THE IMPACT OF THE SURPRISING POSSIBILITIES IMAGINED AND REALIZED THROUGH INFORMATION TECHNOLOGY PROGRAM ON ATTITUDES IN INFORMATION TECHNOLOGY

by

Anna Forssen

A thesis submitted to the Faculty and the Board of Trustees of the Colorado School of Mines in partial fulfillment of the requirements for the degree of Master of Science (Mathematical and Computer Sciences).

Golden, Colorado
Date 8/19/11

Signed: Anna Forssen
Anna Forssen

Signed: Barbara Moskal
Dr. Barbara Moskal
Thesis Advisor

Golden, Colorado
Date 8/22/11

Signed: Tracy Camp
Dr. Tracy Camp
Interim Department Head
Department of Mathematical and Computer Sciences

ABSTRACT

Shying away from STEM-related fields is a common problem for female and minority high school students. The *Surprising Possibilities Imagined and Realized through Information Technology* (SPIRIT) is a NSF funded project designed to increase knowledge of and improve attitudes toward careers in information technology (IT) at the high school level. An IT attitude survey was used to capture participants' changes in general interest in IT and perception of gender stereotypes in IT. This thesis assessed the validity and reliability of the survey, and linear regression analyses were used to examine subgroup differences across teacher, counselor and student participants. A confirmatory factor analysis (CFA) was run on the student version of the IT attitude survey, and an exploratory factor analysis (EFA) was run on the teacher and counselor version of the IT attitude survey. Based on these analyses, results demonstrated that the survey was both valid and reliable for the populations of interest. In addition, the IT attitude survey was able to capture differences in participant subgroups by gender, ethnicity and the year of participation in the SPIRIT program.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to thank Dr. Barbara Moskal, Dr. Amanda Hering and Professor Alka Harriger for helping me complete this thesis. I would also like to thank the National Science Foundation and Colorado School of Mines for providing funding for this research opportunity (DUE-0512062; DRL-0737679). Finally, I would like to thank Nicolaus and all of my friends in the math department who supported me and helped me greatly through this.

CHAPTER 1
INTRODUCTION

The high school years are a critical period in which students decide which career paths they will pursue in college. Stereotypical gender roles can limit students' choices, such as the belief that females are better at liberal arts and biological sciences, and males are better at science, technology, engineering and mathematics fields (STEM). High school is also a time when students decide which subjects they feel most comfortable studying, and which ones they find intimidating. Shying away from mathematics-related fields is a common problem for female and minority high school students, and this problem has been attributed to their low representation in STEM fields.

A gender gap currently exists between females and males in STEM careers. Data available through the NSF indicates that the percentage of females who received bachelor's degrees in computer science (CS) has been steadily decreasing from 28% in 2000 to 17.7% in 2008 [1]. Females often enter introductory programming courses with less experience and confidence than males, and even when females have comparable computing backgrounds to males, they rank their skills lower than their male counterparts [2].

Information Technology (IT) is the field of engineering related to the development, maintenance and use of computer technology that enables the acquisition, storage, processing and distribution of information. Increasing the representation of women and minorities in the IT workforce is an important problem to address because of the benefits that women and minorities bring to the IT field. A diverse workforce is more likely to produce innovative solutions to problems, which is critical for corporate success and global competitiveness [3, 4]. A lack of a diverse workforce can lead to a reduction in creativity and a further strengthening of female and minority stereotypes [5].

Informing high school students of the wide array of possible careers in IT is important for increasing the number of students who pursue college degrees in IT. In a study completed by [6], 130 high school students participated in a week-long seminar with presenters from STEM fields. Ninety percent of the students reported that learning about various career options was the most important component of the program. As part of the first year of the *Surprising Possibilities Imagined and Realized through Information Technology* (SPIRIT), 63 high school students from Indiana completed a survey that examined attitudes toward IT [7]. An exploratory factor analysis indicated that the participating students did not subdivide the field into the same constructs that research typically indicates fields of technology comprise, possibly suggesting a limited understanding of the IT field. This reveals the essential problem at hand: if students do not know what IT is, or what career options are available in IT, then they will be unlikely to choose IT for further study in college or as a career path.

Attracting female and minority students to IT may be a matter of bridging IT with other academic areas such as biological and social sciences. With the wealth of information processing available in today's society, many fields outside of IT or CS are in need of employees that are knowledgeable of IT and programming. Mahmoud describes such cross-disciplinary fields as being highly sought after in the job market due to the level of expertise and skill involved [8]. Students may not be aware that IT can be applied to many different career paths. Providing students with instruction on how IT links to other academic fields may impact students' decisions to pursue IT in college.

High school students' exposure to computers and other technologies in the classroom is important for building students' confidence in using technology as part of their careers. Classroom exposure is governed by high school teachers' knowledge of using technology and the desire to implement technology into their lesson plans. Mahmoud stresses that it is important for teachers to make computers and technology fun for their students, and make it accessible for all students to learn [8].

1.1 Efforts to Attract Students to STEM

Many educational programs have focused on exposing young students to career options in STEM. One example is the K-12 Engineering Education Programs (KEEP) Seminar Series at University of Nevada, Reno, which began in 2003 and ended in 2008. KEEP was targeted at high school juniors and seniors, and tracked career choices over the eight-week seminar series [6]. Weekly sessions engaged students with presenters from STEM fields followed by a social hour, where students were able to engage in dialogue with the speakers. Results indicated that after attending KEEP, 45.3% of students did not make changes in their previous career choices. These students had a firm career decision prior to attending KEEP. For the remaining students, more than half indicated a revised career choice related to one of the presenters [6].

The IT-Adventures program at Iowa State University is another outreach program that motivates high school students to pursue careers in IT by teaching game design programming, robotics and cyber defense [9]. The IT-Adventures program is a year long in duration, and ends with an IT-Olympics event. While the students who participated in the IT-Olympics from the 2007/2008 academic year were predominantly male (89.8%), nearly 98% of them answered "yes" when asked if they planned to attend post-secondary education. In a self-report, two-thirds of the students indicated that they were planning to pursue a degree related to IT [9]. However, neither the KEEP nor the IT-Adventures program examined how students' attitudes changed in the short term with regard to STEM careers before and after participating in the program.

At the University of Michigan, Dearborn, researchers Maxim and Elenbogen hosted one-day computing events with the goal of attracting K-12 students to CS and IT [10]. The computing events were interactive, where students were able to use their programming skills to create multimedia content. Students completed Kay's computer attitude scale [11] before and immediately after the events. Results indicated that students displayed a significantly positive attitude change toward computing [10]. One limitation of this study was the use of a computer attitude scale which may have been outdated (Kay's computer attitude scale was first published in 1989). A second limitation is that the computing events lasted one day, and did not provide time for students to fully explore computing as a field of study.

1.2 SPIRIT

The *Surprising Possibilities Imagined and Realized through Information Technology* (SPIRIT) was a three year project sponsored by the National Science Foundation [NSF, DRL-0737679] under the direction of Alka Harriger at Purdue University [5]. SPIRIT was designed to increase knowledge of and improve attitudes toward careers in information technology (IT) at the high school level. The SPIRIT program targeted high school teachers, counselors and students, with an emphasis on female students. SPIRIT addressed the problem of the gender gap in IT careers by exposing all participants to computers and technology. One of the goals of SPIRIT was to increase the interests of female students in pursuing IT or IT-related fields in college through a summer experience with IT.

SPIRIT sought to demonstrate to high school students that IT can be fun and useful, while leading to a vast array of career options. SPIRIT emphasized the positive impact that IT had on society by providing high school teachers, counselors and students with a diverse array of speakers who were successful in the IT field. Many of the speakers had careers that combined IT with other STEM fields, and they emphasized how IT increased both the quality and efficiency of their work. Career presentations consisted of two designs: career speeches, which were administered in the style of a lecture; and hands-on demonstrations, which illustrated how IT was used in specific careers [12].

SPIRIT introduced participants to Alice as a tool to support visual and animated displays of information. Alice is a three dimensional programming environment that uses a drag and drop editor to create animations for story telling and games, and provides a solid introduction to object-oriented concepts that can be applied to other code-oriented programming software. High school teachers and counselors were able to use the animated nature of Alice storyboards as a tool for instruction. Alice storyboards were believed to make learning more interactive and fun for students, and students also learned to use Alice for their homework assignments and presentations. By learning how to display information with Alice and implementing it in the school environment, high school teachers and counselors were better prepared to advise potentially interested students to pursue IT or other IT-related fields. By positively influencing students year after year, high school teachers and counselors who implement Alice in their schools also have the potential of a multiplier effect, or impacting more students every year [12].

The SPIRIT approach provided different programs for teachers, counselors, and students, respectively. High school teachers participated in a two-week professional development program where Alice was used to create lesson plans that were implemented in the following academic year. High school counselors participated in a one-week professional development program, and high school students participated in a one-week educational summer camp [13]. During the summer of 2010, counselors joined the teachers in the professional development program. All programs featured reinforcing presentations and hands-on activities that related to IT. All SPIRIT participants learned about various careers in IT and how to use Alice to display information.

Participants in the SPIRIT program completed Concept Exams, which evaluated what the participants learned from the Alice tutorials; IT attitude surveys, which determined whether or not the participants' attitudes changed regarding the field of IT; and End of Program Evaluations, which provided an opportunity for participants to share their comments and suggestions. Anonymous Focus Group interviews were conducted by the research team at Colorado School of Mines. The Focus Group activities ensured that participants had a chance to express their viewpoints of the IT field and the SPIRIT program.

The five objectives of the SPIRIT program were:

1. "Show teachers how IT can support instruction in the STEM disciplines,"
2. "Share strategies for sparking interest in computing among female students,"
3. "Communicate career opportunities in many disciplines supported by IT specialists,"
4. "Provide opportunities for teachers to apply their new skills in student workshops," and
5. "Generate useful resources that will facilitate implementation of this technology and strategy in high school classrooms" [5].

The five anticipated outcomes of the SPIRIT program were:

1. "Improve female students' understanding of the wide ranging career opportunities in IT,"
2. "Improve the attitudes of female students regarding the computing disciplines,"
3. "Increase the understanding of counselors regarding the wide ranging career opportunities in IT for women,"
4. "Improve high school teachers' content and pedagogical knowledge on applying IT across STEM disciplines," and
5. "Improve the attitudes of high school teachers regarding the IT career opportunities for women" [5].

1.3 Investigation

This thesis explored data collected from the SPIRIT program from summer 2008 to spring 2011, including three academic years. The primary focus of this thesis was on the IT attitude survey, which was originally developed by Moskal, Dann, Cooper and Guzdial in 2005. Based on statistical and qualitative analyses, along with suggestions from Harriger, the IT attitude survey was further refined. The IT attitude survey was designed to measure changes in attitudes regarding the field of IT and careers in IT. For teachers, counselors and students, the IT attitude survey was administered before the summer workshop began and immediately after the workshop ended. Additionally, the IT attitude survey was administered one year later for teachers and counselors after they implemented Alice in their lesson plans and practices during the academic year. The original IT attitude survey consisted of 47 statements for teachers and counselors and 52 statements for students. All statements were assessed on a four point Likert scale. The IT attitude survey concluded with four free response questions.

For this thesis, various analyses and statistical tests were used to examine three years of IT attitude survey data. A factor analysis of the IT attitude survey examined how the survey statements correlated amongst one another to form latent factors. Multiple linear regressions were used to compare participants' factor scores. An implementation analysis examined differences in workshops across the three years, and a document analysis examined changes in the IT attitude survey across the three years. The statistical tests were calculated in R and Mplus, and all code is contained in the appendix.

Research findings from this thesis will aid in the development and measurement of educational programs similar to SPIRIT. Based on the results of this study, further IT educational programs can be refined and implemented on a broader scale.

1.4 Research Questions

This thesis evaluated three years of SPIRIT data from summer 2008 to spring 2011, including three academic years. The research questions addressed were:

1. Are the established factors for the student version and the theoretical factors for the teacher and counselor version of the IT attitude survey valid and reliable?
2. Based on factors underlying the IT attitude survey, was there a measurable difference in teachers', counselors' and students' attitudes with respect to IT immediately after, and for teachers and counselors, one year after completing the SPIRIT workshop?
3. Based on this same instrument, did attitudes differ across gender, ethnic groups and/or the three years of project implementation?
4. Based on this same instrument, how did regression analyses of the SPIRIT data differ across teachers, counselors and students?

CHAPTER 2
LITERATURE REVIEW

This chapter begins with a review of prior studies on the Alice Curriculum. Next, this chapter discusses the research that has been completed on the IT attitude survey prior to the summer of 2008. Finally, the statistical techniques that are proposed for this investigation are discussed with a focus on theory.

2.1 Prior Studies on the Alice Curriculum

The Alice curriculum was developed by Dr. Stephen Cooper of Saint Joseph's University and Wanda Dann of Ithaca College, and is comprised of instructional material, a textbook and implementation guidelines for teaching programming with the Alice software. Dr. Cooper has since moved to Purdue University and Dr. Dann to Carnegie Mellon University. The intent of the Alice curriculum is to teach introductory computer programming while maintaining or improving teachers' and students' positive attitudes toward the subject [14]. The curriculum is designed to assist students in succeeding rather than "weeding out" students who struggle academically [14].

The Alice program uses a design-first approach to objects which provides a simple, visual approach to understanding programming. Alice uses a drag-and-drop editor which allows students to focus on understanding concepts rather than on language-specific syntax [15].

As a preliminary analysis, the Alice team incorporated the Alice curriculum at Ithaca College and Saint Joseph's University during the 2001-2002 academic year. A total of 11 CS majors participated in the study and were compared to a control group of 10 CS majors completing the traditional programming course without Alice, CS1. All 21 students were considered "at risk" with no previous programming knowledge and little or no knowledge of calculus [15]. Student development was observed in a variety of areas [15]. In later Alice classes, students were observed writing down their thoughts before beginning to program on the computer, indicating a strong sense of design. Also, the Alice software appeared to encourage student collaboration, as students were witnessed combining worlds and assisting each other [15].

In [16], the goal for using the Alice curriculum was to better retain students in CS1 and to increase the enrollment of students in the follow-up computer programming course, CS2. This work was built from the observations previously discussed. Since sample sizes were small, a Kruskal-Wallis (KS) one-way ANOVA was used to examine differences across groups [16]. The KS ANOVA is a non-parametric technique in which no assumptions are made about the underlying distribution of the data. Students in the Alice course displayed significantly greater retention rates in CS1 than did the control group. Eighty-eight percent of the students in the Alice course continued and completed CS2, while only 15% of the students in the traditional CS1 course continued and completed CS2. At-risk students who completed the Alice course performed as well in CS2 as students in the traditional CS1 course, indicating that Alice might "level the playing field" for ill-prepared students [16].

2.1.1 JABRWOC

The *Java-based Animation: Building Virtual Worlds for Object-oriented Programming in Community Colleges* (JABRWOC) was a three year project funded by NSF that used the Alice curriculum in introductory programming courses. Three community colleges participated and used the Alice curriculum: Camden County College, Community College of Philadelphia, and Tompkins Cortland Community College. Due to differences in course lengths, the implementation of the Alice curriculum varied slightly at each institution. Students were given an opportunity to be part of the treatment group, which learned introductory programming with the Alice program, or the control group, which learned introductory programming in the traditional manner. A quasi-experimental design was used since random assignment was not possible.

Students completed three different assessments to measure the differences between treatment and control groups. The Basic Concepts exam tested introductory programming concepts that were platform independent. The Alice Specific exam tested introductory programming concepts specific to the Alice program. The Loyd and Gressard Computer Attitude Survey was used to explore changes in students' attitudes toward programming [17]. The Loyd and Gressard attitude survey employed a Likert scale which was easy for participants to understand and complete. Both pre and post assessment scores were collected, and all assessments were administered online [18].

Results from JABRWOC were collected for three semesters: fall 2004, spring 2005 and fall 2005. Due to the high drop-out rate at community colleges, the number of students who completed both pre and post versions of assessments was low. Pre and post data was paired in order to minimize the assumption that students who dropped out performed the same in knowledge and attitude as students who completed the courses [19]. T-tests were used to compare differences in pre and post assessment scores. Analysis of covariance (ANCOVA) was used to compare differences between men and women.

Across all semesters, the majority of students in the treatment group scored significantly better on the Basic Concepts exam than students in the control group, indicating that the Alice intervention had a stronger impact on the development of conceptual programming knowledge than the traditional classroom approach. When examining all courses offering the Alice curriculum, including those that did not have a control with which to compare, male students had higher post scores on both the Basic Concepts and Alice Specific exams than females. Results did not differ across institution or course length [20].

On average, neither the treatment group nor the control group displayed a significant change in attitude with respect to programming. When students were divided by gender, the results from the attitude survey did not differ. At the time of implementation, the Loyd and Gressard attitude survey had not been updated in more than ten years and contained outdated questions, indicating that a new attitude survey may be necessary to capture changes [20].

2.1.2 Tufts University Alice Implementation

During the 2005-2006 academic year, the Alice curriculum was implemented in an introductory CS course at Tufts University. Students who participated in this study were from a "bi-modal" population, where some of the students had a strong CS background and the others had no previous programming experience [21]. Two approaches were used in the Alice curriculum implementation. The first approach was to initially teach Alice and then transition to a high-level language (HLL) such as C++. The second approach was to mesh Alice and a HLL together throughout the semester by introducing a topic with Alice and then transitioning to a HLL [21].

