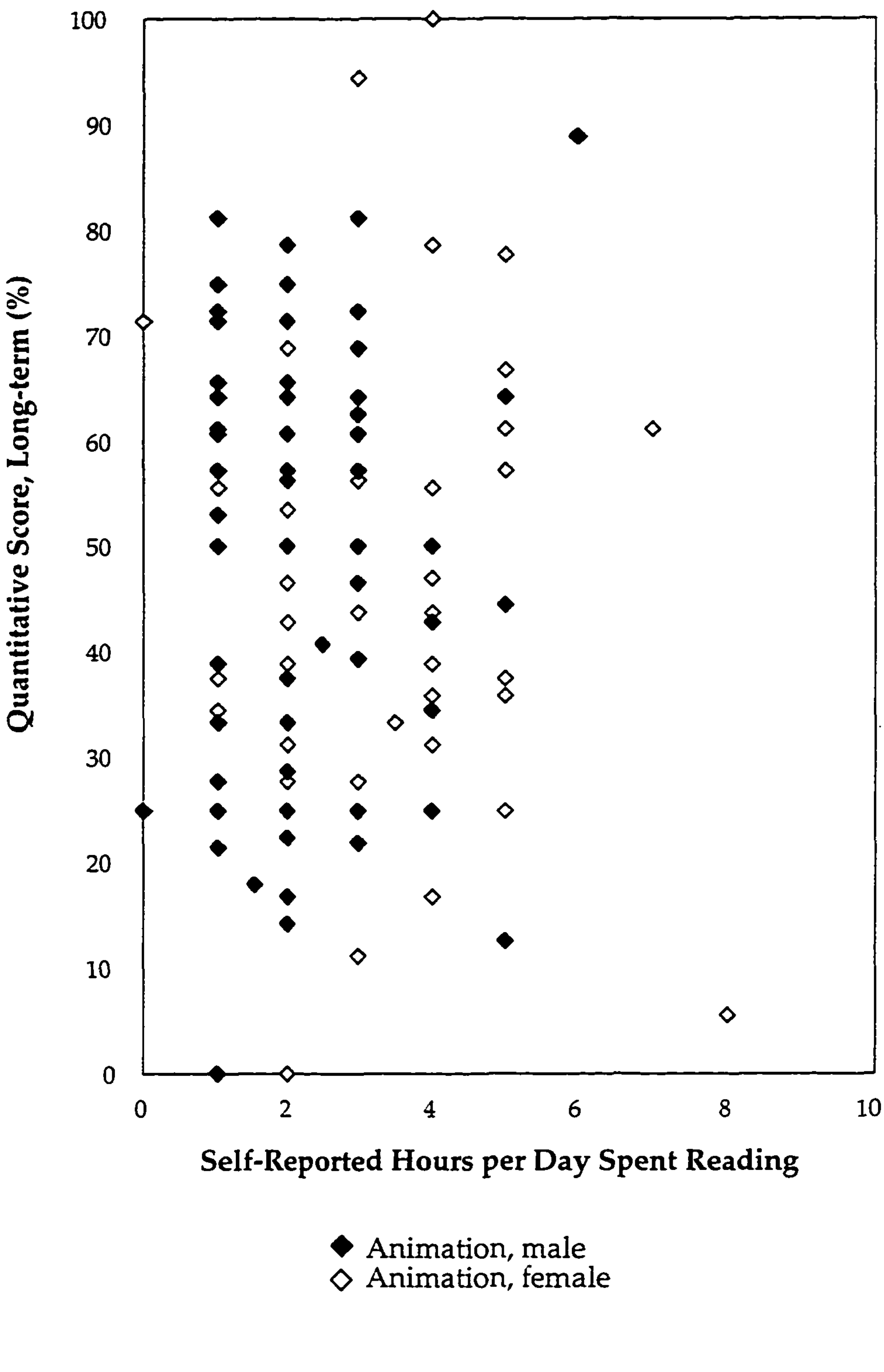

Relationship between self-reported hours per day spent watching television and scores on long-term testing in animation presentations, subgroup A.



