SELF-EFFICACY, CONFIDENCE, AND OVERCONFIDENCE AS

CONTRIBUTING FACTORS TO SPREADSHEET DEVELOPMENT ERRORS

A DISSERTATION SUBMITTED TO THE GRADUATE DIVISION OF THE
UNIVERSITY OF HAWAI'I IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

COMMUNICATIONS AND INFORMATION SCIENCES

AUGUST 2005

By
Steven T. Takaki

Dissertation Committee:

Raymond R. Panko, Chairperson
Morris K. Lai
Diane M. Nahl
Shuqiang Zhang
Frederick T. Bail

Acknowledgments

I would like to acknowledge the support I received from several individuals in the preparation of this dissertation. Dr. Raymond Panko was an excellent advisor in directing me on issues relating to the formulation and reporting of this research. Dr. Morris Lai was an invaluable mentor, whose opinion is highly respected and valued over the many years; he was instrumental in guiding me to continue graduate studies and research. My other committee members (Drs. Diane Nahl, Shuqiang Zhang, and Frederick Bail) and proxy (Dr. Rebecca Knuth) also generously provided invaluable guidance, time, and advice throughout the dissertation process. Finally, I would like to acknowledge the support and love of my wife Kathleen, who believed in me, encouraged me, and never complained about my neglect of chores. Also, the support and understanding of my children, Kyle, Stacey, and Sherrie, for all those lost nights and weekends.

While grateful for all their assistance, the responsibility for this dissertation and its contents, including errors and omissions, remains solely with me.

Abstract

Spreadsheet programs are widely used in business and government. Unfortunately, there is strong evidence that many spreadsheets contain errors. In spite of the importance of spreadsheets in decision-making, studies have shown consistently that end-user spreadsheet developers rarely test their models thoroughly after development in the manner that professional programmers test software.

One contributing factor to both error rates and the lack of post development testing may be that spreadsheet developers are overconfident in the correctness of their spreadsheets. Overconfidence is a widespread human tendency, and it has been demonstrated among spreadsheet developers. When people are overconfident, their "stopping rules" for error detection during and after development may be premature, causing them to stop checking before they should. This may contribute to the number of errors.

At the same time, a research construct that appears to be closely related is self-efficacy, which has been shown that high self-efficacy is positively related to computer task performance, including spreadsheet performance (although not specifically to error reduction performance).

The findings from this research concluded that people with high self-efficacy and high confidence make fewer errors than those with low self-efficacy and high confidence. Also, a "think-aloud" protocol analysis of a subset of subjects observed a lack of system design and analysis effort and a minimal amount of testing during the development of spreadsheet tasks.

Table of Contents

List of Tables

List of Figures

Chapter 1. Introduction

Spreadsheet usage is widespread in both business and government, as evidenced by spreadsheet programs being used by a large majority of all managers surveyed in various organizations (U. S. Department of Labor, Bureau of Labor Statistics, 2001). Spreadsheet usage has undeniably altered business procedures and how non-information technology professionals conduct their operations (Floyd, Walls, & Marr, 1995; Troutt, 2005). We do ad hoc analyses in making almost instantaneous decisions and in conducting our daily business with the routine use of this ubiquitous tool. With the advent of computers, many users are seemingly capable of utilizing electronic spreadsheets on personal computers with little or brief foreknowledge (Chan & Storey, 1996; Panko, 2000). Previously, analysis was lengthy and labor intensive, requiring skills of professionals, specifically, trained accountants, who as experts in their field were able to adapt procedural idiosyncrasies to the generally accepted principles of their profession (Hendry & Green, 1994).

While investigating the omnipresence of spreadsheet usage in organizations, the experiments and field audits of spreadsheets from industry have shown a high incidence of errors in spreadsheets (Panko, 2005b). In the past two decades, over 90% of the real-world spreadsheets that were studied were found to have errors (Butler, 2000; Cragg & King, 1993; Davis & Ikin, 1987). Butler (R. R. Panko, personal communications, August, 1996, September, 1996, and August, 1997) found spreadsheet errors in a 1992 tax audit. A lower percentage of errors were found in studies that used only cursory error checking procedures. The cell error rate (CER), the percent of cells containing errors in a cell-by-cell review, showed 2% to 5% error rate (Panko, 2005b). This number is amplified if we consider the total number of errors per spreadsheet and the cascading

1

effects of errors in a spreadsheet. Substantial error rates have been seen by both novices and experienced users working professionally.

A majority of the spreadsheets used by business and government organizations are developed by the non-IT professional end-users (and not the IT professional programmers) for their personal ad hoc use and by their functional organization. Because spreadsheets appear easy to develop and to use, the end-users argue that the lengthy lead times required of their respective IT organization, make professional development impractical (Lacher, 1997).

Panko and Halverson (1996) observed that spreadsheet development by these end-users requires a similar skill to that used in software development to accomplish a finished product. However, many of the testing methodologies employed in software development are not adhered to in the spreadsheet development arena (Galletta, Abraham, El Louadi, Lekse, Pollailis, & Sampler, 1993; Galletta, Hartzel, Johnson, & Joseph, 1997; Hall, 1996; Hendry & Green, 1994; Nardi & Miller, 1991; Panko, 2000; Panko & Spague, 1998; Schultheis & Sumner, 1994). Code inspection and data testing, two common techniques that are used in software testing, are not customarily incorporated into corporate policies (Cale, 1994; Cragg & King, 1993; Floyd et al., 1995; Speir & Brown, 1996). Yet, evidence of informal testing procedures was occasionally found (Cragg & King, 1993; Nardi & Miller, 1991; Panko, 2000). A lack of consistent testing of spreadsheets could possibly be due to the perception that the spreadsheet program appears to be such a simple tool.

1.1 Confidence and Overconfidence

Confidence is a "faith or belief that one will act in a right, proper, or effective way" (Merriam-Webster's Collegiate Dictionary — 11th Edition, 2003). In this research, confidence in spreadsheet development is defined as an affective and cognitive self-assessment of not having committed an error in performing a spreadsheet development task, and overconfidence is the condition whereby the estimation of being error free exceeds the accuracy of the spreadsheet development. A confidence calibration compares actual performance of a task against the estimation of performance; overconfidence occurs when our estimation or prediction of performance exceeds our actual performance.

The literature shows that overconfidence in spreadsheet development is prevalent (Brown & Gould, 1987; Davies & Ikin, 1987; Floyd et al., 1995; Panko, 2000; Panko & Halverson, 1997; Panko & Sprague, 1998; Reithel, Nichols, & Robinson, 1996). However, research also shows a possibility that the confidence calibration can improve and overconfidence can be unlearned. Experiments by Arkes, Christensen, Lai, and Blumer (1987), although not involving spreadsheet development, attained an increase in accuracy of deceptively difficult problems with a continued high level of confidence in the results by giving feedback to the subjects. In spite of improvements due to feedback, errors still remained. Spreadsheet development experiments by Panko and Featherman (2000) were able to reduce overconfidence; they also increased accuracy and reduced errors significantly.

Overconfidence is a widespread human tendency (Barber & Odean, 1999; Plous, 1993; Pulford & Colman, 1996; Panko, 2000). Overconfidence in spreadsheets is no exception with experiments conducted in spreadsheet research at the University of Hawaiʻi demonstrating high error rates in spreadsheet development. These studies

3

include taxonomy of spreadsheet errors (Panko & Halverson, 1996), overconfidence in spreadsheet development (Panko & Halverson, 1997), overconfidence in novice and expert spreadsheet developers (Panko & Sprague, 1998), influence of general feedback on overconfidence (Panko & Featherman, 2000), and external feedback on confidence calibration in spreadsheet development (Goo, 2002).

The research conducted suggests that "stopping rules" for error detection provide a false belief that an effort is complete and correct. Perhaps overconfidence plays a negative role in the lack of continued spreadsheet testing to attain a high quality final product. Banaji, Bazerman, and Chugh (2003) showed that successful people were positive about themselves and considered themselves above average on measures ranging from driving ability to intelligence tasks. This possibly implies overconfidence and might also serve as a catalyst for unconscious over-claiming behavior, or maybe successful people are better at their niche area. Banaji et al. (2003) posited that this unconscious over-claiming is not too different from the claims that all of the fictitious children in Lake Wobegon are above average (Keillor, 1987). Only through a better understanding of the human causal agents leading to spreadsheet errors can we explain and predict the actions of human behavior.

1.2 Self-efficacy

In another research stream, the self-efficacy construct, as cognitive and affective determinants of behavior, has been successfully applied in numerous domains (Stajkovic & Luthans, 1998). Zimmerman (2001, p. 20) describes self-efficacy as "the perceived ability to implement actions necessary to attain a designated performance level." The body of self-efficacy literature shows the widespread applicability of this

construct ranging from clinical problems such as phobias and eating disorders, to athletics and organizational related issues (Bandura, 1997). Educational research also shows a positive relationship between self-efficacy and students; similarly, a positive relationship between self-efficacy and teachers exists (Pajares, 1997). Bandura (1997) posits that self-efficacy is a good predictor of human behavioral changes. Bandura's social cognitive framework provides a triadic reciprocity among behavior, cognition, and environment elements to enable people to alter controllable events within their environment. Thus, high self-efficacy people will set high goals for themselves, remain strongly committed to them, and are able to achieve better results than low self-efficacy people.

Stajkovic and Luthans (1998) conducted a meta-analysis of self-efficacy studies and found that self-efficacy positively affects work related performance. They divided the task complexity into categories of high, medium, and low levels based on the number of dimensions of the task complexity reviewed, such as required knowledge, cognitive ability, memory capacity, persistence, and physical effort. Their study showed performance was moderated by task complexity and whether the experiment was conducted in a simulated or classroom environment versus a naturalistic environment; the relationship between self-efficacy and performance was positive across a varied and vast range of activities. These studies provided a fertile ground to expand the construct into specific areas such as various disciplines in IT (information technology).

As a part of the expanding self-efficacy research stream, computer self-efficacy was researched for IT usage (Compeau & Higgins, 1995, 1999) and specifically for spreadsheet usage (Maracas, Yi, & Johnson, 1998; Johnson & Maracas, 2000). Computer self-efficacy had a positive effect on job performance and a mediating effect on outcome expectancy. General computer self-efficacy was also shown to be an

antecedent to specific computer self-efficacy and IT performance (Agarwal, Sambamurthy, & Stair, 2000; Looney, Valachich, & Akbulut, 2004); however, error reduction performance was not specifically addressed.

Based on the results of self-efficacy research, a high computer self-efficacy person is able to perform better and to achieve better results than a low computer self-efficacy person. In the spreadsheet research arena, this would seem to imply the existence of either fewer or no spreadsheet errors, but this has not been studied empirically.

Another problem is there is a paucity of research distinguishing the constructs — confidence, overconfidence, and self-efficacy. Computer self-efficacy is possibly related to confidence and overconfidence. Sources that this researcher reviewed were Nahl's (1996) and Nahl and Harada's (1996) research positing that self-confidence and self-efficacy are positively related in the information search arena.

The research streams for self-efficacy, confidence, and overconfidence constructs are viewed differently, which may send mixed messages to those unfamiliar with these research streams and cause confusion to the layperson and often the academician. So, are they different? Self-efficacy theory views high self-efficacy as good as these individuals engage in activities that continue to expand on their competencies. Confidence theory, however, views high confidence as problematic, usually leading to overconfidence. Overconfident individuals can be inferred to prematurely apply "stopping rules" for error correction, applying the least amount of effort, or not setting high goals for themselves. The overconfidence condition indicates that the prediction of task accomplishments exceeds the actual results. This leads to less-than-successful error detection performance. Meanwhile, self-efficacy theory indicates that high self-efficacy individuals set higher goals with every successful performance.

6

1.3 Protocol Analysis

This research used "think-aloud" methodology, a protocol analysis procedure, by recording the spreadsheet developer's thought throughout the task to better understand the error detection processes while the spreadsheets were developed and to determine if these processes relate to the subject's responses to the computer self-efficacy and confidence instruments (Ericsson & Simon, 1993). The taxonomy outlined by Panko and Halverson (1997) was utilized to categorize errors encountered and corrected in the problem-solving tasks as well as those left unresolved. A coding scheme, as exemplified in qualitative research by Tesch (1990), incorporated this taxonomy as it related to the software development life cycle (SDLC) and with specific focus on the design and testing phases.