In the first approach, the researchers observed that students enjoyed working with Alice, and would often laugh out loud when seeing the animation of mistakes. Students that had no previous programming experience felt confident when using Alice, but easily became frustrated when working with a HLL. Some students felt intimidated by the textual language and syntax of a HLL and had difficulty seeing how Alice related to the HLL [21].

In the second approach, the transitions from Alice to a HLL that occurred throughout the semester were troubling for students. Students with no previous programming experience quickly lost confidence, while students with strong CS backgrounds resisted switching back to Alice after making the transition to a HLL. Many students felt discouraged when they were able to program something in Alice, but not in a HLL, and "concluded that they were inadequate programmers" [21]. Also, many of these same students felt that their success with Alice was not "real programming," but was rather a game meant for younger audiences [21].

2.1.3 Integrating Alice into Middle and High School Curriculums

Implementation of the Alice program into middle and high school curriculums in the U.S. has displayed promising results. At Princess Anne High School in Virginia, Alice was used in an introductory programming course taught by John Harrison during the 2007/2008 academic year [22]. Harrison observed that students' mistakes in Alice were easily spotted and corrected, unlike mistakes in traditional programming languages. Also, Alice appealed to both male and female students by featuring a diverse array of objects and environments. Subsequent enrollment in the introductory programming courses taught by Harrison increased and became more diverse in both gender and ethnicity.

In North Carolina, the Adventures in Alice Programming project aimed to integrate Alice into middle and high schools throughout the state, with a focus on middle schools. During the past summers, several beginner and advanced workshops were held for teachers to learn Alice. Susan Rodgers, a professor at Duke University who works with Adventures in Alice Programming, held Alice summer camps at Duke for middle school students during 2008. Rodgers observed that the students were very engaged in the Alice software, and often did not want to stop working with Alice, even when the day's lessons were done [23].

The implementation of Alice into the high school curriculum has also had positive results in Taiwan [24]. In Taiwan, an introductory programming course is part of the mandatory high school computing curriculum. During the fall of 2008, tenth grade students were randomly selected to participate in either the programming course taught with Alice or the traditional programming course taught with C++. An "experience" questionnaire was administered to both the treatment and control group, and consisted of 20 questions that were Alice specific and C++ specific, respectively. Both versions of the questionnaire were designed to test the same concepts. The students in the Alice group scored significantly better than the students in the control group, indicating that students in the Alice group may have gained a better understanding of fundamental programming concepts [24].

2.2 IT Attitude Survey

The *Collaborative Research: Assessing Concept Knowledge and Attitudes in Introductory CS courses* (NSF, DUE-0512062) was a project led by Dr. Barbara Moskal and a team from Colorado School of Mines, Ithaca College, Saint Joseph's University and Georgia Tech that developed two assessment instruments to evaluate student attitudes and learning outcomes in CS [25]. This project was necessitated by the lack of up-to-date, valid and reliable assessment instruments in CS that could be used to thoroughly assess students' attitudes with respect to CS. The focus of this research was the development of two assessment instruments. One was a CS attitude survey and the other a fundamental CS concepts exam.

The constructs of interest to the IT attitude survey were:

1. *Confidence* in students' ability to learn CS skills
2. Perceptions of CS as a male field (*gender*)
3. Beliefs in the *usefulness* of learning CS
4. *Interests* in CS
5. Beliefs about *professionals* in CS

Five constructs, based on the goals listed above, were identified before a preliminary set of 55 attitude survey statements were developed. The statements were written based on research in technology with an emphasis on CS, and were developed to match the desired construct [26].

The attitude survey was initially developed to assess attitudes with respect to CS. When the attitude survey was further altered for use with IT, the phrase "computer science" in the survey statements was replaced by "information technology."

2.3 Factor Analysis

In the social sciences, surveys are often used to evaluate human responses to experiments or observations. Factor analysis is a statistical method that examines the underlying structure of a set of questions (*indicators*) in a survey, and determines the number of latent variables (*factors*) that account for the correlation matrix between the original set of indicators. The goal of factor analysis is to produce factor scores, or scores that would have been observed if the latent factor was measured directly.

Factor analysis originated in psychometrics, a field of study concerning educational and psychological measurement. The ideal psychometric instrument tests what it intends to test; this is known as validity. Additionally, the ideal psychometric instrument performs in a consistent manner across indicators and administrations; this is known as reliability. Jöreskog's measure of scale reliability in used in this thesis rather than Cronbach's alpha, which often over or underestimates reliability if the model contains correlated measurement errors [27, 28]. Evidence to support an instrument's validity can be acquired through factor analysis. Knowledge of factor loadings and interdependencies of observed variables allow for the reduction of insignificant variables in a dataset.

2.3.1 Factor Analysis with Categorical Indicators

Factor analysis belongs to the structural equation modeling (SEM) family. The goal of SEM is to test the hypothesis that the observed covariance matrix for a set of questions, called *indicators*, is equal to the covariance matrix implied by a hypothesized model. In notational form:

$$
\begin{aligned}
H_0 &: \mathbf{\Sigma} = \mathbf{\Sigma}(\boldsymbol{\theta}) \\
H_1 &: \mathbf{\Sigma} \neq \mathbf{\Sigma}(\boldsymbol{\theta})
\end{aligned}
$$

where $\mathbf{\Sigma}$ represents the observed covariance matrix and $\mathbf{\Sigma}(\boldsymbol{\theta})$ represents the covariance matrix implied by $\boldsymbol{\theta}$, a vector of model parameters. In factor analysis, $\mathbf{\Sigma}(\boldsymbol{\theta})$ is a function of latent factor variances and covariances, factor loadings and measurement errors. In notational form:

$$\mathbf{\Sigma}(\boldsymbol{\theta}) = \mathbf{\Lambda}\mathbf{\Psi}\mathbf{\Lambda}^t + \mathbf{\Theta}_e$$

where $\mathbf{\Lambda}$ is a matrix of factor loadings, $\mathbf{\Psi}$ is a matrix of variances and covariances among the latent factors and $\mathbf{\Theta}_e$ is a matrix of measurement errors. This model assumes that the model is properly specified, $\mathbf{\Theta}_e$ is independent of the latent factors, and the measurement errors are independent of one another (ie. $\mathbf{\Theta}_e$ is a diagonal matrix).

Due to the categorical nature of the indicators, the maximum likelihood (ML) method cannot be used to estimate the model parameters in $\boldsymbol{\theta}$. In particular, the ML method based on the sample product-moment covariance matrix does not perform well when the number of categories within indicators is five or less [29]. Using the ML method with categorical indicators can lead to an inflated χ^2 test statistic, underestimated parameters, and negatively biased standard errors [29].

As an alternative, a matrix of polychoric correlations is used. A polychoric correlation estimates the linear relationship between two unobserved continuous variables, given two observed categorical variables. For factor analysis, this assumes that the survey indicators measure an unobserved theoretical continuous construct, and that the observed responses are discrete realizations based on the number of categories within the indicators. The calculation of the polychoric correlation is then based on the assumption of an underlying continuous distribution, y^*, known as the latent response distribution. The relationship between y, the observed discrete distribution, and y^* is $y = c$ if $\tau_c < y^* < \tau_{c+1}$, where τ is a threshold parameter defining the categories $c = 0, 1, 2, ..., C-1$. As such, y changes when τ is exceeded by y^*.

Polychoric correlations can be calculated in two steps: 1) estimate the threshold parameters for each latent response variable and 2) use the estimated threshold parameters along with the observed bivariate contingency table to estimate the latent correlation between the observed variables [30]. The two steps are outlined in more detail below.

Step 1: In order to estimate the threshold parameters, the proportion of observations in each category of an indicator are used. For observed variables y_1 and y_2, the thresholds are:

$$\begin{aligned}\tau_1 &= \Phi_1^{-1}(P_{i,\cdot})\\ \tau_2 &= \Phi_1^{-1}(P_{\cdot,j})\end{aligned}$$

where τ_1 and τ_2 are the thresholds for y_1 and y_2 respectively, P_{ij} is the observed proportion in cell (i, j), $P_{i\cdot}$ and $P_{\cdot j}$ are observed cumulative marginal proportions of the contingency table of y_1 and y_2, and Φ_1 is the univariate standard normal cumulative distribution function (CDF).

Step 2: In order to estimate the correlation between the two latent response variables, the estimated threshold parameters are used along with the observed univariate and bivariate contingency tables. For binary variables y_j and y_k regressed on covariate vector $\boldsymbol{x}$, the log likelihood of the univariate-response probit regression for individual i with variable y_j is:

$$\ell_{ij} = y_{ij} log P(y_{ij} = 1|\boldsymbol{x}_i) + (1 - y_{ij}) log P(y_{ij} = 0|\boldsymbol{x}_i)$$

The log likelihood of the bivariate-response probit regression for individual i with variables y_j and y_k is:

$$\begin{aligned}\ell_{ijk} &= y_{ij}y_{ik}logP(y_{ij}=1,y_{ik}=1|\boldsymbol{x}_i)\\&\quad+y_{ij}(1-y_{ik})logP(y_{ij}=1,y_{ik}=0|\boldsymbol{x}_i)\\&\quad+(1-y_{ij})y_{ik}logP(y_{ij}=0,y_{ik}=1|\boldsymbol{x}_i)\\&\quad+(1-y_{ij})(1-y_{ik})logP(y_{ij}=0,y_{ik}=0|\boldsymbol{x}_i)\end{aligned}$$

Then $\boldsymbol{s}$, a vector of polychoric correlations for a set of p binary variables, $\boldsymbol{y}$, regressed on a set of q covariates, $\boldsymbol{x}$, is the solution to:

$$\mathbf{0}=\sum_{\mathbf{i=1}}^{\mathbf{n}}\begin{pmatrix}\ell_{i1}/\partial\tau_1\\\partial\ell_{i1}/\partial\boldsymbol{\pi}_1\\\partial\ell_{i2}/\partial\tau_2\\\partial\ell_{i2}/\partial\boldsymbol{\pi}_2\\\vdots\\\partial\ell_{ip}/\partial\tau_p\\\partial\ell_{ip}/\partial\boldsymbol{\pi}_p\\\partial\ell_{i21}/\partial\rho_{21}\\\vdots\\\partial\ell_{ipp-1}/\partial\rho_{pp-1}\end{pmatrix}$$

where $\boldsymbol{\pi}_j$ is the q-dimensional vector of probit slopes for variable y_j, and ρ_{jk} is the residual correlation between variables y_j and y_k [31].

Maximum likelihood (ML) estimates for the τ and $\boldsymbol{\pi}$ elements are obtained from the univariate-response probit regressions. Next, holding the estimated values of τ and $\boldsymbol{\pi}$ fixed, the ML solution for ρ is obtained from bivariate-response probit regressions. [32] found that the estimation of polychoric correlations is robust to modest violations of underlying normality, such as skewness in the distribution of categorical responses. As such, $\boldsymbol{s}$ can then be used with a weighted least squares (WLS) approach to model fitting. The fitting function is:

$$F_{WLS}=[\mathbf{s}-\boldsymbol{\sigma}(\boldsymbol{\theta})]^{\mathbf{t}}\mathbf{\Gamma}^{-\mathbf{1}}[\mathbf{s}-\boldsymbol{\sigma}(\boldsymbol{\theta})]$$

where $\mathbf{s}$ is the vector of polychoric correlations, $\boldsymbol{\sigma}(\boldsymbol{\theta})$ is the model-implied vector of population elements in $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ that attempt to match $\mathbf{s}$, and $\boldsymbol{\Gamma}$ is the asymptotic covariance matrix of $\mathbf{s}$. For the specific formulation of $\boldsymbol{\Gamma}$, see [31].

The disadvantages of using WLS estimation for factor analysis with categorical indicators is that the dimensions of $\boldsymbol{\Gamma}$ are large and increase rapidly with an increase of indicators in the model. As such, $\boldsymbol{\Gamma}$ is often nonpositive-definite and cannot be inverted [33]. Also, it is important that $\boldsymbol{\Gamma}$ be simple since it has to be inverted for each iteration [31].

A solution to this problem is robust WLS, presented by [31]. For robust WLS, $\boldsymbol{\Gamma}$ in the WLS fitting function is replaced with $\boldsymbol{W}$, a diagonal matrix filled with the diagonal elements in $\boldsymbol{\Gamma}$. When calculating the χ^2 test statistic and standard errors in the factor model, the full $\boldsymbol{\Gamma}$ is used but is not inverted.

In factor analysis with categorical indicators, factor coefficients are logistic regression coefficients. Squaring a completely standardized factor coefficient provides an estimate of its *communality*, or how much variability in a latent response variable is explained by the latent factor. The residual variance in factor analysis with categorical indicators is not estimated because the latent response variable, y^*, does not have a metric. As such, the measurement error of a latent response variable is not a freely estimated parameter and is equal to 1 minus its squared completely standardized factor loading(s).

Depending on the field of study, a factor coefficient's cutoff level of significance can vary. A factor coefficient of greater than or equal to 0.40 in magnitude is considered significant in the social science field [34], while factor coefficients of less than 0.30 in magnitude are undesirable in the social science field [35]. An insignificant factor coefficient indicates that an indicator is not strongly correlating with the other indicators in the factor, and that an additional factor may need to be considered [34]. A negative factor coefficient indicates that an indicator is negatively correlating with the other indicators in the factor. Deleting indicators that have low factor coefficients may increase the reliability of an instrument [36].

Commonly, the output from a factor analysis contains both a pattern matrix and a structure matrix. The pattern matrix contains the factor loadings, while the structure matrix is equivalent to the pattern matrix multiplied with the factor correlation matrix. In applied research, the pattern matrix is most often interpreted and reported.

2.3.2 Exploratory Factor Analysis with categorical indicators

In Exploratory Factor Analysis (EFA), there are no a priori model assumptions, and all indicators freely load onto all factors. EFA with categorical indicators is performed in four steps: i.) developing the correlation matrix, S, ii.) extracting the initial factors, iii.) deciding how many factors to retain, and iv.) rotating to a final solution [36, 37]. The goal of EFA is to find a sufficient number of factors that explain the correlations between the indicators. When extracting the initial factors, the unique variance of a variable is separated from the common variance, and factor analysis is performed on only the common variance [34]. The common variance between variables is equal to the correlation coefficient squared [38]. Unique variance is thus equal to the total variance minus the common variance.

Choosing the number of factors to retain requires a balance between having a parsimonious model and a plausible model [39]. Selecting too few factors is generally considered worse than selecting too many factors because too few factors can distort the results of the analysis by combining two factors into one. Selecting too many factors can lead to an unnecessarily complex model with unstable factors [40].

One factor selection method is the Scree test, which involves examining a plot of the descending order of the eigenvalues from the correlation matrix. The eigenvalues of the correlation matrix are the values of λ that solve the equation $\boldsymbol{\Sigma x} = \lambda \boldsymbol{x}$ for nonzero $\boldsymbol{x}$ [41]. The sum of the squared factor loadings within a factor is equivalent to the factor's eigenvalue. The point in which the curve of eigenvalues flattens out or breaks is considered to be the cutoff, and the number of data points that are above the cutoff is equal to the number of factors that remain [34]. The disadvantage of the Scree test is that it is vulnerable to human error and subjectivity.

A different factor selection method is Parallel Analysis, which is a Monte Carlo simulation technique. Parallel analysis seeks to minimize underextraction, which occurs when too few factors are extracted, and overextraction, which occurs when too many factors are extracted [40]. In Parallel Analysis, random, uncorrelated Gaussian samples are simulated based on the existing data, and the mean of the eigenvalues are plotted against the eigenvalues of the original data. A factor is thus considered significant if the original eigenvalue is greater than the eigenvalue obtained from the Monte Carlo simulation.

Factor rotation aids in the interpretation of an EFA. The rotated factor pattern matrix, $\boldsymbol{\Lambda}$, is equal to $\boldsymbol{AT}$, where $\boldsymbol{A}$ is the input factor loading matrix and $\boldsymbol{T}$ is a matrix $\boldsymbol{A}$ is postmultiplied with to minimize the complexity function, $f(\boldsymbol{\Lambda})$. For the *geomin* rotation, which is used in this thesis, the complexity function is calculated as:

$$f(\boldsymbol{\Lambda}) = \sum_{i=1}^{p} (\prod_{j=1}^{m} \lambda_{ij}^2)^{1/m}$$

where p is the sample size, and m is the number of indicators [42]. Factor rotation aids interpretability of results by maximizing factor loadings close to 1.0 and minimizing factor loadings close to 0.0 [43]. As such, rotation does not affect indicators' communalities.

Two types of data rotation for the correlation matrix exist. Varimax rotation is the most common, and belongs to a family of orthogonal rotations that produce uncorrelated factors [34]. The factor loadings resulting from orthogonal rotations are easy to interpret, but may oversimplify the factor solution. As [34] argue, behaviors naturally correlate in the social sciences, and expecting to partition behaviors into separate factors is a flawed assumption. Oblique, or promax, rotation allows factors to correlate among each other. The disadvantage of using an oblique rotation is that factor loadings may not represent correlation coefficients due to possible inflation caused by covariance between factors [43]. If all correlations are close to zero for a set of data, oblique rotation will have similar results to orthogonal rotation [39]. *Geomin* rotation is a type of oblique rotation.

The ideal number of factors will lead to the "cleanest" factor structure, where all indicators load significantly, few or none of the indicators crossload, and no factor has fewer than three indicators [34]. A crossloading indicator is a question that loads significantly onto two or more factors. A researcher has two choices when presented with a crossloading indicator: either remove it from the instrument, or change the indicator as the researcher sees fit. A factor with three or fewer indicators is considered unstable, and an ideal factor has five or more indicators [34].