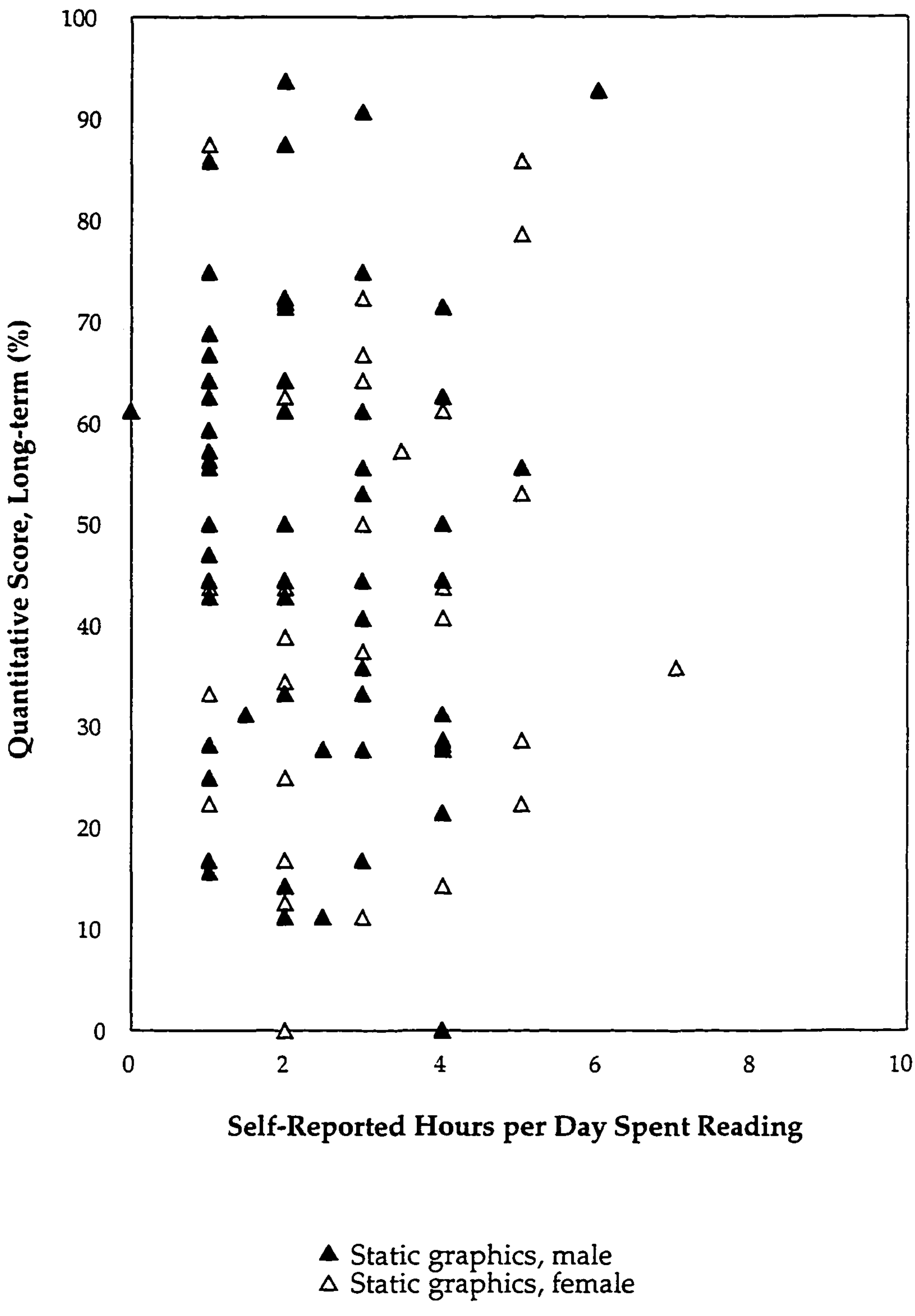

Relationship between self-reported hours per day spent watching television and scores on long-term testing in static graphic presentations, subgroup A.