1.4 Research Approach

This research study used experimental research to better understand these constructs, specific to the arena of errors in spreadsheet development. The experimental research design used two word problems involving spreadsheet tasks, one requesting a pro-forma income statement and the other a domain-independent project bid proposal. Responses were solicited on confidence and computer self-efficacy with instruments using Likert scales during the spreadsheet development of each task. An estimate of the number of errors was also asked to calculate the overconfidence condition. The spreadsheet errors were measured after the spreadsheet developer decided the task was successfully completed.

1.5 Research Contribution

Organizations empower their knowledge-based end-users to use spreadsheets in their operational environments. Spreadsheet usage is influenced by confidence and computer self-efficacy, which are affective variables; therefore, affective dynamics strongly influence their cognitive operations (Nahl, 1995, 1996). Many errors are detected in the resultant spreadsheets. Both confidence and computer self-efficacy, as antecedents of human errors, are critical to our understanding in the information systems (IS) and the information technology (IT) domains. This research bridges this gap in our knowledge.

Chapter 2. Theory and Research

The objective of this research is to extend our understanding of human error performance in spreadsheet usage. With the advent of the personal computer and with the introduction of commercially available computer spreadsheet software in the late 1970s and the early 1980s, the ubiquitous electronic usage of spreadsheet software permeated every facet of business as well as our everyday personal life.

This chapter begins with a review of general human error research and spreadsheet error research. The confidence literature is reviewed next, with a specific focus on errors due to overconfidence. The literature review concludes with a review of Bandura's self-efficacy theory and its relationship to confidence in the development of spreadsheets and the resultant errors in spreadsheets. Spreadsheet errors can have a significant impact on an organization's decision-making and productivity.

2.1 Widespread Use of Spreadsheets

The electronic spreadsheet became a "killer application" for the personal computer when it was introduced in the early 1980s and provided a base for the success of the personal computer (Horowitz, 2004). This end-user tool transformed the way business was conducted. Although appearing to have large benefits from its capabilities, the spreadsheet is a potentially dangerous tool. The development of spreadsheets requires substantial cognitive activity and problem-solving capabilities. Organizations empower their end-users to utilize the tool and to accomplish cognitive oriented and often cognitive intensive tasks in our knowledge-based society.

Earlier versions of spreadsheet analyses were created for mainframe computers

in the late 1970s, prior to the economic popularity of personal computers (e.g., Dynaplan, 2004, http://www.wrkgrp.com/dynasoft). Due to the cost of hardware and software associated with these platforms, the usage of spreadsheets was limited to larger organizations, mainly those organizations that had access to the hardware platforms or the software developers who provided their users with these capabilities (Ronen, Palley, & Lucas, 1989). According to one of the original authors, John Cifonelli (personal communications, June 20, 2004), Dynaplan is still used today on IBM's AS/400 midrange systems.

Nardi and Miller (1991) posited that spreadsheet developers expended only a small investment of time to produce a functioning program of real use. The simple table format of electronic spreadsheets provided an effective visual form for data presentation. The fast and early successes in spreadsheet usage encouraged the continued use of the spreadsheet software. In Nardi and Miller's (1991) ethnographic study, the spreadsheet users were knowledgeable in their respective functional areas. Thus, this seemingly neutral offering with large productivity benefits laid the base for a potentially hidden and dangerous environment for human error. As time progressed, both size and complexity of spreadsheets introduced added dimensions, which further lend themselves to susceptibility to error.

The original versions of spreadsheets for the personal computer introduced in the early 1980s, like Visicalc and Lotus 1-2-3, had a maximum of 64 columns and 256 rows, or a total of 4096 cells (Pender, 1998). More recent versions greatly exceed 4096 rows and columns each, enabling larger spreadsheets; size is now limited by the amount of addressable computer memory available for use. More recently, new added functionalities were incorporated; for example, Microsoft's Excel includes:

- Macros, where several functions, other computer languages, and

external data can be combined and referenced;

- Multiple spreadsheets comprising worksheets within a workbook allowing individual spreadsheets to reference each other to produce a common result;

- Display of charts and graphs based on retrieved data; and

- Retrieval of query data from external sources.

An implicit characteristic of this evolution of spreadsheet capabilities revolved around the personnel skills needed and the training required for the effective utilization of this tool (Nardi & Miller, 1991). Based on the ubiquity of the tool today, limited computer skills and associated accounting and finance skills lend credence that specialized training and skills are not prerequisite requirements for spreadsheet developers (Floyd et al., 1995; Hendry & Green, 1994).

2.1.1 Human Error Research

A theme emerged in the early human error literature where errors were attributed to the system or the environment. It was felt that externally related factors causing the accidents could be altered, thereby reducing errors (Petersen, 1996). These studies examined accidents that ranged from car incidents to power plant mishaps where safety issues were the primary concern. Examples of occurrences of some of the worst catastrophes were the Three Mile Island near meltdown (1979), the Bhopal chemical leak (1984), the Chernobyl nuclear power plant accident (1986), and the Challenger explosion (1986).

2.1.1 Early Studies

A great number of these major incidents resulted in corrective actions to traditional safety management that assisted the operator in controlling these complex environments (Petersen, 1996). Through safety training programs for both employees and supervisors, accident investigations and inspections attempted to reduce accidents through safety media and proper paperwork keeping.

In spite of the continued focus on externally related factors, since 1931, some safety practitioners continue to hold a fundamental belief that the single largest reason for the cause of accidents is people (Petersen, 1996). Senders and Moray (1991) estimated that "between 30% to 80% of serious incidents are due, in some way, to human error" (p. 2).

2.1.1.1 Transition to a Cognitive Focus

With the advent of cognitive psychology to complement behavioral psychology, there has been a much discussed topic between environment factors (Petersen, 1996) and human factors or errors (Norman, 1981; Reason, 1990) as causes of accidents. Earlier error studies focused on behavioral processes (e.g., Rasmussen, 1986, 1990; Rumar, 1990); more recent studies focus on cognitive processes (e.g., Panko, 1996). A paradigm shift away from systems and equipment as the sole source cause of the accident started to occur in safety management. A human error can occur in the design, operation, management and maintenance of complex systems (Petersen, 1996). These complex systems are the product of advances in modern society created by technology.

Very few systematic studies were done on errors prior to the aforementioned major catastrophes. A conference was convened in 1983 with attendees from various

disciplines to examine the role of human error (Senders & Moray, 1991). Errors were considered a result or measure of some other variable, and not a phenomenon of its own. Previously, an error was considered the same as an accident.

2.1.2 Types of Errors

In the human error research literature, there were various taxonomies of errors relating to the cognitive processes of human errors. Reason (1990) developed the Generic Error Modeling System (GEMS) in defining error types based on Rasmussen's skill-rule-knowledge classification of human performance levels. Rasmussen's (1986) skill-based behavior represented sensorimotor performance based on an automatic and highly coordinated pattern of behavior with occasional error feedback; his rule-based behavior occurred in familiar and consciously controlled work situations derived from previous experiences stored as a rule and the skills developed through training. Reason's taxonomy of error types in GEMS categorized skill-based slips and lapses, rule-based mistakes, and knowledge-based mistakes. Previously, slips and mistakes were treated as separate errors; Reason's GEMS framework attempted to integrate these two areas with error operating at all three levels.

In undertaking a study of human error, research categorized errors into slips, lapses, and mistakes. Slips (and lapses), as defined by Reason and Norman, referred to the unintended error of execution of correctly intended or planned action (Senders & Moray, 1991). Mistakes were due to failure of planning itself to achieve its objectives (Reason, 1990).

Reason (1990) provided working definitions of errors for slips and mistakes. Slips and lapses were errors resulting from some failure in the execution stage of an action

13

sequence, regardless of being planned or not; errors of slips and lapses were mainly associated with a failure to properly monitor. Mistakes were deficiencies or failures in judgment and/or inferential processes in the selection or execution in achieving the goal, regardless of the actions directed towards the plan; subsequent errors, rule-based mistakes and knowledge-based mistakes, were mainly attributed to problem-solving failures. Skill-based errors were usually detected with fair rapidity, while the rule-based and knowledge-based mistakes were often difficult to detect and required external intervention for detection and resolution (Reason, 1990).

One of the key features of the GEMS framework was the assertion that humans were biased in searching for a previous solution for a rule before activating the knowledge-based performance level. This pattern-matching characteristic was a common human behavior (Reason, 1987). This approach of matching aspects of the local information sources and states to a situation suggested potential areas where errors caused by overconfidence occurred. Reason offered overconfidence as one source of failure at the knowledge-based level. Human variability was also a base part of adaptation and manifested itself in human error. Problem solvers were likely to be overconfident by justifying their actions on evidence that favored their position while disregarding contrary but valid information.

While examining statistical problem solvers using "think-aloud" protocol analysis, Allwood (1984) identified five categories of errors: execution errors, skip errors, solution method errors, higher level mathematical errors, and other errors. Execution errors were the low level errors such as copying and keying errors, while skip errors occurred when the problem solver forgot or was unable to complete a step. Solution method errors were the erroneous specification of a solution method, while higher mathematical errors involved more advanced mathematical functions. Other types of errors included the

14

remaining errors not included in the other categories.

Specific to examining spreadsheet errors, Panko and Halverson (1996) refined the error types from Allwood's (1984) classification scheme. This categorization was valid for reviewing errors after their occurrence, which suited analyses of the results for spreadsheets errors. A simpler classification of error types categorized errors into mechanical errors, logical errors, and errors of omission. Mechanical errors were simple physical miscues; for example, referencing the wrong cell or mistyping a number was categorized as mechanical errors. Logical errors were errors in using a wrong formula. Finally, omission errors were facts left out from the problem statement. This taxonomy, which closely parallel's Allwood's (1984) categorization, is used as the basis for a portion of this research using "think-aloud" protocol analysis.

2.1.3 The Existence of Spreadsheet Errors

Electronic spreadsheets have a number of salient characteristics that inherently contribute to errors. In many organizations, the spreadsheet development responsibility often resides in the functional or end-user organization, where function knowledge is most prevalent; the other software development effort targeted for corporate functions belong in the IT organization, where software development skills are emphasized for corporate development efforts. A common practice in use by both software developers and spreadsheet developers is referencing previous calculations and prior cell locations to reduce the number and complexity of formulae in use. This approach assumes that the previous work performed is correct. This practice is commonly employed in the structured software programming methodology today and has a ripple effect if early error detection testing is not performed. Spreadsheets, like other software programs, are

15

continuously enhanced and modified based on new requirements. Without extensive testing, these changes contribute to the spreadsheet errors. The growth in size of these spreadsheet matrices further adds to the errors, as well as functional enhancements and corrections to the spreadsheet software product.

It has been well documented that spreadsheet models developed by end-users have a high error rate (Panko, 2005b). Many errors go undetected and can have a deleterious and magnifying effect on the performance and decision-making processes of organizations and individuals. Yet, the information from the spreadsheets continued to be used by both businesses and personal users.

An early documented incident relating to a spreadsheet error occurred during a construction company's bid in Fort Lauderdale, Florida on a $3 million office complex (Stone & Black, 1989). The controller for the construction company of James A. Cummings, Inc. entered an overhead cost element of $254,000 in a spreadsheet cell that fell outside the summed cells, and this overhead cost was consequently omitted from the bid. James A. Cummings, Inc. won the bid. The project resulted in a huge loss to the firm, who subsequently sued the creator of the spreadsheet package for this error. This lawsuit was subsequently dropped (Stone & Black, 1989). This early example illustrated the grave consequences of a seemingly minor spreadsheet error.

Experiments and field audits in error research studies also have shown the existence of human errors. A current and complete list of research efforts in spreadsheet error research is maintained at the Spreadsheet Research Website (Panko, 2005b), and an extract is exhibited in Figure 2.1. The field audits, development experiments, and code inspection experiments all showed continued errors. Specifically, the field audits of operational spreadsheets reviewed in organizations found errors in 24% – 91% of the spreadsheets examined prior to 1995. Since 1995, most of the spreadsheets examined

16

Sources	Participant	Cell Error Rate (CER)	Percent of Models with Errors
Field Audits			
Hicks (1995)	Real company spreadsheets	1.2%	26%
Coopers & Lybrand (1997)	23 spreadsheets from industry		91%
KPMG (1997)	22 spreadsheets from industry		91%
Butler (2000)	Tax submission		86%
Total	367 spreadsheets (weighted average)		24%
Since 1997	54 spreadsheets (weighted average)		91%
Development Experiments			
Janvrin & Morrison (1996, 2000)	Upper & masters business students	7%–14%*	84-95%
Janvrin & Morrison (1996, 2000)	Senior level accounting students	8%–17%	
Panko & Halverson (1997)	Business students	5.4%	80%
Panko & Halverson (1997)	Undergraduate students	2.0%	42%
Panko & Sprague (1999)	Upper and masters business students	2.0%	35%
Unpublished	Business students working alone and in triads	4.6%, 1.0%	86%, 27%
Code Inspection Experiments			
Galletta et al. (1993)	MBA students and CPA accountants	34%–54% **	
Galletta et al. (1997)	MBA students	45%–55% **	
Panko & Sprague (1998)	Business students (with seeded errors)	81% **	
Panko (1999)	Business students (individually & in groups)	40%, 17% **	

* Errors per inter-spreadsheet link   ** Percent of seeded errors undetected

Figure 2.1:   Spreadsheet-error research

contained errors (over 86% in the majority of the audits), indicating that errors have not

abated. Many of these studies attempted to replicate earlier studies. In the cell entry

experiments, a cell-by-cell review was conducted to obtain a cell error rate (CER) metric,

a percentage of unlabeled cells containing errors. This metric is similar to the error rate

per thousands of lines of code (KLOC) metric, commonly used in software programming.

