An Intelligent and Interactive Simulation and Tutoring Environment for Exploring and Learning Simple Machines

by

Lakshman Sundeep Myneni

A dissertation submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Auburn, Alabama
December 12, 2011

Keywords: interactive simulation, intelligent tutoring system, virtual experiments, educational technology

Approved by

N. Hari Narayanan, Chair, Professor of Computer Science and Software Engineering
Dean Hendrix, Associate Professor of Computer Science and Software Engineering
Cheryl Seals, Associate Professor of Computer Science and Software Engineering
Margaret E. Ross, Professor of Educational Foundations, Leadership, and Technology

Abstract

Students in middle school science classes have difficulty mastering physics concepts such as energy and work, taught in the context of simple machines. Moreover, students' naive conceptions of physics often remain unchanged after completing a science class. To address this problem, I developed an intelligent tutoring system, called the Virtual Physics System (ViPS), which coaches students through problem solving with one class of simple machines, pulley systems. The tutor uses a unique cognitive based approach to teaching simple machines, and includes innovations in three areas. (1) It employs a teaching strategy that focuses on highlighting links among concepts of the domain that are essential for conceptual understanding yet are seldom learned by students. (2) Concepts are taught through a combination of effective human tutoring techniques (e.g., hinting) and simulations. (3) For each student, the system identifies which misconceptions he or she has, from a common set of student misconceptions gathered from domain experts, and tailors tutoring to match the correct line of scientific reasoning regarding the misconceptions. ViPS was implemented as a platform on which students can design and simulate pulley system experiments, integrated with a constraint-based tutor that intervenes when students make errors during problem solving to teach them and to help them. ViPS has a web-based client-server architecture, and has been implemented using Java technologies.

ViPS is different from existing physics simulations and tutoring systems due to several original features. (1). It is the first system to integrate a simulation based virtual experimentation

platform with an intelligent tutoring component. (2) It uses a novel approach, based on Bayesian networks, to help students construct correct pulley systems for experimental simulation. (3) It identifies student misconceptions based on a novel decision tree applied to student pretest scores, and tailors tutoring to individual students based on detected misconceptions. ViPS has been evaluated through usability and usefulness experiments with undergraduate engineering students taking their first college-level engineering physics course and undergraduate pre-service teachers taking their first college-level physics course. These experiments demonstrated that ViPS is highly usable and effective. Students using ViPS reduced their misconceptions, and students conducting virtual experiments in ViPS learned more than students who conducted experiments with physical pulley systems. Interestingly, it was also found that college students exhibited many of the same misconceptions that have been identified in middle school students.

To

My wife
Divya

and

My advisor
Dr. N. Hari Narayanan

Acknowledgments

I consider completing this dissertation to be the greatest accomplishment of my life thus far. This is a result of sacrifices and encouragement by many individuals. Although it would not be possible for me to list them all, I would like to mention a handful without whom this accomplishment would have remained a dream.

It is with deep sense of gratitude that I acknowledge my indebtedness to my Ph.D. committee members; in particular, my advisor Dr. N. Hari Narayanan. He has been a wise and dependable mentor and an exemplary role model in helping me achieve my professional goals. Dr. Narayanan has always given me invaluable guidance, support and enthusiastic encouragement. Heartfelt thanks are also extended to other committee members, Dr. Dean Hendrix, Dr. Cheryl Seals, and Dr. Margaret Ross for their suggestions and guidance which have greatly improved the quality of my work. Dr. Sanjay Rebello (Physics Education Research Group, Department of Physics, Kansas State University) graciously served as the domain expert for this research. He provided valuable advice, and along with his graduate students helped me conduct experiments. This research would not have been possible without his guidance. I would also like to thank Dr. Sadhana Puntambekar, Professor of Educational Psychology, and graduate students of education (Garret, Dana, Anu) at the University of Wisconsin Madison, for arranging and assisting with a usability study with middle school students. I would like to thank Dr. Chaudhury for being the external reader for my dissertation.

Special thanks goes to all the students who participated in the evaluation experiments. I would also like to thank all the professors/teachers who have taught me and under whom I have worked as a Teaching Assistant at Auburn. The inspiration I have drawn from my long list of friends, right from my childhood to this date, deserves a special acknowledgement. From the bottom of my heart, I want to thank my parents, my in laws and my extended family for their love and support. Lastly, I would also like to thank my wife Mrs. Divya for her love, support and unstinting faith.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ViPS	Virtual Physics System
ITS	Intelligent Tutoring System
CAI	Computer-Aided Instruction
M1	Misconception 1 – More pulleys easy to pull
M2	Misconception 2 – Longer string easier to pull
M3	Misconception 3 – Pulling upwards is harder than pulling downwards
M4	Misconception 4 – Having more pulleys reduces the amount of work done
M5	Misconception 5 – Size of pulley affects amount of work done
M6	Misconception 6 – Improper interpretation of force and work done

CHAPTER 1: INTRODUCTION

Intelligent tutoring systems are computer systems designed for the support and improvement of learning and teaching in different domains. Computers have been used in education since the sixties (Martin, 2001). Teachers and students use computers in all aspects of education such as researching, preparing study plans and organizing lecture notes, collecting grading information and doing homework. At the present time, it is hard to imagine a modern education without computers. The use of computers can be beneficial for teachers and learners. As (Molnar, 1997) points out, “Research shows that educational technology, when properly applied, can provide an effective means for learning”. In support of this viewpoint, the research reported in this dissertation addresses the relationships between computers and education, the advantages of using computers in education and the pitfalls of using them, in the context of Intelligent Tutoring Systems (ITS).

According to Katie Hafner, “Broadly defined, an intelligent tutoring system is educational software containing an artificial intelligence component. The software tracks student’s work, tailoring feedback and hints along the way. By collecting information on a particular student’s performance, the software can make inferences about strengths and weaknesses, and can suggest additional work.” (Hafner, 2000).

An ITS can be used to enable students to work independently, to improve their understanding of concepts within the domain of discourse, and to improve their problem solving ability. ITS can assist not only students but also their teachers in developing and managing courses. Intelligence involves mental capabilities such as reasoning ability, planning, solving problems, thinking abstractly, comprehending ideas, and learning. Ideally, tutoring systems must have the capability of real teachers, acting as human tutors do in a class. This remains an unrealized dream, however, the field has made tremendous progress in the last three decades.

In the 1970s, taking advantages of computer technology, researchers began looking for new educational paradigms to produce improvements in student learning and problem solving. Combining artificial intelligence, cognitive science and advanced technologies, Intelligent Computer-Aided Instruction (ICAI) came on the scene (Molnar, 1997).

Although Intelligent Tutoring Systems began within the field of Computer-Aided Instruction (CAI), they differ from CAI approaches. Firstly, CAI interfaces are static, and the information presented to each student is exactly the same as the information presented to all other students. ITSs on the other hand use knowledge about the student and the pedagogical process so that the system tries to determine what the student knows or does not know. Contrary to ITSs, CAIs make apriori assumptions about what the student knows. Therefore, the same material is presented to all students by CAIs.

Another difference between CAI and ITS is the feedback system. Some CAIs have the capability of asking questions to students. However, the feedback system is restricted to an indication of whether the student answer was correct or wrong. ITSs instead try to determine the students' weaknesses (knowledge gaps) on a topic using domain and student models. The theory of Intelligent Tutoring Systems began to evolve and separate from Computer-Aided Instruction in the early 1970s.

The general architecture of an ITS consists of four different components as described below.

- Domain and Expert Module: It represents expert knowledge of a domain and the ability to solve problems within the domain.
- Student Module: It embodies a theory of student behaviors. It tries to determine student's knowledge states by developing, on the fly, a model of each individual student from all interaction information it can glean from the student's actions and responses to the system. This model includes information such as what the student knows or does not know, his or her reasoning skills, etc.
- Pedagogical Module: This is also known as the Tutor Module. It is responsible for the instructional competence of the system, and provides implementations of different tutoring strategies.
- Communication Module: It implements the human computer interface of the ITS.

1.1. Motivation

Pulleys are one of the fundamental topics in learning about simple machines in the middle school curriculum. In addition, this context presents several questions that are interesting to explore in terms of how students' learning is supported by virtual experimentation through simulation and intelligent tutoring in comparison with real world experimentation using physical manipulatives. Intelligent tutoring systems for simple machines do not seem to have been studied yet in terms of student's learning, understanding and retention ability.

1.2. Research Questions

Our primary aim is to design and develop a Virtual Physics system (ViPS) combining simulation and intelligent tutoring capabilities that can diagnose students' knowledge, structure, and misconceptions, and tutor them in real time. This dissertation will examine and answer the following research questions:

1. Can students learn better using Virtual Physics System (ViPS) when compared to learning from physical manipulatives?
2. Can ViPS detect and clear the misconceptions students may have?
3. Do experimenting with physical followed by virtual manipulatives (i.e., working with real pulley systems followed by working with ViPS) or vice-versa provide different support for students' conceptual understanding?
4. How satisfied are students using ViPS (i.e., how usable is the system)?

1.3. Organization

The rest of the dissertation is organized as follows. Chapter 2 describes the research background - how ITS and simulations are used in helping students learn better, pros and cons of replacing physical manipulatives with virtual experimentation and the literature on intelligent tutoring systems. Chapter 3 presents the architecture of ViPS.

Chapters 4 and 5 explain the technical methodology that was applied in building the simulation, tutor and feedback modules of ViPS. Chapter 6 covers the evaluation studies and their results. Chapter 7 states our conclusions and future work.

CHAPTER 2: LITERATURE REVIEW

Learning about pulley systems is a challenging task for many students. Research by Hegarty (1988) has shown that the rules a person uses to relate the attributes of a machine to its function reflect his level of understanding of the machine. There are several different levels of understanding of machines that a person might have acquired through experience. Different levels of understanding of machines can be illustrated in the understanding of a simple machine such as a pulley system. According to Wikipedia, pulleys were first used in eighth century B.C., considerably later than other simple machines, such as lever and wedge. Like all simple machines, pulleys were used long before the mathematical relationships between loads and displacements in these machines were formally described by Archimedes (third century B.C.). The analysis of pulley systems in terms of balance of force in the systems depend on principles of Newtonian physics, formalized about 2000 years later. Thus, it is possible to have a practical understanding of pulley systems without understanding the physics principles that underlie their operation. People who may use simple machines in everyday life may understand how these machines work but are unable to extrapolate this knowledge to understand an unfamiliar machine (Hegarty, 1982).

Clement (1983), diSessa (1983), and White (1983) conducted studies on novice understanding of machines and suggested that as a result of everyday experience with machines, people develop intuitive physical laws, which are typically qualitative, situation-specific, and involve misconceptions.

The physics concepts in pulley systems are rich and complex, so that many students get confused. Many students often struggle to solve problems after an instructor explains what they need to know, and keep making the same problem solving errors after teachers tell them the right answers. Meanwhile, students having difficulties may not want to admit they are having problems or may have difficulties explaining their problems to an instructor. These situations happen for many reasons: 1) it is difficult to explain the problem from the student‘s perspective—one must first understand what the student knows and does not know; 2) it is hard to trace how many times a student commits similar errors and to observe repetition of problem solving patterns in a classroom setting; 3) it is hard to remedy each individual student‘s deficient problem solving patterns and encourage sound ones in a class with many students; and 4) an instructor may not know who is having difficulties, may not be able to tell why the student is having these difficulties, and may simply not have enough time to look into every student‘s needs in a large class.

Many researchers have described the affordances and limitations of problem solving using physical manipulatives and computer simulations in science education research (De Jong & Van Joolingen, 1998; Finkelstein, et al., 2005; Triona & Klahr, 2003, Triona, et al., 2005). Zacharia and Anderson (2003) investigated the effects of interactive computer-based simulations that are presented prior to inquiry-based laboratory experiments on the students' conceptual understanding of mechanics and found out that the use of simulations improved students' ability to make acceptable predictions and explanations of the phenomena in the experiments. Triona, et al., (2005) investigated how physical and virtual manipulatives affect student learning about mousetrap cars. Students used either physical or virtual manipulatives to design their cars. The physical and virtual treatments showed the same effectiveness in helping students design cars. Finkelstein, et al., (2005) looked at how students learned about circuits differently with virtual or physical manipulatives. The simulations used by the students were similar to the physical materials, except that the simulations showed electron flow within the circuit, which the physical materials could not. Finkelstein reported that the students who had used the virtual manipulatives, i.e. the simulations, scored better on an exam and were able to build physical circuits more quickly than students who have used physical manipulatives.

Zacharia, Olympiou and Papaevipidou (2008) looked at physical and virtual manipulatives in the context of heat and temperature. One group of students used physical manipulatives, while the other group of student used physical manipulatives followed by virtual manipulatives. Students who used physical followed by virtual manipulatives performed better on a conceptual test than students who used just the physical manipulatives. The authors concluded that the simulation could be manipulated more quickly than the physical manipulative, increasing student learning.

Ploetzner, et al., (2006) conducted two experimental studies to investigate whether students' difficulties in understanding line graphs in kinematics could be improved by means of dynamic representations. After the first study, they reported that the employed representations not only did not improve learning – but even impeded it. This finding was underpinned by the second study, in which almost no sign was found of students having internalized the relevant aspects of the dynamic representations. The authors argue that the use of different dynamic and interactive visualizations has the potential to improve physics learning, but that it also places various demands on the student. "For instance, students need to understand how information is encoded in each representation, how each representation is related to the physical world, and how information in one representation can be related to and transformed into information in other representation" (Ainsworth, 1999). Dynamic representations require students to process continuously changing information (e.g., Lowe, 2003) and if external representations are not only dynamic but interactive, they require students to carefully prepare, execute and evaluate their interactions.

Lindstrom, et al., (1993) investigated how computer microworlds and simulations contributed to the development of intuitive and conceptual understanding of physics concepts. For this study they selected two central concepts in mechanics, potential energy and moment of inertia. Lindstrom reported that on the concept of potential energy no differences were found in the students' intuitive and conceptual understanding, probably due to the relative short working time in relation to mathematical abstractness of the concept. On the concept of moment of inertia, on the other hand, a significant difference was found between the experimental and control groups. They interpreted this as a result of the intense work prompted by the computer simulation, especially the promotion of a student-student dialogue.

Stieff (2005) presented a novel modeling environment for the chemistry classroom known as Connected Chemistry. He argues that the software currently available for chemistry education is of limited scope. A significant constraint on these learning environments is that many are designed as "black-boxes". Such software offers few interactive features for students to explore the simulated world because each model contains only a narrow range of activities for teachers and restricted outcomes for students. Designs of this sort are often poor simulations of chemical phenomena and have a finite ability to adapt to spontaneous student questioning. To overcome these limitations, Stieff designed Connected Chemistry from the perspective that chemistry concepts are best understood as macro-level patterns that result from the specific interactions of molecules on the molecular level. Connected Chemistry supports students' personal exploration and construction of chemistry concepts by employing a "glass-box" design that provides users with immediate and uncomplicated access to the rules that govern the individual behavior of simulated molecules. Connected Chemistry comprises several molecular simulations embedded in the Net Logo modeling software, a multi-agent modeling language that has been employed in a number of different subject areas, including biology, physics, and environmental science, to help students discover how macro-level concepts emerge from micro-level interactions.

Reiser, Anderson, and Farrell (1985) reported that students working with private tutors could learn given material four times faster than students who attended 12 traditional classroom lectures, studied textbooks and worked on homework alone. Bloom (1984) also reported that students have a better grasp of material working with a private tutor than attending traditional classroom lectures. When a qualified private human tutor is not available, the next best option could be an intelligent tutoring system. An intelligent tutoring system (ITS) is a computer-based instructional system that has knowledge of instructional content and teaching strategies (Dağ & Erkan, 2003). It attempts to use a student‘s level of mastery of topics to dynamically adapt instruction. reported that an ITS is a cost-effective means of one-on-one tutoring that provides novices with the individualized attention needed to overcome learning difficulties. Intelligent tutoring systems are not only being used in academia to augment classroom teaching but have also penetrated various industries where companies are using ITSs to train employees to perform their job functions. As a result, ITSs have been built for various domains such as mathematics, medicine, engineering, public services, computer science, etc.

Gertner and VanLehn (2000) designed an Intelligent Tutoring System for introductory college physics know as Andes. The fundamental principles underlying the design of Andes are: (1) encourage students to construct new knowledge by providing hints that require them to derive most of the solution on their own, (2) facilitate transfer from the system by making the interface as much like a piece of paper as possible, (3) give immediate feedback after each action to maximize the opportunities for learning and minimize the amount of time spent going down wrong paths, and (4) give the student flexibility in the order in which actions are performed, and allow them to skip steps when appropriate. Andes interacts with students using coached problem solving (VanLehn, 1996), a method of teaching cognitive skills in which the tutor and the student collaborate to solve problems.

Roll, et al., (2010) designed a system known as Invention Lab using a hybrid of model tracing and constraint-based modeling to offer intelligent support in enquiry environments. It is a system that combines the benefits of an exploratory learning environment, an environment in which a learner attempts to uncover underlying scientific models, and an intelligent tutoring system by offering adaptive support in a relatively constrained environment. Invention Lab accesses the students' knowledge at domain and inquiry levels and uses this information to design new tasks in real time, thus adapting to students' needs and maintaining the features of inquiry process. An evaluation study conducted by the authors showed that Invention Lab helped students develop sophisticated mathematical models and improve their scientific behavior.

Graesser, et al., (2005) designed an intelligent tutoring system known as AutoTutor that helps students learn Newtonian physics, computer literacy, and critical thinking topics through tutorial dialogue in natural language. The experiments conducted using AutoTutor showed significant positive impact on learning gains when compared to the textbook. In an evaluation of physics tutoring, they reported that the learning gains produced by accomplished human tutors in computer-mediated communications were equivalent to the gains produced by AutoTutor.

Mitrovic, (1998) presented SQL-Tutor, an intelligent teaching system for SQL programming. It is designed as a guided discovery learning environment and supports problem solving, conceptual learning and meta-learning. SQL-Tutor is a problem-solving environment that supports knowledge acquisition of domain knowledge in a declarative form (i.e. constraints). The system tailors instructional sessions to the needs, knowledge, learning abilities and general characteristics of its students.

CHAPTER 3: A TUTORING SYSTEM FOR PULLEY SYSTEMS

3.1. Motivation

This chapter describes ViPS (Virtual Physics System), an Intelligent Tutoring System in the domain of pulley systems. As described earlier, pulleys are one of the most important simple machine concepts taught in middle schools. I chose to develop ViPS for several reasons.

- Pulleys serve to illustrate fundamental concepts in learning about simple machines.
- People who use simple machines in everyday life may understand how these machines work at an intuitive level but are unable to extrapolate this knowledge to understand an unfamiliar machine (Hegarty, 1982).
- People may have a practical understanding of pulley systems without understanding the underlying physics principles.
- As a result of everyday experience with machines, people develop intuitive physics laws that are typically qualitative, situation-specific and involve misconceptions (Clement, 1983).
- Middle school and even college students retain the following common misconceptions related to pulley systems (Rebello, S. (2011), personal communication)
    - The more pulleys there are in a pulley system, the easier it is to pull (M1).

- A longer string is easier to pull than a shorter string in a pulley system (M2).
- Pulling upwards is harder than pulling downwards (M3).
- Having more pulleys reduces the amount of work done (M4).
- The size of the pulley affects the amount of work done (M5).
- Improper interpretation of force and work done (M6).

3.2. System Architecture

The ViPS system provides a student with an interactive environment where pulley systems can be created and simulated. Components required for the pulley systems (different sizes and types of fixed and movable pulleys, strings and loads) can be created and manipulated using a drag and drop interface. Students are asked by the system to solve problems in this environment, such as to create and run a pulley simulation or perform more complex operations. As a student is working towards a solution, the system provides feedback to help the student make progress.