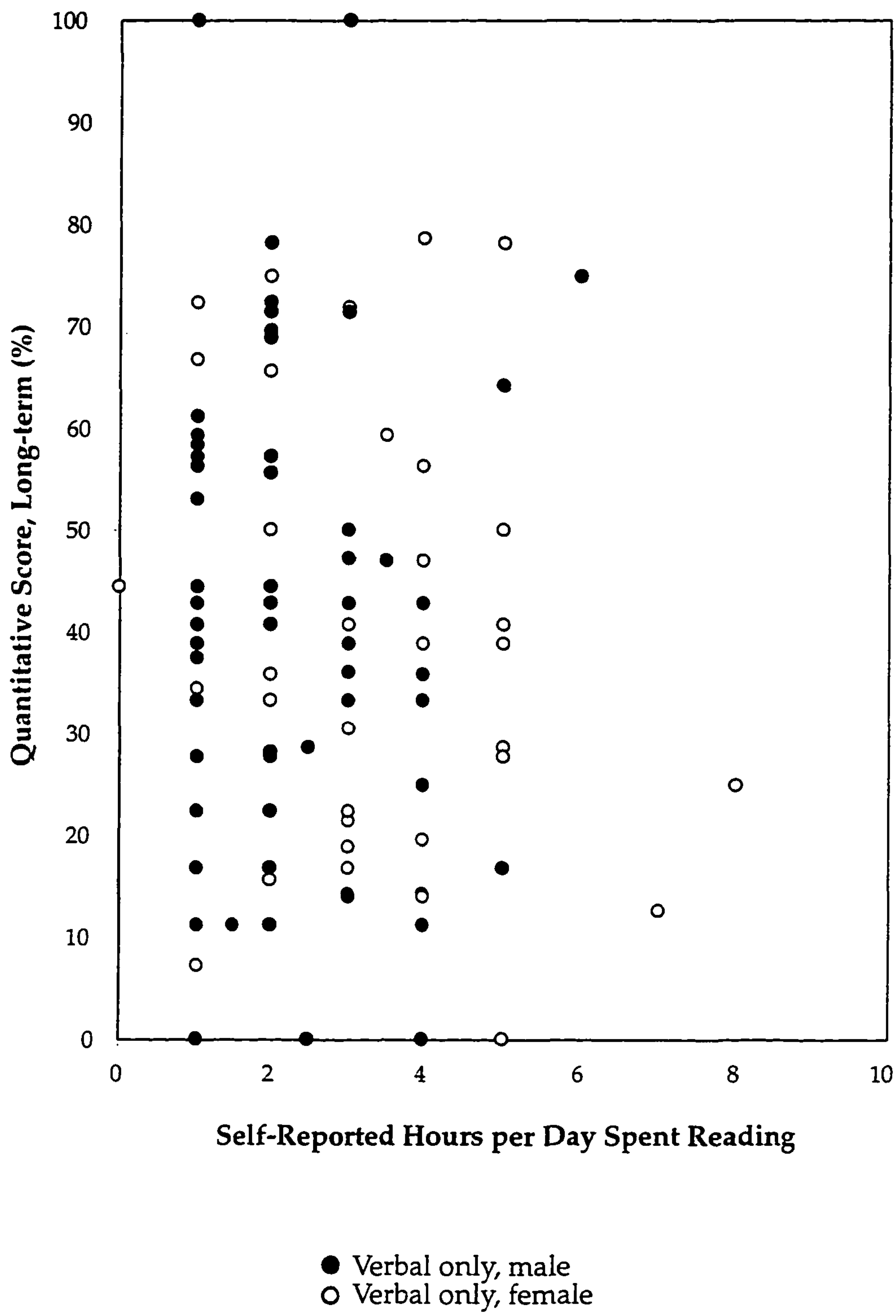

Relationship between self-reported hours per day spent watching television and scores on long-term testing in verbal instruction only presentations, subgroup A.




Relationship between self-reported hours per day spent reading and scores on long-term testing in animation presentations, subgroup A.




Relationship between self-reported hours per day spent reading and scores on long-term testing in static graphic presentations, subgroup A.




Relationship between self-reported hours per day spent reading and scores on long-term testing in verbal instruction only presentations, subgroup A.

# APPENDIX E

## Soil Acidification and Forest Hydrology Animations

Animations enclosed on computer disk.

# Please Note

The diskette is not included in this material. It is available for consultation at this author's graduate school library.

UMI