The early code inspection studies, where spreadsheets reviewed by other developers,

also showed a significantly high incidence of errors in spreadsheets even after reviews

(Galletta et al., 1997; Panko, 1999).

2.1.4 Spreadsheet Development as Programming Development

Fagan (1976) outlined the code inspection methodology used in the software

engineering's testing and development processes. Panko and Halverson (1996) posited

that spreadsheet development required a similar set of skills to accomplish a finished

product. However, many of the testing methodologies employed in software

development have not been adhered to in the spreadsheet development area. Hicks'

field audits (R. R. Panko, personal communications, June 21, 1995) used a methodology

similar to code inspection as used in software programming.

Although spreadsheet developers took many informal precautions during the

development of a spreadsheet (Hendry & Green, 1994; Nardi & Miller, 1991), they had

not taken many of the formal precautions nor used testing methodologies that IT

software programmers performed (Cragg & King, 1993). The spreadsheet developer

was often not given the formal software development training to develop skills such as

extensive and recursive testing used in software engineering, which were needed to

18

ensure a high product quality.

The Software Development Life Cycle (SDLC), adhered to in software engineering, required extensive software testing processes, often ranging from 30% to 45% of the total project effort (NASA, 1992; Panko, 2000). Many of the spreadsheet developers arrived at their position by being functionally knowledgeable. In organizations, the spreadsheet development responsibility often resided in the functional or end-user organization rather than the IT organization, where the software development skills were emphasized and resided. However, many organizations were reluctant to disclose or acknowledge the extent of the existence of errors in their spreadsheets (Kruck, Maher, & Barkhi, 2003).

Panko (1999) argued that spreadsheet development needed to follow many of the software development methodology phases if spreadsheets were to be reliable; specifically, following a more extensive testing effort was needed in spreadsheet development. The error rates in spreadsheet development were similar to software development error rates (Panko, 2005b). The structured walk through approach in software development, which was followed in the SLDC methodology, involved multiple people to check and to validate the design and the code itself. Since individual error detection rates were too low, inspection and review by groups rather than individuals were done. Group code inspection usually followed two phases, individuals usually studied the modules working alone, followed by a review of the modules collectively. Discoveries by pools of individuals yielded acceptable error detection. Panko's experiment showed that teamwork improved error detection rates, especially errors that the individual found difficult to detect. Although the group had not discovered new errors, undetected errors remained. Panko argued that an extensive testing phase was needed to reduce the error rate to more acceptable levels. A concern in spreadsheet code

19

inspection research was overconfidence (Panko, 1999). Panko (1999) argued that we

tend to act on metacognitive beliefs using our rules and cease to check for errors too

soon, as if we are overconfident.

Besides code inspection, the SDLC approach also included execution testing,

auditing, and other phases to increase the quality of the code. Detection of errors in

inspection research has shown that commission and detection rates vary by error types

(Allwood, 1984). Allwood's (1984) investigation into statistical problem-solving behavior

included evaluative activities to determine whether the solution was correct or wrong. His

assertion was that negative evaluations triggered a set of error detection processes,

ranging from initiation, to discovery, and to the elimination of errors. He developed three

error detection categories: standard check, direct error hypothesis formation, and error

suspicion. The standard check was comprised of a general check that was not

associated with a directly diagnosed error and independent of any specific condition,

while the direct error hypothesis formation was initiated by a presumed detection of an

error. Error suspicion, on the other hand, was initiated when the solution appeared

strange or unexpected to the subject and associated with a specific error diagnosis.

Utilizing Allwood's taxonomy, Panko and Halverson (1996, 1997) categorized

error types into logical, mechanical, and omission errors for spreadsheet analyses.

Incorporating these categories of error in extending our understanding of errors, a

protocol analysis effort using the "think-aloud" methodology attempted to study the

incidence of errors and how subjects detected their own errors, as many errors were

made while developing spreadsheets but were caught and corrected before a workable

solution was completed and submitted. The "think-aloud" protocol was used to

investigate salient characteristics of subjects' problem-solving skills and strategies while

developing the experiment spreadsheets. The detection of errors was viewed as part of

20

the larger context of the problem-solving process; the detection of errors was the end

result of certain kinds of evaluation process (Allwood, 1984).

2.2 Confidence

In the domain of sport psychology, Vealey (1986) defines sport-confidence as

"the belief or degree of certainty individuals possess about their ability to be successful

in sport" (p. 222). Vealey, Hayashi, Garner-Holman, and Giacobbi (1998) focused on the

mediating effects of confidence on cognition, affect, and behavior specific to sport and

motor performance. This estimation of performance expectation or correctly responding

to a situation or question is an assessment of a person's knowledge and/or ability within

a specific domain environment of the task. This assessment is an indication of a

person's judgment about successfully performing a task, or a person's confidence or

belief in their abilities to succeed.

In athletic competition events, the superior athlete often fails to succeed in an

event, while losing to a competitor of equal or less physical ability. The winning athlete

had higher confidence in the event and a stronger belief that victory was achievable

(Vealey, 1986, 1988). In end-user computing outside of the IT department, high

confidence was expressed by the end-users in either developing or using spreadsheets

(Floyd et al., 1995).

If the estimation or prediction fails to match the actual performance of the task,

then an error in calibration occurs. Calibration measures the association between the

objective (accuracy) and subjective (confidence) occurrence of an event (Weber &

Brewer, 2004). When the prediction estimate exceeds the actual performance, an

overconfidence condition occurs. Similarly, when the actual results exceed the predicted

estimate, then an underconfidence condition occurs. Although both conditions exist and have negative effects, the consequences of missed opportunities of underconfidence are often recognized but are often unnoticed. The focus in this research is on the overconfidence end of the calibration spectrum.

2.2.1 Overconfidence

The overconfidence phenomenon is reflected in three general theoretical perspectives, labeled as Brunswikian, Thurstonian, and cognitive bias (Ayton & McClelland, 1997). First, the probabilistic functionalism of Brunswik claims that overconfidence is a methodological illusion caused by a large number of misleading questions in conducting the experiments. Gigerenzer, Hoffrage, and Kleinbolting (1991) reformatted a number of the probability questions with frequency terms and many biases were reduced, although not eliminated. They argued that overconfidence bias is due to the methodological illusion, a misunderstanding of probability information. Secondly, the Thurstonian perspective suggests that overconfidence is a reflection of an error component of judgment affecting the calibration calculation. A third perspective, cognitive bias, suggests bias in a probabilistic judgment due to a strategic evaluation, as is illustrated in the works of Tverskey and Kahneman (1982b, 1982c). This is further discussed later. This research focuses on this latter perspective, cognitive bias.

Overconfidence appears as a general human tendency based on the literature and is prevalent in various domains (Reason, 1990; Plous, 1993). In heuristics and in the judgment and human decision-making fields, this human tendency was found in driving skills (Rumar, 1990), in personal lives, in response to general knowledge questions (Lichtenstein & Fishhoff, 1980), in performance of card games (Reason, 1990), in

performance of experts (Shanteau, 1992), and in business planning (Koriat, Lichtenstein, & Fischoff, 1980). An overconfidence condition in driving skills caused late detection of errors, as drivers were not taking the necessary precautionary measures, which resulted in a higher number of accidents (Rumar, 1990).

Problem solvers have a marked tendency to be overconfident when evaluating their usage of their knowledge base (Koriat et al., 1980). They also tended to justify their course of action by focusing on actions that favor their selection criteria, while disregarding alternative and contradictory information (Reason, 1987). The overconfidence condition was one of the most consistent findings in behavioral research (Pulford & Colman, 1996) and pervades all aspects of human cognition.

Bandura (1997) posits that people err in self-appraisals of their performance because they overestimate their capabilities. He argues that the accuracy of self-appraisals, which uses behavior as the standard to evaluate personal efficacy, exceeds performance and reflects an overconfidence judgment. Zimmerman's (2000) cyclic self-regulatory model posits that self-regulation incorporates self-efficacy in its forethought phase. Bandura (1997) also argues that the "performance is usually confounded with interacting motivational, self-regulation and affective non-ability determinants" (p. 71). Bandura (1997) further posits "for familiar activities that must be performed regularly to achieve desired results, it is their perceived self-regulatory efficacy, rather than perceived efficacy for the activity per se, that is most relevant" (p. 64). Although similar to previously experienced tasks, familiar tasks also have differences and variances. Thus, the ambiguity of similar and familiar task demands contributes to insufficient allowance for impediments and may also contribute to an overconfidence judgment. Bandura continues to argue that overconfidence is a misjudgment of task requirement and demands as well as personal cognition (pp. 63-72).

23

Mayer (1992, 2003), while researching learning issues, segregated knowledge into four categories: factual knowledge, semantic or syntactic knowledge, schematic knowledge, and strategic knowledge. Errors can occur within each of these areas of knowledge. Factual knowledge was the basic domain information. Knowledge of the concepts underlying a given situation was the semantic or syntactic knowledge, while the schematic knowledge was the understanding of the distinction between surface information and structural similarities. Strategic knowledge was how humans aggregate, generate, and implement plans in solving problems to address a situation. Mayer also differentiated between the knowledge of novice and expert learners and their approaches to solving problems. Experts clearly have more extensive and up-to-date specific domain knowledge than novices; this extensive knowledge leads to a loss of expertise outside one's domain (Mayer, 1992, 2003; Shanteau, 1992; Sternberg, 2003). It is the expert's chunking or the mentally grouping of pieces of information into meaningful clusters that enable experts to develop heuristics to recall situations (Mayer, 2003; Sternberg, 2003). The thinking of experts relied more on automatic responses and relied less on deductive thinking and more on pattern-matching thinking (Shanteau, 1992). This cognitive shortcut was not without its shortcomings as errors of oversight and overconfidence occurred if details differed.

To further examine experts and novices, Chase and Simon (1973) combined with de Groot (1965) in a classic study reviewed expert chess players and compared the cognitive functioning of chess experts and novices. Since chess experts clearly have more skill in chess than novice chess players, Chase and Simon postulated that if chess expertise were a general skill, then one would expect these experts to perform better than novices in other kinds of memory tests. Based on his studies, Chase and Simon found that both experts and novices performed at similar levels on the standard memory

24

tests; however, experts outperformed novices in remembering positions of chess pieces on the board of an actual chess game. In a similar test but randomly positioning the chess pieces, novices performed as well as experts. Chase and Simon concluded that chess experts outperformed novices on domain specific tests of memory by developing skills to logically relate chess positions in this domain specific area of knowledge (Mayer, 2003; Sternberg, 2003, Vicente & de Groot, 1990).

In the spreadsheet error research arena, Brown and Gould (1987) obtained a median score of "very confident" in their spreadsheet development tasks. Davies and Ikin (1987) and Floyd et al. (1995) found similar results. In an interesting study on overconfidence during their examination of larger spreadsheets, which should contain more errors than the smaller ones, Reithel et al. (1996) rated large well-formatted spreadsheets higher in their confidence for the correctness of their spreadsheets than small minimally formatted spreadsheets.

Experiments utilizing feedback found that people who were initially overconfident could learn to become better calibrated. Arkes et al. (1987) found that confidence calibration correction could be learned; however, this was done in a very limited environment as very little feedback was given.

In an experiment conducted by Panko and Featherman (2000), overconfidence in spreadsheet development was decreased using feedback. By providing feedback about error rates of others developing a similar spreadsheet, overconfidence only moderately decreased; however, the accuracy of the spreadsheets increased slightly. This result provided encouragement that overconfidence can be reduced. Goo's (2002) research on task error feedback attempted to further uncover the effects of negative feedback on errors in spreadsheet development. Overconfidence also occurred in his experiment, but, there were no statistically significant results, raising the question of whether some

25

other confounding variable, or experimental contamination, affected the results.