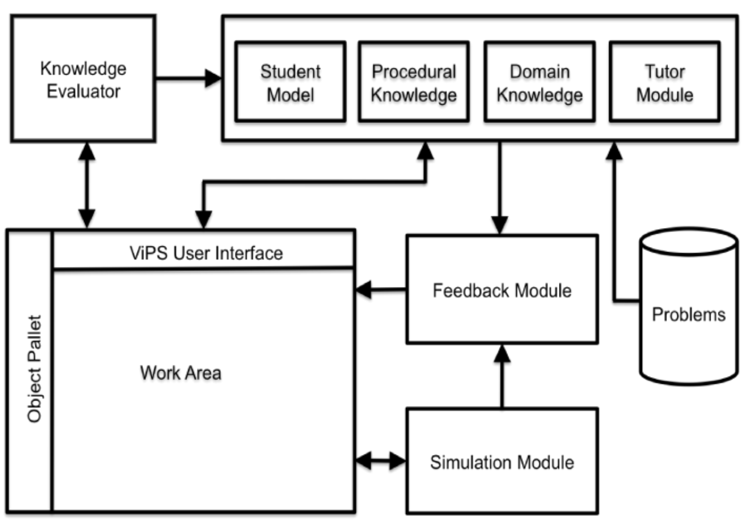


Figure 3.1: Architecture of ViPS

The architecture of ViPS is shown in figure 3.1. It consists of a graphical user interface that manages the interaction with a student; a simulation module that simulates the pulley setups created by the student, a feedback module that generates appropriate messages for the student; a knowledge evaluator that evaluates the prior knowledge of the student; a tutor module that tutors the student for misconceptions; a problem model that represents the problems to be solved; a student model that includes the history of student interactions and various measures of student performance; and a procedural knowledge model that represents the solution paths within individual problems.

3.3. Graphical User Interface

The graphical user interface is responsible for all the interactions with the student. The interface is divided into two main parts: a tabbed work area for creating pulley setups and solving problems; and an object pallet for selecting the components required to create a pulley setup. A snapshot of the interface can be seen in figure 3.2. Using this interface, students can create a pulley setup by dragging the required components from the object pallet to the work area and clicking on the thread button. Students can also interactively manipulate various parameters of the components, like the size of the pulley, value of the load etc.

A problem is given to the student in the form of textual and pictorial representations (figure 3.3). The student is asked to solve the problem by creating the pulley setups required to answer the question, running the simulations and comparing the simulation outputs of the setups created. The problems used in ViPS were designed and created by experienced physics educators. There are in total 30 problems, five for each misconception, stored in a database. A web-based interface is provided to the teachers and educators to add or delete problems. The goal of these problems is to challenge the students' misconceptions regarding pulley systems.

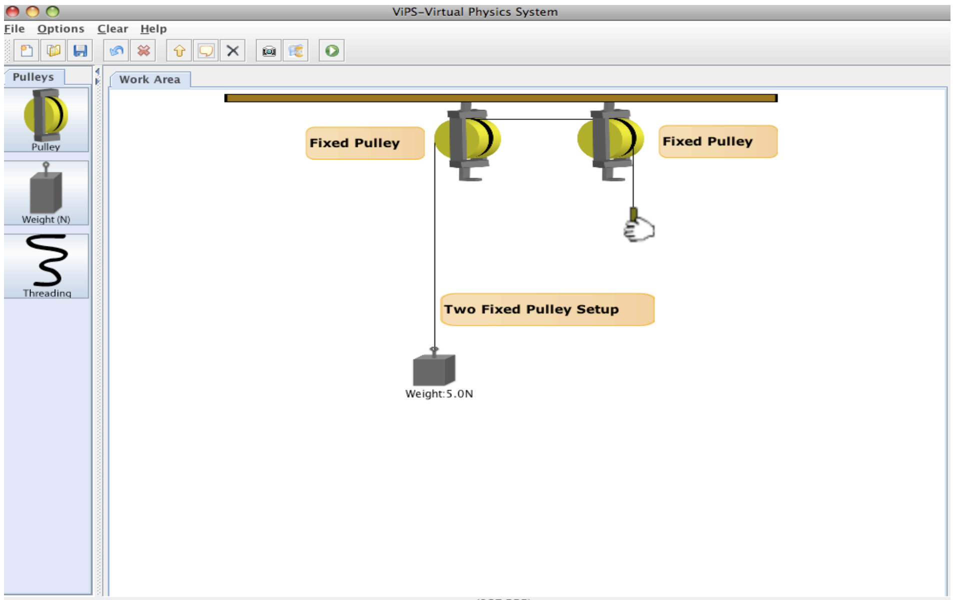


Figure 3.2: Screenshot of ViPS user interface

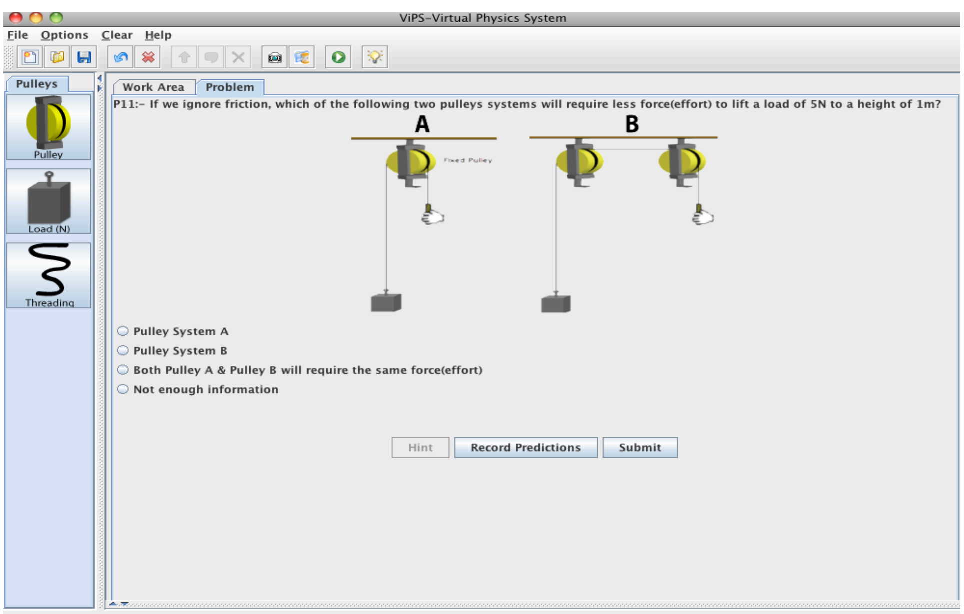


Figure 3.3: Screenshot of ViPS Problem View

3.4. Knowledge Evaluator

When the student first initiates ViPS, a prior knowledge test is given to them. Once he/she finishes the test, the answers are submitted to a knowledge evaluator that evaluates the student's initial knowledge level, determines the misconceptions he/she might have and generates a sequence of problems for the student to solve. Similarly, a post knowledge test is given to the student after the completion of the tutoring and the answers are submitted to the knowledge evaluator to determine the post knowledge level of the student and the status of each originally detected misconception.

3.5. Feedback Module

The feedback module is responsible for generating feedback messages to students. Currently, ViPS can give the following types of feedback.

1. The student creates a setup by dragging the required components onto the work area and clicking the thread button. If the system detects any constraint violations in the setup creation, it generates a feedback known as *setup feedback* (figure 3.4).
2. The student submits his/her problem solution to be evaluated by the tutor module. The system then evaluates his/her solution and based on any constraint violations it generates a feedback message knows as *problem feedback* (figure 3.6).
3. The student tries to create a setup on the work area, but the student sometimes has no idea of what to do next. In those circumstances, ViPS can deliver feedback about the next move the student has to make. This is known as *threading hint feedback* (figure 3.7 and figure 3.8).

4. ViPS can assist students when needed during the process of problem solving, and this is known as *problem hint feedback* (figure 3.5).

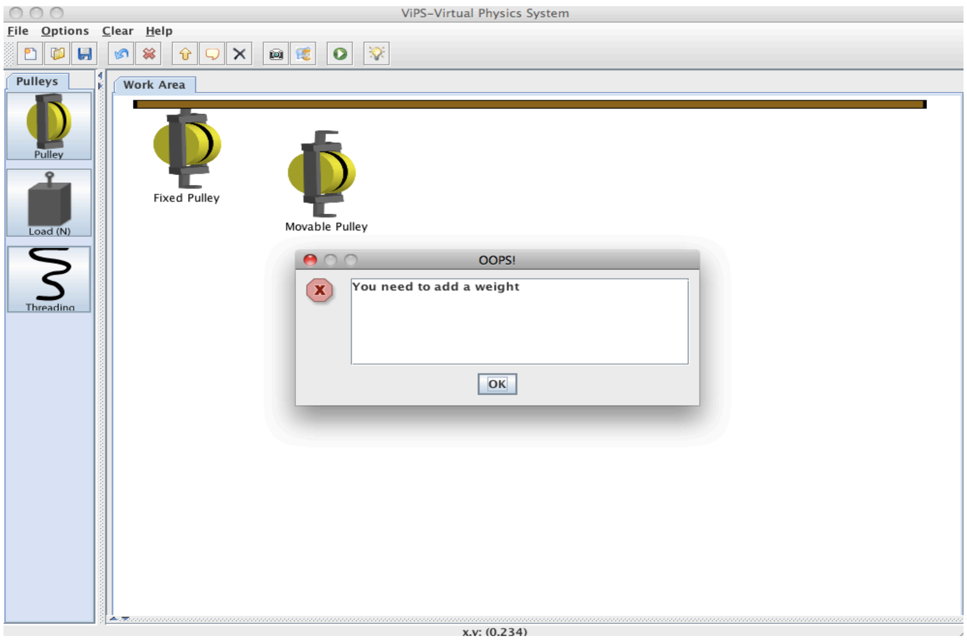


Figure 3.4: Screenshot of setup feedback

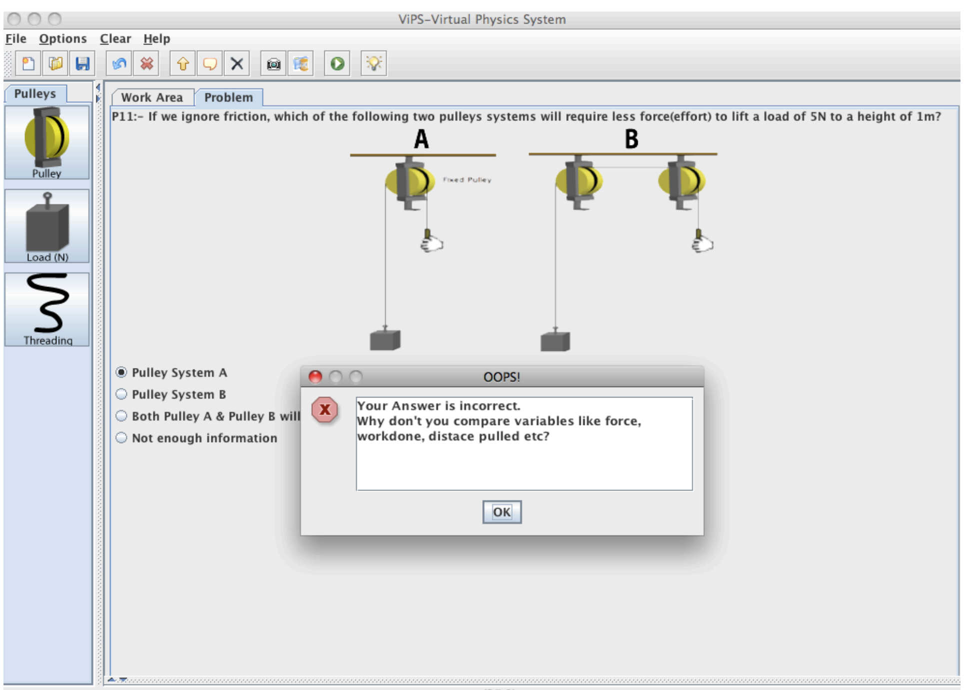


Figure 3.5: Screenshot of Problem hint feedback message

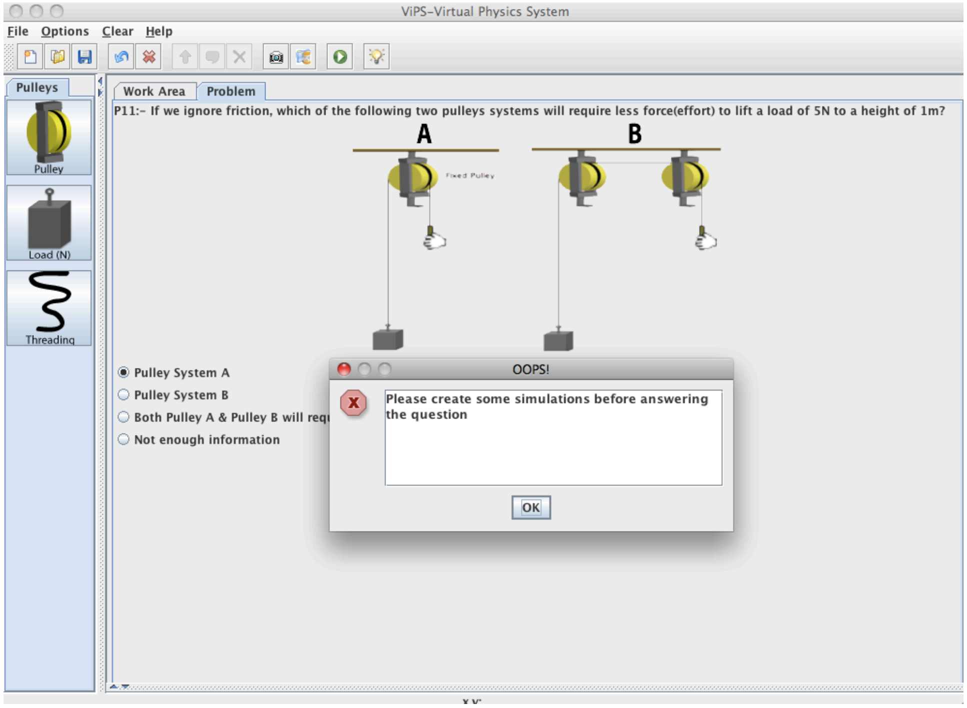


Figure 3.6: Screenshot of Problem feedback message

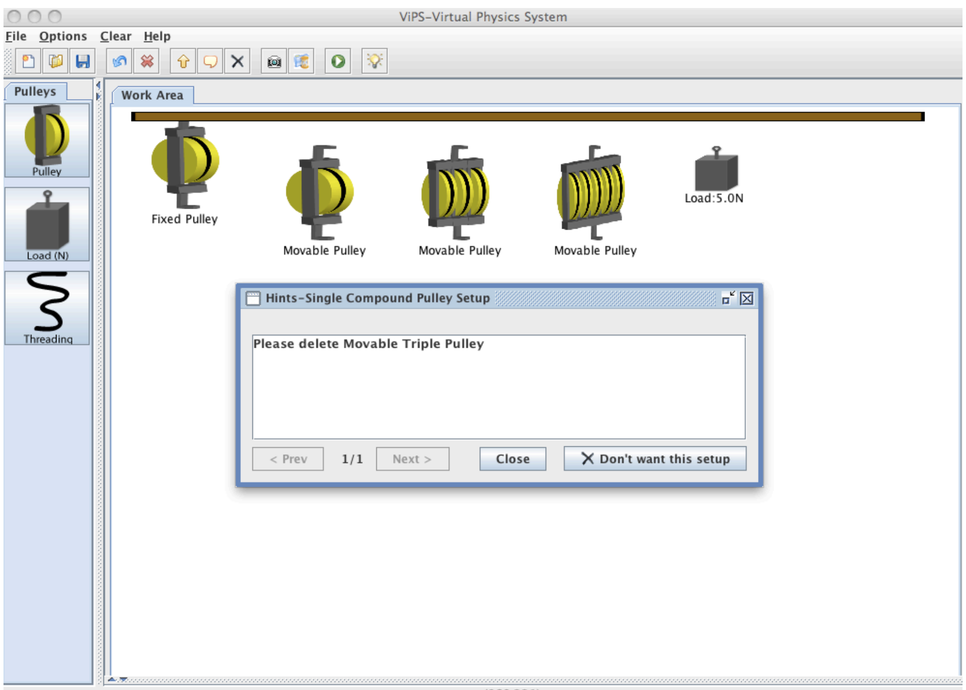


Figure 3.7: Screenshot of Threading hint feedback message

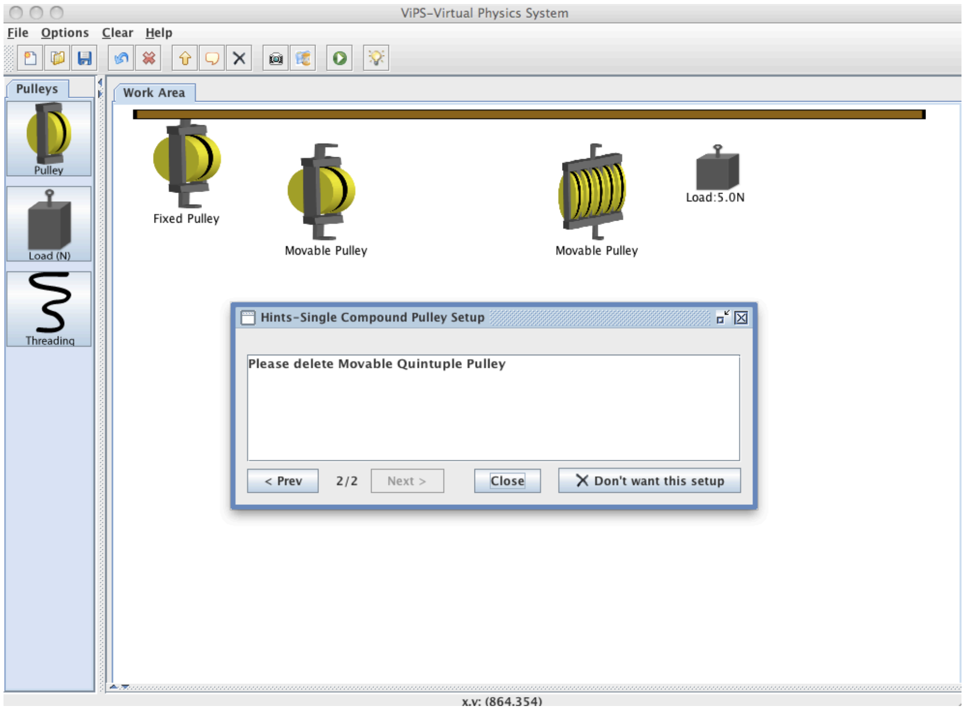


Figure 3.8: Screenshot of Threading Hint feedback message

3.6. Simulation Module

The simulation module is responsible for simulating the pulley setups created by the student. In particular, it provides a platform for running simulations, which may be very hard to create in the physical world, such as complex setups with several multi-grooved fixed and movable pulleys or a setup with zero friction. The outputs generated by the simulation include graphs and real time values of variables like force, work done, potential energy, friction, mechanical advantage and distance pulled. More descriptions about the simulation module is given in Chapter 4.

3.7. Student Model

The student model includes information about an individual students' interactions with the system, pre and post knowledge levels, misconceptions, problem solving behavior and content retention percentage after tutoring or solving the problems. A Bayesian inference network is used to update the student model (Mislevy, 1996; Conati et al, 1997).

A classical approach characterizing how people forget is based on research conducted by Herman Ebbinghaus, and appears in a reprinted form in (Ebbinghaus, 1998). Ebbinghaus' empirical research led him to create a mathematical formula that calculates an approximation of how much may be remembered by an individual in relation to the time from the end of learning: *equation 1*

$$b = \frac{100 * k}{(logt)^c + k} \qquad (1)$$

Where:

- t: the time in minutes counting from one minute before the end of the learning,
- b: the equivalent of the amount remembered from the first learning, and
- c and k: two constants with the following calculated values k=1.84 and c=1.25.

In the student model of ViPS, the Ebbinghaus calculations have been used as the basis for finding out how much tutoring content or content learned from problem solving is retained by the student. After solving the problems related to each misconception or after tutoring the student for that particular misconception, a follow-up test with three questions is given to the student and based on the responses, the students' initial memory state is calculated using *equation 2*.

$$X\% = \frac{b}{100} * RF \qquad (2)$$

Where b is the Ebbinghaus' power function result calculated using equation-1 with t=0. However, equation-2 has a new factor, which is called the Retention Factor (RF). This is used to individualize this equation to the particular circumstances of each student by taking into account his initial Response Quality (RQ) for the follow-up questions. The retention factor is calculated using the initial response quality, which is the number of correct responses to the follow-up questions asked after the completion of problem solving or tutoring for a particular misconception. The initial retention factor is calculated using *equation 3*

$$RF = \frac{100}{RQ} \qquad (3)$$

Once the student is tutored for all the misconceptions he/she might have or after solving all the problems, the student is given a post knowledge test to calculate the final retention factor, post knowledge level, final response quality and to evaluate the status of each misconception. Using the final values of retention factor and response quality, the final memory state of the student is calculated for each misconception using *equation 2*. The difference in initial and final memory states gives us the percentage of tutoring content or content learned during the process of problem solving that is retained by the student. The value of time (t) in *equation 1* would be the time elapsed from the end of tutoring or end of problem solving for that particular misconception to the completion of the post knowledge test. The final retention percentage (X) of each misconception is also used by the system to decide whether to re-tutor the student for that particular misconception or not.