2.3.3 Confirmatory Factor Analysis with categorical indicators

Confirmatory Factor Analysis (CFA) requires a strong theoretical or conceptual background of an instrument in order to guide the adjustments and evaluation of the factor model. The researcher must pre-specify all aspects of the factor model prior to the analysis: the number of factors, the pattern of factor loadings, and the corresponding error theory. For stability in model solutions, there should be a minimum of three indicators assigned to each factor [43].

In CFA, there is no factor rotation due to most or all indicator crossloadings being fixed to zero. Through iterations, CFA attempts to reproduce the observed relationships between indicators with fewer parameter estimates than in EFA. Due to this, CFA models are more parsimonious than EFA solutions, and factor coefficients resulting from CFA are of larger magnitude than from EFA. Another advantage of

CFA over EFA is that in CFA, correlated measurement error can be modeled [43]. An example of correlated measurement error is two indicators that covary for reasons other than the shared influence of the latent factor, such as being similarly worded or addressing the same topic.

2.4 Model Evaluation

Evaluation of the fitted factor analysis solution is based on three aspects: overall goodness-of-fit, assessing the presence of localized areas of strain, and the interpretability/ significance of parameter estimates.

2.4.1 Goodness-of-Fit Indices

Goodness-of-fit indices evaluate how well a parsimonious model can reproduce an original set of data. In factor analysis, goodness-of-fit indices are based on the unstandardized model parameters.

Common goodness-of-fit indices which are used in factor analyses fit into three categories: absolute fit, parsimonious model adjusted fit and comparative fit. At least one index from each category should be used when evaluating model fit [43].

The mean- and variance-adjusted χ^2 test statistic, G_{MV}, is an absolute fit index that evaluates the hypothesis that $\mathbf{\Sigma}(\boldsymbol{\theta})$ equals $\mathbf{\Sigma}$ [31]. G_{MV} is defined as follows:

$$G_{MV} = [d/tr(\boldsymbol{U\Gamma})^2]nF(\boldsymbol{\theta})$$

where $\boldsymbol{U}$ is equal to $\boldsymbol{W}^{-1} - \boldsymbol{W}^{-1}\mathbf{\Delta}(\mathbf{\Delta}^t\boldsymbol{W}^{-1}\mathbf{\Delta})^{-1}\mathbf{\Delta}^t\boldsymbol{W}^{-1}$, d is the integer closest to $(tr(\boldsymbol{U\Gamma}))^2/tr((\boldsymbol{U\Gamma})^2)$, ,$\mathbf{\Delta}$ is equal to $\partial\boldsymbol{\sigma}(\boldsymbol{\theta})/\partial(\boldsymbol{\theta})$, $\mathbf{\Gamma}$ is the asymptotic covariance matrix of $\boldsymbol{s}$, n is the sample size and $F(\boldsymbol{\theta})$ is the minimum of the fitting function, F_{WLS}.

The root mean square error of approximation (RMSEA) is a parsimony correction of the χ^2 index of fit and assesses the extent to which $\mathbf{\Sigma}(\boldsymbol{\theta})$ is reasonably close to

$\mathbf{\Sigma}$. RMSEA is equal to $\sqrt{d/df}$ where $d = (\chi^2 - df)/n$, n is the sample size and df is the model degrees of freedom. The range of RMSEA is from 0 to positive infinity. Smaller values of RMSEA indicate better model fit.

Bentler's comparative fit index (CFI) evaluates the fit of F_{WLS} in relation to the fit of a "null" model in which the covariances among the input variables are fixed to zero. Bentler's CFI is calculated as follows:

$$CFI = \frac{1 - max[(\chi_T^2 - df_T), 0]}{max[(\chi_T^2 - df_T), (\chi_B^2 - df_B), 0]}$$

where χ_T^2 is the χ^2 value of the target model, or the model being evaluated, and χ_B^2 is the χ^2 value of the baseline "null" model. Bentler's CFI can range from 0.0 to 1.0, with larger values indicating better model fit.

Simulation studies by [44] indicate that good fit is obtained when RMSEA is close to or below 0.6 and Bentler's CFI is close to or greater than 0.95.

2.4.2 Localized Areas of Strain

While goodness-of-fit statistics provide an overall descriptive evaluation, they do not reveal localized areas of strain. In CFA, modification indices can be used to evaluate specific relationships in the factor solution. A modification index is an estimate of how much the overall model's χ^2 would decrease if the constrained parameter was freely estimated, and is roughly equivalent to comparing two nested models with 1 degree of freedom difference. At an alpha level of 0.05, the critical value of χ^2 with 1 degree of freedom is 3.84. Modification indices that are greater than or equal to 3.84 indicate that model fit can be significantly improved [43].

Modification indices can be calculated across all constrained parameters, and as such, model respecifications should be based in theory. Two parameters should only be freely estimated if they theoretically covary. Parameters should not be freely estimated for the sole purpose of improving goodness-of-fit indices. Adding a parameter estimate to a CFA model that had a borderline modification index can result in unstable factor loadings [43]. Also, it is important to be consistent in decision rules that are used to add a freely estimated parameter [43].

2.4.3 Interpretability and Significance of Parameter Estimates

In applied research, completely standardized parameter estimates are commonly reported. However, the unstandardized solution is used to calculated standard errors and goodness-of-fit indices. Parameter estimates are defined as "salient" when they load at 0.30 or above [43]. Too large or too small standard errors are problematic and can result in large parameter estimates that are not significant.

Latent factor parameter estimates can be negative due to the factors being inversely related. Small or statistically insignificant latent factor covariances are typically retained in the solution because they provide evidence of discriminant validity. If latent factor covariance are 0.80 or above, this indicates poor discriminant validity, and the researcher should consider combining the latent factors [43].

In CFA, a problematic factor loading may indicate one of several model misspecifications: the indicator may have been specified to load onto the incorrect factor; the indicator may have been specified to load onto only one factor, but actually should load onto two or more; or the indicator may have been specified to load onto one factor, but has no relationship to any factor. The solution is to either respecify the CFA model or to remove the indicator [43].

2.4.4 Jöreskogs's Estimate of Scale Reliability

Scale reliability, denoted ρ, is equal to $var(T)/var(Y)$ where var(T) is the true score variance, or shared variance among indicators, and var(Y) is the total variance of the measure, which is equal to the true score variance plus the error variance of the measure. Using parameters from the factor analysis model, the scale reliability can be calculated as follows:

$$\rho = (\sum_{i=1}^{p} \lambda_i)^2 / [(\sum_{i=1}^{p} \lambda_i)^2 + \sum_{i=1}^{p} \theta_{ii} + 2 \sum_{i=1}^{p} \theta_{ij}]$$

where λ_i are the unstandardized factor loadings, θ_{ii} are the unstandardized measurement error variances, and θ_{ij} are the measurement error covariances.

A reliability coefficient of at least 0.80 for an established instrument is considered satisfactory, while a coefficient of at least 0.70 is considered satisfactory for a newly established instrument [36, 45].

Scale reliability can be affected by certain traits in an instrument. If a survey has too few indicators, the coefficient may be low. Increasing the number of indicators in a survey may improve the coefficient, but the disadvantage of adding indicators may outweigh any gains. A long survey is a burden on participants, and the longer a survey is, the less likely participants will complete the survey [36].

2.5 Other Statistical Methods

This section contains information about the other statistical methods used in this thesis.

2.5.1 Student's t Test

The Student's t test is a commonly used inferential statistic that is used to compare the means of two groups. In the case of repeated measures, where the same participants are tested twice on the same instrument, a paired difference test is used. The variance of the pre data and the variance of the post data are assumed to be equal. The t statistic for dependent, paired data is:

$$t = \frac{\overline{x} - \mu_D}{s/\sqrt{n}}$$

where $\overline{x}$ is the difference between the pre and post sample means, μ_D is the true difference between the population means, s is the standard deviation of the difference, and n is the sample size. The two-sided hypothesis is:

$$H_0 : \mu_D = 0$$
$$H_1 : \mu_D \neq 0$$

The null hypothesis is rejected when the p-value is less than α. For a two-sided test under the assumption that the null hypothesis is true, the p-value is equal to:

$$2 * P(T_{n-1} \geq |t|)$$

2.5.2 Tukey's Honest Significant Differences

Tukey's Honest Significant Differences (HSD) is one method that can be used to control the familywise error rate among comparisons of treatment means. The test statistic for comparing two treatment means is:

$$t = \frac{\overline{y}_i - \overline{y}_j}{s\sqrt{2/n}}$$

where $\overline{y}_i$ is the sample mean for treatment i, $\overline{y}_j$ is the sample mean for treatment j, s is the standard deviation, and n is the sample size. The test statistic is compared to the Studentized range distribution, and μ_i is significantly different from μ_j if:

$$|t| > q_{a,(N-a),1-\alpha}/\sqrt{2}$$

where a is the total number of treatments, N is the total sample size and α is the allowable error rate.

2.5.3 Multiple Linear Regression

Multiple linear regression (MLR) is a statistical method that can be used to linearly model one quantitative variable as a mix of qualitative and quantitative variables. Since MLR examines correlations between variables, the resulting regression model does not imply causation. The assumptions for an MLR are i.) observations are independent, ii.) the mean of the response variable, Y_i, is a linear function of the explanatory variables, x_i, iii.) the Y_i's have a common variance of σ^2 for all values of x_i, and iv.) the random errors, ϵ_i, are normally distributed.

The goal of regression analysis is to find the most parsimonious model that adequately explains the data. The test statistic used for testing if a reduced model is adequate compared to the full model is:

$$F = \frac{SSE_R - SSE_F}{(df_R - df_F)/MSE_F}$$

where SSE is the sum of squared error, *df* is the degrees of freedom and MSE is the mean squared error. The formulas for SSE and MSE are as follows:

$$SSE = \sum_{i=1}^{n}(Y_i - \overline{Y})^2$$

$$MSE = \text{SSE/(n-p)}$$

where n is the sample size and p is the number of estimated parameters. The regression degrees of freedom are (p - 1).

2.5.4 Wilcoxon Signed Rank Test

The Wilcoxon signed rank test is a nonparametric median test that is used for small sample sizes. The two-sided hypothesis is as follows:

$$H_0 : Q(0.50) = Q_0(0.50)$$
$$H_1 : Q(0.50) \neq Q_0(0.50)$$

where $Q_0(0.50)$ is the benchmark median value being tested. The sample data, which is assumed to be symmetric, is ranked from smallest to largest, ignoring negative signs. Then, the sum of the ranks of the positive differences, denoted w^+, are compared to the sum of the ranks of the negative differences, denoted w^-. The smaller of these two sums is compared to the probability of attaining such a sum from a symmetrically distributed population centered around the benchmark median value. The null hypothesis is rejected when w^+ or w^- is sufficiently large.

2.5.5 Shapiro-Wilks Test of Normality

The Shapiro-Wilks test of normality tests whether or not a sample of data comes from a normal distribution. More specifically, the following hypothesis is tested:

$$H_0 : x_1, x_2, ..., x_n \text{ comes from a normal distribution}$$
$$H_1 : x_1, x_2, ..., x_n \text{ does not come from a normal distribution}$$

The test statistic is calculated as follows:

$$W = \frac{(\sum_{i=1}^{n} a_i x_{(i)})^2}{(\sum_{i=1}^{n} (x_i - \overline{x})^2)}$$

where a_i is a constant and $x_{(i)}$ is the i^{th} order statistic. When W is sufficiently large, the null hypothesis is rejected.

2.5.6 Fligner-Killeen Test

The Fligner-Killeen test of homogeneity of variance is a non-parametric test that is robust to deviations from normality in data, and tests the following hypotheses:

$H_0 : k$ samples from $x_1, x_2, ..., x_n$ have homogeneous variance

$H_1 : k$ samples from $x_1, x_2, ..., x_n$ do not have homogeneous variance

For the specific formulation of the Fligner-Killeen test statistic, see [46].

CHAPTER 3

METHODS

This chapter provides an overview of information concerning the SPIRIT program's sponsoring university, the evaluation site, the subjects involved in the study and the recruitment process. Next, this chapter discusses the IT attitude survey, the Focus Group interviews, and the types of variables and statistical analyses used in this investigation.

3.1 Sponsoring University

The department of Computer and Information Technologies (CIT) at Purdue University teaches students how to apply IT to problem solving in industry and business. The SPIRIT program was administered by the CIT department at Purdue. All SPIRIT data was collected at Purdue University under the direction of Alka Harriger, professor and associate department head in CIT.

3.2 Evaluation Site

All data collected during the SPIRIT program was sent to Colorado School of Mines (CSM) for evaluation. Participants in SPIRIT remained anonymous to the evaluation team at CSM. The data was analyzed under the direction of Dr. Barbara Moskal, director of the Center for Assessment in STEM at CSM.

3.3 SPIRIT

The SPIRIT program was designed to increase knowledge of and improve attitudes toward careers in IT at the high school level. The SPIRIT program targeted high school teachers, counselors and students, with an emphasis on female students. During the summer workshops, SPIRIT sought to demonstrate to high school students that IT can be fun and useful for problem solving, while showing that an IT college degree can support a vast array of career options. SPIRIT used Alice versions 2.0 and 2.2 as a tool to support visual and animated displays of information. In addition, high school teachers used Storytelling Alice during the summer of 2008. Due to problems with the software, Storytelling Alice was not used in subsequent years.

During the summer workshops, pre and post IT Attitude Surveys and Concept Exams were collected, along with End of Program Evaluations. Focus Groups were solely administered during the summer workshops. Academic year data included post-post IT Attitude Surveys, post-post Concept Exams, and post-post End of Program Evaluations.

3.3.1 Subjects

Table 3.1 contains a breakdown of the number of participants who fully completed the SPIRIT program per year. Counselors were combined with teachers in the summer 2010 workshops, and were thus treated as teachers in subsequent analyses. Demographic data was also collected from participants in SPIRIT. Due to small sample sizes, statistical analyses comparing differences based on teachers' and counselors' ethnicity were not considered in this thesis.

Table 3.1: Demographic Data for SPIRIT Participants

Year	Participant	Total	Male	Female	Non-Minority	Minority
Summer 2008	Teachers	18	5	13	15	3
	Counselors	8	2	6	7	1
	Students	59	23	36	39	20
Summer 2009	Teachers	18	7	11	17	1
	Counselors	9	1	8	6	3
	Students	71	21	50	44	27
Summer 2010	Teachers	16	8	8	16	0
	Students	52	16	36	37	15

3.3.2 SPIRIT Recruitment

Although female high school students were the primary target audience of the SPIRIT program, SPIRIT was open to both genders and approximately one-third of student participants were male. Students in Indiana and neighboring states were recruited through the use of postal and electronic mail, the SPIRIT website and personal visits. The ideal student participant had not yet considered IT as a potential career path. The SPIRIT project sought to persuade students to attend SPIRIT through incentive programs, such as earning daily monetary stipends. Advertisements for SPIRIT emphasized the "fun" nature of the program [5].

High school teachers were recruited through publicity efforts, conferences and other forms of communication. SPIRIT sought teachers from different locations that instructed a variety of subjects and had the potential of reaching many students, especially female students [5].

3.4 Instrument

The Attitude Survey, which was created and validated under NSF grant DUE-0512064, was originally developed to assess changes in attitudes regarding CS [25]. For SPIRIT, a slightly modified version of the attitude survey was implemented. In order for the survey statements to reflect IT rather than CS, the term "computer science" was replaced with the term "information technology" throughout the instrument.

For teachers and counselors, the IT attitude survey consisted of 47 statements. For students, the IT attitude survey consisted of 52 statements in the summer of 2008, but was reduced to 20 statements based on a factor analysis by [47], which is discussed in further detail in the Document Analysis of the Results section. All statements in the IT attitude survey were assessed on a 4-point Likert scale with the following categories: strongly agree (SA), agree (A), disagree (D), or strongly disagree (SD). In order to force participants to make a positive or negative decision, a neutral category was not included. Statements in the survey were both positively and negatively worded. In order to quantitatively score the survey results, individual responses to a survey statements were coded from 0 to 3, where a higher number reflected a more positive attitude to IT or gender neutrality. For a positively worded statement, SD was coded as 0, D was coded as 1, A was coded as 2 and SA was coded as 3. For a negatively worded statement, SD was coded as 3, D was coded as 2, A was coded as 1 and SA was coded as 0. A positive attitude was always reflected through a higher score. A factor score within the IT attitude survey was found by summing all statement scores that fell into a single factor. All versions of the IT attitude survey can be found in the appendix.

All participants in SPIRIT completed a pre-attitude survey before the program began. Teachers completed a post-attitude survey following the two week summer workshop. Counselors and students completed a post-attitude survey following the one week summer workshop. Additionally, teachers and counselors completed a post-post-attitude survey at the end of the academic year, after following a year of implementing Alice in high schools.

3.5 Focus Groups

The focus group interviews served as a qualitative assessment of changes in attitudes and program implementation across the years. Interviews were administered by external evaluators during a three day period near the end of the summer workshops, and each interview lasted for approximately thirty minutes. Participants' names and any personal identifying information was blinded from the evaluators, and no workshop personnel were permitted to administer or be present at any of the interviews. Participant responses were captured through note taking.

Responses to the focus group interviews were evaluated through emergent themes where an evaluator grouped similar responses into categories, and an independent evaluator matched a sub-sample of the responses to the categories to ensure that there was consistency in the grouping.