Panko and Sprague (1998) conducted spreadsheet tests on both undergraduates and MBA students. This experiment used a simple and domain free knowledge model to measure the error rates of the subjects. Utilizing the procedures followed by Galletta et al. (1997) in separating the MBA students into experienced and inexperienced developers based on hours of experience, even the MBAs with over 250 hours of working experience had a 24% error rate. There was no significant difference in errors per model between the other MBAs and undergraduates. In this experiment, the MBA sample contained a mean of 2,269 hours of experience and a median of 635 hours; this was well over the sample separation point used in the experiment (250 hours). This led to the conclusion that there was no significant difference between a novice group and a more experienced group in the domain of spreadsheet modeling. Panko and Sprague (1998) concluded that making a spreadsheet error was not a novice-level phenomenon.

In the decision-making research literature, a miscalibration in the confidence calibration has been reported by Shanteau (1992). He also reported that experts used less information in decision-making than novices. The discarding of both relevant and irrelevant information used in making decisions by experts was evident. A conclusion derived from this line of research was that "experts" often were inadequate decision-makers. Shanteau (1992) delineated several psychological features possessed by experts including their keen perceptual and attention abilities, a sense of what was relevant in making their decisions, simplification of complex problems, and strong self-confidence. This latter point also contributed to the expert's overconfidence condition.

Keren (1992) distinguished between good and bad decisions by contrasting the optimal or ideal decision from the actual decision. These human constraints assumed that the decision-maker was confined by natural human limitations including limits on

26

memory and processing capability; this was substantiated by the information processing framework paradigm. Keren (1992) argued that humans were error prone and vulnerable to emotional and motivational states influencing their performance. Keren (1992) posited that the normative or ideal and optimal approach of the "superrational" or "idealized" human being was used as fundamental assumptions. He proposed a prescriptive approach, which still targeted for optimization but was more tuned to the human limitations of information processing and took into account all not rational effects such as emotion and motivation. This differentiation between good and bad decisions could also be attributed to overconfidence.

2.2.2 Debate on General Knowledge Questions

The debate surrounding the validity of general knowledge questions was in part related to overconfidence and biased responses. The research conducted used known and correct responses to general knowledge questions. "Whether New York City or Rome is further north" is an example of a general knowledge question that is debated. The veracity of a weather forecast would not be known until sometime in the future; hence, it would be argued as not being a general knowledge question. This debate surrounds the validity of general knowledge questions used.

Tversky and Kahneman (1982c) cited representativeness, availability, and anchoring and adjustment as heuristics, where many judgment decisions were based and inferred from events and experiential information that were derived from limited data validity. Representativeness, or similarity, as an example of a heuristic rule we often followed, lead us to a biased position and subject us to serious errors; this was usually employed when asked to judge the probability that an object or event belonged to a

27

class or event. An example often used as a general knowledge question was based on a description of an individual's personality within a prescribed environment; a female bank teller and her participation in a political rally, as an example used by Tversky and Kahneman (1982c), appeared as incongruous events. Without obtaining the necessary additional factual information, the conclusion derived resulted in a biased position and was not specific to the actual situation. A base-rate frequency or prior probability should not be affected and should be independent of sample size; yet, were often biased with a representativeness description, which results in a biased decision. Tversky and Kahneman (1982b) referred to this as the base-rate neglect. In the tossing of a fair coin, as another example, we should have an equal probability of a head as well as a tail. As a misconception of chance, or a lingering belief in what Tversky and Kahneman (1982a) facetiously called the "law of small numbers," a run of heads may lead us to believe that the random process may not be temporarily valid, making us believe that a head will occur on the next flip, even with a fair coin.

Another heuristic commonly encountered is the number of instances or occurrences that can be readily brought to mind. Based on the premise that instances of large classes would usually be reached better and faster than instances of less frequent classes, the most current and recent occurrence most likely would be recalled when the subject is asked to assess the plausibility of an event. This retrievability of instances often resulted in a biased cognitive view, causing a potential error condition.

Another example used in general knowledge questions was whether London was further north than Chicago. In this situation, an initial estimate was developed from preliminary information and was adjusted to the final answer. This anchoring and adjustment heuristic was biased by the initial values or starting points. When evaluating compound events, people tended to overestimate the probability of conjunctive or a

combination of events, while underestimating the probability of disjunctive or separately

occurrences of events. Bar-Hillel (1973) referred to this as the conjunction effect. The

conjunction rule of P(A ∩ B) ≤ P(A), where an additional specification (B) can only

reduce the probability of a combination of both events (A and B), is one of the basic laws

of probability and was illustrated in studies by Tversky and Kahneman (1982b). These

heuristics, while highly economical and efficient from a cognitive perspective, led to

systematic and predictable errors.

The classic approach is that general knowledge can be fully and easily

transferred from one domain to another domain. This classic belief was contradicted as

research evidence demonstrated that a preponderance of specific skills learned in one

domain was successfully used mainly in that domain (Mayer, 1992, 2003). This in part

explained the limitations of general knowledge questions when asked of individuals that

did not possess specific domain knowledge.

In the 1980's, the question of people being biased in their confidence judgment

appeared settled with people being overconfident on all but the easiest question

(Klayman, Soll, Gonzalez-Vallejo, & Barlas, 1999). This debate was reopened in the

1990's where opponents argued that the choices of questions were biased and people

are imperfect with unbiased judges of confidence. The arguments continue today with

differences occurring in overconfidence between domains and between individuals

(Klayman et al., 1999)

2.3 Self-efficacy

Bandura (1977a) first introduced the theory of self-efficacy in *Self-efficacy:*

Toward a unifying theory of behavioral change (*Psychological Review*, 1977). He later

incorporated the self-efficacy construct into his *Social Learning Theory* (1977b) and

subsequently into his larger Social Cognitive Theory of human behavior (Pajares, 1997).

In Bandura's (1986) Social Cognitive Theory, people operate on a model of

triadic reciprocity in which behavior, environmental events, and personal factors in terms

of cognitive, affective and biological events, all interact as determinants of each other

(Figure 2.2). The interdependency of internal and external stimuli is collectively the

driving force, and not as separate forces. Individuals possess a self-system enabling

them to exercise control over thoughts, feelings, motivation, and actions. The desire for

control translates into their action. This self-referent thought mediates the relationship

between knowledge and action (Bandura, 1986).
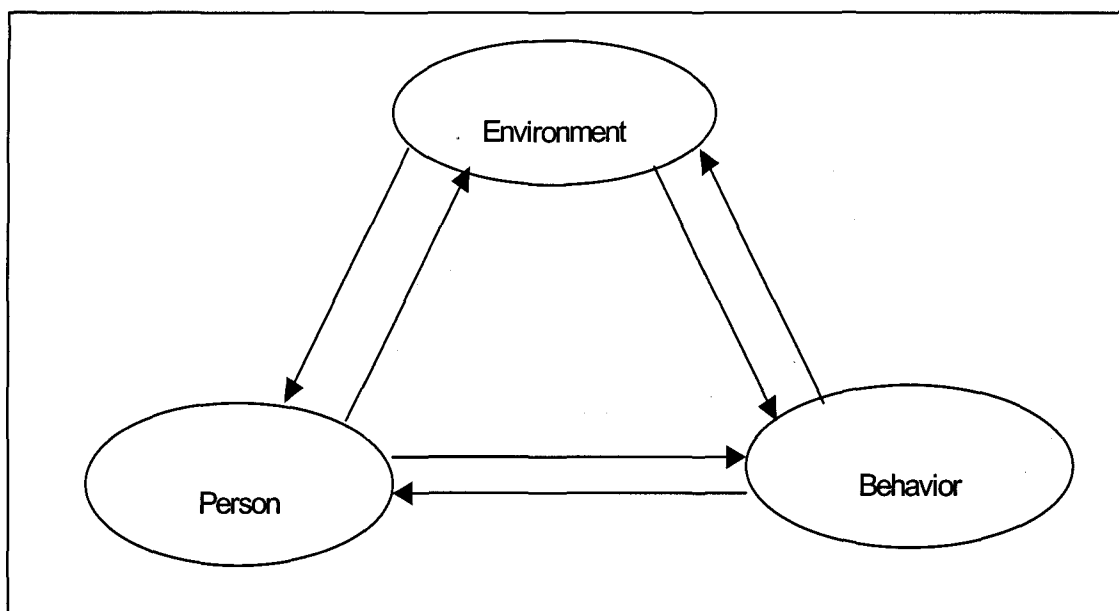
Figure 2.2:   Triadic reciprocity in Social Cognitive Theory

The triadic reciprocity and interaction among behavior, cognition, and

environment elements within the Social Cognitive Theory framework serves to provide

30

the self-regulatory function. This enables people to alter their environment, specifically controllable events, such as the selection of an undertaken task, rather than the physical milieu itself. This triadic reciprocity is the foundation of Bandura's view of human behavior as products and producers of their environments and social systems (Pajares, 1997).

Bandura (1986) posits that people with a strong sense of self-efficacy set high goals for themselves and remain strongly committed to them even in adversity. Social Cognitive Theory addresses the development of competencies and a self-system enabling an individual to exercise control over their actions. The regulations of action develop through knowledge structures; the sources for these structures are formed through acquired knowledge, observational learning, exploratory efforts, and emotional arousals, along with verbal instructions (see Section 2.3.1 on sources of self-efficacy). The rules and strategies for complex human behavior patterns define the appropriate skills that must be selected, sequenced, and integrated to fit a particular purpose. Skilled action and corrective adjustments provide the behavior execution to suit the dynamics requirements of each particular situation.

Through continued practice, proficient behavior become routine and no longer require extensive cognitive control. The execution of the task is relegated to a lower level sensory-motor system for recurrent tasks; people's perceived efficacy is not required for continued directed thought. As skills became routine, people's behavior is performed in accordance to what they believe without much further thought. However, whenever novel events or behavior patterns occur, the learning practice continues. Thus, Social Cognitive Theory encompasses a large variety of factors acting as regulators and motivators for an individual's cognitive, social, and behavioral skills (Bandura, 1986, 1997).

31

Subsequently, Bandura (1997) published *Self-efficacy: The exercise of control*, in which he further situates self-efficacy within a theory of personal and collective agency operating in concert with other sociocognitive factors that regulate human well-being and attainment. People always strive to control events that affect their lives (Bandura, 1997). Agency, as referred to by Bandura, is acts done intentionally. People contribute to what happened to them, rather than the event being the sole determiner.

2.3.1 Sources of Self-efficacy

Bandura (1986, 1997) delineates four sources of self-efficacy. First, the most influential source is an individual's mastery experience or enactive attainment. Enactive learning, or mastering the experience of learning by doing, depends heavily on the consequences of an individual's actions (Schunk, 2001). Individuals gauge the effects of their actions and the interpretations of these effects on their action, which help create their self-efficacy beliefs (Pajares, 1997). When an action is interpreted as successful, self-efficacy rise. When interpreted as a failure, self-efficacy is lowered especially early in the course of events; self-efficacy is not reflective of a lack of effort or due to an external event. However, after a strong sense of self-efficacy is developed through repeated successes, an occasional failure is unlikely to have a negative effect on an individual's capabilities (Bandura, 1986). Schunk and Zimmerman (1997) include self-efficacy as a self-influencing function in formulating their social cognitive model of the development of self-regulatory competence. The focus of this research study is on self-efficacy; the broader concept of self-regulation is not elaborated.

Capability is needed to produce the desired outcome; however, there are many other factors that affected the level of performance that have little to do with capability.

Empty encouragement without capability would not compensate for a lack of mastery experience. Factors influencing the assessment of an individual's self-efficacy include difficulty of task, amount of effort expended, amount of external aid received, circumstances of the performance act, and the temporal pattern of success or failure. Appraisal of an individual's self-efficacy contributes to performance success or failure (Bandura, 1997). Performance itself is not the primary criterion for high or low self-efficacy. Changes in perceived self-efficacy are the result of cognitive processing of diagnostic information that the performance conveys and not the performance itself (Bandura, 1997). Hence, the impact of performance on self-efficacy is not dependent on the results of the performances themselves.