3.8. Tutor Module

The tutor module is responsible for overseeing the process of tutoring the student for the misconceptions he/she might have, and it's also responsible for overseeing the process of problem solving by using the information generated by the student model. A detailed description of the tutor module is provided in Chapter 5.

3.9. Procedural Knowledge

The procedural knowledge model is a probabilistic model that includes information about a student's problem solving behavior. The model is used by the tutor module to keep track of students' progress towards a solution, and intervene with appropriate feedback messages when necessary.

CHAPTER 4: SIMULATION MODULE

The simulation module is responsible for allowing the student to conduct virtual experiments, i.e., to create various pulley setups by creating and connecting components (pulleys, strings and loads) on the work area and to simulate the created setup. One challenge in providing virtual experimentation capability is for the system to be able to determine valid pulley setups based on the components that a student creates, and to enable the student to select and simulate one of the many possible valid setups given a set of components. ViPS employs a Bayesian network (Gertner and VanLehn (2000)) for this purpose.

A Bayesian network is a mathematical model to represent the casual relationships and hence infer probabilistic outcomes in a domain. Since tutoring knowledge is full of uncertainty and characterized by causal relationships and hierarchical structures (Millán & Pérez-de-la-Cruz, 2002; VanLehn, Niu, Siler, & Gertner, 1998), Bayesian networks are increasingly popular in designing and (Gertner and VanLehn (2000)) student models. ViPS uses a Bayesian network to model valid experimental setups that a student may want to create from the components that he/she has selected.

A Bayesian network is a probabilistic graphical model that represents a casual relationship between independent variables and dependent variables via a directed acyclic graph. The nodes of a Bayesian network represent observable variables, unknown parameters or hypotheses. Edges represent conditional dependencies; nodes which are not connected represent variables which are conditionally independent of each other. Each independent node is represented by a pair of probability values and each dependent variable is represented by a probability function that takes as input a particular set of values for the node's parent variables and gives the probability of the variable represented by the node.

Since it is not possible to examine all possible occurrences of specific events exhaustively, probability theory is used to model the real world in Bayesian networks. A probability is about the belief of an event occurring based on the observed occurrences of other events so far. Therefore, belief probabilities can change after receiving more evidence. Before the acquisition of any evidence, the initial probability can be set to any value or be obtained from a small size of sample data. This probability is called a prior probability and needs to be refined with more evidence. After the acquisition of new evidence, the updated probability is called posterior probability.

A full joint probability distribution of all the random variables can be obtained through the products of conditional probabilities of each random variable given all parent nodes in the Bayesian network. Therefore, any probabilistic question about the random variables can be answered by the full joint probability distribution represented in the Bayesian network. The procedure of answering questions is called probabilistic inference.

The inference procedure in a Bayesian network requires prior probabilities of root nodes and conditional probability tables for the non-root nodes. For instance, a Bayesian network for a single fixed pulley setup has two independent variables single fixed pulley and load as shown in figure 4.1. Each of these can be true (the component is present in the work area) or false (the component is absent from the work area).

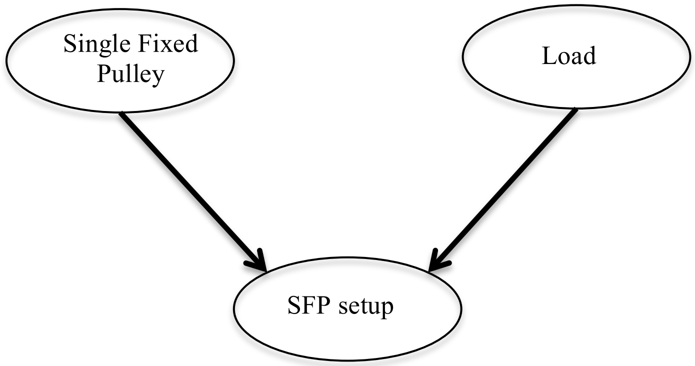


Figure 4.1: A Simulation Setup Bayesian Network Example

Single Fixed Pulley (SFP)	
True	False
0.5	0.5

Table 4.1: Prior probabilities for root node single fixed pulley (SFP)

Load	
True	False
0.5	0.5

Table 4.2: Prior probabilities for root node (Load)

Single Fixed Pulley (SFP)	Load	P (SFP setup | Single Fixed Pulley, Load)
True	True	0.99
True	False	0.85
False	True	0.3
False	False	0.01

Table 4.3: Conditional probability table for non-root node in SFP setup

If a student creates a single fixed pulley setup on the work area, the prior probability of SFP setup is set to true and the probability of single fixed pulley is updated to posterior probability:

$$\mathrm{P(SFP = true \mid SFP\ setup = true)} = \frac{P\ (SFP = true, SFP\ setup = true)}{P\ (SFP\ setup = true)}$$

P (SFP setup = true) can be calculated through P (SFP = true, SFP setup = true) + P (SFP = false, SFP setup = true) = 1, and P (SFP = true, SFP setup = true) can be calculated from full joint probability distribution of the Bayesian network through the summation of other random variables.

$$\mathrm{P\ (SFP = true, SFP\ setup = true)} = \sum_{Load} P\ (SFP = true, SFP\ setup = true, Load)$$

The value in the full joint probability distribution, P (SFP = true, SFP setup = true, Load=true) can be obtained from the products of conditional probabilities of each node in the Bayesian network as: P (SFP=true, SFP setup=true, Load=true) =P (SFP setup = true |SFP = true, Load=true) * P (SFP = true) * P (Load=true).

When a node is a root, it has no parent and its conditional probability becomes the prior probability. All the conditional probabilities and prior probabilities can be found from the tables associated with the Bayesian network. The other values in the full joint probability distribution P (SFP = true, SFP setup = true, Load = false) can be calculated similarly, and the other posterior probability P (SFP = false | SFP setup = true) can also be calculated with a similar method.

A Bayesian network can eliminate the relationship between a random variable and its non-ancestors and provides a precise way to calculate the values in the full joint probability distribution of the random variables, and hence can answer any probabilistic queries about the variables. However, it requires exponential computational time, which is represented by $O(2^n)$ where n is the number of random variables in the network. Therefore, the computational time explodes as number of random variables increase.

The domain knowledge of the simulation model regarding possible or valid pulley setups is represented in the form of Bayesian network as shown in figure 4.2.

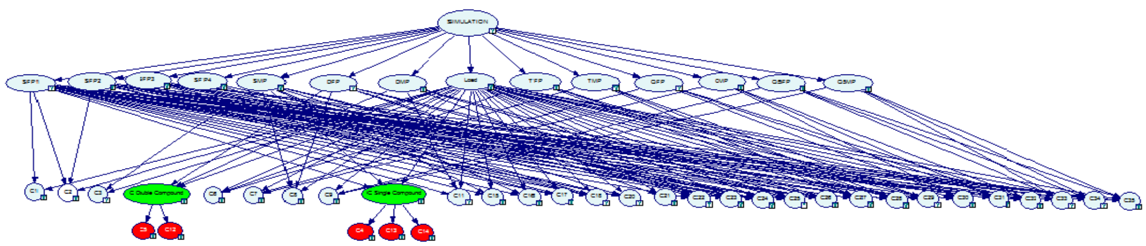


Figure 4.2: Simulation Model Bayesian network

The simulation Bayesian network is used by ViPS to 1) find all possible setups that can be created using components that an individual student has created on the work area, 2) find missing components for creating a setup that the student selects from all possible setups, and 3) generate dynamic hints regarding pulley setups based on student actions.

4.1. Inferring setups from Bayesian network

Based on the components present in the work area, ViPS can generate all the possible setups. The setup inference process for a single compound pulley setup with extra fixed pulley (C5) is shown in figure 4.3 as an example. The initial probabilities of all the components required to create a single compound pulley are set to false in the Bayesian network (figure 4.3(a)).

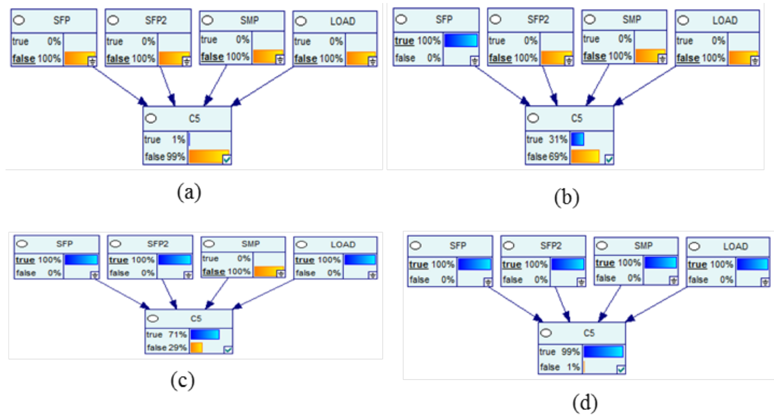


Figure 4.3: Setup inference using Bayesian belief network

As the student creates a fixed pulley in the work area, the evidence is updated in the Bayesian network as shown in figure 4.3(b). The resulting update leads to an increase in the probability for the creation of the setup C5. There is a substantial increase in the probability of the setup C5 when the student in the next step adds a second pulley and a load (probability increases from 31% to 71%) (figure 4.3(c). The probability for C5 increases to 99% upon the addition of a movable pulley to the existing setup (figure 4.5 (d)). This example showcases the inference process in VIPS, if all the required components for a particular setup are present in the work area.

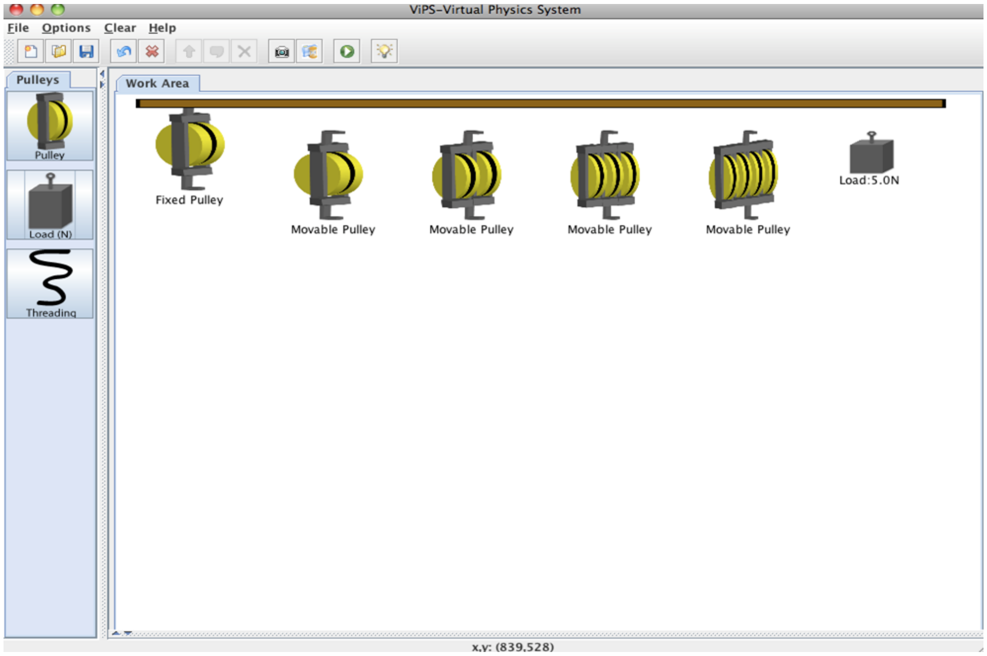


Figure 4.4: Work area with different components

If a student initiates the process of threading with x number of components in the work area (figure 4.4) with which no unique setup is possible, the system displays a list of possible setups based on the probabilities computed by the Bayesian network and ranked using a rank order algorithm. To determine the rank order of all possible setups, the system calculates and uses a complexity factor of each setup. The complexity factor is composed of five attributes.

1) Number of missing components
2) Number of grooves in each pulley
3) Total number of components in the setup
4) Number of times each setup was created (past interaction data)
5) Random selection (if no past interaction data is available)

The student is then asked to choose one of the possible setups (figure 4.5). Once a setup is chosen, the simulation module allows the student to simulate the pulley setup and it produces graphical and textual outputs showing the parameters of the simulation such as energy and work (figure 4.6).

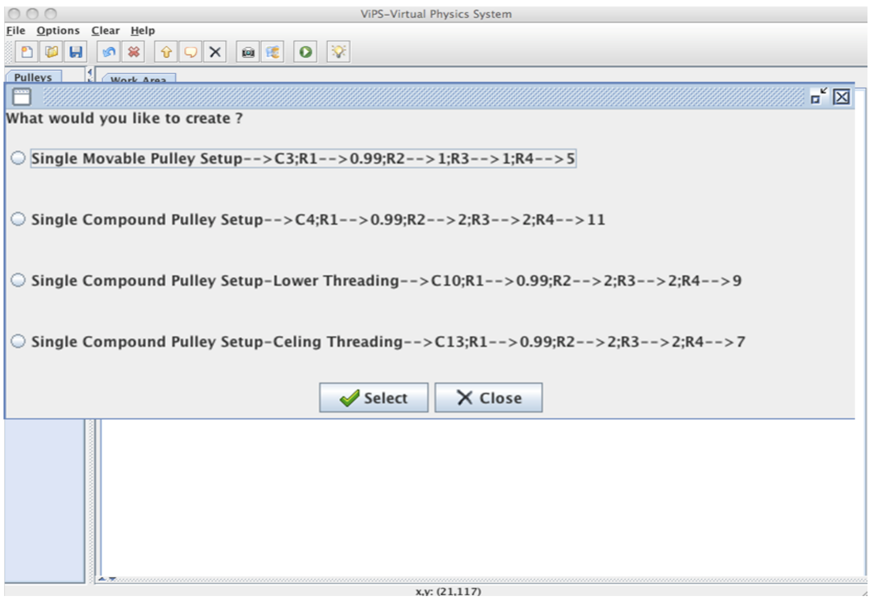


Figure 4.5: Possible setups inferred using Rank Order Algorithm

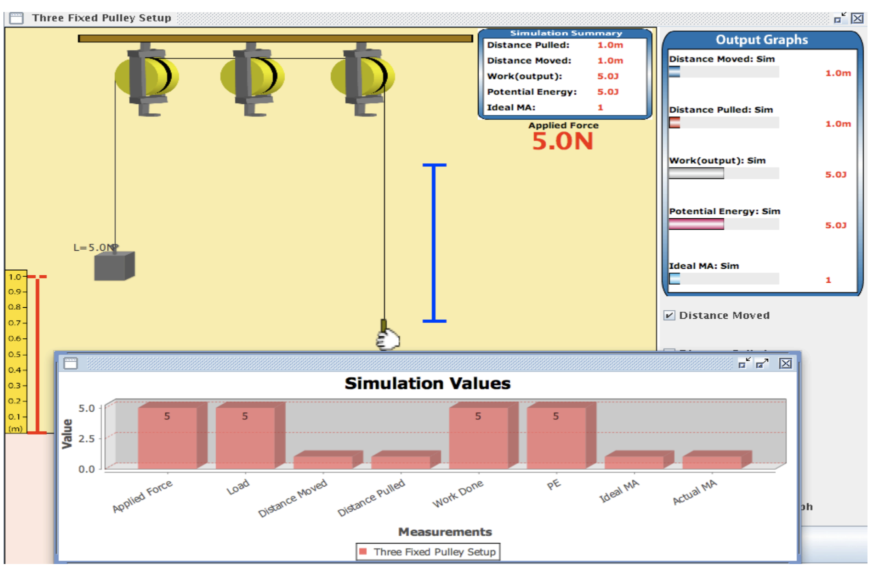


Figure 4.6: Output parameters of the simulation

CHAPTER 5: TUTOR MODULE

The Tutor module of ViPS is based on the cognitive theory of multimedia learning and Vygotsky's theory of learning. The tutor module is responsible for overseeing the process of tutoring the student for the misconceptions he/she might have, and it is also responsible for overseeing the process of problem solving by using the information generated by the student model of ViPS. The process of designing the content for the tutor adheres to the principles stated in the theory of multimedia learning and the feedback generated by the tutor module is based on the Zone of Proximal Development (ZPD) component of Vygotsky's theory of learning.

5.1. Cognitive Theory of Multimedia Learning

The cognitive theory of multimedia learning draws on Paivio's (1986; Clark & Paivio, 1991) dual coding theory, Baddeley's (1992) model of working memory, Sweller's (Chandler & Sweller, 1991) cognitive load theory, Wittrock's (1989) generative theory, and Mayer's (1996) SOI model of meaningful learning. According to these theories, the learner possesses a visual information processing system and a verbal information processing system. Information presented by animations etc. goes into the visual system, whereas information presented as printed or spoken text goes into the verbal system.

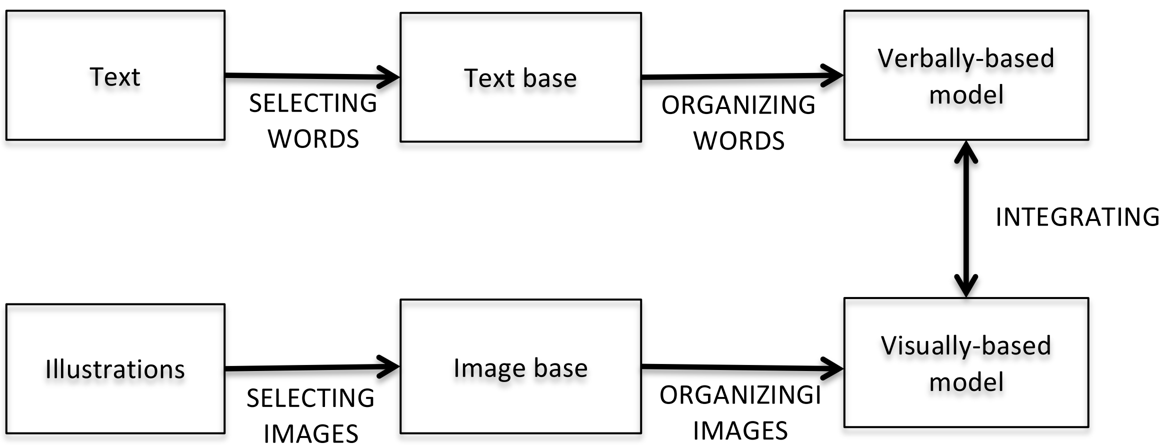


Figure 5.1: A generative model of multimedia learning

Figure 5.1 summarizes some cognitive conditions for the construction of meaningful learning in a multimedia environment used in generative theory (Wittrock, 1989): selecting words and selecting images from the presented material, organizing words and organizing images into coherent mental representations, and integrating the resulting verbal and visual representations with one another.

The cognitive theory of multimedia learning is based on the following assumptions: (a) working memory includes independent auditory and visual working memories (Baddeley, 1992); (b) each working memory store has limited capacity (Chandler & Sweller, 1992); (c) humans have separate systems for representing verbal and non-verbal information (Clark & Paivio, 1991); and (d) meaningful learning occurs when a learner selects relevant information in each store, and organizes the information into a coherent representation in each store (Mayer, 1997). Figure 5.2 depicts the cognitive theory of multimedia learning with these assumptions (Moreno & Mayer, 1999).