Since SPIRIT participants objected to the wording of the gender-biased survey statements during summers 2008 and 2009, the focus group interviews during the summer of 2010 provided an opportunity to ask participants whether or not they felt pressured or influenced to provide the answer they should, rather than an honest answer.

3.6 Qualitative and Quantitative Variables

In this thesis, variables were categorized as being qualitative or quantitative, and explanatory or response in nature. Table 3.2 contains the variables and the statistical analyses used to summarize the variables. For example, participants' gender was a qualitative explanatory variable that was incorporated into the linear regression analyses.

Table 3.2: Qualitative and Quantitative Variables

Methodology	Variable	Data	Analysis
Qualitative	Explanatory	Gender Ethnicity Year in Program	Linear Regression
	Response	Focus Groups Program Evaluations	Emergent Category Analysis
Quantitative	Explanatory	Concept Exam Scores	Linear Regression
	Response	Attitude Survey Scores	Factor Analysis Linear Regression

3.7 Statistical Analysis Techniques

This section provides the background information used for the statistical analyses in this thesis. The research that supports the use of these techniques is discussed in the Literature Review.

Inconsistent responses from participants were not included in any of the analyses in this thesis. Participants' inconsistency was determined by comparing responses to positively- and negatively-worded equivalent statements in the IT attitude survey. After assigning higher scores to more positive or gender neutral responses, if a response to a positively-worded statement differed by more than 1 Likert scale point to the negatively-worded equivalent, the participant was removed.

3.7.1 Factor Analysis

For the teacher and counselor version of the IT attitude survey, five pre-existing factors based on previous theory were defined:

- Confidence: confidence in their own ability to learn computing skills
- Interest: interests in computing
- Gender: perceptions of computing as a male field
- Usefulness: beliefs in the usefulness of learning computing
- Professional: beliefs about professionals in computing

For the student version of the IT attitude survey, two pre-existing factors based on previous theory were defined:

- IT: Beliefs about IT in general
- Gender: Beliefs about gender stereotypes in IT

For the teacher and counselor IT attitude survey, EFA was used to explore the factor structure. For the student IT attitude survey, CFA was used to confirm the established factors.

3.7.2 Linear Regression Analysis

For this thesis, the response variable of interest was change in factor scores from the IT attitude survey, and the explanatory variables of interest were gender, ethnic group, year of implementation and change in Concept Exam scores. In order to answer the second research question which assessed whether or not there were significant changes in attitude for participants, paired data t-tests were used to test for the difference in mean IT attitude survey factor scores.

For the third research question, which assessed whether or not subgroup differences such as gender or ethnicity were significant, separate MLRs were run for teachers, counselors and students, respectively. For teachers and counselors, two different linear models were selected, where the first linear model examined pre to post differences, and the second linear model examined pre to post-post differences. The pre-to-post linear model equation for teachers was:

$$
\begin{aligned}
Y_i = \beta_0 &+ \beta_1 * gender_i + \beta_2 * year1_i \\
&+ \beta_3 * year2_i + \beta_4 * concept1_i \\
&+ \beta_5 * (gender * year1)_i + \beta_6 * (gender * year2)_i \\
&+ \beta_7 * (gender * concept1)_i + \beta_8 * (year1 * concept1)_i \\
&+ \beta_9 * (year2 * concept1)_i + \epsilon_i
\end{aligned}
\tag{3.1}
$$

where Y_i was the change in factor score from the IT attitude survey for the ith teacher participant, the β_i's were the linear coefficients, and the ϵ_i's were the normally distributed random errors. The variable *concept1* measured the difference between pre and post Concept Exam scores. Only second order interactions were included in this model due to small sample sizes and for simplicity in interpretation. Gender and year of implementation were indicator variables in this model, and were defined as follows:

$$gender_i = \begin{cases} 1 & \text{if male} \\ 0 & \text{if female} \end{cases}$$

$$year1_i = \begin{cases} 1 & \text{if attended 2008} \\ 0 & \text{otherwise} \end{cases}$$

$$year2_i = \begin{cases} 1 & \text{if attended 2009} \\ 0 & \text{otherwise} \end{cases}$$

The pre-to-post-post linear model equation for teachers is:

$$\begin{aligned} Y_i &= \beta_0 + \beta_1 * gender_i + \beta_2 * year1_i \\ &+ \beta_3 * year2_i + \beta_4 * concept1_i \\ &+ \beta_5 * concept2_i + \beta_6 * (gender * year1)_i \\ &+ \beta_7 * (gender * year2)_i + \beta_8 * (gender * concept1)_i \\ &+ \beta_9 * (gender * concept2)_i + \beta_{10} * (year1 * concept1)_i \\ &+ \beta_{11} * (year1 * concept2)_i + \beta_{12} * (year2 * concept1)_i \\ &+ \beta_{13} * (year2 * concept2)_i + \epsilon_i \end{aligned} \tag{3.2}$$

where the variable *concept2* measured the difference between pre and post-post Concept Exam scores.

For counselors, the pre-to-post linear model equation was:

$$\begin{aligned} Y_i &= \beta_0 + \beta_1 * gender_i + \beta_2 * year1_i \\ &+ \beta_3 * concept1_i + \beta_4 * (gender * year1)_i \\ &+ \beta_5 * (gender * concept1)_i + \beta_6 * (year1 * concept1)_i \\ &+ \epsilon_i \end{aligned} \tag{3.3}$$

The indicator variables in this model were defined the same as above. The pre-to-post-post linear model equation for counselors was:

$$
\begin{aligned}
Y_i = \beta_0 &+ \beta_1 * gender_i + \beta_2 * year1_i \\
&+ \beta_3 * concept1_i + \beta_4 * (concept2)_i \\
&+ \beta_5 * (gender * year1)_i + \beta_6 * (gender * concept1)_i \\
&+ \beta_7 * (gender * concept2)_i + \beta_8 * (year1 * concept1)_i \\
&+ \beta_9 * (year1 * concept2)_i + \epsilon_i
\end{aligned} \tag{3.4}
$$

The full linear model equation for students was:

$$
\begin{aligned}
Y_i = \beta_0 &+ \beta_1 * gender_i + \beta_2 * ethnicity_i \\
&+ \beta_3 * year1_i + \beta_4 * year2_i + \beta_5 * concept1_i \\
&+ \beta_6 * (gender * ethnicity)_i + \beta_7 * (gender * year1)_i \\
&+ \beta_8 * (gender * year2)_i + \beta_9 * (gender * concept1)_i \\
&+ \beta_{10} * (ethnicity * year1)_i + \beta_{11} * (ethnicity * year2)_i \\
&+ \beta_{12} * (ethnicity * concept1)_i + \beta_{13} * (year1 * concept1)_i \\
&+ \beta_{14} * (year2 * concept1)_i + \epsilon_i
\end{aligned} \tag{3.5}
$$

Ethnicity was an indicator variable in this model and was defined as follows:

$$
ethnicity_i = \begin{cases} 1 & \text{if minority} \\ 0 & \text{if non-minority} \end{cases}
$$

Nested F tests were used to simultaneously test the significance of multiple explanatory variables in the regression models.

In order to assess the fourth research question, which involved examining differences in linear model coefficients across participant type, the IT attitude survey for teachers and counselors was reduced to an equivalent version of the student IT attitude survey. A separate EFA was run for this student-equivalent version in order to examine the factor structure. The linear regression for teachers and students was modeled as in Equation 3.1. The linear regression for counselors was modeled as in Equation 3.3.

The explanatory variables were standardized in order to compare linear regression model coefficents across different models.

3.7.3 Implementation Analysis

SPIRIT was not a controlled experiment, where influencing variables were carefully monitored and measured. Due to this limitation, there may have been additional variables that were not considered or measured that affected IT attitude survey scores.

Analyzing and recording differences in SPIRIT across the years may provide possible explanations for deviations or irregular results in the data that may have not been previously considered. In particular, each summer program had different presenters and activities. Additionally, part of this analysis consisted of examining emergent themes from the Focus Group interviews. Examining differences in emergent themes across the years may indicate how changes in program implementation affected participants, and may provide another layer of evidence to support the results in this thesis.

3.7.4 Document Analysis

Analyzing and recording changes made to the IT attitude survey may provide possible explanations for deviations or irregular results in the data that may have not been previously considered. The reduction of the student version of the IT attitude survey by [47] is discussed, and well as the effects this reduction may have had on the student participants.

3.8 Study Limitations

Participants in the SPIRIT program were volunteers and as a consequence, results from this thesis may not generalize to the broader population. The SPIRIT program design did not include a control group with which SPIRIT participants could be compared. Due to potential volunteer bias and the absence of a control group, it is difficult to measure the extent to which attitudes regarding IT change throughout this study, and causality cannot be inferred. Also, since the SPIRIT program had a clear goal of improving attitudes with respect to IT, this expectation may have influenced the participants' answers to the IT attitude survey.

3.9 Summary

This thesis examined the IT attitude survey data from SPIRIT with a factor analysis, multiple regressions, an implementation analysis, and a document analysis. The research questions of interest were:

1. Are the established factors for the student version and the theoretical factors for the teacher and counselor version of the IT attitude survey valid and reliable?
2. Based on factors underlying the IT attitude survey, was there a measurable difference in teachers', counselors' and students' attitudes with respect to IT immediately after, and for teachers and counselors, one year after completing the SPIRIT workshop?
3. Based on this same instrument, did attitudes differ across gender, ethnic groups and/or the three years of project implementation?
4. Based on this same instrument, how did regression analyses of the SPIRIT data differ across teachers, counselors and students?

A factor analysis of the IT attitude survey was run across three years of implementation for teachers, counselors and students, respectively. Given that an EFA had already been run on the student version of the IT attitude survey by [47], a CFA was used to confirm the latent factors. An EFA was run on the teacher and counselor version of the IT attitude survey. The factors that were extracted during factor analysis were further used in the regression analyses.

Linear regressions and paired t-tests were performed on IT attitude survey factor scores across three years of implementation for teachers, counselors and students, respectively. This analysis examined the second research question by determining if there was a measurable difference across teachers', counselors' and students' attitudes with respect to IT before the SPIRIT program and immediately after the SPIRIT program. Additionally, teachers and counselors were tested for differences in the IT attitude survey factor scores resulting from administration of the survey one year later.

The third research question was addressed by testing for the significance of variables containing gender, ethnic group and year of implementation in SPIRIT for each regression analysis with nested F tests. For the fourth research question, an additional factor analysis was performed on the reduced form of the students' IT attitude survey for teachers and counselors. After confirming that the factors were loading as expected, a regression analyses were performed, and results for teachers, counselors and students were compared.

The implementation analysis examined differences in program implementation across three years. The document analysis examines differences in the IT attitude survey across three years. These differences were considered when interpreting the factor analysis and subsequent regression analyses.

CHAPTER 4

RESULTS

This section contains the results of a document analysis of the IT attitude survey and an implementation analysis of the SPIRIT program, followed by the statistical analysis of student, teacher and counselor participant data.

4.1 Document Analysis

A promax EFA by [7] was performed on the pre student IT attitude survey with data from summer 2008. Only statements belonging to the Gender factor loaded as expected. None of the statements belonging to the Professional factor loaded on a single factor, indicating that students were unable to distinguish the defined factor with respect to IT [7]. Statements belonging to the Professional factor were removed and the EFA was re-run. Two factors emerged from this analysis, one with Gender statements, and the other with Confidence, Usefulness and Interest statements. Based on this result, the student IT attitude survey was reduced to 20 statements consisting of two factors: general interest in IT and gender stereotypes in IT. The reduced 20 statement IT attitude survey, which is displayed in Table 4.2 and divided into its factor constructs, was administered to student SPIRIT participants during summer 2009 and summer 2010.

The teacher and counselor survey remained at 47 statements throughout the SPIRIT program, as there was insufficient data to run an EFA during program administration. Table 4.1, Table 4.2, and Table 4.3 contain the original student version, the reduced student version, and the original teacher and counselor version of the attitude survey, divided into its factors. The appendix of this thesis contains the SPIRIT-administered versions of the IT attitude survey.

Table 4.1: 2008 Student IT Attitude Survey, divided by construct

Confidence Construct (C)	
C1	I am comfortable with learning computing concepts.
C2	I have little self-confidence when it comes to computing courses.
C3	I do not think that I can learn to understand computing concepts.
C4	I can learn to understand computing concepts.
C5	I have a lot of self-confidence when it comes to computing courses.
C6	I can achieve good grades (C or better) in computing courses.
C7	I am confident that I can solve problems by using computer applications.
C8	I am uncertain that I can achieve good grades (C or better) in computing courses.
C9	I am not comfortable with learning computing concepts.
C10	I doubt that I can solve problems by using computer applications.
Interest Construct (I)	
I1	I would not take additional information technology courses if I were given the opportunity.
I2	I think information technology is boring.
I3	I hope that my future career will require the use of information technology concepts.
I4	The challenge of solving problems using information technology does not appeal to me.
I5	I like to use information technology to solve problems.
I6	I do not like using information technology to solve problems.
I7	The challenge of solving problems using information technology appeals to me.
I8	I hope that I can find a career that does not require the use of information technology concepts.
I9	I think information technology is interesting.
I10	I would voluntarily take additional information technology courses if I were given the opportunity.
Gender Construct (G)	
G1	I doubt that a woman could excel in computing courses.
G2	Men are more capable than women at solving computing problems.
G3	Computing is an appropriate subject for both men and women to study.
G4	It is not appropriate for men to study computing.
G5	Women are more capable than men at solving computing problems.
G6	Women are more likely to excel in careers that involve computing than men are.
G7	Women produce higher quality work in computing than men.
G8	Women and men can both excel in careers that involve computing.
G9	I doubt that a man could excel in computing courses.
G10	It is not appropriate for women to study computing.

Table 4.1 – continued from previous page

Gender Construct (G)–continued	
G11	Men produce higher quality work in computing than women.
G12	Men are more likely to excel in careers that involve computing than women are.
G13	Women produce the same quality work in computing as men.
G14	Men and women are equally capable of solving computing problems.
G15	Men and women can both excel in computing courses.
Usefulness Construct (U)	
U1	Developing computing skills will not play a role in helping me achieve my career goals.
U2	Knowledge of computing will allow me to secure a good job.
U3	I use computing skills in my daily life.
U4	My career goals do not require that I learn computing skills.
U5	Developing computing skills will be important to my career goals.
U6	Knowledge of computing skills will not help me secure a good job.
U7	I do not use computing skills in my daily life.
U8	I expect that learning to use computing skills will help me achieve my career goals.
Professional Construct (P)	
P1	Doing well in information technology does not require a student to spend most of his/her time at a computer.
P2	A student who performs well in information technology will probably not have a life outside of computers.
P3	To do well in information technology, a student must spend most of his/her time at a computer.
P4	A student who performs well in information technology is likely to have a life outside of computers.
P5	Being good at information technology is a negative quality.
P6	Students who are skilled at information technology are less popular than other students.
P7	Being good at information technology is a positive quality.
P8	Students who are skilled at information technology are just as popular as other students.
P9	Students who are skilled at information technology are more popular than other students.

Focus Group interviewees during the summer 2010 program indicated that the gender-related statements in the IT attitude survey were misleading. For negatively worded statements, such as "Men produce higher quality work in computing than women," participants would have liked to respond neutrally, but there was no neutral category in the Likert scale. Also, participants commented that the gender-related statements seemed "old-fashioned." A few students mentioned that the IT attitude survey raised concern about gender-inequality in IT of which they were not previously aware. Many participants admitted to not knowing how to respond to the gender-related statements, resulting in blank responses in the survey. As a result, these surveys had to be discarded.

Table 4.2: Reduced Student IT Attitude Survey, divided by construct

General Interest Construct (I)	
I1	I have a lot of self-confidence when it comes to computing courses.
I2	I am confident that I can solve problems by using computer applications.
I3	I hope that my future career will require the use of information technology concepts.
I4	I like to use information technology to solve problems.
I5	I do not like using information technology to solve problems.
I6	The challenge of solving problems using information technology appeals to me.
I7	I think information technology is interesting.
I8	I would voluntarily take additional information technology courses if I were given the opportunity.
I9	Developing computing skills will be important to my career goals.
I10	I expect that learning to use computing skills will help me achieve my career goals.
Gender Construct (G)	
G1	Women are more capable than men at solving computing problems.
G2	Women are more likely to excel in careers that involve computing than men are.
G3	Women produce higher quality work in computing than men.
G4	I doubt that a man could excel in computing courses.
G5	It is not appropriate for women to study computing.
G6	Men produce higher quality work in computing than women.
G7	Men are more likely to excel in careers that involve computing than women are.
G8	Women produce the same quality work in computing as men.
G9	Men and women are equally capable of solving computing problems.
G10	Men and women can both excel in computing courses.