A second source of self-efficacy information is vicarious experience, which is the appraising of an individual's capabilities in relation to the attainment of others. The effect produced by the action of others is weaker than the mastery experiences (Bandura, 1977a). However, when people are uncertain about their own abilities or have limited prior experience, they become more sensitive to the modeling effects, or seeing or hearing of others perform a similar activity or task (Pajares, 1997). Visualizing other similar people perform successfully raises their belief that they too possess the capability; this persuades them that if others could do it, they could be able to achieve at least some level of performance. Similarly, seeing failure of others perceived to be of similar competence, despite high effort, would lower the observer's judgment of his or her capabilities and undermine his or her efforts (Brown & Inouye, 1978). Vicarious experience and peer modeling are powerful influences on developing self-perceptions on competence (Schunk, 1983). If a model is taught a better way of doing things, then his or her self-efficacy would escalate.

Another source of self-efficacy is verbal persuasion. Although a weaker source of

efficacy than both mastery and vicarious experience, persuaders can be an important source in the development of his or her self-beliefs (Zeldin & Pajares, 2000). Effective persuasion is not to be confused with extreme and erroneous praise or empty inspiration, but if applied within realistic bounds, it could contribute to successful performance (Bandura, 1986). Just as positive verbal persuasion could work to encourage and empower, negative persuasion could work to weaken self-beliefs. Raising unrealistic beliefs of personal competence only invited failure, which undermine the persuaders' credibility while the recipient's self-efficacy is jeopardized. In fact, it is usually easier to weaken self-efficacy beliefs through negative appraisals than to strengthen such beliefs through positive encouragement (Bandura, 1986).

The fourth source of self-efficacy is physiological. Physiological states such as anxiety, stress, and mood provide information about self-efficacy beliefs. In stressful and taxing situations, emotional arousal conveys ominous signs of dysfunction. When fear provokes thoughts about ineptness, people elevate their levels of distress, which produce the very dysfunction(s) they want to avoid. Treatments to eliminate emotional arousal of subjective threats such as phobias provide improvements in performance and self-efficacy (Bandura, 1977a).

2.3.2 Computer Self-efficacy

Computer self-efficacy in spreadsheet development refers to personal perceptions of an individual's capabilities to use IT technologies to perform spreadsheet-related tasks within the Social Cognitive Theory framework. Compeaux and Higgins (1995, 1999) developed a domain specific and task specific computer self-efficacy construct consistent with Bandura's definition of self-efficacy. However, they found that

34

computer self-efficacy was not significant to outcome expectancy as the authors had

hypothesized. Figure 2.3 extracts a portion of their model positing that outcome

expectancy was an antecedent to performance.
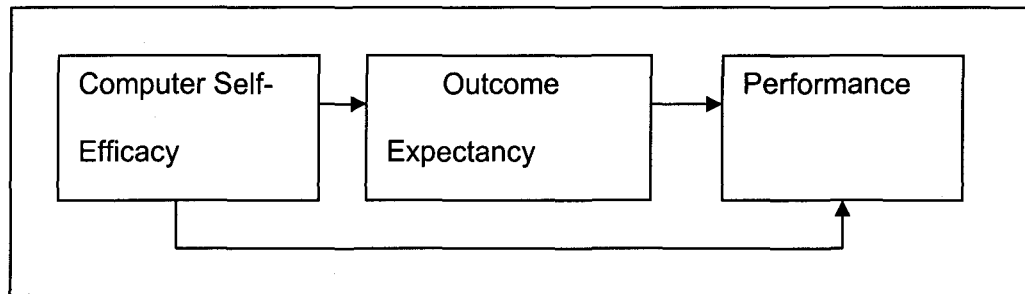
Figure 2.3:   Compeaux and Higgins' proposed performance-outcome
expectancy relationship

Subsequently, Maracas et al. (1998) and Johnson and Maracas (2000) refined

Compeaux and Higgins' (1995, 1999) model utilizing spreadsheets in their development

of computer self-efficacy. Agarwal et al. (2000) also tested Windows 95 self-efficacy

within the computer self-efficacy framework and suggested that initial general self-

efficacy beliefs will positively influence subsequent specific computer self-efficacy

beliefs. Recently, Looney et al. (2004) research showed general self-efficacy as an

antecedent to computer self-efficacy. Their analyses showed that computer self-efficacy

mediated general self-efficacy, while investigating online investment self-efficacy,

another domain's specific task.

In self-efficacy research, the dimensions of efficacy beliefs varied on three

dimensions: level of task demands, generality, and strength and perseverance of

efficacy beliefs. The level of task demands represented the various challenges or

impediments for successful performance, which reflected a person's perceived

capabilities. The efficacy belief was not context or situation independent; otherwise, its

predictability was weak.

As a second dimension, the generality of the efficacy belief varied based on similarity of task, situation, and characteristic of environment. According to Bandura (1997), "self-efficacy for a specified domain of functioning is usually assessed at the intermediate level of generality because the self-regulative demands in some particular settings may be unrepresentative of the demands under the conditions that exist when people usually perform the activity" (p. 49).

Finally, the strength of efficacy beliefs' dimension reflected the perseverance and likelihood that an activity be performed successfully. A weak efficacy belief resulted in failure based on negative experiences, whereas a tenacious efficacy belief in a person's capability persevered through numerous difficulties and obstacles (Bandura, 1997).

Johnson and Maracas' (2000) scale for computer self-efficacy measured the level, the generality, and the strength of the individual's computer self-efficacy belief. This research uses a modified instrument from the Johnson and Maracas (2000) study and is detailed in the Computer Self-efficacy section 3.2.1.

2.3.3 Self-efficacy Applicability

The self-efficacy component of Social Cognitive Theory is widely tested in various disciplines and settings in diverse fields. Self-efficacy beliefs is found to be positively related to clinical problems such as phobias (e.g., Bandura, 1977a), addiction (e.g., Hays & Ellickson, 1990), and eating disorders (e.g., Schneider, O'Leary, & Agras, 1987) as well as in education research in academic motivation and self-regulation (e.g., Pintrich & Schunk, 1996) and in information search behavior (e.g., Nahl, 1996; Nahl &

Harada, 1996). These studies show that high self-efficacy positively influences behavior, while low self-efficacy has a negative effect. In addition, health promoting behavior (e.g., DiClemente, Prochaska, Fairhurst, Velicer, Velasquez, & Rossi, 1991), athletics (e.g., Gould & Weiss, 1981) and organizational related issues (e.g., Onglatco, Yuen, Leong, & Lee, 1993) show self-efficacy applicability in a similar fashion as the previously referenced studies. Self-efficacy's broad application across various domains of behavior accounts for its popularity in motivation research as self-efficacy proves to be a much more consistent predictor of behavior and behavior change than past performance or any of the other closely related expectancy variables (Bandura, 1986; Graham & Weiner, 1996).

Bandura (1986) defines perceived self-efficacy as "people's judgments of their capabilities to organize and execute courses of action required to achieve designated types of performances. It is not concerned with the skills one has, but with judgments of what one can do with whatever skills one possesses" (p. 391). This is not meant that people could accomplish tasks beyond their capabilities simply by believing that they could, but rather that their self-perception of their capabilities has helped determine what the individual could do with their knowledge and skills they possess. Self-efficacy beliefs are critical determinants of how knowledge and skill are acquired both initially and continuously (Pajares, 1997).

Self-efficacy requires sub-skills to be organized into integrated courses of action to serve innumerable purposes. Self-efficacy entails more than simply knowing what to do in a fixed manner, but involves the act(s) to be performed, how it is to be varied, and all the dynamics that were entailed. After generating and testing alternative behaviors, the developed strategies are the behaviors successfully achieved. This generative capability differentiates between possessing a sub-skill and using that sub-skill in diverse

37

situations; this is often evidenced in athletic competition where one of superior ability often loses to another of lesser ability. Hence, perceived self-efficacy is actively concerned with this generative capability of human behavior, and not with the component act itself.

Strong self-efficacy shows predictive coping behavior in dynamic situations, while weakly held self-efficacy beliefs are highly vulnerable to changes. People may possess a skill needed to accomplish a task, but a strong self-efficacy is needed to successfully execute the activities. When action is interpreted as successful, self-efficacy rise; when interpreted as failures, self-efficacy is lowered, especially early in the course of events, and is not reflective of a lack of effort or due to an external event. However, after a strong sense of self-efficacy is developed through repeated successes, an occasional failure is unlikely to have a negative effect on an individual's capabilities (Bandura, 1986). To illustrate, Bandura (1986) provides an example in measuring driving self-efficacy when people are asked not to judge whether they could turn on the car's ignition key, accelerate or stop a car, or to change to a passing lane, but rather whether they are able to judge their strength of their self-efficacy skills in navigating through congested traffic, onrushing traffic, or maneuver twisting mountainous roads.

Incentives and resources are also needed to act on behalf of an individual's beliefs of self-efficacy. Consistent with Social Cognitive Theory, people with a strong sense of self-efficacy set high goals for themselves and remained strongly committed to them even in face of adversity and perform well. A person with a weak self-efficacy would not establish high goals and when encountering adversity, would not perform well.

Reasonably accurate appraisal of a person's capabilities is valuable to successful functioning. Decisions on what activities to engage in are partly determined by the individual's judgment of self-efficacy (Bandura, 1986). People tend to undertake

38

tasks and situations they believe themselves capable of handling, while avoiding those they believe exceed their capabilities. High self-efficacy people engage in activities that foster active engagement that expand and grow their competencies. In contrast, low self-efficacy lead people to avoid enriching activities and environments that retard development or potential capabilities. Efficacy judgments that are most beneficial are those that slightly exceed what an individual can do. These self-appraisals lead people to undertake realistic and challenging tasks out of motivation for self-development of their capabilities (Bandura, 1986). However, faulty self-efficacy assessment due to a misjudgment of personal capabilities, or ambiguity of task demands or requirements may contribute to an overconfidence judgment (Bandura, 1997). Capability is needed to produce the desired outcome; however, there are many other factors that affect the level of performance that has little to do with capability; for instance, a bad experience or performance can have a negative effect on subsequent self-efficacy and performance.

The vicarious experience of seeing failure of others of similar perceived competence and high effort lowers the observer's judgment of their capabilities and undermines their efforts (Brown & Inouye, 1978). Vicarious experience and peer modeling are powerful influences on developing self-perceptions on competence (Schunk, 1983). If models taught one a better way of doing things, then self-efficacy escalates.

Judgments of self-efficacy also determine how much effort is expended and how persistent people are when faced with obstacles (Bandura, 1986). The stronger the perceived self-efficacy, the more vigorous and persistent would be the effort. Strong perseverance often pays off in higher performance attainments. When difficulties are consistently encountered, self-doubt enters causing the subjects to lessen their effort or to cease altogether resulting in a lower self-efficacy.

39

The effects of self-efficacy are differentiated during the learning phase and during execution phase of establishing skills. In executing a learned task in applying skills already acquired, a strong belief in his or her self-efficacy intensifies and sustains the effort needed to achieve difficult and challenging performances. This would be difficult to achieve for an individual who has a low self-efficacy. During the learning phase, it is usually easier to weaken self-efficacy beliefs through negative appraisals than to strengthen such beliefs through positive encouragement (Bandura, 1986). For instance, raising unrealistic beliefs of personal competence only invites failure, which undermines the persuaders' credibility while the recipient's self-efficacy is jeopardized.

Cognitive processing of these sources of information, which are acquired enactively, vicariously, persuasively, or physiologically, is integrated into self-efficacy judgments. This cognitive processing of efficacy information either boosts or deflates self-efficacy. Boosting self-efficacy in a particular situation produces positive results; deflated self-efficacy produces negative results. This research focuses on the individual's perceived computer self-efficacy, as most spreadsheet development is done individually rather than collectively.

2.3.4 Relationship between Self-efficacy and Confidence?

Stajkovic and Luthans (1998) conducted a meta-analysis of self-efficacy studies. They found that self-efficacy directly affected work related performance, which was moderated both by task complexity and by the environment in which the experiment was conducted. The task complexity was measured on the number of dimensions and divided into high, medium, and low categories, while the environment was examined based on whether the experiment occurred in a simulated or classroom environment or a

naturalistic environment. The generative requirement of self-efficacy, where generating and testing alternative behaviors and strategies that successfully achieved the desired behavior, facilitates an individual's understanding of the construct.

Bandura (1997) argues overconfidence is a misjudgment of personal capabilities and task requirements and demands. "Insufficient allowance for likely impediments may yield overconfidence judgment" (pp. 64-65). The result is a process that has not been understood, has been ill defined, has no social consequences of task demands, has a lack of adequate resources, or has some other external constraints (Bandura, 1997, pp. 63-72). Bandura (1997) offers a case whereby a common finding in academics is an inflated expectation among disadvantaged students with deficient academic preparation and achievement. An inadequate level of knowledge for college entrance or success in college could stem from an overestimation of their personal capabilities. This form of judgment disparity can lead to an overconfident judgment.