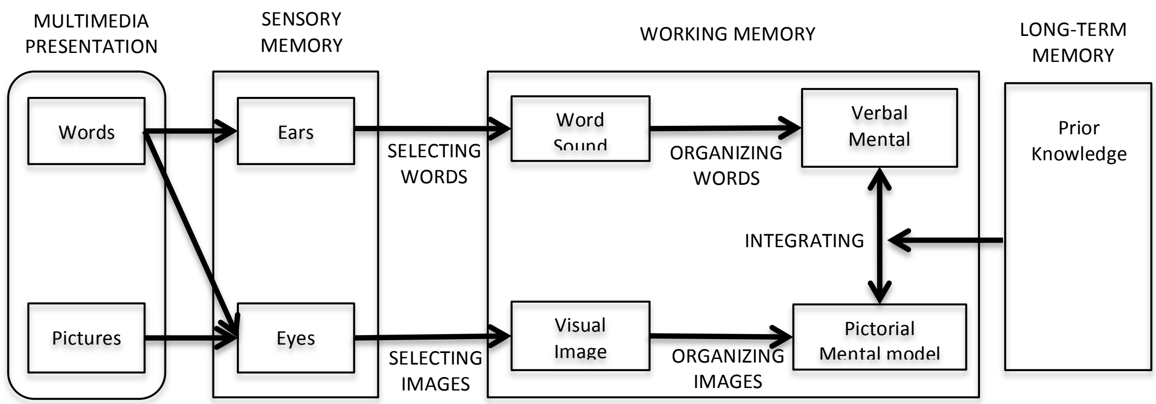


Figure 5.2: Depiction of a cognitive theory of multimedia learning

In multimedia learning, the learner engages in three important cognitive processes. The first cognitive process, selecting, is applied to incoming verbal information to yield a text base and is applied to incoming visual information to yield an image base. The second cognitive process, organizing, is applied to the image base to create a visually-based model of the to-be-explained system. Finally, the third process, integrating, occurs when the learner builds connections between corresponding events in the verbally-based model and the visually-based model. The experiments conducted by Mayer (1997) to evaluate several predictions derived from his theory of multimedia learning concerning (a) whether or not multimedia instruction is effective, (b) when multimedia instruction is effective, and (c) for whom multimedia instruction is effective yielded five major principles of how to use multimedia to help students understand a scientific explanation.

First Principle

Multiple Representation Principle: It is better to present an explanation in words and pictures than solely in words.

The first principle is simply that it is better to present an explanation using two modes of representation rather than one. For example, an experiment conducted by Mayer & Anderson (1991) revealed that students who listened to a narration explaining how a bicycle tire pump works while also viewing a corresponding animation generated twice as many useful solutions to subsequent problem-solving transfer questions than did the students who listened to the same narration without viewing any animation. Similarly, students who read a text containing captioned illustrations placed near the corresponding words generated about 65% more useful solutions on a subsequent problem-solving transfer test than did students who simply read the text (Mayer & Gallini, 1990).

Second Principle

Contiguity Principle: When giving a multimedia explanation, present corresponding words and pictures contiguously rather than separately.

The second principle is that students better understand an explanation when corresponding words and pictures are presented at the same time than when they are separated in time. For example, students who read a text explaining how tire pumps work that included captioned illustrations placed near the text generated about 75% more useful solutions on problem-solving transfer questions than did students who read the same text and illustrations presented on separate pages (Mayer, 1989).

Third Principle

Split-Attention Principle: When giving a multimedia explanation, present words as auditory narration rather than visual on-screen text.

The third principle is that words should be presented auditorily rather than visually. For example, students who viewed an animation depicting the formation of lightning while also listening to a corresponding narration generated approximately 50% more useful solutions on a subsequent problem-solving transfer test than did students who viewed the same animation with corresponding on-screen text consisting the same words as the narration (Mayer, 1991). This results is consistent with the cognitive theory of multimedia learning because the on-screen text and animation can overload the visual information processing system whereas narration is processed in the verbal information processing system and animation in processed in the visual information processing system.

Fourth Principle

Individual Differences Principle: The foregoing principles are more important for low-knowledge than high-knowledge learners, and for high-spatial rather than low-spatial learners.

The fourth principle is that multimedia effects, contiguity effects, and split-attention effects depend on individual differences in the learner. For example, students who lack prior knowledge tended to show stronger multimedia effects and contiguity effects than students who possessed high levels of prior knowledge (Mayer & Gallini, 1991). According to a cognitive theory of multimedia learning, students with high prior knowledge may be able to generate their own mental images while listening to a narration or reading a verbal text so having a contiguous visual presentation is not needed. Additionally, students with high spatial ability are able to hold the visual image in visual working memory and thus are less likely to benefit from contiguous presentation of words and pictures.

Fifth Principle

Coherence Principle: When giving a multimedia explanation, use few rather than many extraneous words and pictures.

The fifth principle is that students learn better form a coherent summary that highlights the relevant words and pictures than from a longer version of the summary. For example, students who read a passage explaining the steps in how lightning forms along with corresponding illustrations generated 50% more useful solutions on a subsequent problem-solving transfer test than did students who read the same information with additional details inserted in the materials(Harp & Mayer, 1997). This result is consistent with the cognitive theory of multimedia learning, in that a shorter presentation primes the learner better to select relevant information and organize it productively.

The graphical interface of ViPS and textual information presented to students during tutoring adhere to principles one, two and five. ViPS does not incorporate auditory narration.

5.2. Vygotskian Theory of Learning

Vygotsky's theory is one of the foundations of constructivism. It asserts three major themes:

1. Social interaction plays a fundamental role in the process of cognitive development. In contrast to Jean Piaget's understanding of child development (in which development necessarily precedes learning), Vygotsky felt that social learning precedes development. He states: "Every function in the child's cultural development appears twice: first, on the social level, and later, on the individual

level; first, between people (interpsychological) and then inside the child (intrapsychological).” (Vygotsky, 1978).

2. The More Knowledgeable Other (MKO). The MKO refers to anyone who has a better understanding or a higher ability level than the learner, with respect to a particular task, process, or concept. The MKO is normally thought of as being a teacher, coach, or older adult, but the MKO could also be peers, a younger person, or even computers.
3. The Zone of Proximal Development (ZPD). The ZPD is the distance or gap between a student’s ability to perform a task under adult guidance and/or with peer collaboration and the student’s ability of solving the problem independently. According to Vygotsky, learning occurrs in this zone.

Tutoring by ViPS occurs in the ZPD because it is initiated only when the student is unable to independently solve a set of problems, and it allows the student to instead solve problems under the system’s guidance.

5.3. Tutoring Process

The tutor’s decision to tutor or not depends on the student’s response to the problems he or she has been given to solve. For every problem, the student has to enter his prediction, actual answer and answer to a follow-up question. Based on these answers (correct: T and wrong: F) the problem is classified into one of the categories as shown in Table 5.1.

Prediction(P)	Answer(A)	Follow-Up(F)	Classification
T	T	T	R+
T	T	F	R-
T	F	T	W+
T	F	F	W
F	T	T	R-
F	T	F	R
F	F	T	W+
F	F	F	W-

Table 5.1: Student's Problem Solving Classification Truth Table

The problem is classified as successfully solved (true), if the outcomes are R+, R-, or R or else it is classified as incorrectly solved (false) (W+, W-, W). The tutor module presents two or three problems per misconception to determine whether a student has that misconception or not. It uses these problem outcomes to decide whether to tutor the student for the current misconception or move on to evaluate the next misconception using another set of three problems. Table 5.2 shows the tutor action truth table. For example, if the student solves the first two problems correctly, then he or she is determined not to have the corresponding misconception, so the tutor will move on to the next misconception (Table 5.2, row 1). If the student solves the first problem correctly but errs in the second one, the tutor will present a third problem and depending on its outcome will either move to the next misconception (Table 5.2, row 2) or start tutoring actions to clear the current misconception (Table 5.2, row 3). Tables 5.1 and 5.2 together illustrate the tutor module decision tree.

Problem 1	Problem 2	Problem 3	Tutor Action
T(R+,R-,R)	T(R+,R-,R)	N/A	Next Misconception
T(R+,R-,R)	F(W+,W-,W)	T(R+,R-,R)	Next Misconceptio
T(R+,R-,R)	F(W+,W-,W)	F(W+,W-,W)	Tutor Action
F(W+,W-,W)	T(R+,R-,R)	T(R+,R-,R)	Next Misconceptio
F(W+,W-,W)	T(R+,R-,R)	F(W+,W-,W)	Tutor Action
F(W+,W-,W)	F(W+,W-,W)	N/A	Tutor Action

Table 5.2: Tutor Action Truth Table

Figure 5.3 shows the tutoring process followed by the tutoring module to clear any misconceptions students might have with respect to pulley systems. The interaction between the tutor module and the student begins with the student attempting a "pre-knowledge test". This test helps the tutor module to find out any misconceptions the students might have about pulley systems. After detecting and recording misconceptions that are present, the tutor module helps the student resolve these misconceptions by encouraging them to solve particular "misconception related" problems. In order to clear every misconception, the tutor module provides three problems (two initial problems are given and then if the misconception is not cleared, a third one). If the student solves the first problem, the system moves ahead to the next problem. Students have four attempts to solve each problem. If for some reason, the problem is not solved within the four attempts, the system assumes that the student doesn't have the necessary knowledge to solve the current problem and moves ahead to the next problem. After the first problem is resolved, the module moves to the next problem where again, the student can make up to four attempts to solve the problem. Failure to solve the problem in four attempts leads to a third problem being posed by the tutor module. If the student is able to successfully answer the first two problems, the tutor module advances ahead to clear other misconceptions the students might have. Students who fail to solve all the three problems posed by the module to clear the misconception are guided by the system to a tutoring session in which the student is tutored for the current misconception. If the student doesn't have any misconceptions about pulley systems, no tutoring will be given to the student. After all the misconceptions are cleared by the problem solving and guidance sessions of the tutor module, the student can exit the tutor module with enhanced knowledge about pulley systems.

5.4. Knowledge Retention

After helping students to clear a particular misconception that the system is currently addressing through problem solving or tutoring, a follow up test is given in order to estimate the knowledge acquired by the students. After the students exits the tutor module, a post knowledge test is given to evaluate the status of all detected misconceptions. The results from the post knowledge test along with the follow up test are used to determine whether a student retained the acquired knowledge through the end of the session or not.

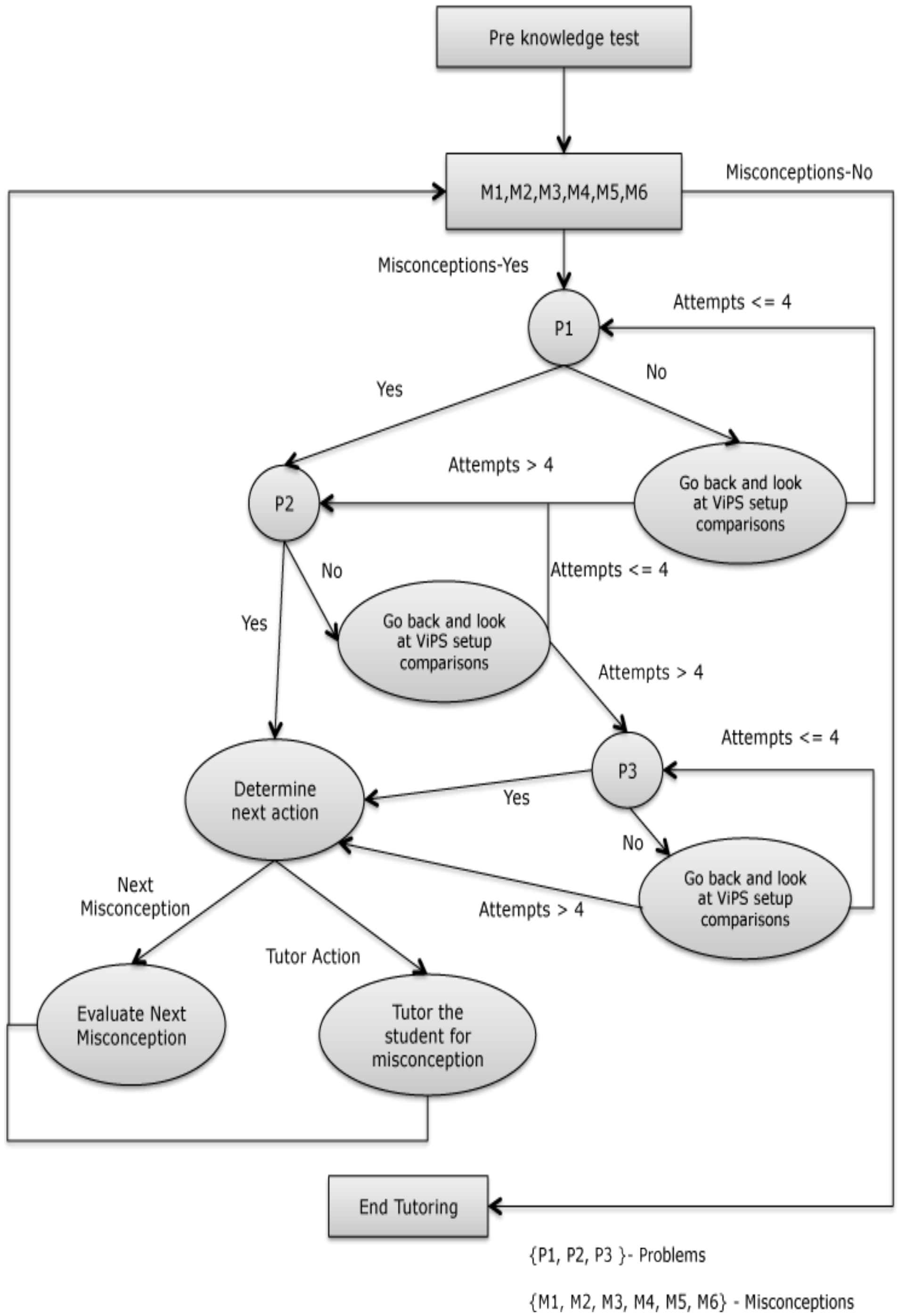


Figure 5.3: Tutoring Process Flowchart

CHAPTER 6: SYSTEM EVALUTION

Empirical studies were conducted at Auburn University, Kansas State University, and middle schools in the Wisconsin school district to evaluate the effectiveness and usability of ViPS. Participants used ViPS and physical pulleys to solve various problems with varying degrees of complexity.

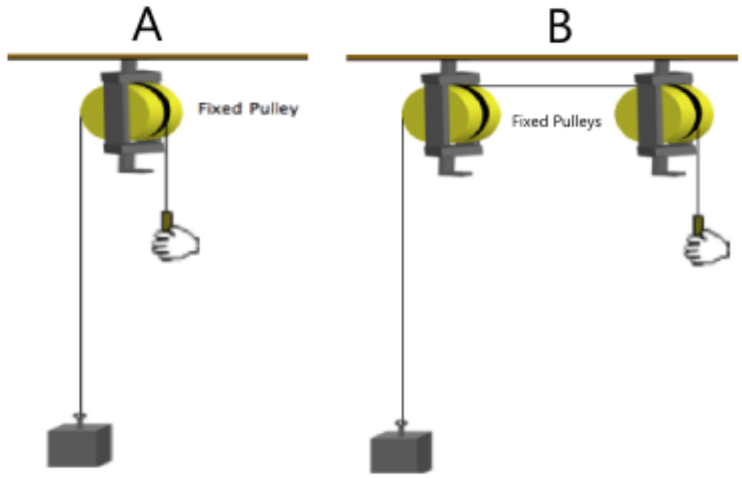
6.1 Subjects

Two hundred student volunteers from Kansas State University, 30 student volunteers from Auburn University, and 36 student volunteers from the Madison (WI) school district participated in the studies conducted to evaluate the effectiveness and usability of ViPS.

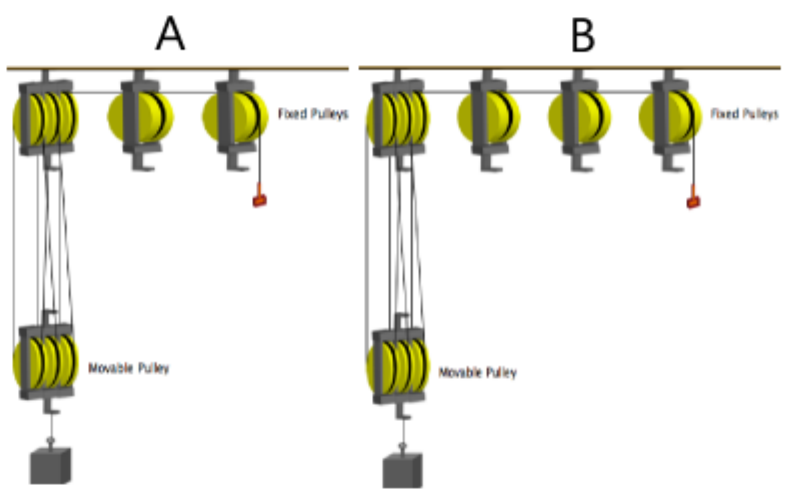
6.2 Experimental Tasks

The subjects were asked to solve various problems in pulley systems using either ViPS or actual pulleys. Sample problems are shown below.

Problem 1: If we ignore friction, which of the following two pulleys systems will require less force (effort) to lift the load to the same height?



Problem 2: Bill lifts a box 1m using pulley system A. Paul lifts an identical box to the same height using pulley system B. What can you tell about the work done, if you ignore the friction?



6.3 Hypothesis

H0$_1$: Students will learn the same using ViPS as physical pulleys.

Ha$_1$: Students will learn better using ViPS than physical pulleys.

H0$_2$: ViPS cannot clear the misconceptions students might have.

Ha_2: ViPS can clear the misconceptions students might have.

$H0_3$: Working with physical pulleys first and then with ViPS or vice-versa will not provide different levels of support for students' conceptual understanding.

Ha_3: Working with physical pulleys first and then with ViPS or vice-versa will provide different levels of support for students' conceptual understanding.

$H0_4$: Students are not satisfied using ViPS.

Ha_4: Students are satisfied using ViPS.

6.4 Experimental Procedure for Evaluation Studies

6.4.1 Virtual Only Condition

- *Consent Process:* At the beginning of the study, the researcher provided an informed consent form to each volunteer, explained it, answered any questions and had him/her sign and return it. The researcher also collected the following demographic data: age, gender and major.
- *Pre-Test:* In a pre-test all the participants were asked to answer 18 questions related to pulley systems individually on-line in order to measure their knowledge of pulley systems and to allow detection of existing misconceptions regarding pulley systems.
- *Problem Solving:* Participants solved problems related to six misconceptions individually using ViPS.
- *Post-Test:* In a post-test, all the participants were asked to answer 18 questions related to pulley systems individually on-line in order to measure their knowledge

of pulley systems and to see whether any existing misconceptions have been cleared.

- *Usability Survey:* All participants were asked to fill out a usability survey individually to measure their overall satisfaction with using ViPS.

6.4.2 Virtual-Physical (VP) and Physical-Virtual (PV) Sequence

- *Consent Process:* At the beginning of the study, the researcher provided an informed consent form to each volunteer, explained it, answered any questions and had him/her sign and return it. The researcher also collected the following demographic data: age, gender and major.
- Pre-*Test:* In a pre-test all the subjects were asked to answer 18 questions related to pulley systems individually on paper or on-line in order to measure their knowledge of pulley systems.
- *Group Assignment:* The experiment was conducted in groups of two (students were grouped randomly) and all the volunteer groups were randomly assigned to either "Physical-Virtual Group (PV)" (in which they worked with actual pulleys to solve the problems followed by ViPS) or "Virtual-Physical Group (VP)" (in which they used ViPS followed by actual pulleys).
- *Problem Solving:* Volunteer pairs solved three problems related to one specific misconception (the more pulleys there are in a pulley system, the easier it is to pull) using either actual pulleys (PV group) or ViPS (VP group).
- *Mid-Test:* In a mid-test that followed the above problem solving session, all the subjects were asked to answer 18 questions related to pulley systems individually

on paper in order to measure their knowledge after solving the problems using either actual pulleys or simulation.

- *Problem Solving:* Volunteer pairs solved three problems related to the same specific misconception using either actual pulleys (VP group) or ViPS (PV group) .
- *Post-Test:* In a post-test all the subjects were asked to answer 18 questions related to pulley systems individually on paper in order to measure their knowledge after solving the problems using both actual pulleys and simulation. The questions used in the three tests were the same but their order was changed in each test.
- *Usability Survey:* All the subjects were asked to fill out a usability survey individually to measure their overall satisfaction of using the ViPS system.

Figure 6.1 summarizes the evaluation study procedure.