Table 4.3: Teacher and Counselor IT Attitude Survey, divided by construct

Confidence Construct (C)	
C1	I am comfortable with learning computing concepts.
C2	I have little self-confidence when it comes to computing courses.
C3	I do not have a good understanding of computing concepts.
C4	I have a lot of self-confidence when it comes to teaching computing courses.
C5	I am confident that I can solve problems by using computer applications.
C6	I doubt that I can solve problems by using computer applications.
Interest Construct (I)	
I1	I would not take additional information technology courses if I were given the opportunity.
I2	I think information technology is boring.
I3	The challenge of solving problems using information technology does not appeal to me.
I4	I like to use information technology to solve problems.
I5	I do not like using information technology to solve problems.
I6	The challenge of solving problems using information technology appeals to me.
I7	I think information technology is interesting.
I8	I would voluntarily take additional information technology courses if I were given the opportunity.
Gender Construct (G)	
G1	I doubt that a woman could excel in computing courses.
G2	Men are more capable than women at solving computing problems.
G3	Computing is an appropriate subject for both men and women to study.
G4	It is not appropriate for men to study computing.
G5	Women are more capable than men at solving computing problems.
G6	Women are more likely to excel in careers that involve computing than men are.
G7	Women produce higher quality work in computing than men.
G8	Women and men can both excel in careers that involve computing.
G9	I doubt that a man could excel in computing courses.
G10	It is not appropriate for women to study computing.
G11	Men produce higher quality work in computing than women.
G12	Men are more likely to excel in careers that involve computing than women are.
G13	Women produce the same quality work in computing as men.
G14	Men and women are equally capable of solving computing problems.
G15	Men and women can both excel in computing courses.

Table 4.3 – continued from previous page

Usefulness Construct (C)	
U1	My career requires the use of information technology concepts.
U2	Developing computing skills has not played a role in helping me achieve my career goals.
U3	Knowledge of computing has allowed me to secure a good job.
U4	I use computing skills in my daily life.
U5	My career does not require that I have computing skills.
U6	Developing computer skills will be important to my career goals.
U7	Knowledge of computing skills has not helped me secure a good job.
U8	I do not use computing skills in my daily life.
U9	Learning to use computing skills has helped me achieve my career goals.
Professional Construct (P)	
P1	Doing well in information technology does not require a student to spend most of his/her time at a computer.
P2	A student who performs well in information technology will probably not have a life outside of computers.
P3	To do well in information technology, a student must spend most of his/her time at a computer.
P4	A student who performs well in information technology is likely to have a life outside of computers.
P5	Being good at information technology is a negative quality.
P6	Students who are skilled at information technology are less popular than other students.
P7	Being good at information technology is a positive quality.
P8	Students who are skilled at information technology are just as popular as other students.
P9	Students who are skilled at information technology are more popular than other students.

4.2 Implementation Analysis

The following section contains a qualitative analysis of differences and similarities in SPIRIT implementation across the years.

4.2.1 Differences Across Years

During the first week of the 2008 SPIRIT summer program, high school teachers attended two presentations that were omitted in 2009 and 2010: neuro-linguistic programming and grant writing. During the second week of the 2008 SPIRIT summer program, all participants attended four Critical Thinking exercises where they learned how scientific processes were able to be used to solve everyday problems. These Critical Thinking exercises were omitted in 2009 and 2010. In addition, three presentations in 2008 were not repeated in 2009 or 2010: IT in the government, Powerpoint animation and the use of IT in videogames.

Four presentations in the 2009 SPIRIT summer programs were unique to 2009: IT in manufacturing, IT and Renewable Energy, Digital Manipulation with Photoshop, and movie making with Adobe After Effects. These presentations were not offered in 2008 or 2010.

Based on the student identified concerns in 2008, the length of lectures was reduced and hands-on activities were increased in 2009. Also in 2009, students gave negative feedback for one session which was delivered by graduate students who did not adequately prepare and had problems arise.

Five presentations in the 2010 SPIRIT summer program were unique to 2010: Music and IT, Electric Vehicles, Smartphones, Fashion and IT, and Racing and IT. These presentations were not offered in 2008 or 2009.

In 2010, both teachers and counselors participated in the first week of the SPIRIT summer program. Due to the same treatment effect, counselors were combined with teachers in 2010 analyses.

4.2.2 Similarities Across Years

Across all three years, there were common presentations. In 2008 and 2009, forensics-related IT was covered in three different presentations and two different presentations, respectively. In 2010, forensics-related IT was only presented once. Healthcare-related IT was covered in three different presentations in 2008, but was only covered in one presentation in 2009 and 2010. Robotics and Social Networking were covered in equal intensity throughout the years. Career-panels and business-related IT presentations were covered in four presentations in 2008, two presentations in 2009, and three presentations in 2010.

Some topics were covered in only a fraction of the summer workshops. In 2009 and 2010, two sessions and one session were devoted to Pico Crickets, respectively. In 2008, Pico Crickets was not offered. In 2008 and 2009, a presentation on search-engines was offered, but was not repeated in 2010.

4.3 Student Results

This section discusses the construction of the IT attitude survey model and the subsequent CFA that was run for pre data. Running the CFA on pre data allowed for examination of how the factors were loading before participants were influenced by participating in the SPIRIT program. Following the CFA are statistical analyses examining differences between student group participants. The resulting IT attitude survey structure is displayed in the appendix. Three student participants responded to the IT attitude survey in an inconsistent manner and were removed from this analysis. Recall that participants' inconsistency was determined by comparing responses to positively- and negatively-worded equivalent statements in the IT attitude survey. If a response to a positively-worded statement differed by more than 1 Likert scale point to the negatively-worded equivalent, the participant was removed.

4.3.1 Factor Analysis

Based on previously discussed theory and evidence from the Document Analysis, an initial two-factor model was specified, with statements I1 through I10 loading onto factor I, and statements G1 through G10 loading onto factor G. Factor I and factor G were permitted to correlate, as well as statements I4 and I5. Figure 4.1 contains a visual specification of the two-factor model.

Table 4.4 contains the two-factor structure results for the pre IT attitude survey. Also listed in Table 4.4 are the five largest modification indices after constructing the model. Note that statements G4 through G9 did not load significantly onto factor G, and that the correlation between statements I4 and I5 was not significant. Since these statements, I4 and I5, were oppositely worded but otherwise equivalent, this parameter was not removed.

Table 4.4: CFA Results, pre Student IT Attitude Survey

Parameter	Estimate	Parameter	Estimate
λ_{I1}	0.647*	λ_{G1}	0.921
λ_{I2}	0.701	λ_{G2}	0.898
λ_{I3}	0.720	λ_{G3}	0.924
λ_{I4}	0.770	λ_{G4}	0.605*
λ_{I5}	0.732	λ_{G5}	0.680*
λ_{I6}	0.868	λ_{G6}	0.552*
λ_{I7}	0.775	λ_{G7}	0.562*
λ_{I8}	0.844	λ_{G8}	0.454*
λ_{I9}	0.785	λ_{G9}	0.627*
λ_{I10}	0.779	λ_{G10}	0.786
$\delta_{I4,I5}$	0.611*		
ϕ_{GI}	0.222*		
* indicates insignificance at 0.70 cutoff			
Largest Modification Indices		Goodness of Fit Indices	
Index	Value	Index	Value
G6, G7	192.471	χ^2 value	840.510
General Interest, G10	101.664	RMSEA	0.147
G1, G2	42.289	CFI	0.875
General Interest, G2	37.268		
General Interest, G9	33.145		

The goodness of fit for the pre IT attitude survey factor structure was evaluated with SRMR, RMSEA and CFI, based on recommendations from [43]. Table 4.4 contains the goodness of fit indices for the pre IT attitude survey, all of which do not meet acceptable levels. The critical value for the χ^2 distribution with 168 degrees of freedom was 199.24.

The large modification indices between factor I and statements G9 and G10 raise concern. Statements G8, G9 and G10 are gender-neutral, indicating that students may be interpreting these statements with regard to general interest in IT. In addition, statements G6 and G7 are assessing whether or not a participant believes "men are better than women." A decision was made to remove statements G6 though G10. Even though statement G10 loaded significantly onto factor G, it was removed due to its similarity to statements G8 and G9. Table 4.5 contains the results from this reduced model.

Table 4.5: CFA Results, pre Student IT Attitude Survey, G6-G10 removed

Parameter	Estimate	Parameter	Estimate
λ_{I1}	0.658*	λ_{G1}	0.946
λ_{I2}	0.701	λ_{G2}	0.930
λ_{I3}	0.722	λ_{G3}	0.966
λ_{I4}	0.777	λ_{G4}	0.591*
λ_{I5}	0.733	λ_{G5}	0.469*
λ_{I6}	0.873	$\delta_{I4,I5}$	0.607*
λ_{I7}	0.768	ϕ_{GI}	0.069*
λ_{I8}	0.836		
λ_{I9}	0.782		
λ_{I10}	0.773		
* indicates insignificance at 0.70 cutoff			
Largest Modification Indices		Goodness of Fit Indices	
Index	Value	Index	Value
G4, G5	59.866	χ^2 value	292.067
I9, I10	46.591	RMSEA	0.112
I1, I2	14.775	CFI	0.959
I5, G5	12.740		
I7, G4	12.596		

Comparing Table 4.4 to Table 4.5, all goodness-of-fit indices improved. The coefficient between factor I and factor G decreased. Statements G4 and G5 were still not loading significantly onto factor G, and were removed from the survey. The results from the further reduced survey are displayed in Table 4.6.

Comparing Table 4.5 to Table 4.6, all goodness-of-fit indices improved. Based on a large modification index, one parameter was added to the model: $\partial_{I9,I10}$. Since all other modification indices were small in comparison, no additional parameters were added. Results from the CFA are displayed in Table 4.7. The critical value of the χ^2 distribution with 62 degrees of freedom is 81.38.

Table 4.6: CFA Results, pre Student IT Attitude Survey, G4-G10 removed

Parameter	Estimate	Parameter	Estimate
λ_{I1}	0.659*	λ_{G1}	0.944
λ_{I2}	0.702	λ_{G2}	0.937
λ_{I3}	0.722	λ_{G3}	0.975
λ_{I4}	0.778	$\delta_{I4,I5}$	0.605*
λ_{I5}	0.734	ϕ_{GI}	0.069*
λ_{I6}	0.874		
λ_{I7}	0.765		
λ_{I8}	0.835		
λ_{I9}	0.782		
λ_{I10}	0.772		
* indicates insignificance at 0.70 cutoff			
Largest Modification Indices		Goodness of Fit Indices	
Index	Value	Index	Value
I9, I10	63.064	χ^2 value	183.722
I1, I2	19.616	RMSEA	0.102
I1, I10	15.924	CFI	0.975
I6, I10	10.031		
I2, I5	9.125		

Comparing Table 4.6 to Table 4.7, all goodness-of-fit indices improved. Since the modification indices were small and similar to one another, no further parameters were added to the model. The RMSEA and CFI are close to acceptable levels.

4.3.2 Reliability

In order to further validate the removal of statements G4 through G10 from the IT attitude survey, Jöreskog's measure of scale reliability was used. Table 4.8 contains the reliability coefficients for the original and reduced IT attitude survey, separated

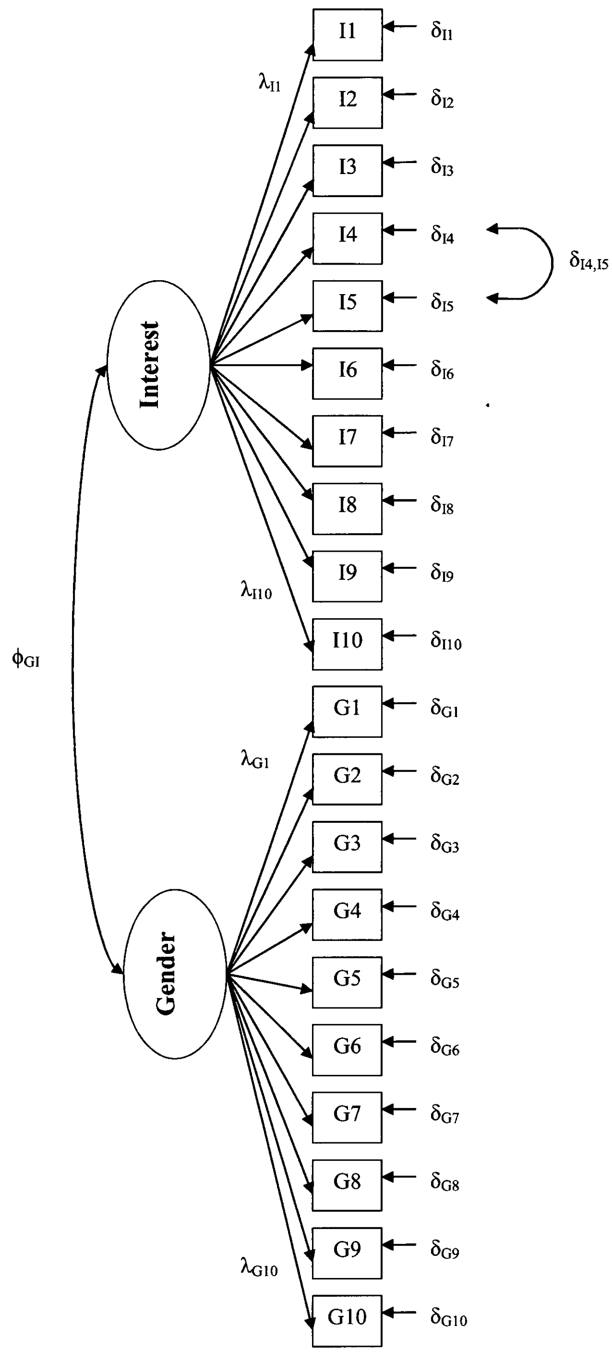

Figure 4.1: CFA model structure, 20 statement survey

Table 4.7: CFA Results, pre Student IT Attitude Survey, G4-G10 removed

Parameter	Estimate	Parameter	Estimate
λ_{I1}	0.666*	λ_{G1}	0.944
λ_{I2}	0.718	λ_{G2}	0.937
λ_{I3}	0.733	λ_{G3}	0.975
λ_{I4}	0.790	$\delta_{I4,I5}$	0.628*
λ_{I5}	0.747	$\delta_{I9,I10}$	0.587*
λ_{I6}	0.878	ϕ_{GI}	0.008*
λ_{I7}	0.781		
λ_{I8}	0.844		
λ_{I9}	0.682		
λ_{I10}	0.646*		
* indicates insignificance at 0.70 cutoff			
Largest Modification Indices		Goodness of Fit Indices	
Index	Value	Index	Value
I1, I2	17.290	χ^2 value	124.137
I3, I9	10.324	RMSEA	0.074
I1, I10	9.107	CFI	0.987
I2, I3	8.833		
I4, I10	7.506		

by factor. The reliability coefficients for both the original 20 statement survey and the reduced 13 statement survey are all at acceptable levels. The reliability coefficent for factor G increased after removing statements G4 through G10, but the reliability coefficient for factor I decreased. However, due to the small size of factor G and the fact that many of the participants did not know how to respond to these statements, no significant differences are predicted to be found in changes in perception of gender stereotypes in IT. Table 4.9 contains the reduced student IT attitude survey, divided into its factors.

Table 4.8: Reliability per Construct, Student IT Attitude Survey

Version	Factor	Reliability coefficient, ρ
Original 20 statement survey	General Interest	0.99489
	Gender	0.99971
Reduced 13 statement survey	General Interest	0.98836
	Gender	0.99981

Table 4.9: Further Reduced Student IT Attitude Survey, divided by construct

General Interest Construct (I)	
I1	I have a lot of self-confidence when it comes to computing courses.
I2	I am confident that I can solve problems by using computer applications.
I3	I hope that my future career will require the use of information technology concepts.
I4	I like to use information technology to solve problems.
I5	I do not like using information technology to solve problems.
I6	The challenge of solving problems using information technology appeals to me.
I7	I think information technology is interesting.
I8	I would voluntarily take additional information technology courses if I were given the opportunity.
I9	Developing computing skills will be important to my career goals.
I10	I expect that learning to use computing skills will help me achieve my career goals.
Gender Construct (G)	
G1	Women are more capable than men at solving computing problems.
G2	Women are more likely to excel in careers that involve computing than men are.
G3	Women produce higher quality work in computing than men.

4.3.3 Linear Regression Analysis for Factor I

In order to answer the third research question, nested F-tests were run on comparable linear models. With this method, a linear regression model was compared to a simpler model, in which all regression coefficents containing the explanatory model of interest were set to zero. For example, in order to examine if gender of participants was significant in explaining the variance in factor I scores for students, the following hypotheses based on equation 3.5 were tested:

$$H_0 : \beta_1 = \beta_6 = \beta_7 = \beta_8 = \beta_9 = 0$$
$$H_1 : \text{at least one of the } \beta_i \neq 0$$

The corresponding nested F-test p value was 0.5408. Since this p value was insignificant, all variables with *gender* were removed from the model, and the model reduced down to:

$$\begin{aligned} Y_i = \beta_0 + \beta_2 * ethnicity_i + \beta_3 * year1_i + \beta_4 * year2_i + \beta_5 * concept_i \\ + \beta_{10} * (ethnicity * year1)_i + \beta_{11} * (ethnicity * year2)_i \\ + \beta_{12} * (ethnicity * concept)_i + \beta_{13} * (year1 * concept)_i \\ + \beta_{14} * (year2 * concept)_i + \epsilon_i \end{aligned}$$

A similar approach was used for the corresponding nested F-tests for *ethnicity*, *year1*, *year2* and *concept*. The model reduced down to:

$$Y_i = \beta_0 + \beta_2 * ethnicity_i + \beta_4 * year2_i + \beta_{11} * (ethnicity * year2)_i + \epsilon_i$$

A similar approach was also used for the student factor G model and for the teacher and counselor factor I and factor G models.

Table 4.10 contains the summary of the resulting factor I model for students. The regression coefficient for *ethnicity* was not significant, but the regression coefficient for the interaction between *ethnicity* and *year2* was significant.