One key feature of Reason's GEMS framework, as previously outlined, was the assertion that humans were biased in using a rule-based search for a previous solution before activating the knowledge-based performance level. One source of failure at the knowledge-based level is this pattern-matching characteristic which was a common human behavior. This approach of matching aspects of the local information sources and states to a situation suggested potential areas where errors occurred as human variability were also a base part of adaptation.

Self-efficacy judgments are task and situation specific to provide predictive power. Many motivational and self-regulatory influences also contribute to level of performance. However, Bandura (1986) distinguishes judgments of self-efficacy from outcome expectations as is illustrated in Figure 2.4.

```
+-------------------------------------------------------------------+
|                                                                   |
|   PERSON--------------->BEHAVIOR--------------->OUTCOME            |
|                    :                       :                      |
|            +--------------+          +--------------+             |
|            | Efficacy     |          | Outcome      |             |
|            |              |          |              |             |
|            | Beliefs      |          | Expectancies |             |
|            +--------------+          +--------------+             |
|                                                                   |
+-------------------------------------------------------------------+
```
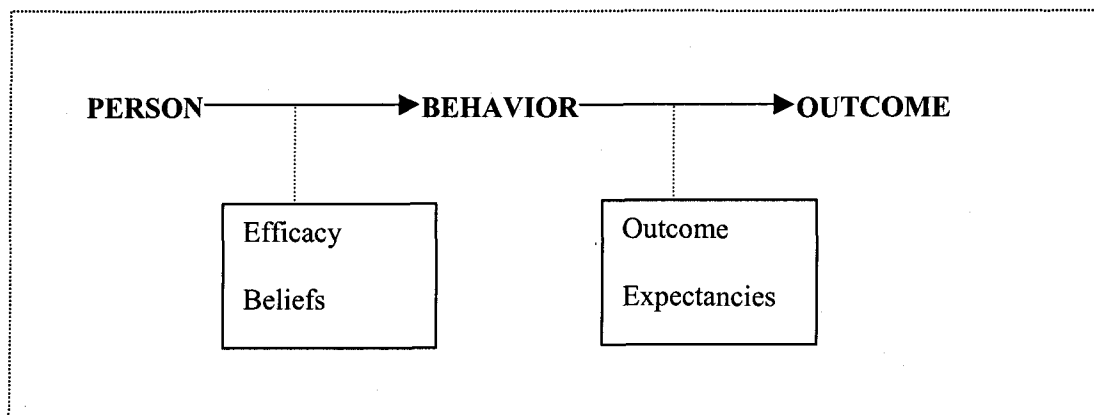
Figure 2.4:   Self-efficacy and outcome expectations

Perceived self-efficacy is a judgment of an individual's capability to accomplish a

certain level of performance and its influence on behavior; an outcome expectation is a

judgment of the likely consequence of such a behavior. An outcome is the consequence

of an act and not the act itself. Outcome become a measure of adequacy of

performance, while the regulation of behavior is governed by an individual's self-efficacy

on deciding which course of action to pursue. Bandura (1986) posits that outcome

expectations and self-efficacy could be separated provided no action is taken or action is

not linked to the level of performance. Yet, outcome expectation still determines the

motivation and the consequences expected to be received from the behavior

(Zimmerman, 2001).

Although Bandura (1986) distinguishes self-efficacy from outcome expectation,

he argues the existence of an overconfidence misjudgment. Bandura (1997) further

posits "the construct of self-efficacy differs from the colloquial term confidence, which is

widely used in sports psychology. Confidence is a nondescript term that refers to

strength of belief but does not necessarily specify what the certainty is about" (p. 382).

Vealey et al. (1998) posits an expanded theory in their sport-confidence model, which

conceptualizes a sport-confidence trait, a sport-confidence state, and a dispositional

(goal oriented) construct called competitive orientation. Their theory posits a dispositional construct or trait (sport-confidence trait) that an individual usually possess about his or her ability to be successful, and a state construct (sport-confidence state) that changes with the time reference at any particular moment (Vealey, 1986). These are coupled with the competitive orientation, which refers to an athlete's tendency to strive toward achieving a certain type of goal when competing in sport (Vealey et al., 1998).

This research further expands and distinguishes among the affective and cognitive constructs and incorporates self-efficacy, confidence, and overconfidence to our understanding of spreadsheet error performance.

Chapter 3. Research Hypotheses and Methodology

The primary objective of this research is to advance our understanding of human error performance in spreadsheet development. In this research, confidence in the accuracy of spreadsheet development refers to an individual's assessment that he/she has successfully performed a spreadsheet task error free. Overconfidence, further, indicates an individual's prediction of task accomplishments that exceed the actual results, i.e., more errors committed. Uncorrected error rates range from 0.5% in simple mechanical tasks, such as typing, to 5% in more complex logical activities, such as writing programs (Panko, 2005a).

High self-efficacy is viewed positively in self-efficacy research because it leads to higher performance. Confidence theory, however, views high confidence problematically, usually leading to overconfidence. If people are overconfident, they tend to do less checking of their work during spreadsheet development and this results in a higher error rate (Panko, 2005b). A research framework for this study is illustrated in Figure 3.1.

The cognitive nature of the task of spreadsheet development results in errors, which are influenced by confidence, overconfidence, and self-efficacy as affective variables. This research is an investigation into the relationship among these affective variables and its effects on spreadsheet errors. This chapter concludes with a discussion of the protocol analysis methodology that was used to investigate incidence of errors during spreadsheet development that are corrected prior to the final product delivery or left uncorrected in the final product delivery.
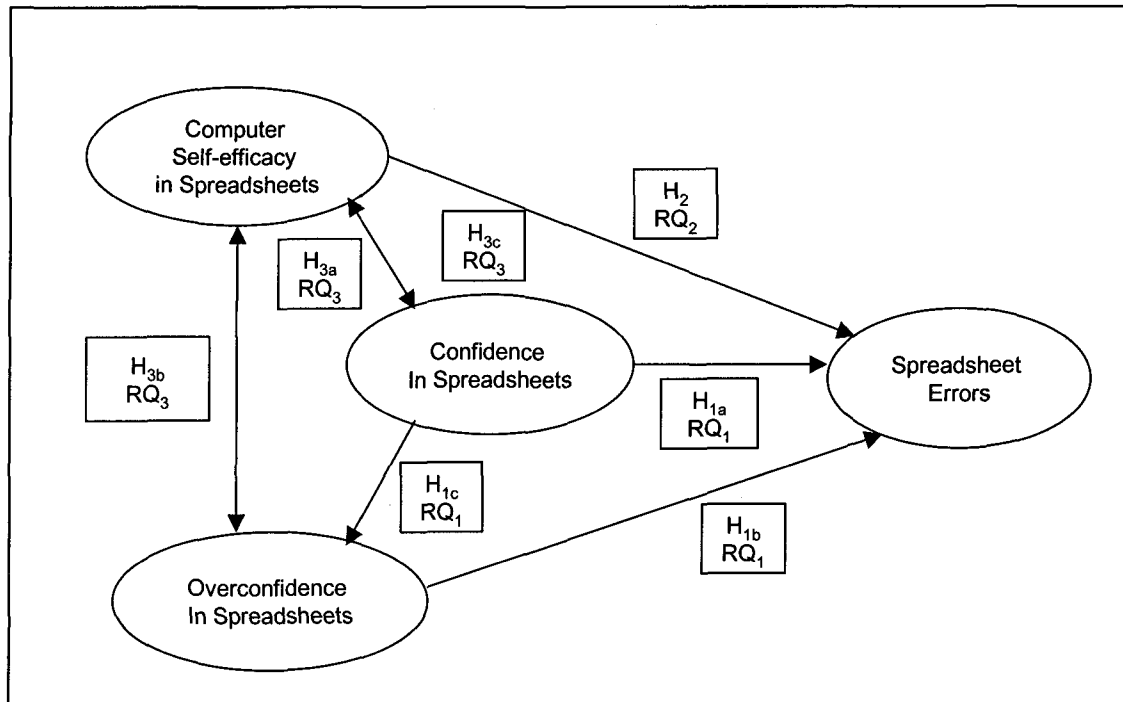
Figure 3.1: Research framework

Legend: RQ_1, RQ_2, and RQ_3 are the research questions.
H_{1a}, H_{1b}, H_{1c}, H_2, H_{3a}, H_{3b}, H_{3c} are the research hypotheses.

3.1 Research Issues

In software engineering, the potential for the existence of errors requires the programs to go through several phases of systematic testing during their software development life cycle (SDLC). As previously outlined in Chapter 2, most of the literature on software development refer to software engineering as a professional specialty; the spreadsheet development that this research focuses on involves the end-user, and is a close kin to the software development process. Yet, in the spreadsheet studies, a consistent lack of testing by spreadsheet developers is contrary to the modus operandi of professional software developers who extensively test their software. A result is errors in the delivered product.

45

3.1.1 Confidence and Overconfidence Constructs

Confidence in spreadsheet development is an assessment of performing a spreadsheet task accurately. Miscalibration in confidence, as referenced in the research literature, is measured by the difference between the estimated or predicted assessment and the actual performance of a task. When the predicted estimate exceeds the actual performance, an overconfidence calibration condition occurs; in this case, the subjective estimate differs from and exceeds the objective results. Similarly, if the actual performance exceeds the predicted estimate, then an underconfidence calibration condition occurs. Underconfidence may result in lost opportunities and the failure to pursue potential areas; this may be far less noticeable than the effects of optimistic ventures of overconfidence. Underconfidence will not be investigated in this research. The first research question (RQ_1) to be confirmed in this study is

RQ_1: Does overconfidence lead to more

errors in spreadsheet development?

As discussed in Chapter 2, the available literature this researcher examined establishes that a general human tendency is overconfidence. Overconfidence is widely prevalent in various domains, ranging from driving skills and card games to the judgment and human decision-making field and business planning. Overconfidence has also been seen in spreadsheet development as reviewed in Chapter 2.

To address the first research question (RQ_1: Does overconfidence lead to more errors in spreadsheet development?), confidence in spreadsheet development is measured by using a summed total of the four rating scale items concerned with spreadsheet development tasks (see 3.2 Scales). Overconfidence is measured as the

actual number of errors minus the estimated number of errors. The following hypotheses are investigated:

H_{1a}: Confidence in the accuracy of spreadsheets, as measured by the four rating scale items, will be positively correlated to the number of errors per spreadsheet.

H_{1b}: Overconfidence in the accuracy of spreadsheets, as measured by the difference between actual and estimated number of errors, will be positively correlated to the number of errors per spreadsheet.

H_{1c}: Confidence, as measured by the four rating scale items, and overconfidence, as measured by the difference between actual and estimated number of errors, in the accuracy of spreadsheets will be positively correlated.

3.1.2 Computer Self-efficacy Construct

As discussed in Chapter 2, over the past three decades, self-efficacy was found positively related to performance in many domains, including clinical and health-related

47

fields, education, athletics, information searching, and organizational-related fields. As also delineated in Chapter 2, this construct was shown to have applicability to the computer domain and IT arena. The second research question (RQ_2) to be investigated in this study is

> RQ_2: Will high computer self-efficacy be
> related to fewer errors in spreadsheet
> development?

Computer self-efficacy is very task and situation specific. The cognitive processing of information will either boost or deflate an individual's self-efficacy (Bandura, 1997). The second research question (RQ_2: Will high computer self-efficacy be related to fewer errors in spreadsheet development?) is designed to validate the existence and strength of computer self-efficacy in spreadsheet development tasks. The computer self-efficacy construct will be measured within the spreadsheet development environment by using a weighted summed total of the eight rating scale items concerned with the spreadsheet development (see section 3.2.1 Computer Self-efficacy Scale). The following hypothesis is investigated:

> H_2: Computer self-efficacy, as measured by
> the eight rating scale items, will be
> negatively correlated to the number of
> errors per spreadsheet.