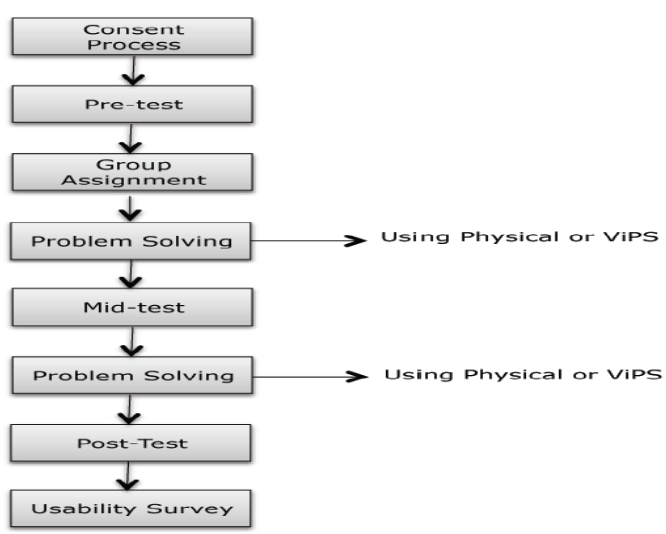


Figure: 6.1: Experimental Procedure of VP-PV Evaluation Study

The design of this evaluation study in shown in figure 6.2.

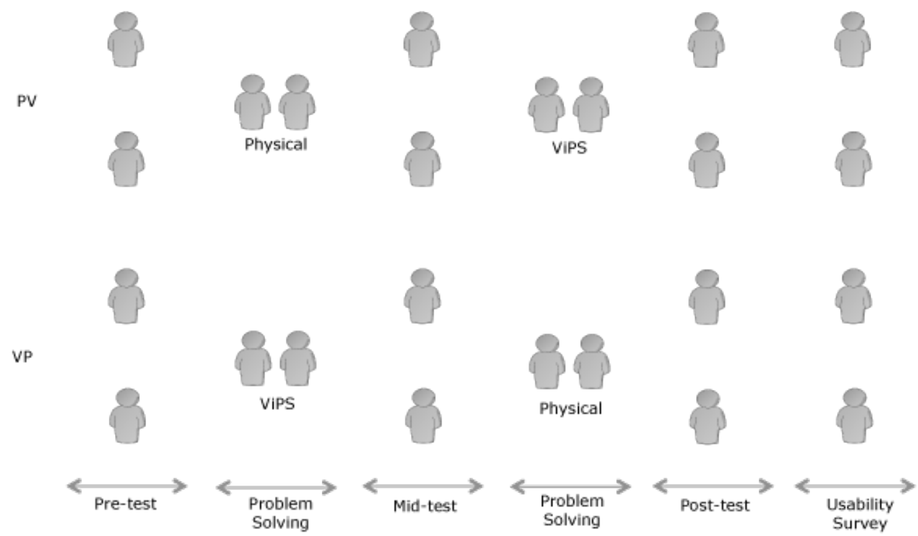


Figure 6.2: Experimental Setup of VP-PV Evaluation Study

6.5 Experimental Procedure for Usability Study

- *Consent Process:* At the beginning of the study, the researcher provided an informed consent form to each volunteer, explained it, answered any questions and had him/her sign and return it. The researcher also collected the following demographic data: age, gender and major.
- *Perform Tasks:* Using ViPS, the volunteers completed a set of representative tasks presented to them in as efficient and timely manner as possible, and provided feedback regarding the user interface and usability of ViPS.
- *System Usability Questionnaire:* The volunteers completed a system usability questionnaire which was used to evaluate different usability aspects of the system.

6.6 Preliminary Usability Study

A preliminary usability study was conducted to evaluate the initial design of ViPS. The objectives of this study were 1) to determine design inconsistences and usability problem areas within the user interface and content areas of ViPS (potential sources of error could include a) navigation errors, b) presentation errors, and c) control usage problems), 2) to test ViPS under controlled test conditions with representative users, and 3) to establish baseline user performance and user-satisfaction levels of the user interface for future usability evaluations.

6.6.1 Methods

6.6.1.1 Participants

The number of students who participated in the usability testing were 36, and they were in the age group of 10- 14 years. The participants completed a set of representative tasks presented to them in as efficient and timely manner as possible using ViPS, and provided feedback regarding the user interface and usability of the system. The students who participated consisted of three groups: a) 13 6th graders, b) 13 7th graders, and, c) 10 8th graders, all from the Madison (WI) school district.

6.6.1.2 Data Collected

The data collected during the process of usability testing include: 1) observer notes, 2) screen recordings (recording the computer screen of the participant), 3) audio tapes of the discussion, 4) feedback survey, and 5) system log files.

6.6.1.3 Procedure

The students were given a demo of the software. All the students were randomly divided into groups of two for the purpose of testing and each group was observed by two people (facilitator and observer). The facilitator interacted with the students and the observer took the notes. Students were given a set of tasks to complete using ViPS, during which the observer made notes of user behavior, user comments, and all the important conversations that happened between the group members related to system actions. The facilitator instructed the participants to 'think aloud' so that a verbal record of their interaction with ViPS could be captured in audio tapes. The facilitator also informed the groups that their computer screens would be recorded during the process of testing. Once the students completed their assigned tasks, they were asked to take a feedback survey. Each group was provided with instructions on how to complete the survey and the surveys were anonymous.

The survey asked seven attribute-based questions scored on a Likert-type scale, which cross-referenced the attributes of usefulness, easiness, and satisfaction. Participants responded to items using a five-point scale as follows:

1=strongly disagree, 2= disagree, 3=neither agree nor disagree, 4=agree, and
5= strongly agree.

The responses were coded and analyzed using SPSS. Usability research questions addressed the attributes of satisfaction, easiness, comfort, and liking (independent variables) as predictors of overall satisfaction (dependent variable) of use. It was assumed that the survey scale was an equal interval scale. The reliability and validity of the survey scale could not be determined in this context as each attribute had less than two questions.

6.6.2 Results

A standard regression analysis was conducted to test the individual effect of easiness, liking and comfort of using the software on overall satisfaction. In analysis 1, I found a strong correlation between the easiness and overall satisfaction attributes with R=0.682, $F(1,35)=29.6$, $p<0.001$. The overall satisfaction (dependent variable) changed by 0.643 ($\beta=0.643$) for unit change in easiness (figure 6.3).

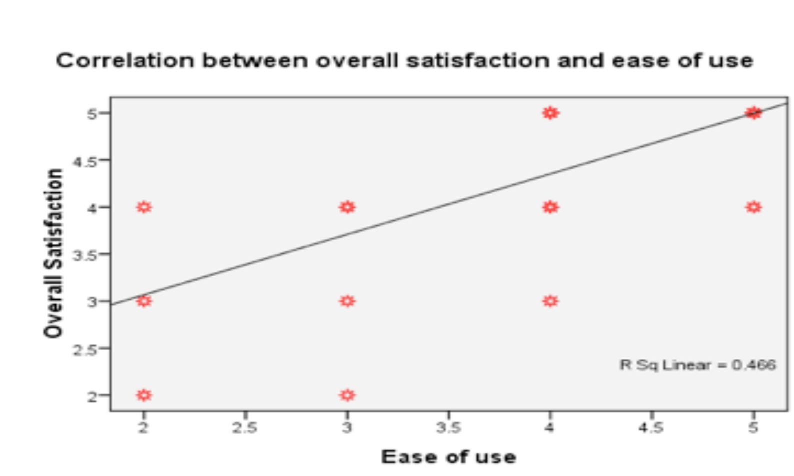


Figure 6.3: Ease of use Vs. Overall Satisfaction scatter plot

In analysis 2, I found a strong positive correlation between liking and overall satisfaction with R=0.861, $F(1,35)=97.103$, $p<0.001$. The overall satisfaction (dependent variable) changed by 0.765 ($\beta=0.765$) for unit change in liking (figure 6.4).

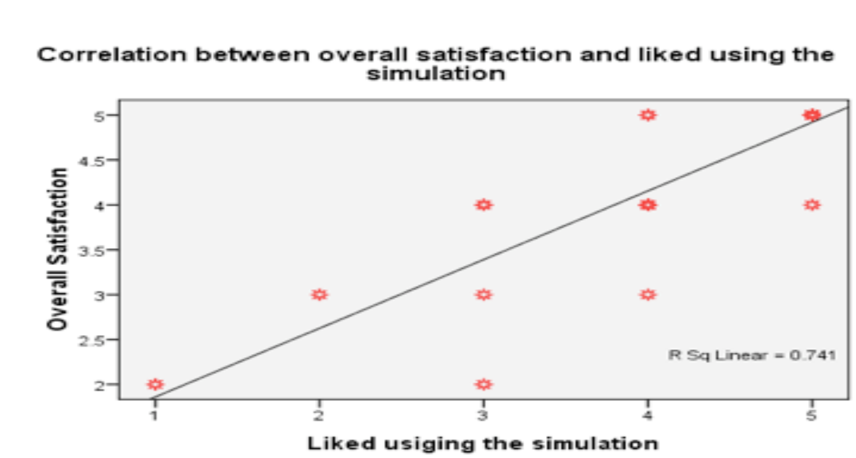


Figure 6.4: Liking vs. Overall Satisfaction scatter plot

In analysis 3, I found a strong positive correlation between comfort and overall satisfaction with R=0.642, $F(1,35)=23.82$, $p<0.001$. The overall satisfaction (dependent variable) changed by 0.615 ($\beta=0.615$) for unit change in comfort (figure 6.5).

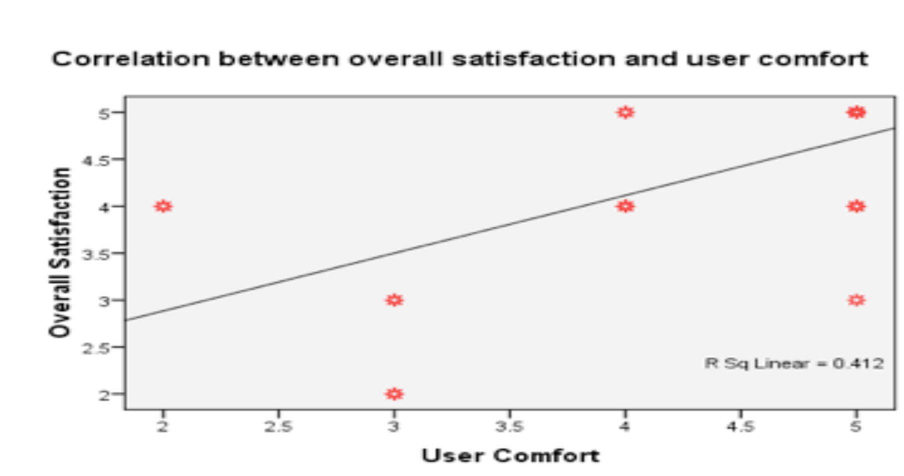


Figure 6.5: User Comfort vs. Overall Satisfaction scatter plot

A hierarchical linear regression was conducted to analyze the effect of independent variables of easiness, liking and user comfort on the dependent variable of overall satisfaction of using the software. In Model 1, easiness alone accounts for 47% (0.466) of the variance in predicting the overall satisfaction with R=0.682. In Model 2, liking accounts for 28% (0.280) of the variance in overall satisfaction after controlling for easiness with R=0.863, R^2 change=0.280. In Model 3, user comfort is not a good predictor of overall satisfaction above and beyond easiness and liking as the variance is not statistically significant ($p=0.68$). Together, the value of R^2 (0.771) indicates that approximately 77% of the variance in predicting overall satisfaction can be accounted for by its linear relationship with easiness, liking and user comfort. Table 6.1 displays the variable means and standard deviations.

	N	Mean	SD
Easiness	36	4.06	0.924
Liking	36	4.32	0.980
User Comfort	36	4.44	0.909
Overall Satisfaction	36	4.39	0.871

Table 6.1: Means and Standard Deviation for Independent and dependent variables

6.6.3 Discussion

The regression analysis indicates that easiness and liking are statistically significant predictors of overall satisfaction and that easiness predicts overall satisfaction better than liking. User comfort was not a statistically significant predictor of overall satisfaction after controlling for both easiness and liking. As the subjects used for this testing were middle school students, we assume that the effect of user comfort on overall satisfaction is low because of more access to and experience with computers. These days students use computers from an early age, so they get comfortable with computers.

This study was used to identify any inconsistences and usability problem areas within the user interface and content areas. It was also used to assess whether usability goals regarding an effective, efficient and well-received user interface were achieved. The results indicate that the students found the ViPS interface to be usable.

6.6.4 Limitations

This study was limited by the relatively small sample size.

6.7 Evaluation Studies

6.7.1 Statistical Test Selection

Statistical tests are of two types: parametric and non-parametric. Each parametric test depends on several assumptions, such as the data follow the normal distribution, the sample size is within a specified range, and there are no outliers in the data. When these assumptions are met, a parametric test is more powerful than its corresponding non-parametric test. Non-parametric methods do not depend on the normality assumption, wok quite well for small samples, and are robust to outliers.

T-test is suitable for smaller sizes (e.g. <30). For polytomous independents (i.e. if the samples are subdivided into many distinct parts) the analysis of variance, ANOVA, tests are more suitable.

Therefore, before I could finalize which statistical tests were most suitable to evaluate the students' performance, I needed to analyze the data to see whether it satisfied the normality assumption and that no outlier properties existed.

I used a Q-Q plot of residuals[1] to test for normality. The Q-Q is a plot of residuals in sorted order (Y-axis) against the value those residuals should have if the distribution of the residuals were normal; i.e., it shows the observation on the X-axis plotted against the expected normal scores (Z-scores, known as quintiles) on the Y-axis. The line shows the ideal normal distribution with mean and standard-deviation of the sample. If the points roughly follow the line, then the sample has normal distribution. I used box plots to identify the outliers, i.e., data points which are numerically distant from the rest of the data. In a box plot, the outliers are indicated using circles.

6.7.2 Study-1 (with Virtual Only group) and Study-2 (PV and VP groups)

6.7.2.1 Test for Normality

Figures 6.6 and 6.7 show the Q-Q plot of residuals for the pretest and posttest scores of the Virtual Only group respectively. The points on the Q-Q plots of residuals lie

[1] The residual of a sample is the difference between the sample and the observed sample mean.

nearly on the straight line, which indicates that both the pretest and posttest scores follow normal distribution.

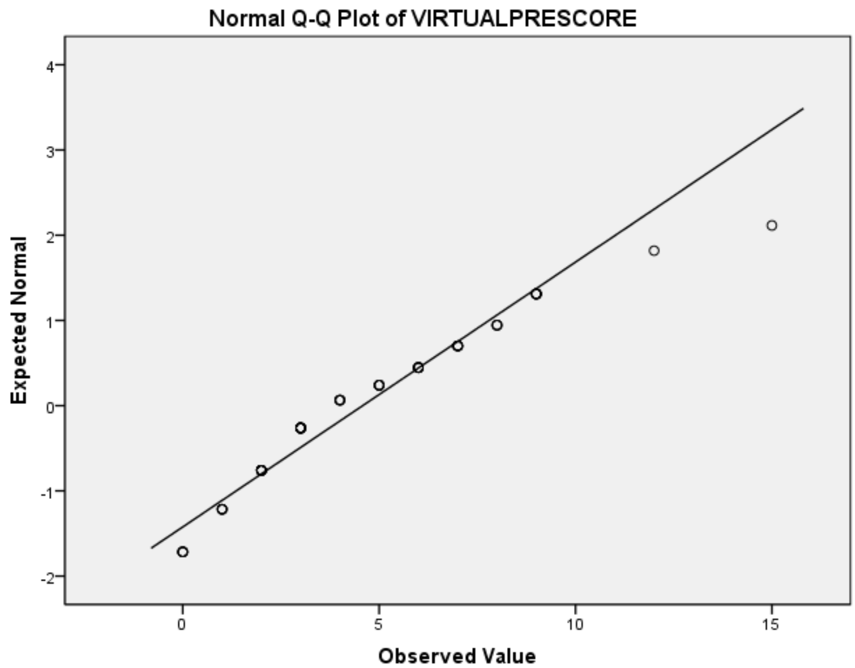


Figure 6.6: Q-Q Plot of Residuals (Virtual Only –Pre Test Scores)

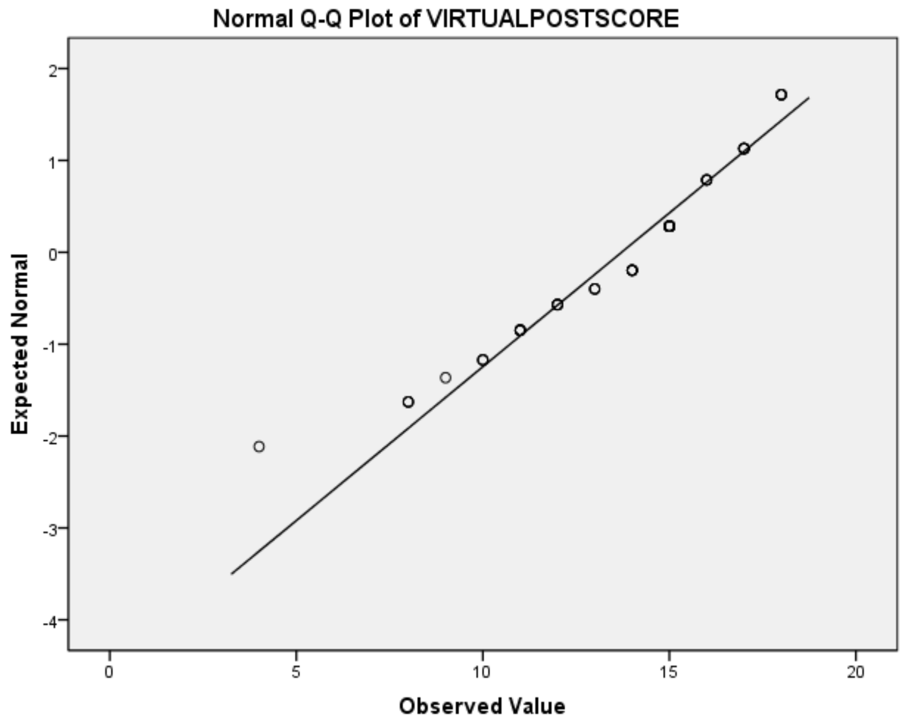


Figure 6.7: Q-Q Plot of Residuals (Virtual Only –Post Test Scores)

6.7.2.2 Outliers

The box plots for the pretest scores and posttest scores of Virtual-Only group are given in Figure 6.7. There are two circles in Figure 6.8, which indicate that there is one outlier each in pretest and posttest scores. These outliers are excluded from our analysis.

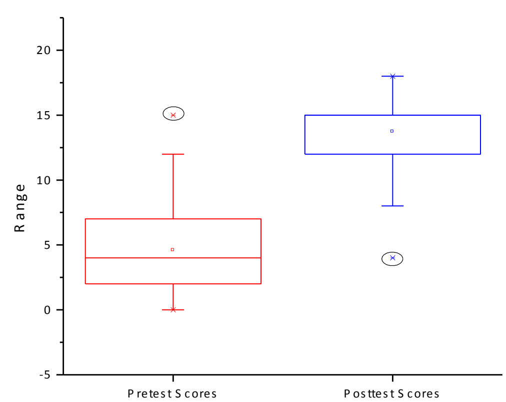


Figure 6.8: Box plot (Pre and Post test scores for Virtual Only group)

The box plots in Figure 6.9 and 6.10 show the pretest, midtest and posttest scores of Virtual-Physical group (VP) and Physical-Virtual group (PV). There are no circles in the Figures 6.8 and 6.9, which indicate that there are no outliers in test scores of VP and PV groups.