Table 4.10: Model Summary for Factor I, Students

Variable	Estimate	Error	p-value
Intercept	2.17	0.59	0.0003*
Ethnicity	-1.00	0.72	0.1659
Year2	-3.47	0.90	0.0002*
Ethnicity*Year2	3.05	1.12	0.0072*
* indicates significance at α of 0.05			

4.3.4 Linear Regression analysis for Factor G

The linear model for changes in factor G scores reduced from equation 3.5 down to:

$$Y_i = \beta_0 + \beta_1 * gender_i + \beta_4 * year2_i + \beta_8 * (gender * year2)_i + \epsilon_i$$

Table 4.11 contains the summary of the resulting linear model. The regression coefficient for *year2* was significant. A nested F-test comparing the resulting model with a reduced model containing only *year2* had a p value of 0.0245, so *gender* and *gender*year2* remained in the final model. Since the interaction between *year2* and *gender* was not significant, the main effects were analyzed separately.

Table 4.11: Model Summary for Factor G, Students

Variable	Estimate	Error	p-value
Intercept	-0.04	0.21	0.8434
Gender	-0.27	0.35	0.4550
Year2	1.08	0.33	0.0012*
Gender*Year2	-0.96	0.58	0.1009
* indicates significance at α of 0.05			

4.3.5 Comparison of Factor I Subgroups

Students' ethnicity and the year in which a student participated in SPIRIT significantly explained differences in factor I scores. Table 4.12 contains a summary of post minus pre differences across ethnicity and year in program, where non-minority was abbreviated "NM" and minority was abbreviated "M". The abbreviation "NM(2008)" refers to non-minority students who participated in 2008. Significant changes in general interest in IT were found for minority students in 2008 and non-minority students in 2010.

Tukey's comparison of multiple means was used to find significant differences. A familywise error rate of 0.05 was used. Table 4.13 contains summary information for the Tukey comparison and Figure 4.2 contains a boxplot summary of the means.

A significant difference was found between minority students' changes in factor I scores between 2008 and 2009. Minority students who participated in SPIRIT in 2008 displayed a positive change in general interest in IT, while minority students who particiapted in 2009 displayed a negative change. A significant difference was also found between minority students in 2009 compared to non-minority students in 2010. Non-minority students in 2010 displayed a positive change in IT interest after participating in SPIRIT while minority students in 2009 displayed a negative change in IT interest.

Table 4.12: Comparison of Factor I scores by Ethnicity and Year in Program

	Number of Participants	Post-Pre Difference	p-value
M(2008)	20	2.35	0.0182*
M(2009)	27	-1.30	0.0548
M(2010)	15	1.93	0.0694
NM(2008)	39	0.82	0.1600
NM(2009)	44	0.75[1]	0.0866
NM(2010)	37	1.54	0.0141*
* indicates significance at α of 0.05			
[1] indicates non-normal distribution with Shapiro-Wilks test; Wilcoxon signed rank test used			

Table 4.13: Tukey's HSD of Factor I scores by Ethnicity and Year in Program

Comparison		Difference	p-value
NM(2009)	NM(2008)	-0.07	0.9999
NM(2010)	NM(2008)	0.72	0.9488
M(2008)	NM(2008)	1.53	0.6161
M(2009)	NM(2008)	-2.12	0.1638
M(2010)	NM(2008)	1.11	0.9047
NM(2010)	NM(2009)	0.79	0.9161
M(2008)	NM(2009)	1.60	0.5466
M(2009)	NM(2009)	-2.05	0.1724
M(2010)	NM(2009)	1.18	0.8722
M(2008)	NM(2010)	0.81	0.9624
M(2009)	NM(2010)	-2.84	0.0216*
M(2010)	NM(2010)	0.39	0.9991
M(2009)	M(2008)	-3.65	0.0076*
M(2010)	M(2008)	-0.42	0.9993
M(2010)	M(2009)	3.23	0.0558
* indicates significance at α of 0.05			

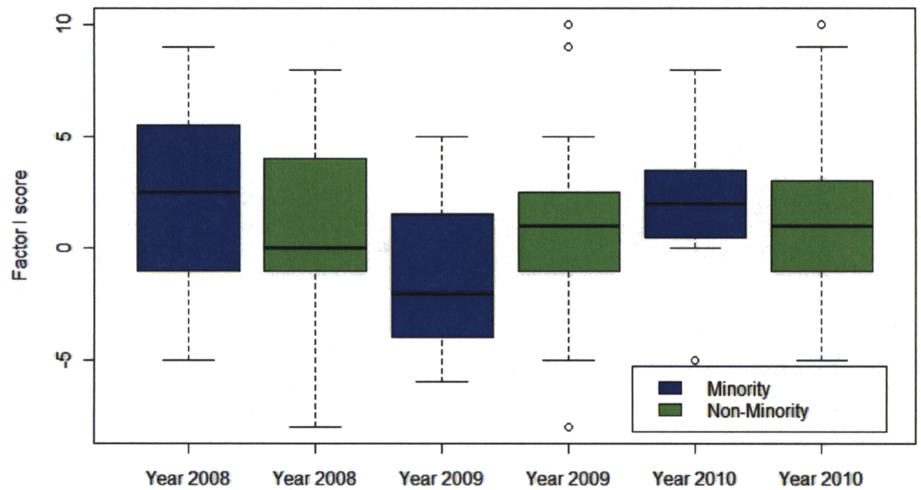

Figure 4.2: Boxplots for mean Factor I scores

4.3.6 Comparison of Factor G Subgroups

Students' gender and the year in which students participated in SPIRIT both significantly explained differences in factor G scores. Since the residuals from the student factor G model were significantly non-normal, caution needs to be taken when interpreting the following results. Table 4.14 contains a summary of the differences in factor G scores across the year of implementation.

Table 4.14: Comparison of Factor G scores by Year in Program

Year	Number of Participants	Post-Pre Difference	p-value
2008	59	-0.17[1]	0.5042
2009	71	0.68[1]	0.0028*
2010	52	-0.10[1]	0.7947
[1] indicates non-normal distribution with Shapiro-Wilks test; Wilcoxon signed rank test used * indicates significance at α of 0.05			

A Wilcoxon signed rank test found that male students' factor G scores were significantly different from female students' factor G scores (p=0.0211). On average, female students demonstrated a more positive change in perception of gender stereotypes than male students. In other words, after attending SPIRIT, female students viewed the IT field as more gender-neutral, where men and women are equal in IT ability, than male students. Tukey's comparison of multiple means was used to find significant differences across year of program implementation. A familywise error rate of 0.05 was used, and a significant difference between changes in factor G scores in 2008 compared to 2009 was found (p=0.0245). Table 4.15 contains a summary of the Tukey comparison, and Figure 4.3 and Figure 4.4 contain boxplot representations of the means.

Table 4.15: Tukey's HSD of Factor G scores by Year in Program

Comparison		Difference	p-value
2009	2008	0.84	0.0245*
2010	2008	0.07	0.9755
2010	2009	-0.77	0.0548
* indicates significance at α of 0.05			

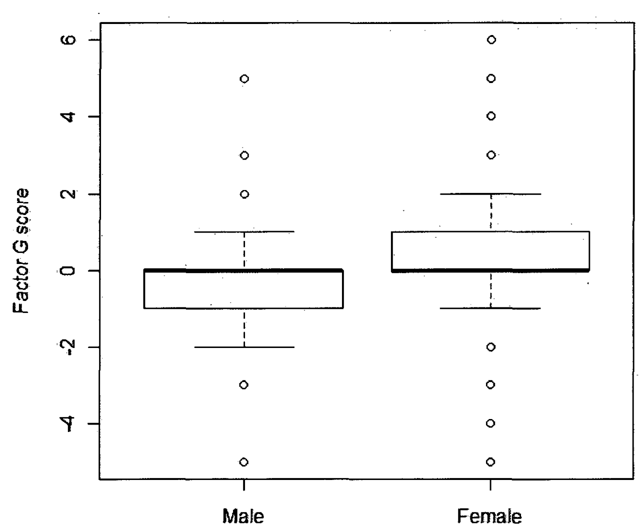

Figure 4.3: Boxplots for mean Factor G scores by Gender

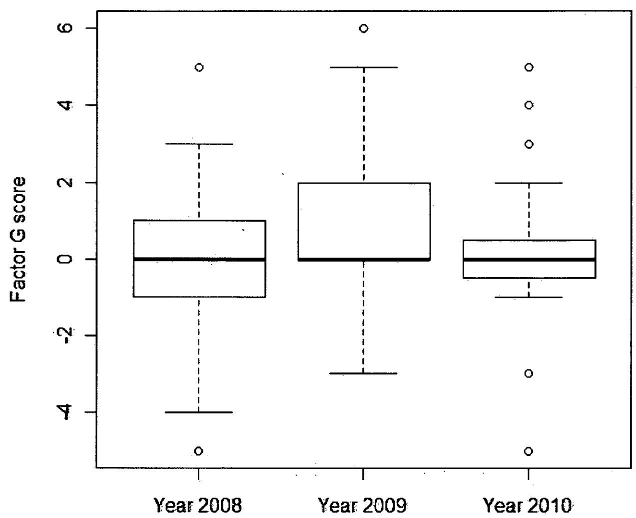

Figure 4.4: Boxplots for mean Factor G scores by Year in Program

4.4 Teacher and Counselor Results

This section discusses the EFA that was run with pre IT attitude survey data for teachers and counselors. An EFA was run on pre data in order to examine how the factors were loading before participants were influenced by the SPIRIT program. Following the EFA are statistical analyses examining differences in teacher subgroups and counselor subgroups. One teacher and one counselor participant were excluded from the following analysis due to inconsistent attitude survey responses. Recall that participants' inconsistency was determined by comparing responses to positively- and negatively-worded equivalent statements in the IT attitude survey. If a response to a positively-worded statement differed by more than 1 Likert scale point to the negatively-worded equivalent, the participant was removed.

4.4.1 Factor Analysis

An initial WLSMV EFA with geomin rotation was run on a five factor solution, and the results are displayed in Table 4.16. Statement I1 double-loaded onto more than one construct and 14 statements did not load significantly onto any factor. All Gender statements loaded onto a single factor, with the exception of statements G13, G14 and G15. Usefulness and Professional statements did not load as expected.

Table 4.16: EFA Loadings, 5 factor solution

Statement	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
C1	0.922				
C2	0.982				
C3	0.941				
C4	0.821				
C5	0.549				
C6	0.456*				
I1	0.543		0.576		
I2			0.592		
I3			0.407*		
I4			0.693		
I5			0.548		
I6			0.787		
I7			0.842		
I8			0.928		
G1		0.516			
G2		0.841			
G3		0.478*			
G4		0.847			
G5		0.828			
G6		0.780			
G7		0.857			
G8		0.603			
G9		0.695			
G10		0.739			
G11		0.876			
G12		0.732			
G13				0.361*	
G14				0.482*	
G15					0.347*

Table 4.16 – continued from previous page

Statement	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
U1				-0.387*	
U2	0.550				
U3	0.760				
U4	0.496*				
U5	0.474*				
U6			0.401*		
U7	0.588				
U8	0.490*				
U9	0.481				
P1				0.511	
P2		0.385*			
P3				0.522	
P4				0.543	
P5		0.526			
P6				0.650	
P7			0.707		
P8				0.406*	
P9				0.573	
* indicates insignificance at cutoff of 0.50					

Closer examination of statements G13, G14 and G15 reveals that these are gender-neutral statements. Due to the conflicting nature of these statements, and the fact that these gender-neutral statements were not loading as expected for the student survey, statements G3, G8, G13, G14 and G15 were removed. Closer examination of statements P5 and P7 revealed that these statements were equivalent; since neither of these statements loaded onto the intended construct, they were also removed. Also worth noting is that many of Usefulness statements loaded onto the same construct as Confidence statements. The EFA was re-run on the reduced survey and the four-factor solution is displayed in Table 4.17. Table 4.18 contains the latent factor covariance matrix for the four factor solution. The largest covariance was between Factor 1 and Factor 2.

All Confidence, Interest and Usefulness statements loaded onto a single factor. The remaining Gender statements loaded onto a single factor. Professional statements did not loading as expected. An EFA was re-run with a three-factor solution, and the results are displayed in Table 4.19.

Due to the instability of factor P, all statements in this factor were removed. The EFA was re-run, and the two-factor solution is displayed in Table 4.20. All statements apart from U1 loaded significantly onto their expected factor. Table 4.21 contains the latent factor covariance matrix.

Table 4.17: EFA Loadings from reduced survey, 4 factor solution

Statement	Factor 1	Factor 2	Factor 3	Factor 4
C1	0.854			
C2	0.837	-0.603		
C3	0.973	-0.513		
C4	0.881			
C5	0.754			
C6	0.575			
I1	0.813			
I2	0.586			
I3	0.617			
I4	0.893			
I5	0.646			
I6	0.809			
I7	0.800			
I8	0.633			
G1			0.487*	
G2			0.853	
G4			0.835	
G5			0.794	
G6			0.718	
G7			0.846	
G9			0.752	
G10			0.785	
G11			0.917	
G12			0.761	
U1	0.515			
U2	0.677			
U3	0.856			
U4	0.656			
U5	0.539			
U6	0.491			
U7	0.634			

Table 4.17 – continued from previous page

Statement	Factor 1	Factor 2	Factor 3	Factor 4
U8	0.466			
U9	0.793			
P1				0.586
P2			0.464	
P3				0.648
P4				0.383*
P6			0.499*	
P8		0.361		
P9				0.530
* indicates insignificance at cutoff of 0.50				

Table 4.18: Latent Factor Covariance Matrix, Original Survey

	Factor 1	Factor 2	Factor 3	Factor 4
Factor 1	1.000			
Factor 2	0.426	1.000		
Factor 3	0.097	0.150	1.000	
Factor 4	0.255	0.332	0.092	1.000

Table 4.19: EFA Loadings from reduced survey, 3 factor solution

Statement	Factor 1	Factor 2	Factor 3
C1	0.834		
C2	0.826		
C3	0.866		
C4	0.858		
C5	0.755		
C6	0.573		
I1	0.803		
I2	0.587		
I3	0.620		
I4	0.899		
I5	0.649		
I6	0.826		
I7	0.813		
I8	0.652		
G1		0.575	
G2		0.846	
G4		0.873	
G5		0.771	

Table 4.19 – continued from previous page

Statement	Factor 1	Factor 2	Factor 3
G6		0.708	
G7		0.830	
G9		0.848	
G10		0.868	
G11		0.948	
G12		0.764	
U1	0.545		
U2	0.673		
U3	0.855		
U4	0.651		
U5	0.574		
U6	0.516		
U7	0.663		
U8	0.669		
U9	0.814		
P1		0.328*	
P2		0.569	
P3		0.498*	
P4		0.275*	
P6		0.593	
P8		0.464*	
P9		0.220*	
* indicates insignificance at cutoff of 0.50			

In addition to the reduced two-factor solutions, a WLSMV EFA based on the reduced 13 statement student equivalent to the IT attitude survey was run for the teacher and counselor attitude survey. The results are displayed in Table 4.22. All statements apart from U1 loaded significantly onto their expected factor. However, the factor loadings differed in magnitude. For example, statement C3 loaded at 0.610 while statement C3 loaded at 0.986, indicating that the variance in responses to statement C3 was better explained by the latent factor than for statement C6. Table 4.23 contains the latent factor covariance matrix. Interestingly, the covariance between the two factors for the 13 statement survey was smaller than for the 33 statement survey.

Table 4.20: EFA Loadings from reduced survey, 2 factor solution

Statement	Factor 1	Factor 2
C1	0.906	
C2	0.968	
C3	0.986	
C4	0.862	
C5	0.743	
C6	0.610	
I1	0.804	
I2	0.591	
I3	0.593	
I4	0.865	
I5	0.670	
I6	0.815	
I7	0.807	
I8	0.652	
U1	0.475*	
U2	0.661	
U3	0.867	
U4	0.673	
U5	0.591	
U6	0.512	
U7	0.684	
U8	0.689	
U9	0.823	
G1		0.549
G2		0.803
G4		0.863
G5		0.696
G6		0.695
G7		0.790
G9		0.890
G10		0.918
G11		0.942
G12		0.770
* indicates insignificance at cutoff of 0.50		

Table 4.21: Latent Factor Covariance Matrix, Original Survey

	Factor 1	Factor 2
Factor 1	1.000	
Factor 2	0.414	1.000

Table 4.22: EFA Loadings from student equivalent survey, 2 factor solution

Statement	Factor 1	Factor 2
C4	0.655	
C5	0.746	
I4	0.878	
I5	0.759	
I6	0.990	
I7	0.910	
I8	0.828	
U1	0.466*	
U6	0.636	
U9	0.836	
G5		0.834
G6		0.799
G7		0.777
* indicates insignificance at cutoff of 0.70		

Table 4.23: Latent Factor Covariance Matrix, Student Equivalent Survey

	Factor 1	Factor 2
Factor 1	1.000	
Factor 2	0.231	1.000

4.4.2 Reliability

In order to further validate the removal of statements from the teacher and counselor version of the IT attitude survey, Jöreskog's measure of scale reliability was used. Table 4.24 contains the reliability coefficients for the original and reduced IT attitude survey, separated by factor. In order to calculate the reliability of a construct from EFA, each statement was assumed to load onto its theoretical factor, regardless of whether it loaded significantly onto a different factor. The reliability coefficients for both the original 47 statement survey and the reduced 33 statement and 13 statement versions are all at acceptable levels, apart from factor G from the 13 statement survey. The reliability coefficent for factor G remained approximately the same when comparing the original 15 statement factor to the reduced 10 statement factor. The reliability coefficents for the 23 statement and 10 statement factor Is were both at acceptable levels, but the reliability for the 10 statement factor I was larger. The 33 statement IT attitude survey and the 13 statement IT attitude survey are displayed in Table 4.25 and Table 4.26, respectively.