3.1.3 Computer Self-efficacy and Confidence Constructs Distinction

Bandura (1997) argues that overconfidence is a misjudgment of personal capabilities and task requirements and demands. Possible consequences of these misjudgments are final outcomes are not understood and are ill defined, there are no social consequences of the task demand, there is a lack of adequate resources, or there are other external constraints. Bandura (1997) also argues "confidence is a nondescript term that refers to strength of belief but does not necessarily specify what the certainty is about" (p. 382). To the academician, self-efficacy, confidence, and overconfidence are terms with specific meaning in their respective research streams. Thus, this research intends to investigate and clarify the relationship among these constructs and addresses the following general research question (RQ_3) as it relates to spreadsheet errors:

> RQ_3: What is the relationship among self-efficacy, confidence, and overconfidence in human performance of spreadsheet development?

Because computer self-efficacy, confidence, and overconfidence may contain commonalities in each research stream, it is possible that they are positively correlated.

> H_{3a}: There will be a positive correlation between the computer self-efficacy, as measured by the eight rating scale items, and the confidence of subjects, as

49

measured by the four rating scale items.

H_{3b}: There will be a positive correlation between the computer self-efficacy, as measured by the eight rating scale items, and the overconfidence of subjects, as measured by the difference between actual and estimated number of errors.

H_{3c}: There will be an interaction between computer self-efficacy, as measured by the eight rating scale items, and confidence, as measured by the four rating scale items, on the total number of errors.

The constructs of computer self-efficacy, confidence, and overconfidence in error rates during spreadsheet development need to be further distinguished. Construct validity tests need to be performed. Convergent and discriminant validity, both subtypes of construct validity, would provide evidence for construct validity.

3.2 Scales

3.2.1 Computer Self-efficacy Scale

As discussed in Chapter 2, Johnson and Maracas (2000) developed a scale for computer self-efficacy specific to spreadsheet tasks, which has subsequently been used by Yi and Davis (2003), and Yi and Im (2004).

The computer self-efficacy scale included an item that the authors felt did not measure the construct as intended. This item was deemed to be measuring a motor-skill

level (Johnson & Maracas, 2000, p. 408). The computer self-efficacy scale, used in this research and administered in the experiment (see Chapter 4), excluded this motor-skill as suggested by Johnson and Maracas (2000). Hereafter, computer self-efficacy refers to this modification to the authors' computer self-efficacy instrument specific to the spreadsheet development domain.

This modified computer self-efficacy instrument contains eight items, which composes the computer self-efficacy scale being used. Each of the eight items range from 0 to 100 with the summed total of the eight items composing the composite computer self-efficacy measure (WtSE). The items, incorporated in Appendixes C, D, and E, are

1. I believe I have the ability to manipulate the way a number appears in a spreadsheet. (Labeled SE#1)

2. I believe I have the ability to use and understand the cell references in a spreadsheet. (Labeled SE#2)

3. I believe I have the ability to use a spreadsheet to communicate numeric information to others. (Labeled SE#3)

4. I believe I have the ability to write a simple formula in a spreadsheet to perform mathematical calculations. (Labeled SE#4)

5. I believe I have the ability to summarize numeric in a spreadsheet. (Labeled SE#5)

6. I believe I have the ability to use a spreadsheet to share numeric information with others. (Labeled SE#6)

7. I believe I have the ability to use a spreadsheet to display numbers as graphs. (Labeled SE#7)

8. I believe I have the ability to use a spreadsheet to assist me in making decisions. (Labeled SE#8)

In the above items, a product of each of the eight item's applicability (indicated by a Yes or No and weighted 1 or 0 respectively) and its strength measured in a range from 0 to 100 is summed to compose the WtSE (composite weighted computer self-efficacy).

51

The product range is 0 to 100. The strength asks how sure the subject feels about each item. A NO response indicates a 0 strength (starting the range at 0), while a YES response indicates the degree of strength ranging from 10 to 100. The WtSE range is 0 to 100 for each of the eight items, while the summed total composite WtSE ranges from 0 to 800.

Since computer self-efficacy is very dependent on domain, task, and situation, the computer self-efficacy instrument in this study was administered before each spreadsheet was developed. According to the literature, the relation between self-efficacy beliefs and action is more accurate when measurements are closer together; hence, any temporal effects are eliminated in this experiment by administering the computer self-efficacy instrument during each development of the spreadsheet tasks. The instructions ask the subjects not to refer back to prior sections of the instruments; this procedural instruction avoids copying of information but requires the subject to reevaluate his or her psychological state when responding to the computer self-efficacy instrument.

3.2.2 Confidence Scale

Most of the confidence measures used in this study were previously utilized in experiments as detailed in Chapter 2. The weighted confidence measure (Wt-Conf) and range is subsequently discussed. This research included the following composite questions, as referenced in Appendix E:

1. What is the probability that YOU made an error in this task? (Qu#1)

2. My confidence in the *accuracy* of my spreadsheet model is (Qu#7)

3. I think that my spreadsheet is error free. (Qu#8)

4. The number of errors I probably made for both spreadsheet tasks. (Qu#9)

The first three questions were asked both before and after each spreadsheet development task to obtain the subject's assessment of being accurate. The fourth item was asked only after both tasks were completed to obtain the subject's combined estimate of his or her number of errors. The fourth item was added to obtain a clearer measurement of the individual subject's overconfidence condition, as previous studies measured the individual's assessment compared with the group as a whole. While the wording used for items #1 through #3 before the spreadsheet task and after the spreadsheet task were very similar, they differed only in syntax to indicate the before and after temporal effects of doing the spreadsheet task.

The confidence instrument was administered a total of five times in the experiment. The first three items were administered once before and once after each of the two-spreadsheet tasks; all four items of the confidence instrument were again administered as a composite response to both spreadsheet tasks at the conclusion of both tasks (see Figures 3.2 and 3.3 for detailed sequencing).

The results from the composite four confidence questions at the conclusion of both spreadsheets were used in the analyses to obtain a composite assessment of the subject's confidence level. The confidence measurement reflecting the subject's confidence level in completing the development of the spreadsheet was calculated by summing the four items each with equal weights. From the list of the four items, Qu#1 asked each subject for his or her probability assessment of making an error in his or her spreadsheet development. Thus, 1 minus their probability assessment was used in the statistical measurement for the subject's confidence level of not making an error. Similarly, Qu#4 asked each subject for his or her assessment on the number of errors made in both spreadsheet tasks. Since no subject estimated more than 10 errors for both spreadsheet tasks, 10 minus their estimated number of errors was used in the

statistical measurement for the estimated confidence level of not making the estimated number of errors. The analyses was performed on the four-item confidence measure, which were summed and equally weighted; all items were adjusted to a 100–point scale to obtain the weighted confidence measure (Wt-Conf) for each subject. The weighted confidence measure (Wt-Conf) is the sum of these four items as discussed above. Based on the aforementioned adjustments, the range of the Wt-Conf measure is from 0 to 400.

The spreadsheets were also evaluated for the actual number of errors and were compared with the subject's estimated number of errors to determine whether an overconfidence condition exists. Since the experimenter knew the solution for the spreadsheet tasks in advance, each spreadsheet was corrected in order to obtain a known solution; this approach enabled the researcher to categorize the type of errors. If the hypothesis were valid, overconfidence would exist and have an impact on the number of errors for most of the subjects.

The instruments administered included many questions that will not be utilized in the final research results. The intent was to get the subject to freely respond to his or her cognitive thoughts, rather than be biased by the experimenter's directed questions or the exact nature of the experiment. Hence, only the confidence, overconfidence, and computer self-efficacy questions were used in the final analyses undertaken by this research. The other questions in the confidence instrument will be used for future research; these are Qu#2 through #4 of the instrument before each spreadsheet task, Qu#2 through #16 and Qu#19 through #30 of the instrument after each spreadsheet task, and Qu#2 through #6 and Qu#10 through #20 in the final composite assessment.

3.2.3 Instrument Administration

Figures 3.2 and 3.3 illustrate the experiment design indicating when each instrument and task was administered. Since the subjects were randomly assigned to a lab computer that was sequentially numbered, the even computer assignments are illustrated in Figure 3.2, while the odd computer assignments are illustrated in Figure 3.3.

Time Sequence	Instrument or Task Administration
T_0	Complete demographic information & Human Subjects Assessment form
T_1	Computer Self-efficacy instrument
T_2	Confidence instrument – 3-items (*before*)
T_3	Wall Task
T_4	Confidence instrument – 3 items (*after*)
T_5	Computer Self-efficacy instrument
T_6	Confidence instrument – 3 items (*before*)
T_7	MicroSlo Task
T_8	Confidence instrument – 3 items (*after*)
T_9	Computer Self-efficacy instrument (composite)
T_{10}	Confidence instrument – 4-items (composite)

Figure 3.2: Instrument administration for even numbered computers

Time Sequence	Instrument or Task Administration
T_0	Complete demographic information & Human Subjects Assessment form
T_1	Computer Self-efficacy instrument
T_2	Confidence instrument – 3-items (*before*)
T_3	MicroSlo Task
T_4	Confidence instrument – 3-items (*after*)
T_5	Computer Self-efficacy instrument
T_6	Confidence instrument – 3-items (*before*)
T_7	Wall Task
T_8	Confidence instrument – 3-items (*after*)
T_9	Computer Self-efficacy instrument (composite)
T_{10}	Confidence instrument – 4-items (composite)

Figure 3.3: Instrument administration for odd numbered computers

3.3 Research Experiment Design and Methodology

3.3.1 Tasks

The experiment was designed to give the subjects two spreadsheets to develop, the Wall Task and the MicroSlo Task. The Wall Task, a simple and domain independent task that does not require problem domain knowledge, is included in Appendix D and asked the subjects to submit project bids for building either a tile wall or a lava (volcanic material) wall. This spreadsheet was previously used in experiments conducted by Panko (2000). The following is the wording of the Wall Task:

> Your own a company that builds walls. Your spreadsheet should allow you to create a bid to build a wall for a customer. The customer will have to choose between brick or lava rock.
> Both walls will be built by crews of two. Crews will work three eight-hour days to build either type of wall. The wall will be 20 feet long, 6 feet tall, and 2 feet thick. Wages will be $10 per hour per person. You will have to add 20% to wages to cover fringe benefits. Lava rock will cost $3 per cubic foot. Brick will cost $2 per cubic foot. Your bid must add a profit margin of 30% to your expected cost.

The MicroSlo Task, the other assigned spreadsheet task, is included in Appendix C and asks the subjects to complete a *pro forma* income statement; this fictitious business is selling microwave slow cookers. This spreadsheet was previously used in experiments conducted by Goo (2002). The following is the wording of the MicroSlo Task:

> Your task is to build a two-year pro forma income statement for a company, MicroSlo.
> The company sells microwave slow cookers, for use in restaurants. The owner will draw a salary of $80,000 per year. There is also a manager of operations, who will draw a salary of $60,000 per year. The corporate income tax rate is expected to be 25% in each of the two years. Each MicroSlo cooker will require $40 in materials costs and $25 in labor costs in the first year. These numbers are expected to change to $35 and $29 in the second year. Unit sales price is expected to be $200 in the first year and to grow by 10% in the second year. There will be three sales people. Their salary is expected to average $30,000 per person in the first year and $31,000 in the second. Factory rent will be $3000 per month. The company expects to sell 3000 MicroSlo cookers in the first year. In the second, it expects to sell 3200.

56

3.3.2 Subjects and Selection Process

All subjects were undergraduate juniors or seniors in the College of Business Administration at the University of Hawai'i. At least two accounting courses were prerequisites for these majors as well as a computer science course ICS101 (Tools for the Information Age), or its equivalent, in which spreadsheet development was introduced. Hence, the subjects should be able to do these tasks. The subjects were given these experiment tasks in the middle of the BUS 311 (Information Systems for a Global Business Environment) curricula. The initial few weeks of the core BUS 311 course itself also required subject's to perform spreadsheet exercises. The subjects were selected only from the BUS 311 courses. In addition, BUS 310 (Statistical Analysis for Business Decisions), another required course for business students, used Excel spreadsheets as an integral part of the course. Many subjects took BUS 310 and BUS 311 concurrently.

Six sections of BUS 311 in the College of Business Administration at the University of Hawai'i were asked to participate as subjects in this experiment. Participating subjects received extra credit amounting to 2.5% of their letter grade. This extra credit was over and above the normal course credit. This approach would not penalize any of the potential subjects. Those subjects who elected not to participate and those subjects who opted to participate in advance but were unable to attend any session of the experiment due to any last minute personal reasons were offered an alternate project of equivalent time and effort so that extra credit was afforded them.