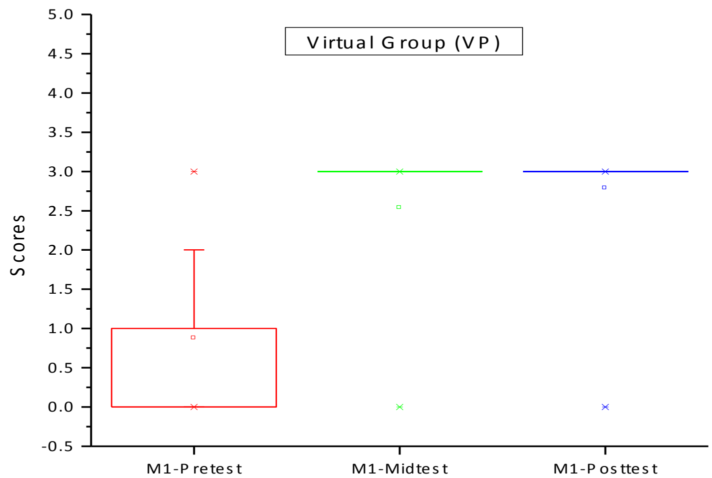


Figure 6.9: Box plot (Virtual group (VP) test scores)

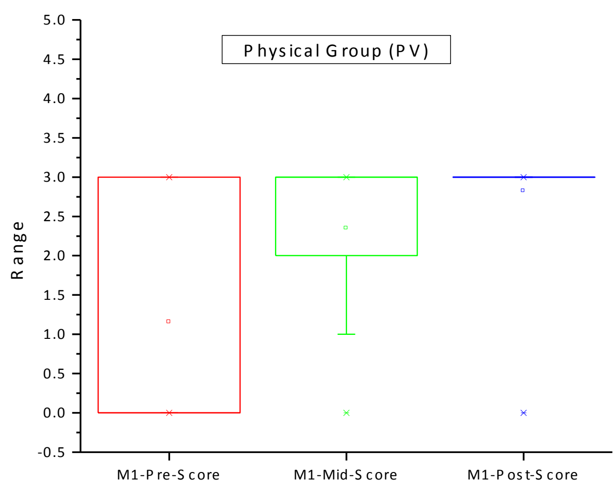


Figure 6.10: Box plot (Physical group (PV) test scores)

6.7.2.3 Statistical Test Determination for Study-1(with Virtual Only group) and Study-2 (Virtual-Physical and Physical-Virtual Sequence)

I used the paired-sample t-Test to compare the pretest to posttest scores of Virtual Only group. The t-Test depends on several assumptions:

- If the sample size is less than 15, then the data for the t-test should be strictly normal.
- If the sample size is between 15 and 40, then the data may be partially normal, but is should not contain outliers.
- When sample size is more than 40, then the data may be markedly skewed.
- Our sample size was 57 for virtual only group, and both pretest and posttest scores followed normal distribution, and the outliers were excluded.

I used repeated measures ANOVA to compare the test scores across the Virtual-Physical (VP) and Physical-Virtual (PV) groups. The repeated measures ANOVA depends on the following assumption.

- The design is based on the assumption of Sphericity, which means that the variance of the population difference scores for any two conditions should be the same as the variance of the population difference scores for any other two conditions. But, this condition is only relevant to the one-way repeated measures ANOVA and in other cases this assumption is commonly violated.

6.7.2.4 Can students learn better using ViPS than physical pulleys? (Hypothesis 1)

A paired-sample t-test was performed on the pre-to-post test scores of the students in Virtual-Only group to evaluate their learning gain (see Table 6.2 for Mean and SD values) after solving problems using ViPS. There was a difference in pre-to-post test scores with statistical significance (see Table 6.3 for t and p values).

	N	Mean	SD	SEM
Pretest-Score	57	4.57	3.21	0.42
Posttest-Score	57	13.71	2.98	0.39
Difference		-9.14		

Table 6.2: Descriptive statistics for paired-sample t-test for Research Hypothesis 1

t Statistic	DF	P
-17.66	56	0.001

Table 6.3: t-statistics of paired-sample t-test for Research Hypothesis 1

The box plot in Figure 6.8 shows the pretest and posttest scores of 57 students in Virtual-Only group. The boxes contain 50% of the data points, and the line between lower border and box contains 25% of the data points. The small square in the plot indicates the mean value and the horizontal line in the middle of the box indicates the median value. The plot indicates that the mean of the posttest score is higher than pretest mean score. The scores went up by 300% from an average score of 4.57 in pretest to 13.71 in posttest. The above results indicate that there was a statistically significant positive effect of ViPS on student learning about pulleys.

A repeated measures analysis of variance test was performed on pretest to midtest scores of both Virtual Group (VP) and Physical Group (PV) (they solved problems related to only one misconception) to compare the learning gain from ViPS with the learning gain from physical pulleys. Results showed that the learning gain (see Figure 6.11) was higher (see Table 6.4 for test means) in VP group who used ViPS first to solve the problems with a statistically significant p value (F (1,156) =4.54, p=0.035, η^2=0.28, and power=0.563). This shows that students who used ViPS to solve the problems gained more knowledge than students who used physical manipulatives.

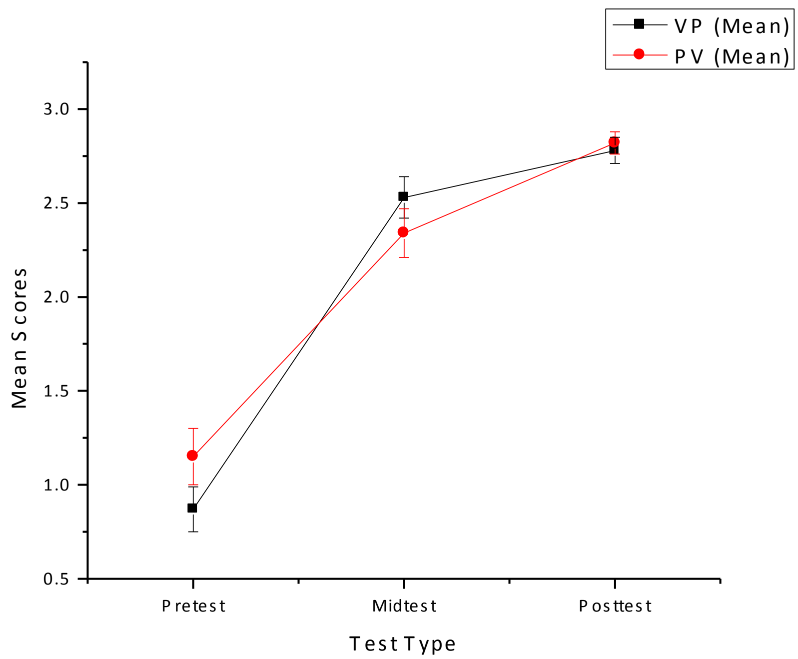


Figure 6.11: Mean Pre, Mid, and Post test scores for Virtual (VP) and Physical (PV) groups

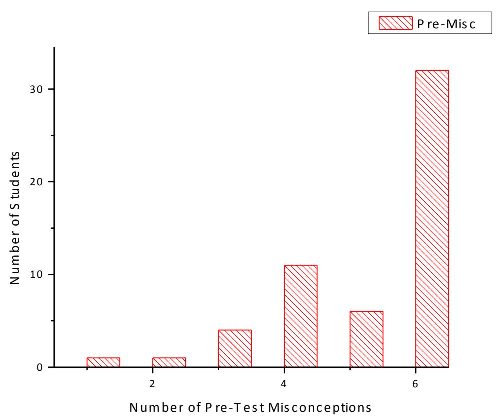
Test Type	Virtual-Physical (VP)			Physical-Virtual (PV)		
	Mean	SD	Std Error	Mean	SD	Std Error
Pretest	0.87	1.15	0.12	1.15	1.35	0.15
Midtest	2.53	1.03	0.11	2.34	1.16	0.13
PostTest	2.78	0.63	0.07	2.82	0.57	0.06

Table 6.4: Descriptive statistics for Pre, Mid, and Post tests in Virtual (VP) and Physical (PV) groups

Decision: Reject $H0_1$ in favor of Ha_1. Thus I have sufficient statistical evidence to conclude that students can learn better using ViPS than physical pulleys.

6.7.2.5 Can ViPS clear the misconceptions students might have? (Hypothesis 2)

A paired-sample t-test was conducted to compare the number of misconceptions in pretest to posttest (see Figure 6.12 for pre and post misconceptions values) in Virtual-Only group. There was a significant difference in number of misconceptions from pretest to posttest with statistical significance (see Table 6.6 for t and p values). On an average, each student exhibited five misconceptions (see Table 6.5 for N, Mean, and SD) after pretest and two misconceptions after posttest. The number of misconceptions decreased significantly after working with ViPS.



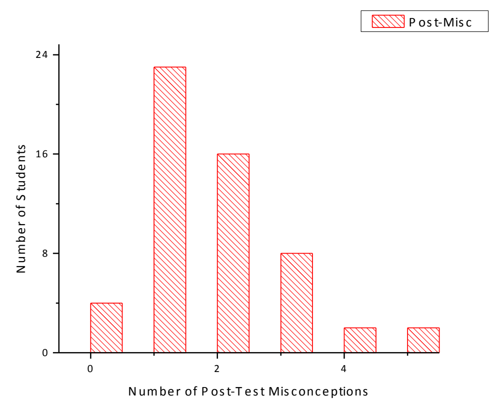


Figure 6.12: Histograms showing number of misconceptions after pretest and posttest

	N	Mean	SD	SEM
Pretest-Misc	55	5.10	1.24	0.16
Posttest-Misc	55	1.76	1.13	0.15
Difference		3.34		

Table 6.5: Descriptive Statistics of Paired Samples t-test for Research Hypothesis 2

t Statistic	DF	P
16.6	54	0.001

Table 6.6: t-statistics of paired-sample t-test for Research Hypothesis 2

The above results indicate that ViPS helped students in learning about pulley systems and also cleared most of their misconceptions about pulley systems.

Decision: Reject $H0_2$ in favor of Ha_2 since p-value < α (α=0.05). Thus we have sufficient statistical evidence to conclude that ViPS can clear misconceptions student might have.

6.7.2.6 Does working with physical pulleys first and then with ViPS or vice-versa provide different levels of supports for students conceptual understanding? (Hypothesis 3)

A repeated measures analysis of variance test was performed on pretest to midtest scores of both the groups (VP and PV- they solved problems related to only one misconception) to compare the learning gains. Results showed that the learning gain (see Figure 6.8) was higher in the VP group who used ViPS to solve the problems with a statistically significant p value ($F(1,156) = 4.54$, $p=0.035$, $\eta^2=0.28$, and power=0.563)). Similarly, repeated measures ANOVA test was performed on midtest to posttest scores and we found that the students in PV group who used ViPS to solve the problems have higher learning gain (see Figure 6.8), but the gain was not statistically significant $F(1,156) = 2.24$, $p=0.137$, $\eta^2=0.014$, and power=0.319).

Repeated measures ANOVA was also performed on pretest to posttest scores of both the groups to see the effect of Virtual-Physical or Physical-Virtual sequence of experimental conditions on learning gain or conceptual understanding of the students. I found that the pretest to posttest increase in scores was significant in both the groups (see Table 6.4 for test means) but there was no difference in learning gain between the groups. This shows that the sequence of experimental conditions (Virtual-Physical (VP) or

Physical-Virtual (PV)) has no effect on the students' conceptual understanding, i.e., both sequences improve students' conceptual understanding by similar levels.

Decision: $H0_3$ cannot be rejected in favor of Ha_3 since p-value $(0.137) > \alpha$ $(\alpha=0.05)$. Thus we do not have sufficient statistical evidence to conclude that the sequence of experimental conditions of learning by conducting physical experiments with pulleys first and then virtual experiments with ViPS or vice-versa can provide different levels of support for students' conceptual understanding of pulleys.

6.7.2.6 How satisfied are the students using ViPS? (Hypothesis 4)

A Post-Study System Usability Questionnaire (PSSUQ) (Lewis 1995) was used to assess the satisfaction of users after interacting with ViPS. PSSUQ consists of 19 items aimed to address the following usability characteristics: quick completion of work, ease of learning, high-quality documentation and online information. An exploratory principal factor analysis of the PSSUQ data indicated that the overall scale could be defined by three subscales: System Usefulness (SYSUSE), Information Quality (INFOQUAL), and Interface Quality (INTERQUAL). These three factors account for 87% of the validity (Fruhling, A & Lee, S., 2005). Questionnaire items 1-8 refer to SYSUSE, items 9-15 refer to INFOQUAL, and items 16-18 refer to INTERQUAL. Items in the questionnaire are 7-point Likert scales, anchored at the end points with the terms "Strongly agree" for 1 and "Strongly disagree" for 7. For this study, we used a slightly modified scale with users scoring on 6-point Likert scales, anchored at the end points with terms "Strongly disagree" for 1 and "Strongly agree for 6 to be consistent with the initial usability study. We assumed that the questionnaire scale was an equal interval scale.

Reliability. Reliability values reported by the authors (Fruhling, A and Lee, S. (2005)) of PSSUQ are the following: reliability for the overall summative scale (OVERALL) was 0.97 and reliability ranged for 0.91 to 0.96 for the three subscales (SYSUSE=0.96, INFOQUAL=0.91 and INTERQUAL=0.91). Therefore, the overall scale and the three subscales have excellent reliability.

A standard regression analysis was conducted to test the individual effect of SYSUSE, INFOQUAL and INTERQUAL on overall satisfaction of using ViPS. In analysis 1, I found a strong positive correlation between SYSUSE and OVERALL (overall satisfaction) with attributes R=0.885, R^2=0.783, F(1,195) =701.30, p<0.001. The overall satisfaction (dependent variable) changed by 0.885 (Standardized Beta=.885) for unit change in SYSUSE (Figure 6.13).

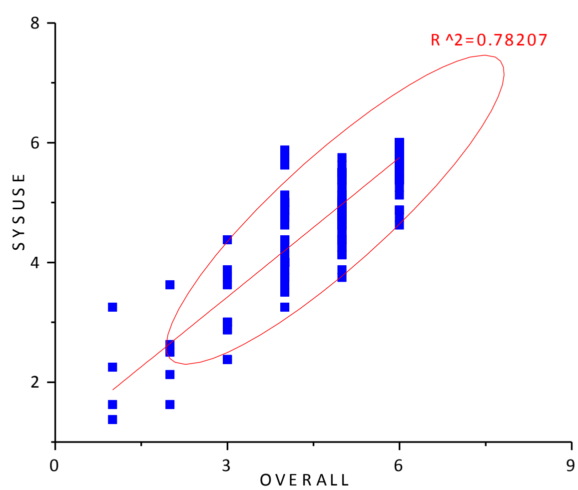


Figure 6.13: Correlation between SYSUSE and OVERALL satisfaction

In analysis 2, I found a strong positive correlation between INTERQUAL and OVERALL satisfaction with R=0.882, R^2=0.778, F(1,195) = 679.27, p<0.001. The overall satisfaction (dependent variable) changed by 0.882 (Standardized Beta=0.882) for unit change in INTERQUAL (Figure 6.14).

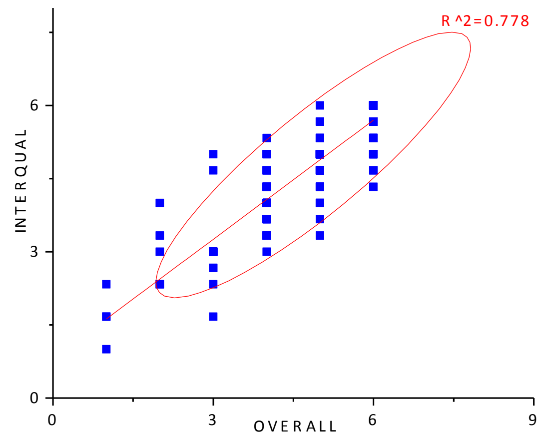


Figure 6.14: Correlation between INTERQUAL and OVERALL satisfaction

In analysis 3, I found a strong positive correlation between INFOQUAL and OVERALL satisfaction with R=0.820, R^2=0.673, F(1,195) = 399.574, p<0.001. The overall satisfaction (dependent variable) changed by 0.933 (Standardized Beta=0.820) for unit change in INFOQUAL (Figure 6.15).

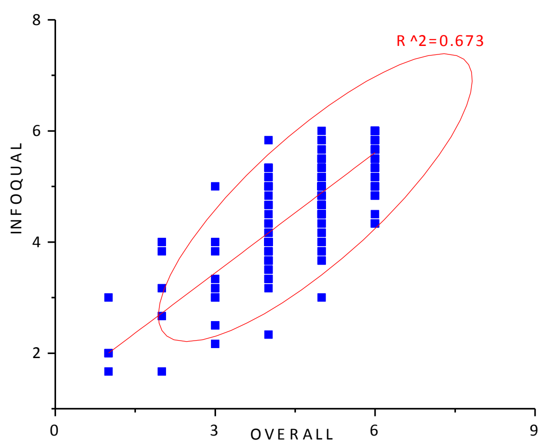


Figure 6.15: Correlation between INFOQUAL and OVERALL satisfaction

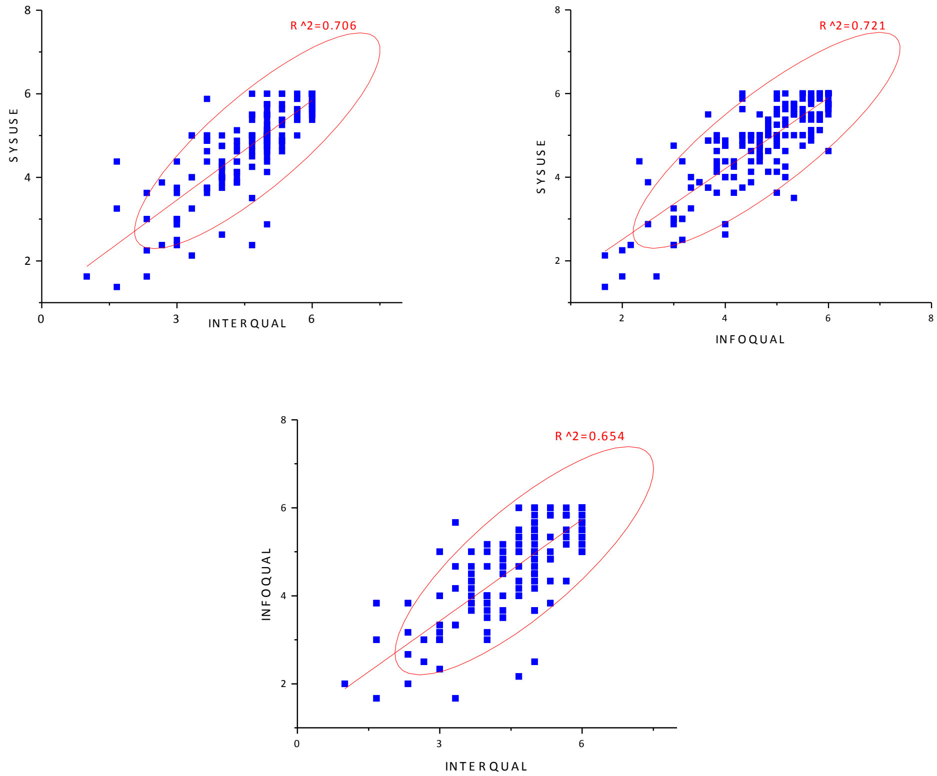


Figure 6.16: Correlations between INTERQUAL, INFOQUAL, and SYSUSE

In analysis 4, I found strong positive correlations between SYSUSE, INTERQUAL, and INFOQUAL independent variables with statistical significance (Figure 6.16). Table 6.7 shows the statistical attributes for the correlations.

CORRELATION	**R**	**F**	**p**	**Std Beta**
SYSUSE-INTERQUAL	0.840	466.471	0.001	0.840
SYSUSE-INFOQUAL	0.849	501.081	0.001	0.849
INFOQUAL-INTERQUAL	0.809	366.57	0.001	0.809

Table 6.7: Statistical attributes for correlations between INTERQUAL, INFOQUAL and SYSUSE independent variables

A hierarchical linear regression was conducted to analyze the effect of independent variables SYSUSE, INTERQUAL, and INFOQUAL on the dependent variable OVERALL satisfaction of using the software. In Model 1, SYSUSE alone accounts for 78% (R^2=0.783) of the variance in predicting the OVERALL satisfaction with R=0.885, p≤0.001. In Model 2, INTERQUAL accounts for 6.5% of the variance in OVERALL satisfaction after controlling for SYSUSE with R=0.921, R^2change=0.065, p≤0.001. In Model 3, INFOQUAL is not a good predictor of OVERALL satisfaction of using the software above and beyond SYSUSE and INTERQUAL as the variance is not statistically significant with R=0.922, R^2change=0.002,p=0.082. Together, the value of R^2 (0.848) indicates that approximately 85% of the variance in predicting OVERALL satisfaction can be accounted by its linear relationship with SYSUSE, INTERQUAL and INFOQUAL. Table 6.2 displays the variable means and standard deviations.