Table 4.24: Reliability per Construct, Teacher/Counselor IT Attitude Survey

Version	Factor	Reliability coefficient, ρ
Original 47 statement survey	Confidence	0.9357
	Interest	0.9487
	Usefulness	0.8598
	Gender	0.9590
	Professional	0.8034
Reduced 33 statement survey	General Interest	0.8731
	Gender	0.9523
Reduced 13 statement survey	General Interest	0.9417
	Gender	0.8460

Table 4.25: Reduced 33-statement Teacher and Counselor IT Attitude Survey

General Interest Construct (I)	
C1	I am comfortable with learning computing concepts.
C2	I have little self-confidence when it comes to computing courses.
C3	I do not have a good understanding of computing concepts.
C4	I have a lot of self-confidence when it comes to teaching computing courses.
C5	I am confident that I can solve problems by using computer applications.
C6	I doubt that I can solve problems by using computer applications.
I1	I would not take additional information technology courses if I were given the opportunity.
I2	I think information technology is boring.
I3	The challenge of solving problems using information technology does not appeal to me.
I4	I like to use information technology to solve problems.
I5	I do not like using information technology to solve problems.
I6	The challenge of solving problems using information technology appeals to me.
I7	I think information technology is interesting.
I8	I would voluntarily take additional information technology courses if I were given the opportunity.

Table 4.25 – continued from previous page

General Interest Construct (I)–continued	
U1	My career requires the use of information technology concepts.
U2	Developing computing skills has not played a role in helping me achieve my career goals.
U3	Knowledge of computing has allowed me to secure a good job.
U4	I use computing skills in my daily life.
U5	My career does not require that I have computing skills.
U6	Developing computer skills will be important to my career goals.
U7	Knowledge of computing skills has not helped me secure a good job.
U8	I do not use computing skills in my daily life.
U9	Learning to use computing skills has helped me achieve my career goals.
Gender Construct	
G1	I doubt that a woman could excel in computing courses.
G2	Men are more capable than women at solving computing problems.
G4	It is not appropriate for men to study computing.
G5	Women are more capable than men at solving computing problems.
G6	Women are more likely to excel in careers that involve computing than men are.
G7	Women produce higher quality work in computing than men.
G9	I doubt that a man could excel in computing courses.
G10	It is not appropriate for women to study computing.
G11	Men produce higher quality work in computing than women.
G12	Men are more likely to excel in careers that involve computing than women are.

Table 4.26: Reduced 13-statement Teacher and Counselor IT Attitude Survey

General Interest Construct (I)	
C4	I have a lot of self-confidence when it comes to teaching computing courses.
C5	I am confident that I can solve problems by using computer applications.
I4	I like to use information technology to solve problems.
I5	I do not like using information technology to solve problems.
I6	The challenge of solving problems using information technology appeals to me.
I7	I think information technology is interesting.
I8	I would voluntarily take additional information technology courses if I were given the opportunity.
U1	My career requires the use of information technology concepts.

Table 4.26 – continued from previous page

General Interest Construct (I)–continued	
U6	Developing computer skills will be important to my career goals.
U9	Learning to use computing skills has helped me achieve my career goals.
Gender Construct (G)	
G5	Women are more capable than men at solving computing problems.
G6	Women are more likely to excel in careers that involve computing than men are.
G7	Women produce higher quality work in computing than men.

4.4.3 Linear Regression Analysis for Factor I, Teachers

The following section examines differences across teacher subgroups for the 33-statement version of the IT attitude survey. One teacher was omitted from the following analysis due to inconsistent attitude survey responses. The pre-to-post linear model was examined first.

For the pre-to-post linear model for factor I, an overall nested F-test had a p value of 0.2862. Therefore, the linear model for pre to post changes in factor I scores reduced from equation 3.1 down to:

$$Y_i = \overline{y} + \epsilon_i$$

For the second model, which compares pre-to-post-post differences, the correlation between the variables *content.1* and *content.2* was large (r=0.640). Figure 4.5 contains a plot of *content.1* versus *content.2* which demonstrates the linear dependence between the variables.

Due to this, all variables with *concept1* were removed, and linear model reduced from equation 3.2 to equation 4.1:

$$\begin{aligned} Y_i = \beta_0 + \beta_1 * gender_i + \beta_2 * year1_i + \beta_3 * year2_i + \beta_5 * concept2_i \\ + \beta_6 * (gender * year1)_i + \beta_7 * (gender * year2)_i \\ + \beta_9 * (gender * concept2)_i + \beta_{11} * (year1 * concept2)_i \\ + \beta_{13} * (year2 * concept2)_i + \epsilon_i \end{aligned} \tag{4.1}$$

After the subsequent nested F-tests, the linear model reduced down from equation 4.1 to the following:

$$Y_i = \beta_0 + \beta_2 * year1_i + \epsilon_i$$

Table 4.27 contains the summary of the resulting linear model. The regression coefficient for *year1* was significant.

Table 4.27: Model Summary for Factor I, Teachers

Variable	Estimate	Error	p-value
Intercept	7.38	1.17	0.0000*
Year1	-5.38	1.99	0.0095*
* indicates significance at α of 0.05			

4.4.4 Linear Regression Analysis for Factor I, Counselors

The following section examines differences across counselor subgroups for the 33-statement version of the IT attitude survey. One counselor was omitted from the following analysis due to inconsistent attitude survey responses. The pre-to-post linear model was examined first.

For the pre-to-post linear model for factor I, an overall nested F-test had a p value of 0.7628. Therefore, the linear model for pre to post changes in factor I scores reduced from equation 3.3 down to:

$$Y_i = \overline{y} + \epsilon_i$$

For the second model, which compared pre-to-post-post differences, the correlation between the variables *content.1* and *content.2* was somewhat large (r=0.549). Figure 4.6 contains a plot of *content.1* versus *content.2* which demonstrates the linear dependence between the variables.

Due to this, all variables with *concept1* were removed, and the linear model reduced from equation 3.4 to equation 4.2:

$$Y_i = \beta_0 + \beta_1 * gender_i + \beta_2 * year1_i + \beta_4 * (concept2)_i + \beta_5 * (gender * year1)_i \\ + \beta_7 * (gender * concept2)_i + \beta_9 * (year1 * concept2)_i + \epsilon_i \tag{4.2}$$

After subsequent nested F-tests, the linear model reduced from equation 4.2 to the following:

$$Y_i = \overline{y} + \epsilon_i$$

4.4.5 Linear Regression Analysis for Factor G, Teachers

The following section examines differences across teacher subgroups for the 33-statement version of the IT attitude survey.

Due to large p values from the overall nested F-tests, both the pre-to-post and pre-to-post-post Factor G model for teachers reduced down to the following:

$$Y_i = \overline{y} + \epsilon_i$$

4.4.6 Linear Regression Analysis for Factor G, Counselors

The following section examines differences across counselor subgroups for the 33-statement version of the IT attitude survey.

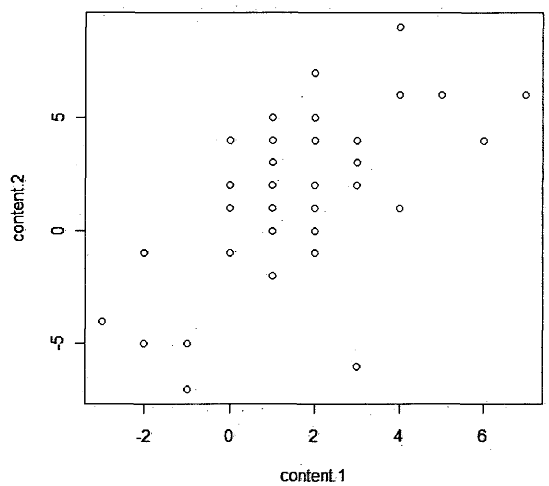

Figure 4.5: Plot of *content.1* versus *content.2*

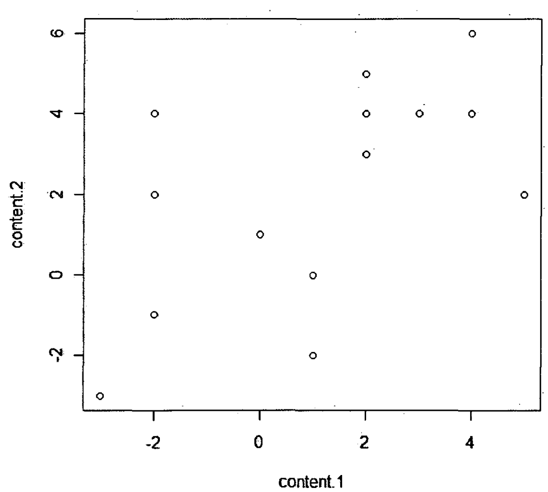

Figure 4.6: Plot of *content.1* versus *content.2*

Due to large p values from the overall nested F-tests, both the pre-to-post and pre-to-post-post Factor G model for counselors reduced down to the following:

$$Y_i = \overline{y} + \epsilon_i$$

4.4.7 Comparison of Factor I Subgroups

The year in which a teacher participated in SPIRIT significantly explained differences in factor I scores for the pre-to-post-post model. No other subgroup differences in factor I scores were found to be significant across teacher and counselor participants. Table 4.28 contains a summary of mean differences in factor I scores for teachers and counselors. Significant pre-to-post-post changes in general interest in IT were found for teachers in 2009 and 2010 and counselors across all three years. No other significant differences were found.

Table 4.28: Comparison of Factor I Scores by Model Type

Pre-to-Post Model			
	Number of Participants	Post-Pre Difference	p-value
Teachers	52	1.92	0.1253
Counselors	17	1.29	0.3566
Pre-to-Post-Post Model			
	Number of Participants	PostPost-Pre Difference	p-value
Teachers, 2008	18	2.00	0.1498
Teachers, 2009	18	6.11	0.0013*
Teachers, 2010	16	8.81	0.0005*
Counselors	17	5.12	0.0000*
* indicates significance at α of 0.05			

Tukey's comparison of multiple means was used to find significant differences among teacher subgroups. A familywise error rate of 0.05 was used. Table 4.29 contains summary information for the Tukey comparison.

Table 4.29: Tukey's HSD of Factor I Scores by Year in Program, Teachers

Pre-to-Post-Post Comparison		Difference	p-value
Teachers, 2009	Teachers, 2008	4.11	0.1776
Teachers, 2010	Teachers, 2008	6.81	0.0149*
Teachers, 2010	Teachers, 2009	2.70	0.4869
* indicates significance at α of 0.05			

A significant difference was found between teachers' pre-to-post-post changes in factor I scores between 2008 and 2010. Teachers who participated in SPIRIT in 2010 displayed a greater positive change in general interest in IT than teachers in 2008. ?? contains a boxplot summary of teachers' changes in IT interest across the years.

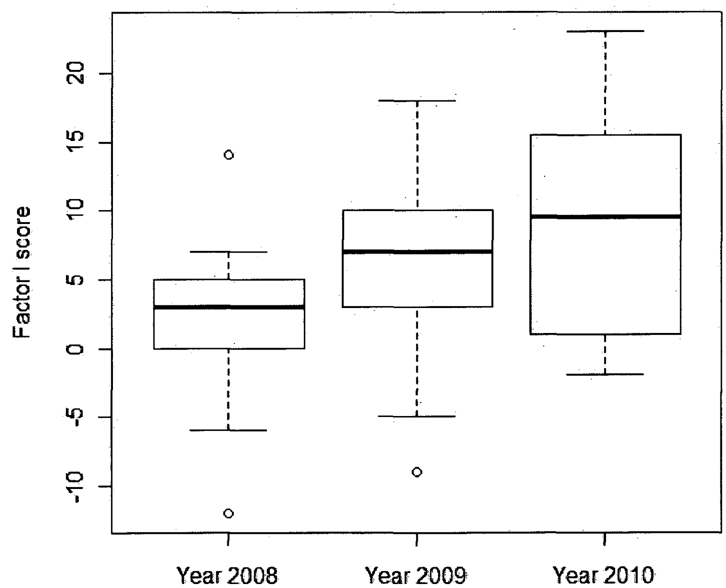

Figure 4.7: Boxplot Summary of Changes in Factor I Scores, Teachers

4.4.8 Comparison of Factor G Subgroups

No subgroup differences in factor G scores were found to be significant across teacher or counselor participants. Table 4.30 contains a summary of mean differences in factor I scores for teachers and counselors. No significant difference were found between pre-to-post or pre-to-post-post changes in factor G scores for either teachers or counselors.

Table 4.30: Comparison of Factor G scores by Model Type

Pre-to-Post Model			
	Number of Participants	Post-Pre Difference	p-value
Teachers	52	0.46	0.3336
Counselors	17	0.06	0.9271
Pre-to-Post-Post Model			
	Number of Participants	PostPost-Pre Difference	p-value
Teachers	52	0.71	0.1035
Counselors	17	-0.29	0.6367
* indicates significance at α of 0.05			

4.5 Comparison of Teachers, Counselors and Students

In order to compare the effect of gender and year of participation across participant group, the reduced 13-statement version of the attitude survey was used. Pre-to-post-post comparisons were not considered in this analysis because post-post scores were not available for students. In addition, one teacher and one counselor were omitted from the following analysis due to inconsistent attitude survey responses. Table 4.31, Table 4.32 and Table 4.33 contain comparisons of standardized regression coefficients for factors I and G for students, teachers and counselors, respectively.

For student participants, participating in 2009 had a negative linear relationship with change in general IT interest and a positive linear relationship with perception of gender stereotypes of the IT-field. In particular, participating in 2009 correlated with viewing the IT field as more gender-neutral. In addition, being a male student participant correlated with a more gender-biased view of the IT field. However, the student factor G model violated the assumption of normality, so results should be interpreted with caution.

Being a male teacher who participated in 2008 correlated with a more gender-neutral view of the IT field. All other model coefficients had large standard errors and were thus not interpreted. The appendix contains a summary of all linear models in this thesis, along with Fligner-Killeen and Shapiro-Wilks p values associated with assumption checks.

Table 4.31: Student Model Coefficients, Factor I and G

	Factor I		Factor G	
Variable	Coefficent	Error	Coefficient	Error
Gender	-0.07	0.07	-0.17*	0.07
Year 2008	0.00	0.08	-0.02	0.09
Year 2009	-0.21*	0.08	0.20*	0.09
Gender*Year 2008	-0.14	0.08	0.14	0.08
Gender*Year 2009	-0.02	0.08	-0.03	0.09
	* indicates significance at α of 0.05			

Table 4.32: Teacher Model Coefficients, Factor I and G

	Factor I		Factor G	
Variable	Coefficent	Error	Coefficient	Error
Gender	0.05	0.15	-0.05	0.14
Year 2008	-0.17	0.18	-0.07	0.17
Year 2009	-0.11	0.18	-0.32	0.16
Gender*Year 2008	0.04	0.18	0.41*	0.17
Gender*Year 2009	-0.06	0.17	0.12	0.16
	* indicates significance at α of 0.05			

Table 4.33: Counselor Model Coefficients, Factor I and G

	Factor I		Factor G	
Variable	Coefficent	Error	Coefficient	Error
Gender	-0.11	0.32	0.20	0.32
Year 2008	0.18	0.29	-0.07	0.30
Gender*Year 2008	-0.17	0.29	-0.06	0.30
	* indicates significance at α of 0.05			

CHAPTER 5
DISCUSSION

This chapter begins with answering each of the four research questions, followed by discussing further details of the findings. Finally, this chapter discusses the implications of this study and the suggested direction of future research.

5.1 Research Questions

This section addresses each research questions in this thesis. Subsequent sections provide more detail about the findings and inferences of this analysis.

1. Are the established factors for the student version and the theoretical factors for the teacher and counselor version of the IT attitude survey valid and reliable?

   Factor I, or general interest in IT, from the student IT attitude survey is both valid and reliable based on a CFA. Factor G, or perception of gender stereotypes in IT, reduced from ten statements to three statements and did not behave well in statistical analyses.

   For the teacher and counselor version of the IT attitude survey, the five predicted factors did not emerge. Rather, Confidence, Usefulness and Interest statements combined onto a single latent factor (factor I). The Gender statements loaded well onto their own latent factor (factor G), and Professional statements did not load well and were removed entirely. The teacher and counselor IT attitude survey showed promising validity and reliability for factor I at the exploratory stage, but a future CFA is desired. Factor G did not behave well in statistical analyses.

2. Based on factors underlying the IT attitude survey, was there a measurable difference in teachers', counselors' and students' attitudes with respect to IT

immediately after, and for teachers and counselors, one year after completing the SPIRIT workshop?

For factor I, significant pre to post differences were found for student, and significant pre to post-post differences were found for teachers and counselors. For factor G, significant pre to post differences were found for students, but no significant differences were found for teachers or counselors.

3. Based on this same instrument, did attitudes differ across gender, ethnic groups and/or the three years of project implementation?

For students, changes in factor I scores differed across ehtnicity and the year of the program. Specifically, significant differences were found between minority students in 2009 compared to 2009 (p=0.0076) and minority students in 2009 compared to non-minority students in 2010 (p=0.0216). For teachers, changes in factor I scores differed across the year of the program. A significant difference was found between teachers pre to post-post change in factor I scores in 2008 compared to 2010 (p=0.0149).

For students, changes in factor G scores differed across gender and the year of the program. Since the interaction between these two variables was not significant, the main effects were interpreted separately. Female students demonstrated a more positive change in perception of gender stereotypes in IT than male students (p=0.0211). A significant difference was found between changes in factor G scores in 2008 compared to 2009 (p=0.0245). For teachers and counselors, no significant subgroup differences were found for factor G.

4. Based on this same instrument, how did regression analyses of the SPIRIT data differ across teachers, counselors and students?

Participating in 2009 correlated with a decrease in IT interest for students. The combination of being a male teacher and participating in 2008 correlated with a negative change in perception of gender stereotypes in IT.

5.2 IT Attitude Survey

The student version of the IT attitude survey demonstrated strong validity based on a CFA and strong reliability based on Jöreskog's estimate of scale reliability. The 13-statement attitude survey explained more variance among the set of survey statements than the 20-statement attitude survey. The factor loadings within each of the factors in the student survey were significantly large and close to one another in magnitude, demonstrating strong convergent validity. The focus group interviews in 2010 further supported the reduction of the gender stereotypes factor. Both teacher and student participants indicated that they were confused with how to interpret statements in factor G. Modification indices from the final model, which estimated how much the χ^2 statistic would improve if a parameter was freely estimated, were small, indicating strong divergent validity. While the reliability for factor G increased when the survey was reduced, the reliability for factor I decreased. However, the changes in the reliability coefficients were small. Given these findings, there is evidence to support that this instrument can be validly used on the participating population of students to examine IT attitudes with respect to general interest. Factor G, however, did not behave as well in statistical analyses. In addition to having too few statements, factor G also led to non-normally distributed results for student participants.

The teacher and counselor version of the IT attitude survey also provided strong validity evidence based on an EFA and good reliability based on Jöreskog's estimate of scale reliability. The factor loadings within each of the factors in the 33-statement survey were significantly large apart from one statement in factor I. However, the factor loading coefficents varied in magnitude within each factor, indicating that the validity is not as strong as desired. The teacher and counselor attitude survey demonstrated discriminant validity since none of the statements crossloaded in the final model. The reliability of factor I was lower than the reliability of factor G. The covariance between factor I and factor G was larger for the 33-statement survey than the 13-statement survey. However, both covariances were small. Future research with additional data can assess the validity and reliability of the survey with greater confidence by running a CFA.

Interestingly, the majority of the factor loadings for the 13-statement student equivalent version of the teacher and counselor IT attitude survey were significant, and none of the statements crossloaded. This further reinforces the fact that the IT attitude survey is successfully capturing two latent factors: general interest in IT and perception of gender stereotypes in IT. However, the 33-statement version of the survey is preferred because it was reduced through an EFA and may be capturing more information than the 13-statement version.

The IT attitude survey may only break down to two latent factors because participants may not fully understand what IT is before participating in SPIRIT. A survey administered to high school students in California and Arizona found that 98% of the students did not have a good grasp of what the CS field entailed [48]. Since IT can be more difficult to define than CS, this effect may have been even larger for participants entering into the SPIRIT program. One concern of this effect is that teachers and counselors often act as "gatekeepers" by directing students to certain fields of study [49]. If a teacher or counselor does not know what IT or CS fully entails, then they may not encourage high school students to take these courses. This may be further exasperated by low-income school districts which cannot afford to offer IT or CS courses [49]. However, after participating in the SPIRIT program, teachers and counselors had a better grasp of what the IT field entails, and are more likely than those who did not participate to encourage high school students to pursue these fields.

5.3 Comparison of Subgroup Differences

The following sections address the second, third and fourth research questions of this thesis. The second and third research questions examined subgroup differences for teachers, counselors and students, and the fourth research question examined differences across all participant types based on equivalent linear models.

5.3.1 Student Subgroups

Significant changes in general IT interest were found for minority students in 2008 and non-minority students in 2010. All other subgroups did not display a significant change in general IT interest. One possible explanation for this is that changes in IT interest take longer than one week to occur, and by implementing the IT attitude survey immediately after the SPIRIT program, these changes in IT interest had not yet fully developed. One possible reason for why there were differences in changes in IT interest across ethnicity is that variations in the SPIRIT program across the years affected minority students differently than non-minority students. However, SPIRIT was an observational study and causation cannot be determined.

Interestingly, research has shown a gender-disparity in IT/computer confidence and course enrollment at the college level, but no self-reported gender-disparity in computer interest, knowledge of CS and plans to major in CS at the college level [50] [51]. The analysis of the SPIRIT program indicates that ethnicity may be an important variable to consider in similar studies. Little research has currently investigated the potential ethnic-disparity in IT interest based on a validated instrument.

Similar results were found for factor G: only female students in 2009 displayed a significant change in perception of gender stereotypes in IT after participating in SPIRIT. No other subgroup differences were found. Finding significant changes was not likely with factor G due to its small size and participants' confusion with answering the statements in factor G. In addition, the model residuals were significantly non-normal.

Based on a study by Weinberg et al. [52], the factor G results for female students in 2009 are promising. In [52], middle and high school female students participated in a robotics educational program. The researchers found that females who believed in traditional gender roles were more likely to have negative self-concepts of ability and lower expectations for success in math and science. With regards to the SPIRIT program, this may mean that a more gender-neutral perception of the IT-field may lead to more positive self-concepts of ability and higher expectations of success in the IT-field.

5.3.2 Teacher and Counselor Subgroups

The IT attitude survey significantly captured positive pre-to-post-post changes in IT interest for teachers in 2009 and 2010 and counselors in 2008 and 2009. No significant pre-to-post changes in IT interest were found, further supporting the hypothesis that changes in IT attitude take longer than a one week workshop to develop.

No significant differences with respect to factor G were found for teachers or counselors. This is likely to be influenced by participants' reported confusion with answering the statements in factor G.

5.4 Comparison of Teachers, Counselors and Students

The effects of gender and the year in which a participant attended SPIRIT were compared across participant type by using equivalent linear regression models and standardizing the regression coefficents. Participating in 2009 correlated with a decrease in IT interest for students. The combination of being a male teacher and participating in 2008 correlated with a gender-neutral view of the IT field. In other words, male teachers in 2008 entered the SPIRIT program with the belief that females were better than males in the IT field and left the program with the belief that female and males were equally skilled in the IT field. Significant coefficents were found for the student factor G model, but the residuals were significantly non-normal. No significant coefficents were found for the counselor models due to large standard errors.

5.5 Implications of Study and Future Research

In summary, the 13-statement student IT attitude survey was able to successfully capture changes in general interest in IT. In addition, the survey was able to capture differences in student subgroups. Given these findings, there is evidence to support that this instrument can be validly used on the participating population of students to examine IT attitudes with respect to general interest. Future research can investigate the effect of including a "neutral" response category to the gender-related attitude survey statements. In this case, all of the original statements in factor G should be included in the administered attitude survey in order to examine if and how the statements load differently. Also, future research can investigate how the wording of the gender-related statements may affect participant responses. For example, gender could be introduced as a scenario in a statement, such as whether or not a participant believes a female IT teacher is more qualified to teach versus a male IT teacher. Another example could involve participants selecting the "most qualified" IT-professional among a set of photographs on a computer screen.

The 33-statement teacher and counselor IT attitude survey loaded well onto a 2-factor EFA structure, but a future CFA is desired before making conclusive statements concerning validity and reliability. The survey was able to capture changes in teachers' general interest in IT across the three years of the SPIRIT program, which is a promising result.

With regards to choosing a factor loading cutoff, future research can examine the effect of plotting the magnitudes of the factor loadings and examining for sudden drops in magnitude. This may be a less arbitrary method for choosing the cutoff level.

Due to the observational nature of this study, causation of the differences between participant subgroups cannot be determined. However, the subgroup differences indicate that participants were responding to changes across the years, however small or large these changes were. This study can provide future direction for tailoring IT educational programs to the type of participant in order to maximize positive changes in attitude with respect to IT. For example, a program can be tailored by presentation topics, such as an engineering-based focus on IT versus a focus of combining IT with medical fields. Interesting findings may result from dividing participants into groups based on presentation topics. In this case, a researcher may bias the educational program, but instruments such as the attitude survey need to remain unbiased. Also, changes in IT attitude may take time to develop, and a future study can examine how IT attitudes change across different time points.

REFERENCES CITED

[1] National Science Foundation. Table 5-1. bachelor's degrees by sex and field: 2000-2008. `http://www.nsf.gov/statistics/wmpd/tables.cfm`, 2008.

[2] W. Duplantis, E. MacGregor, M. Klawe, and M. Ng. Virtual family: an approach to introducing java programming. *SIGCSE Bulletin - Inroads Special Issue: Women and Computing*, **34**(2):40–43, 2002.

[3] N. Ramsey and P. McCorduck. Where are the women in information technology? `http://anitaborg.org/files/abi_wherearethewomen.pdf`, 2005.

[4] N. L. Johnson. Collective problem solving: functionality beyond the individual. `http://www.collectivescience.com/Documents_1/NLJsims_AB_v11.pdf`, 2005.

[5] A. R. Harriger, H. E. Dunsmore, and K. D. Lutes. Surprising possibilities imagined and realized through information technology (spirit). `http://nsf.gov/awardsearch/showAward.do?AwardNumber=0737679`, 2007.

[6] P. Cantrell and J. Ewing-Taylor. Exploring stem career options through collaborative high school seminars. *Journal of Engineering Education*, **July**: 295–303, 2009.

[7] D. Heersink and B. M. Moskal. Measuring high school students' attitudes toward computing, 2010. Proceedings from SIGCSE'10, Milwaukee, Wisconsin.

[8] Q. H. Mahmoud. Revitalizing computing science education. *Computer*, **38**(5): 98–100, 2005.

[9] J. A. Rursch and D. Jacobson. It-adventures: turning high school students "on" to information technology, 2009. Proceedings from 39th ASEE/IEEE Frontiers In Education Conference, San Antonio, Texas: October 18-21.

[10] B. R. Maxim and B. S. Elenbogen. Attracting k-12 students to study computing, 2009. Proceedings from 39th ASEE/IEEE Frontiers In Education Conference,San Antonio, Texas: October 18-21.

[11] R. H. Kay. Practical research tool for assessing ability to use computers: the computer ability survey. *Journal of Research on Computing in Education*, **26** (1):16–27, 1993.

[12] A. Munson, B. Moskal, A. Harriger, T. Lauriski-Karriker, and D. Heersink. Computing in the high school classroom: changing what teachers and students know and believe. *Computers and Education*, **57**(2):1836–1849, 2011.

[13] A. Harriger. Finding success through spirit. *Journal of STEM Education*, **93** (3,4):31–36, 2008.

[14] A. Hutchinson, B. Moskal, W. Dann, S. Cooper, and W. Navidi. *Innovations 2008: International Network for Engineering Education Research.* Begell House Publishing.

[15] S. Cooper, W. Dann, and R. Pausch. Teaching objects-first in introductory computer science, 2003. Proceedings from 34th SIGCSE Technical Symposium on Computer Science Education, Reno, Nevada.

[16] B. Moskal, D. Lurie, and S. Cooper. Evaluating the effectiveness of a new instructional approach, 2004. Proceedings from 35th SIGCSE Technical Symposium on Computer Science Education, Norfolk, Virginia.

[17] B. H. Loyd and C. P. Gressard. Computer attitude scale. *Journal of Computing Research*, **15**(3):241–259, 1984.

[18] A. Hutchinson, B. Moskal, W. Dann, and S. Cooper. Formative assessment: an illustrative example using "alice", 2005. Proceedings from American Society for Engineering Education, Portland, Oregon.

[19] A. Hutchinson. A statistical analysis of outcomes in an educational assessment: impact of the alice curriculum on male and female performances and attitudes at community colleges, 2005. Masters dissertation, Colorado School of Mines.

[20] A. Hutchinson, B. Moskal, W. Dann, and S. Cooper. Impact of the alice curriculum on community college students' attitudes and learning with respect to computer science, 2008. Proceedings from American Society for Engineering Education, Pittsburgh, PA.

[21] K. Powers, S. Ecott, and L. Hirshfield. Through the looking glass: teaching cs0 with alice, 2007. Proceedings from 38th SIGCSE Technical Symposium on Computer Science Education, Covington, Kentucky.

[22] L. L. Briggs. Alice offers gentler intro to programming. `http://thejournal.com/Articles/2008/06/05/Alice-Offers-Gentler-Intro-to-Programming.aspx?sc_lang=en&Page=1`, 2008.

[23] Duke University. Bringing boys and girls to computer science with "alice". `http://www.sciencedaily.com/releases/2009/06/090623112115.htm`, 2009.

[24] T. Wang, W. Mei, S. Lin, S. Chiu, and J. M. Lin. Teaching programming concepts to high school students with alice, 2009. Proceedings from 39th ASEE/IEEE Frontiers In Education Conference. San Antonio, Texas: October 18-21.

[25] B. Moskal, N. Behrens, M. Guzdial, A. Tew, W. Dann, and S. Cooper. Collaborative research: Assessing concept knowledge and attitudes in introductory computer science, 2005. NSF, DUE-0512064.

[26] B. M. Moskal, N. Behrens, M. Guzdial, A. Tew, W. Dann, and S. Cooper. Computer science assessment instrument development: evaluating attitudes and outcomes, 2006. Proceedings from National STEM Assessment Conference, Washington D.C., Oct. 19-21.

[27] K. G. Jöreskog. Statistical analysis of sets of congeneric tests. *Psychometrika*, **36**:109–133, 1971.

[28] T. Raykov. Bias of cronbach's alpha for fixed congeneric measures with correlated errors. *Applied Psychological Measurement*, **25**:69–76, 2001.

[29] B. Muthén and D. Kaplan. A comparison of some methodologies for the factor analysis of non-normal likert variables: A note on the size of the model. *British Journal of Mathematical and Statistical Psychology*, **45**:19–30, 1992.

[30] U. Olsson. Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika*, **44**(4):443–460, 1979.

[31] B. Muthén, S. H. C. du Toit, and D. Spisic. Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes, 1997. Unpublished manuscript.

[32] D. B. Flora and P. J. Curran. An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods*, **9**(4):466–491, 2004.

[33] P. M. Bentler. Eqs structural equations program manual, 1995. Encino, CA: Multivariate Software.

[34] A. B. Costello and J. W. Osborne. Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research and Evaluation*, **10**(7), 2005.

[35] J. C. Nunally. *Psychometric Theory.* New York: McGraw-Hill Book Company, second edition, 1978.

[36] H. Bragadottir and D. Reed. Psychometric instrument evaluation: the pediatric family satisfaction questionnaire. *Pediatric Nursing*, **28**(5):475–482, 2002.

[37] R. Ledesma and G. Molina. Vista-cita: Classical item and test analysis with vista. *Behavior Research Methods*, **37**(4):684–690, 2005.

[38] L. R. Gay and P. Airasian. *Educational Research: Competencies for Analysis and Application.* Merrill: Upper Saddle River, NJ, sixth edition, 2000.

[39] L. R. Fabrigar, D. T. Wegener, R. C. MacCallum, and E. J. Strahan. Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, **4**(3):272–299, 1999.

[40] R. D. Ledesma and P. Valero-Mora. Determining the number of factors to retain in efa: an easy-to-use computer program for carrying out parallel analysis. *Practical Assessment, Research and Evaluation*, **12**(2), 2007.

[41] B. Noble and J. W. Daniel. *Applied Linear Algebra.* Prentice-Hall: Englewood Cliffs, NJ, third edition, 1988.

[42] M. W. Browne. An overview of analytic rotation in exploratory factor analysis. *Multivariate Behavioral Research*, **36**(1):111–150, 2001.

[43] T. A. Brown. *Confirmatory Factor Analysis for Applied Research.* New York: The Guilford Press, 2006.

[44] L. Hu and P. M. Bentler. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, **6**:1–55, 1999.

[45]. J. R. A. Santos. Cronbach's alpha: a tool for assessing the reliability of scales. *Journal of Extension*, **37**(2), 1999.

[46] UCLA SOCR Statistics. Ap statistics curriculum 2007 nonparam varindep. `http://wiki.stat.ucla.edu/socr/index.php`, 2008.

[47] D. Heersink, B. M. Moskal, W. Dann, A. Harriger, and S. Cooper. Investigating high school students' computing beliefs, 2010. Proceedings from Annual Meeting of the American Educational Research Association, Louisville, Kentucky.

[48] L. Carter. Why students with an apparent aptitude for computer science don't choose to major in computer science, 2006. Proceedings from SIGCSE 2006, Houston, Texas: March 1-5.

[49] J. Goode. If you build teachers, will students come? the role of teachers in broadening computer science learning for urban youth. *Journal of Educational Computing Research*, **36**(1):65–88, 2007.

[50] G. H. Rowell, D. G. Perhac, J.A. Hankins, B.C. Parker, C.C. Pettey, and J.M. Iriarte-Gross. Comput,er-related gender differences, 2003. Proceedings from SIGCSE 2003, Reno, Nevada: February 19-23.

[51] S. Beyer, K. Rynes, J. Perrault, K. Hay, and S. Haller. Gender differences in computer science students, 2003. Proceedings from SIGCSE 2003, Reno, Nevada: February 19-23.

[52] J. B. Weinberg, J.C. Pettibone, S. L. Thomas, M.L. Stephen, and C. Stein. The impact of robot projects on girls' attitudes toward science and engineering, 2007. Proceedings from 2007 RSS Robotics in Education Workshop.