	N	Mean	SD
SYSUSE	196	4.87	1.18
INTERQUAL	196	4.88	1.04
INFOQUAL	196	4.78	1.09
OVERALLSAT	196	4.80	1.04

Table 6.8: Means and Standard deviations of independent and dependent variables

The overall satisfaction of the users can be seen in Figure 6.17. Out of all the participants, 88% provided an overall satisfaction rating of more than 4 with an average of 5.21. Almost 95% of the participants were satisfied with the interface quality and system usefulness.

I reviewed the comments given by participants as part of the usability survey. I found that most of the users felt that the system was easy to use, easy to learn, and that it helped them clear the misconceptions. Below is a verbatim comment given by one of the participants:

> *"I think this system helped me to understand the functions of pulleys better, I wish this had been around and available to me while I was learning pulleys in physics. I think this system will help me remember information about pulleys more clearly because of the hands on and visual aspect it provides. However, I feel like it would be easy to use this system to figure out and understand more clearly. This software is intelligently designed and very user-friendly I think it will be used well to increase the understanding of physics concepts."*

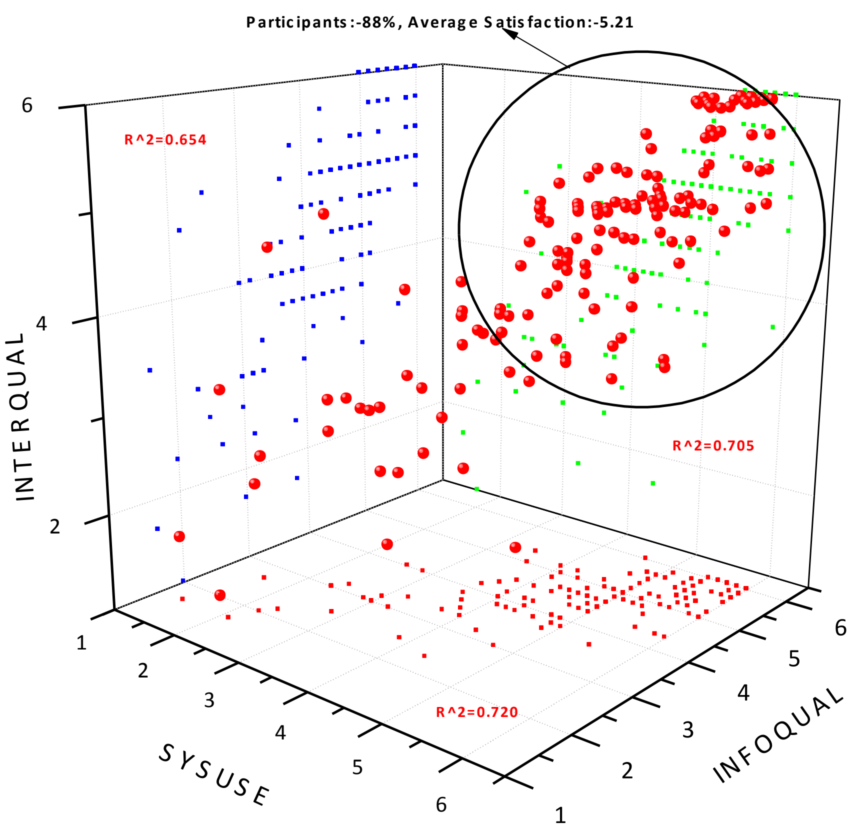


Figure 6.17: Overall satisfaction of the participants after using ViPS

The above results and user comments demonstrate that most of the students who worked with the ViPS system liked it.

Decision: Reject $H0_4$ in favor of Ha_4. We have sufficient evidence to conclude that most students are satisfied after using ViPS.

6.8 Log File Analysis

After demonstrating that ViPS is effective in helping students learn and is also well perceived by students, it is important to understand in more detail the features of the interaction between the students and the system that are highly correlated with learning. The interactions between ViPS and the students have been comprehensively logged. From these log files, several features such as the number of simulations created and time spent working with ViPS were extracted and compared with learning gain using linear regression. Only data from college students at Auburn and Kansas State are included in these analyses.

6.8.1 Pre-test scores

Linear regression found a significant negative correlation (see Table 6.9) between pre-test score and learning gain in Virtual Only, Virtual (VP), and Physical (PV) groups. It is not surprising that these correlations are strong as many of the students have low pre-test scores.

	N	R	R^2	p	Std Beta
Virtual Only	57	0.664	0.441	0.001	-0.664
Virtual (VP)	80	0.663	0.400	0.001	-0.633
Physical (PV)	78	0.620	0.384	0.001	-0.620

Table 6.9: Correlation between pre-test score and learning gain

6.8.2 Problems solved

Linear regression found a significant positive correlation (N=57, R=0.756, R^2=0.571, p=0.03, Standardized Beta=0.792) between learning gain and number of problems solved in the Virtual Only group. On average, each student solved 8 problems while working with the ViPS tutor. The other two groups (PV and VP) were excluded from this analysis as they solved a fixed number of problems related to only one misconception (3 problems).

6.8.3 Number of Simulations Created

Linear regression found a positive correlation between learning gain and the number of simulations created, but the value of p is not statistically significant ($N=57$, $R=0.039$, $R^2=0.002$, $p=0.830$). On average, each student created 14 simulations.

6.8.4 Problem Time

Figure 6.18 shows the average time taken to solve the three problems in each misconception category (see Table 6.10). Repeated measures ANOVA revealed an overall significant difference in the average time taken to solve the three problems while working with ViPS ($F(1,140)=9.1$, $p<0.02$). The time required to solve a problem decreased significantly as students moved towards subsequent problems in the same misconception category. This shows that students took more time to solve the first problem in every misconception category as they were seeing a problem related to that misconception for the first time, but that they were faster at solving subsequent problems.

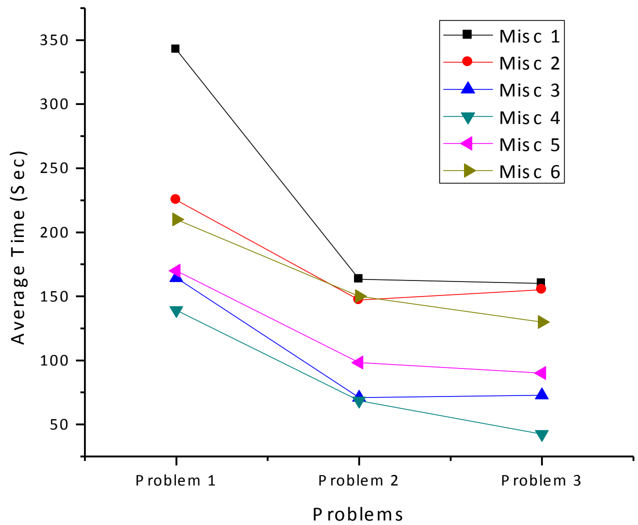


Figure 6.18: Average time taken to solve three problems in each misconception category

Misconceptions	Definition
Misc 1	More pulleys easy to pull
Misc 2	Pulling upwards is harder than pulling downwards
Misc 3	Longer string easier to pull
Misc 4	Size of the pulley affects the amount of work done
Misc 5	Having more pulleys reduces the amount of work done
Misc 6	Improper interpretation of force and work done

Table 6.10: Different misconceptions tutored by ViPS

6.8.5 Misconceptions

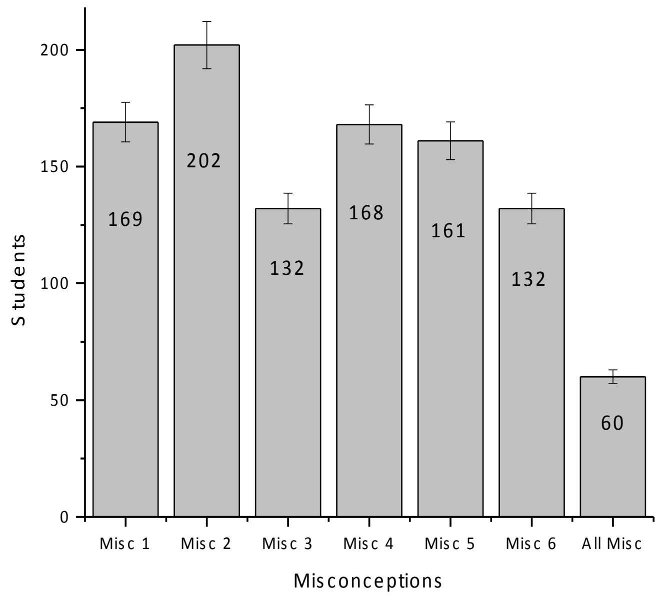


Figure 6.19: Frequency of misconceptions

Figure 6.19 shows the detected frequency of each misconception. The most common misconception among all the students who participated in the evaluation experiments is Misc 2 (see table 6.10) followed by Misc 1 and Misc 4. Out of all the students, 60 exhibited all the six misconceptions. That misconceptions persist in college students is an interesting finding, given that these misconceptions point to a naïve understanding of physics concepts such as work, force and energy, and that science curricula in middle through high school cover topics such as simple machines to help students better understand these concepts.

6.9 Summary

Despite the difficulties that arose form differences in student populations between Auburn and Kansas State students and other unanticipated variables, the evaluation studies provided evidence supporting three of our four hypotheses. The studies showed that ViPS is a useful system that helps students learn about pulley systems and also clear any misconceptions students might have. It was found to be more effective in helping students learn than actual pulleys. Students liked working with ViPS and found it usable.

CHAPTER 7: CONCLUSIONS AND FUTURE WORK

7.1 Contributions

This dissertation presents a number of significant research contributions. Also, this work further substantiates the validity of the development methodology of Intelligent Tutoring System grounded on empirical evidence of effective tutorial strategies extracted from a careful analysis of previous work in the field of ITS.

The intelligent simulation and tutoring system developed in this research, ViPS, is innovative in many ways. First of all, ViPS is the first system that tutors for misconceptions in pulley systems; it is also a new tool for creating, exploring, and simulating various pulley systems which are hard to build in the physical world. The graphical interface of ViPS is designed to help bring the abstract and difficult concept of simple machines to a more tangible level. Moreover, ViPS brings together the activities of creating, simulating and tutoring for pulleys systems in one platform. I believe the most innovative features of ViPS are its Bayesian network based dynamic inference of possible pulley systems from the components that a student selects and places on the workspace, and the dynamic hint generation from student actions. Other novel features are the problem based misconception detection approach and the associated truth table based determination of student misconceptions. To date, ViPS is the only system that can do such inference and misconception detection, and is the first system that brings simulating and tutoring under one roof.

The evaluation of ViPS was conducted at two institutions of higher education, where over 200 students worked with the system in different experiments. The results showed that ViPS is effective in helping students learn about pulleys and also in clearing any misconceptions students might have. Students liked working with the system and found it both usable and useful. We found that the dynamic pulley setup inference by the simulation module and hints generated by the tutor module of ViPS effectively guide students towards the right solution paths.

7.2 Future Work

This research offers many possibilities for future work. The following list includes some of the interesting directions.

- The ViPS system could be released for public use, so that the community at large could benefit from it. I am planning to release ViPS under open source licensing. In this way, the research communities could contribute to its further development. For example, physics educators could easily write new problems that could be integrated into ViPS.
- The ViPS system could be extended to many other simple machines like inclined planes, wedges, and screws.
- Currently, the feedback messages generated by ViPS are "*proactive*" and they are automatically generated based on student's actions. Exploring the concept of "*on demand*" hints in ViPS and comparing the relative benefits of these two types of hinting would be an interesting idea to pursue in future.

REFERENCES

Anderson, J. R., & Skwarecki, E. (1986). The automated tutoring of Introductory Computer Programming. Communications of the ACM, 29, 842-849.

Ainsworth, S. (1999) The functions of multiple representations, Computers & Education, Volume 33, Issues 2-3, September 1999, Pages 131-152.

Bloom, B. S. (1984). The 2 Sigma Problem: The Search for Methods of Group Instruction as Effective as One-to-One Tutoring. Educational Researcher, 13, 3-16.

Conati, C., Gertner, A., VanLehn, K., & Druzdzel, M, (1997). On-line student modeling for coached problem solving using Bayesian Networks. In A. Jameson, C. Paris, & C. Tasso (Eds.), User Modeling: Proceedings of the Sixth International Conference, UM97, Berlin: Springer, pp. 231-242.

Clement, J. (1983), A conceptual model discussed by Galileo and used intuitively by physics students. In D. Gentner & A. L. Stevens (Eds.), Mental Models. Hillsdale, NJ: Erlbaum.

Dağ, F., & Erkan, K., (2003). Realizing of optimal curriculum sequences for a web based general purpose intelligent tutoring system. Proceedings Turkish Symposium of Artificial Intelligence and Neural Networks, 1, pp. 1-3.

De Jong, T., van Joolingen, W.R. (1998). Scientific discovery learning with computer simulations of conceptual domains. Review of Educational Research, 68, 179-202.

DiSessa, A., A. (1983). Phenomenology and the evolution of intuition. In D. Gentner & A. L. S evens (Eds.), Mental Models. Hillsdale, NJ: Erlbaum.

Ebbinghaus, H. (1998). “Classics in Psychology, 1885: Vol. 20, Memory”, R.H. Wozniak (Ed.), Thoemmes Press, 1998.

Finkelstein, N. D., Adams, W. K., Keller, C. J., Kohl, P. B., Perkins, K. K., and Podolefsky, N. S. (2005). When learning about the real world is better done virtually: A study of substituting computer simulations for laboratory equipment. Physical Review Special Topics – Physics Education Research, 1(1), 010103.

Fruhling, A and Lee, S. (2005). Assessing the Reliability, Validity and Adaptability of PSSUQ. In Proceedings of the 9th Americas Conference on Information Systems, Omaha, Nebraska, USA, August 2005.

Gertner, A., & VanLehn, K. (2000). Andes: A coached problem solving environment for physics. In Gauthier, Frasson, and VanLehn (Eds.), Intelligent Tutoring Systems: 5th International Conference, Vol. 1839, Lecture Notes in Computer Science (pp. 133-142). Berlin & Heidelberg: Springer-.

Hafner K. (2000), Intelligent tutoring systems (a subtopic of education).

Hegarty, M., Just, M. A., & Morrison, I. R. (1988). Mental models of mechanical systems: Individual differences in qualitative and quantitative reasoning. Cognitive Psychology, 20, 19.

Lewis, J. R. (1995). IBM computer usability satisfaction questionnaires: Psychometric evaluation and instructions for use. International Journal of Human-Computer Interaction, 7(1), 57-78.

Lindstrom, B., Marton, F., Ottoson, T., & Laurillard, D. (1993). Computer-Simulation as a Tool for Developing Intuitive and Conceptual Understanding in Mechanics. Computers in Human Behavior, 9(2-3), 263-281.

Lowe, R. K. (1999) Extracting information from an animation during complex visual learning. European Journal of Psychology of Education, 14 (1999), pp. 225–244.

Martin B., & Mitrovic A. (2001), Easing the ITS bottleneck with constraint-based modelling. New Zealand Journal of Computing, 8(3), 38–47.

Molnar A. (1997). Computers in education: A brief history. THE Journal, 24, retrieved from http://www.thejournal.com/articles/13739.

Mislevy, R.J., & Gitomer, D. H. (1996). The role of probability-based inference in an intelligent tutoring system. User-Modeling and User Adapted Interaction, 5, 253-282.

Moreno, R., & Mayer, R. (1999). Cognitive principles of multimedia learning: The role of modality and contiguity. Journal of Educational Psychology, 91, 358–368.

Ploetzner, R., Lippitsch, S., Galmbacher, M., & Heuer, D. (2006). Students' difficulties in learning physics from dynamic and interactive visualizations. Paper presented at the International Conference of the Learning Sciences 2006, Bloomington, IN, pp. 550-556.

Reiser, B. J., Anderson, J. R., & Farrell, R. G. (1985). Dynamic student modeling in an intelligent tutor for LISP programming. Proceedings of the International Joint Conferences on Artificial Intelligence, pp. 8-14.

Roll, I., Aleven, V., & Koedinger, K. R. (2010). The invention lab: Using a hybrid of model tracing and constraint-based modeling to offer intelligent support in inquiry environments. In V. Aleven, J. Kay & J. Mostow (Eds.), Proceedings of the International Conference on Intelligent Tutoring Systems, pp. 115-124.

Stieff, M. (2005). Connected Chemistry – A novel modeling environment for the chemistry classroom. Journal of Chemical Education, 82(3), 489-493.

Triona, L.M., Klahr, D., & Williams, C. (2005). Point or click or building by hand: Comparing the effects of physical vs. virtual materials on middle school students' ability to optimize an engineering design, 27th Annual Conference of the Cognitive Science Society, Stresa, Italy.

Triona, L. M., & Klahr, D. (2003). Point and click or grab and heft: Comparing the influence of physical and virtual instructional materials on elementary school students' ability to design experiments. Cognition and Instruction, 21, 149-173.

VanLehn, K. (1996). Conceptual and meta learning during coached problem solving. In Frasson, C., Gauthier, G., and Lesgold, A., eds., Proceedings of the 3rd International Conference on Intelligent Tutoring Systems ITS '96. Berlin: Springer, pp. 29–47.

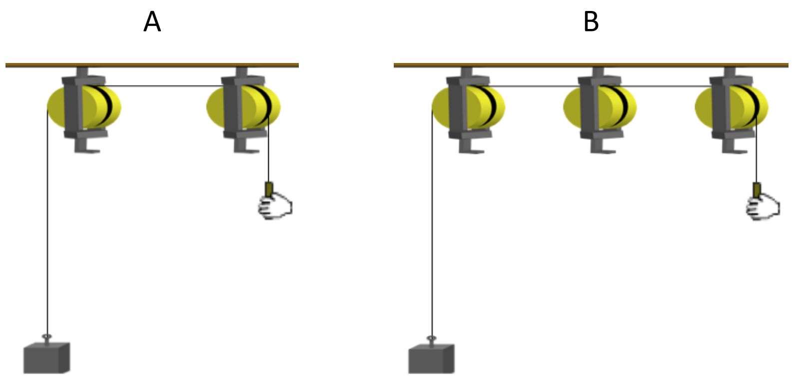
White, B.Y. (1983). Sources of difficulty in understanding Newtonian dynamics. Cognitive Science, 7, 41-65.

Zacharia, Z., & Anderson, O. R. (2003). The effects of an interactive computer-based simulation prior to performing a laboratory inquiry-based experiment on students' conceptual understanding of physics. American Journal of Physics, 71(6), 618-629.

APPENDIX A

ViPS Evaluation Study – Problems used in Virtual-Physical and Physical-Virtual sequence experiment

P1. If we ignore friction, which of the following two pulleys systems will require less force (effort) to lift the load to the same height?

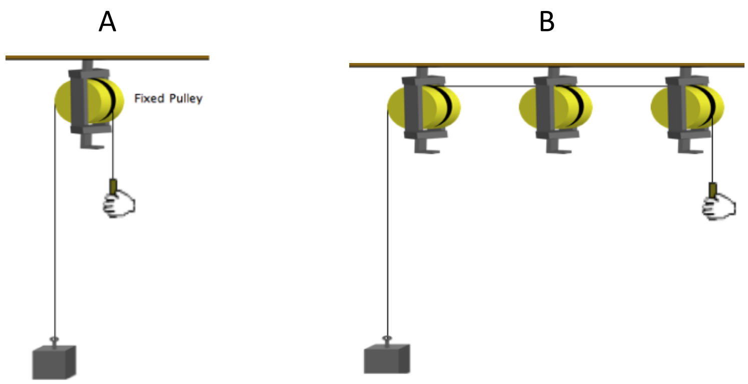


***Please record you prediction first**

Options	Prediction*	Actual Answer
Pulley System A		
Pulley System B		
Both Pulley A & Pulley B will require the same force (effort)		
Not enough information		

Pulley System A	Values		Pulley System B	Values
Weight:			Weight:	
Distance Moved (m):			Distance Moved (m):	
Distance Pulled (m):			Distance Pulled (m):	
Force Applied (N):			Force Applied (N):	
Work Done (J):			Work Done (J):	

P2. If we ignore friction, which of the following two pulleys systems will require less force (effort) to lift the load to the same height?

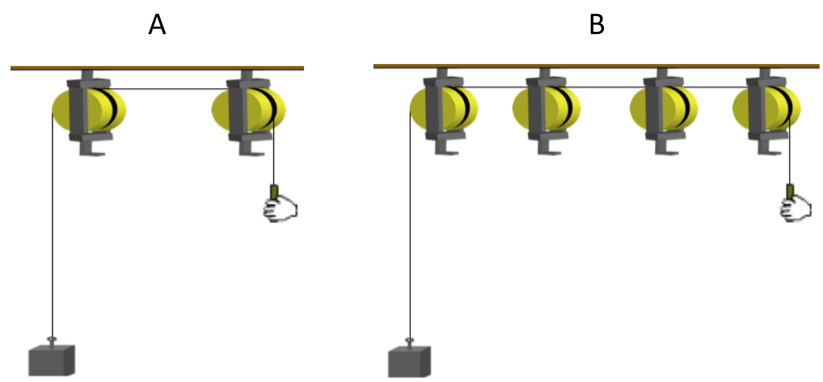


***** Please record your prediction first

Options	**Prediction ***	**Actual Answer**
Pulley System A		
Pulley System B		
Both Pulley A & Pulley B will require the same force (effort)		
Not enough information		

Pulley System A	**Values**		**Pulley System B**	**Values**
Weight:			Weight:	
Distance Moved (m):			Distance Moved (m):	
Distance Pulled (m):			Distance Pulled (m):	
Force Applied (N):			Force Applied (N):	
Work Done (J):			Work Done (J):	

P3. Jack is using pulley system A and Mary is using B. What can you tell about the work done and force required to lift the same load to the same height by each of them, if you ignore the friction?



***** Please record your prediction first

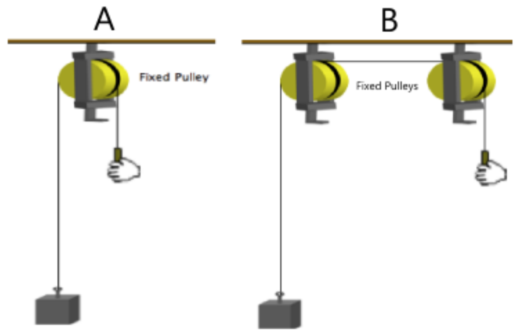
Options	Prediction *	Actual Answer
Pulley System A		
Pulley System B		
Both Pulley A & Pulley B will require the same force (effort)		
Not enough information		

Pulley System A	Values		Pulley System B	Values
Weight:			Weight:	
Distance Moved (m):			Distance Moved (m):	
Distance Pulled (m):			Distance Pulled (m):	
Force Applied (N):			Force Applied (N):	
Work Done (J):			Work Done (J):	

APPENDIX B

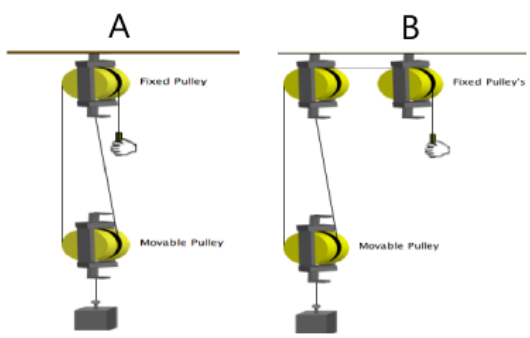
Pretest/ Midtest/ Posttest Questionnaire

Q1. If we ignore friction, which of the following two pulleys systems will require less force(effort) to lift the load to the same height?



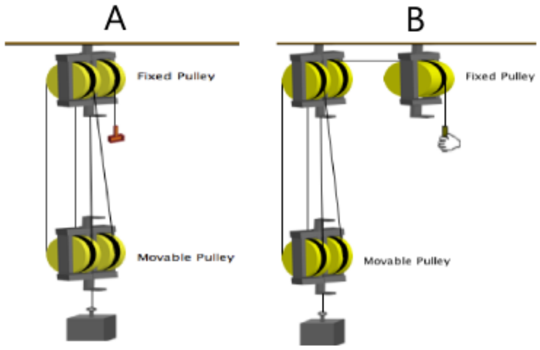
- Pulley System A
- Pulley System B
- Both Pulley A & Pulley B will require the same force(effort)
- Not enough information

Q2. If we ignore friction, which of the following two pulleys systems will require less force(effort) to lift the load to the same height?



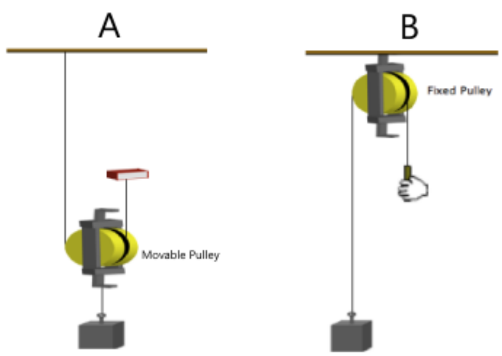
- Pulley System A
- Pulley System B
- Both Pulley A & Pulley B will require the same force(effort)
- Not enough information

Q3. If we ignore friction, which of the following two pulleys systems will require less force(effort) to lift the load to the same height?



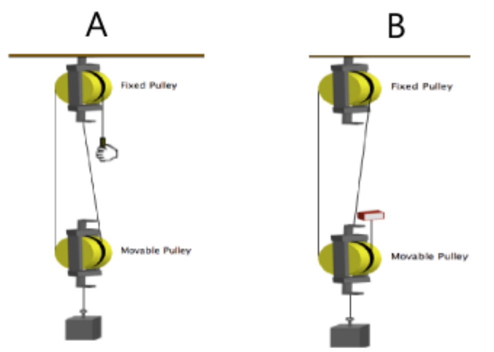
- Pulley System A
- Pulley System B
- Both Pulley A & Pulley B will require the same force(effort)
- Not enough information

Q4. If we ignore friction, which of the following two pulleys systems will require less force(effort) to lift the load to the same height?



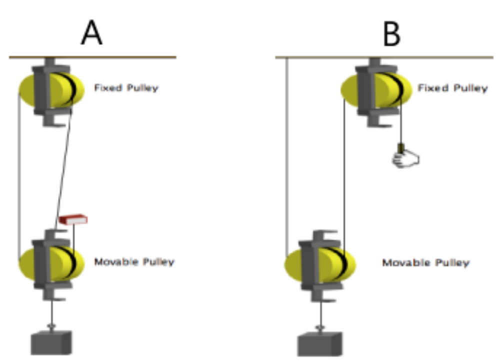
- Pulley System A
- Pulley System B
- Both Pulley A & Pulley B will require the same force(effort)
- Not enough information

Q5. If we ignore friction, which of the following two pulleys systems will require less force(effort) to lift the load to the same height?



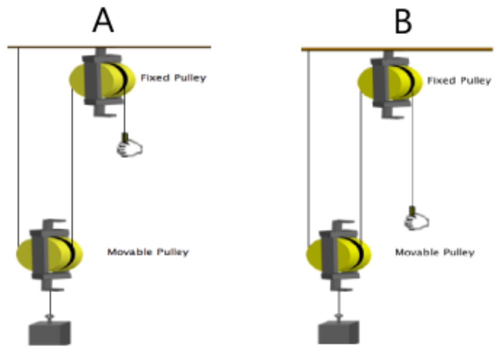
- Pulley System A
- Pulley System B
- Both Pulley A & Pulley B will require the same force(effort)
- Not enough information

Q6. If we ignore friction, which of the following two pulleys systems will require less force(effort) to lift the load to the same height?



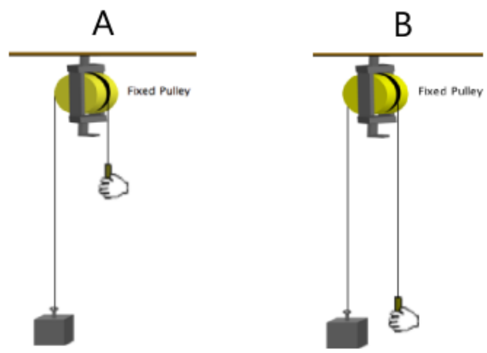
- Pulley System A
- Pulley System B
- Both Pulley A & Pulley B will require the same force(effort)
- Not enough information

Q7. If we ignore friction, which of the following two pulleys systems will require less force(effort) to lift the load to the same height?



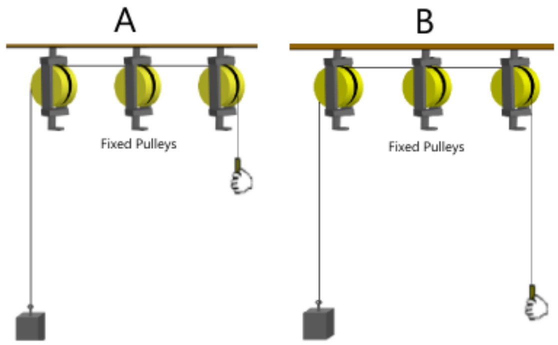
- ◯ Pulley System A
- ◯ Pulley System B
- ◯ Both Pulley A & Pulley B will require the same force(effort)
- ◯ Not enough information

Q8. If we ignore friction, which of the following two pulleys systems will require less force(effort) to lift the load to the same height?



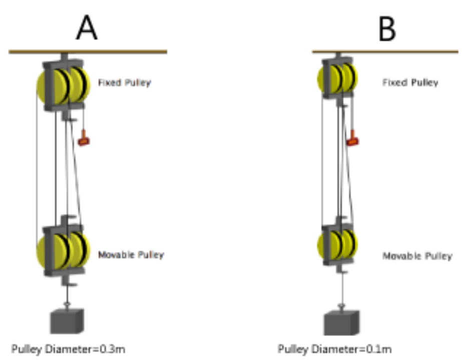
- ◯ Pulley System A
- ◯ Pulley System B
- ◯ Both Pulley A & Pulley B will require the same force(effort)
- ◯ Not enough information

Q9. If we ignore friction, which of the following two pulleys systems will require less force(effort) to lift the load to the same height?



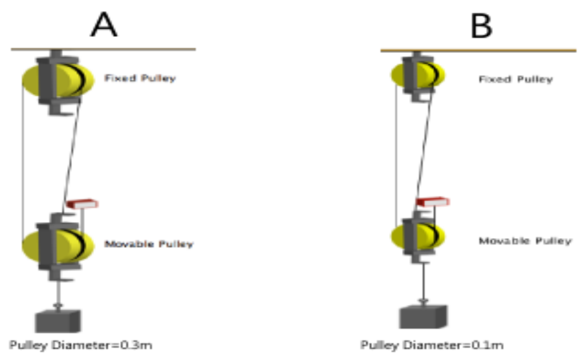
- Pulley System A
- Pulley System B
- Both Pulley A & Pulley B will require the same force(effort)
- Not enough information

Q10. If we ignore friction, which of the following two pulleys systems will require less force(effort) to lift the load to the same height?



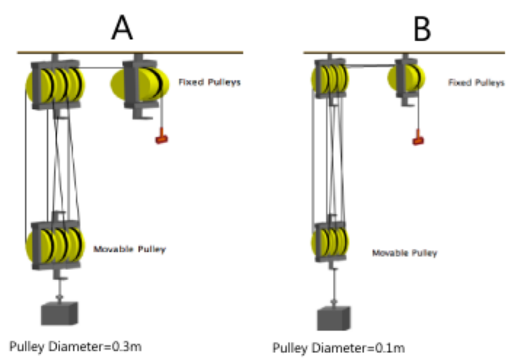
- Pulley System A
- Pulley System B
- Both Pulley A & Pulley B will require the same force(effort)
- Not enough information

Q11. If we ignore friction, which of the following two pulleys systems will require less force(effort) to lift the load to the same height?



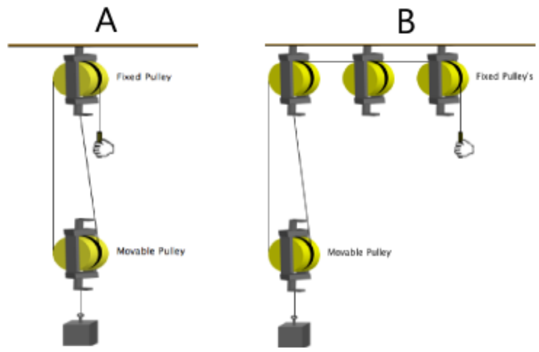
- Pulley System A
- Pulley System B
- Both Pulley A & Pulley B will require the same force(effort)
- Not enough information

Q12. If we ignore friction, which of the following two pulleys systems will require less force(effort) to lift the load to the same height?



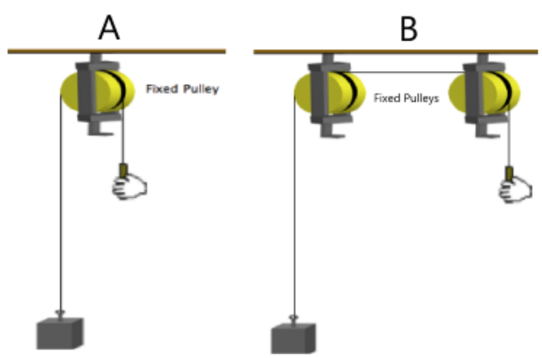
- Pulley System A
- Pulley System B
- Both Pulley A & Pulley B will require the same force(effort)
- Not enough information

Q13. Tom is using pulley system A and Sam is using B. What can you tell about the work needed to lift the same load to the same height by each of them, if you ignore the friction?



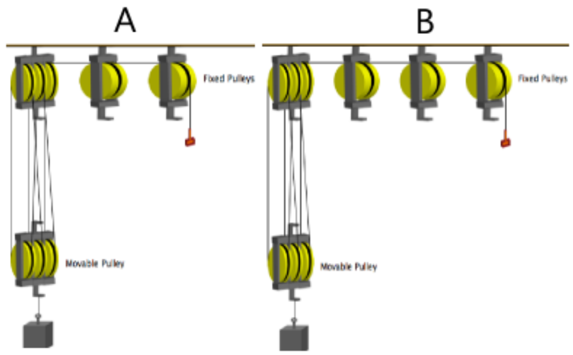
- Tom (using pulley system A) is doing more work
- Sam (using pulley system B) is doing more work
- Work done in both the situations is same
- Not enough information

Q14. Jack lifts a box 1m using pulley system A. Mary lifts an identical box to the same height using pulley system B. What can you tell about the work done, if you ignore the friction?



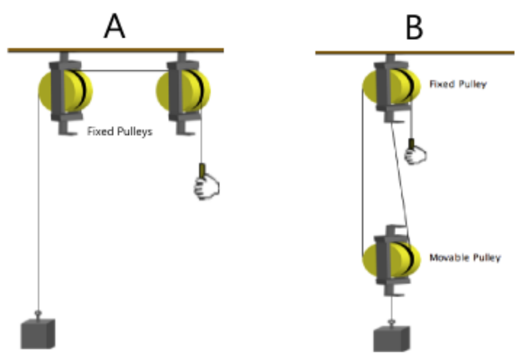
- Jack (using pulley system A) is doing more work
- Mary (using pulley system B) is doing more work
- Work done in both the situations is same
- Not enough information

Q15. Bill lifts a box 1m using pulley system A. Paul lifts an identical box to the same height using pulley system B. What can you tell about the work done, if you ignore the friction?



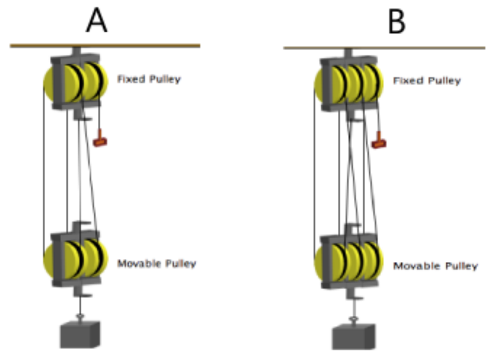
- Bill (using pulley system A) is doing more work
- Paul (using pulley system B) is doing more work
- Work done in both the situations is same
- Not enough information

Q16. Tom is using pulley system A and Sam is using B. What can you tell about the work done and force required to lift the same load to the same height by each of them, if you ignore the friction?



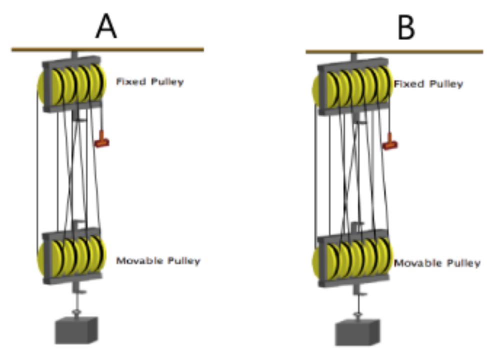
- Both require same force, but work done is different
- Both require same force and work done is same
- Both require different force, but work done is same
- Not enough information

Q17. Jack is using pulley system A and Mary is using B. What can you tell about the work done and force required to lift the same load to the same height by each of them, if you ignore the friction?



- Both require same force, but work done is different
- Both require same force and work done is same
- Both require different force, but work done is same
- Not enough information

Q18. Bill is using pulley system A and Paul is using B. What can you tell about the work done and force required to lift the same load to the same height by each of them, if you ignore the friction?



- Both require same force, but work done is different
- Both require same force and work done is same
- Both require different force, but work done is same
- Not enough information

APPENDIX C

System Usability Questionnaire

System Usability Questionnaire

Username*:

1. Overall, I am satisfied with how easy it is to use this system. *

	1	2	3	4	5	6	
STRONGLY DISAGREE	○	○	○	○	○	○	STRONGLY AGREE

2. It was simple to use this system. *

	1	2	3	4	5	6	
STRONGLY DISAGREE	○	○	○	○	○	○	STRONGLY AGREE

3. I could effectively complete the tasks and scenarios using this system. *

	1	2	3	4	5	6	
STRONGLY DISAGREE	○	○	○	○	○	○	STRONGLY AGREE

4. I was able to complete the tasks and scenarios quickly using this system. *

	1	2	3	4	5	6	
STRONGLY DISAGREE	○	○	○	○	○	○	STRONGLY AGREE

5. I was able to efficiently complete the tasks and scenarios using this system. *

	1	2	3	4	5	6	
STRONGLY DISAGREE	○	○	○	○	○	○	STRONGLY AGREE

6. I felt comfortable using this system. *

	1	2	3	4	5	6	
STRONGLY DISAGREE	○	○	○	○	○	○	STRONGLY AGREE

7. It was easy to learn to use this system. *

	1	2	3	4	5	6	
STRONGLY DISAGREE	○	○	○	○	○	○	STRONGLY AGREE

8. I believe I could become productive quickly using this system. *

	1	2	3	4	5	6	
STRONGLY DISAGREE	○	○	○	○	○	○	STRONGLY AGREE

9. The system gave error messages that clearly told me how to fix problems. *

	1	2	3	4	5	6	
STRONGLY DISAGREE	○	○	○	○	○	○	STRONGLY AGREE

10. Whenever I made a mistake using the system, I could recover easily and quickly. *

	1	2	3	4	5	6	
STRONGLY DISAGREE	○	○	○	○	○	○	STRONGLY AGREE

11. The information (such as in-line help, on-screen messages and other documentation) provided with this system was clear. *

	1	2	3	4	5	6	
STRONGLY DISAGREE	○	○	○	○	○	○	STRONGLY AGREE

12. It was easy to find the information I needed. *

	1	2	3	4	5	6	
STRONGLY DISAGREE	○	○	○	○	○	○	STRONGLY AGREE

13. The information provided for the system was easy to understand. *

	1	2	3	4	5	6	
STRONGLY DISAGREE	○	○	○	○	○	○	STRONGLY AGREE

14. The information was effective in helping me complete the tasks and scenarios. *

	1	2	3	4	5	6	
STRONGLY DISAGREE	○	○	○	○	○	○	STRONGLY AGREE

15. The organization of the information on the system screen was clear. *

	1	2	3	4	5	6	
STRONGLY DISAGREE	○	○	○	○	○	○	STRONGLY AGREE

16. The interface of the system was pleasant. *

	1	2	3	4	5	6	
STRONGLY DISAGREE	○	○	○	○	○	○	STRONGLY AGREE

17. I liked using the interface of this system. *

	1	2	3	4	5	6	
STRONGLY DISAGREE	○	○	○	○	○	○	STRONGLY AGREE

18. The system has all the functions and capabilities I expect it to have. *

	1	2	3	4	5	6	
STRONGLY DISAGREE	○	○	○	○	○	○	STRONGLY AGREE

19. Overall, I am satisfied with this system. *

	1	2	3	4	5	6	
STRONGLY DISAGREE	○	○	○	○	○	○	STRONGLY AGREE