Subjects in each BUS 311 section were asked to sign up in advance for an experiment session that best met the subject's schedule. Each section was limited to five seating positions initially, as the lab had a maximum of 30 computers. Hence, each BUS 311 section had an equally opportunity to select a particular experiment session. Once

57

each BUS 311 section had the initial opportunity to elect their preferred experiment session, the signup was then open to all BUS 311 sections. This overall approach provided an equal and fair opportunity to participate in obtaining extra credit. This scheme also provided a good mixture of subjects from each section, and minimized the subjects' familiarity with each other.

The spreadsheets used in this experiment were previously used as discussed above. The sequence of the spreadsheet tasks alternated by subject and detailed below. Instructions provided to individual subjects were tailored to the sequence of the spreadsheets to be performed (Appendix A). A complete set of the four (4) sub-sections of the experiment (Appendixes B through E) was also provided to each subject.

The subject's demographic information, spreadsheet development experience, and native language were collected (Appendix B). The forms as required by the Committee on Human Studies in compliance with the Institutional Review Board for experiment participation was also included (Appendix B) with one copy to be returned to the experimenter, while the other copy to be retained by the subject.

The sequence of spreadsheets alternated the MicroSlo Task (Appendix C) and the Wall Task (Appendix D) between adjacent subjects in the computer lab; this controlled for order effects and minimized cheating by the subjects during the development of the spreadsheet task. The final sub-section (Appendix E) contained the composite assessment of confidence and computer self-efficacy questions and asked for any feedback on the experiment that may assist the researcher in problems encountered that may have affected the experiment and for future experiments.

The measurements of the confidence and computer self-efficacy constructs are detailed in Section 3.2, Scales. The instruments were administered as a complementary part of the spreadsheet tasks (Appendixes C through E) with the sequence of

administration as illustrated in Figures 3.2 and 3.3. The computer self-efficacy instrument was administered three times, once before to the start of each spreadsheet exercise and again at the conclusion of the spreadsheet exercises. The computer self-efficacy instrument attempted to measure the subject's perceived ability to successfully develop a spreadsheet.

The confidence items were asked on five different occasions, once before each of the two spreadsheet exercises and again after the completion of each spreadsheet exercise, and finally as a composite response to both spreadsheet tasks at the conclusion of the experiment. In addition, the final composite responses asked the subject to estimate the number of errors on both spreadsheet tasks. To address the confidence issue in this research, the final composite response was used for the analyses. The confidence instrument attempted to measure the subject's prediction to accurately perform the given task error free.

3.3.3 Conducting the Experiment

During the Fall 2004 semester, 103 participants, or approximately 51% of the subjects in the six BUS 311 classes volunteered for this experiment. The computer lab at the College of Business Administration (CBA) at the University of Hawai'i was utilized for this experiment. Only one computer lab of the several computer labs at the CBA was used at all times for proper experimental control. Only subjects from the CBA were selected for this experiment.

The testing was held over a seven-week period with ten group sessions and thirteen individual sessions offered to provide the greatest opportunity for subject participation. The individual sessions were targeted for the data gathering used for the

protocol analysis portion of this experiment (see Section 3.4). The computer lab used Microsoft's Windows 2000 as the operating system along with Microsoft's Excel as the spreadsheet tool that the subjects used to develop the experiment spreadsheets. Since all subjects regularly used the CBA computer lab during the normal course of their curricula, each had his or her logon identification and was familiar with the computer environment. Overall, no operational problems were encountered. A lab assistant was available during all experiment sessions to assist in resolving any computer issues in its general operation, but not questions pertaining to spreadsheet usage.

At each experiment session, each subject was asked to report to the sign in desk to verify their participation and to enable proper credit for their participation. Preference was given to those subjects who had signed up from their respective class sections; others who were not on the signup list were given an opportunity to participate on a first come space available basis. By a random drawing from a package containing labels for all available pre-numbered computers in the lab, each subject was assigned to a specific computer in the lab.

When the experiment started, the experimenter explained the nature of the experiment and provided procedural and administrative instructions via a verbally scripted narrative (Appendix F). After the verbal instructions were issued, each subject received an envelope packet containing the two-spreadsheet tasks, a diskette, the instructions, and the survey instruments (Appendixes A through E). Each subject was requested to mark the provided envelope with the assigned computer number for any follow up questions or issues, if it was necessary. One handout (Appendix B) included forms as required by the Committee on Human Studies in compliance with the Institutional Review Board. Also, demographic information was requested (Appendix B). After receiving the verbal instructions (Appendix F), each subject read the instructions

and was allowed to begin work. A comment section on the survey instrument (Appendix E) was requested from each subject on all aspects of the experiment; this information was used as feedback if subjects encountered any difficulties. The comments did not reflect any problems.

The spreadsheet tasks were distributed alternatively to adjacent subjects in the computer lab to control for order effects; this procedure also limited the opportunity for cheating. Further, the experiment administrator monitored the subjects during the experiment to preclude cheating.

The experiment administrator verbally suggested that the subjects either take a break between each spreadsheet development, or take a break at their convenience. Written suggestion on taking a break after each spreadsheet exercise was also provided (see Appendixes C and D). Candy was provided for the subject's consumption; previous research has shown that candy created a more favorable mood environment (Jones & Rogers, 2003). The subjects were requested to record the start and end times for each spreadsheet development task. The allocated 3 hours for the experiment period proved to be more than adequate, as the majority of the subjects completed both spreadsheet tasks within a 2-hour timeframe. Each subject completed the computer self-efficacy instrument once and the confidence instrument twice (before and after each task) for each spreadsheet task as outlined in Figures 3.2 and 3.3; the subjects were asked to save each of their spreadsheet models on the diskette provided and to print a hardcopy version using the computer lab's printer facilities.

At the conclusion of both spreadsheet tasks, each subject was asked to complete the composite computer self-efficacy and confidence instruments (Appendix E), to provide his or her overall assessment of the experiment, and to solicit comments on any difficulties encountered in the experiment. When each subject completed his or her

61

assigned tasks, the experimenter confirmed that a printed copy of their spreadsheet results was submitted along with a diskette containing the saved spreadsheet models. All material was returned in the envelope provided and the subjects were dismissed. Again candy was offered as they departed.

In tabulating the results of the experiment, 10 out of the 103 subjects were excluded from the statistical analyses due to either missing or incomplete information in their responses. Of the 10 excluded subjects, five subjects performed the spreadsheet calculations either manually or external to the spreadsheet results; consequently, no formulae or computations could be ascertained to determine the source of their errors. Another five subjects provided incomplete solutions in their responses to either one or both spreadsheet tasks; either problem statement data or partial calculations were submitted. These 10 subjects were eliminated from the final analyses. The final analyses used these 93 valid subjects.

3.4 Protocol Analysis Investigation

3.4.1 Protocol Analysis Design

In previous spreadsheet experiments conducted at the University of Hawai'i, the errors analyzed were at the conclusion of the experiment. The research literature suggests that errors occur during the task involvement and are subsequently corrected before the final solution. As an example in another domain, driving errors occur but corrections are made continuously; otherwise, accidents would be more prevalent.

Cognitive load theory (Sweller & Chandler, 1994), posits that an individual's limited processing capacity become a potential barrier to effectively assimilating a problem or task. Spreadsheet development, a high demand cognitive task, exhibits

learning difficulties and is limited by the number of elements to be handled individually and simultaneously.

Using the taxonomy outlined by Panko and Halverson (1996, 1997), the "think-aloud" protocol analysis technique, as delineated by Ericsson and Simon (1993), was used to better understand errors encountered and corrected in the problem-solving period as well as those left unresolved. Utilizing the "think-aloud" protocol analysis, a better insight and understanding of the types of errors — mechanical, logical, and omission — can be gleaned; specifically, how a person develops spreadsheets and when errors are examined.

3.4.2 Conducting the Protocol Analysis

In conducting the protocol analysis using the "think-aloud" methodology, individual sessions were used for the subjects to verbalize aloud their thoughts. The articulation of the subject's problem-solving approach required working with one individual per session. A pilot experiment for the "think-aloud" protocol with three subjects was undertaken in the Fall 2003 to understand the methodology and procedures. Equipped with the information from the pilot experiment, this experiment involved a pool of 13 subjects. Due to the limited sample size, formal hypotheses are not submitted for these data; observations are submitted.

Ericsson and Simon (1993) suggested practice exercises to familiarize each subject with the process. The experimenter reviewed each subjects' expectations in advance of performing the same spreadsheet development tasks as in the group sessions. Appendix G outlines the practice session employed until each subject became

comfortable articulating his or her thoughts. While performing the practice session, the experimenter encouraged the subject to verbalize as much of their surroundings entailed in their thought patterns, as the experimenter would not have the benefit of knowing about the subject's background nor experiences of his or her problem-solving approaches. At the conclusion of the practice session, the subject was again asked whether he or she felt uncomfortable with this procedure of "think-aloud" methodology and the recording of the session, as these were crucial components in this portion of the experiment.

A recorder was used to capture the problem-solving thoughts of each subject. Each subject performed the same spreadsheet tasks and instruments as completed by other subjects in their group sessions (Appendix A through E). A microphone was affixed to the subject to obtain the best vocalization of his or her problem-solving thoughts. The experimenter also listened to the recording through a separate earpiece to verify the proper volume and capture of the recording. Each subject then continued with the spreadsheet experiment in the same fashion, as had the other group session subjects.

Other than the individual session format and familiarization with the practice verbalization exercises, the other aspects of the experiment required of each subject were the same as the group session subjects.

Of the 13 subjects participating in the individual sessions, three were excluded. One subject was not comfortable verbalizing his thoughts and was excused; however, this subject continued with the remaining portions of the experiment as though he was participating in a group session, the only difference being the subject was alone in the computer lab while completing the spreadsheet tasks. Another subject was excluded as his spreadsheets contained only a restatement of the problem data; a minimum of effort was observed while solving both spreadsheet tasks. The recording of a third subject was

64

lost due to a mechanical error. The results of the remaining 10 subjects were included as

part of the 93 valid subjects used in the data analyses for this research.

Chapter 4. Results

The experiment was conducted in the Fall 2004 semester to investigate the hypotheses outlined in Chapter 3 for this research. The usability of the instruments with the two-spreadsheet tasks, and the procedural and administrative aspects for the experiment were tested previously in a pilot test conducted in the Fall 2003. This chapter presents the findings of the experiment.

4.1 Descriptive Statistics

Table 4.1 provides the background information of the valid subjects (n = 93) participating in this experiment. Of the subjects having previous experience, the mean experience entering data was 1.6 years based on 33 subjects. Of these 33 subjects, 17 subjects had a mean of 1.8 years of experience entering, reviewing, and training others in spreadsheet tasks. Over 80% of the subjects did not have any working experience using spreadsheets and only academic classroom exercises.

Table 4.2 contains the means, medians, and standard deviations of these subjects on several salient measures. The mean and median difference in the total number of errors and the estimated number of errors is large. This suggests the existence of an over-claiming condition. Similarly, a large variance is seen in the total number of errors, estimated number of errors, and actual less estimated number of errors.

Table 4.1: Background information

	Male	Female	Combined
Percent participation	40.9%	59.1%	100%
English as primary language	84.2%	69.1%	70.9%
Grade point (average)	3.19	3.34	3.21
Working experience using spreadsheets (outside of classes attended)	31.6%	38.2%	35.5%
Previous working experience in years (Entering data) [a]	1.4 yrs/subj	1.7 yrs/subj	1.6 yrs/subj
Previous working experience in hrs worked per week (entering data) [a]	11.5 hrs/wk/ subj	9.8 hrs/wk/ subj	9.1 hrs/wk/ subj
Previous working experience in years (creating, checking, or training others) [b]	1.8 yrs/subj	1.8 yrs/subj	1.8 yrs/subj
Previous working experience in hrs worked per week (creating, checking, or training others) [b]	10.2 hrs/wk/ subj	8.0 hrs/wk/ subj	8.9 hrs/wk/ subj

[a] Based on 33 subjects having previous experience entering data into spreadsheets.

[b] Based on 17 subjects having previous experience creating, checking models of others, or training others in spreadsheet usage.

Table 4.2: Means, medians, and standard deviations of subjects for key construct measurements

	Mean	Median	Standard Deviation
Wt-Conf	202.69	194.29	102.22
WtSE	532.04	590.00	193.76
Total # of Errors	3.08	2.00	2.28
Estimated # Errors	5.53	5.00	2.93
Act – Est # Errors	-2.45	-2.00	3.29

Figures 4.1 and 4.2 compare the total actual number of errors and the estimated number of errors, respectively, during the development of both spreadsheet tasks for each subject. Most (43%) of the subjects made 1 or 2 errors, while 13% estimated making 1 or 2 errors, but 22% estimated making 3 or 4 errors.
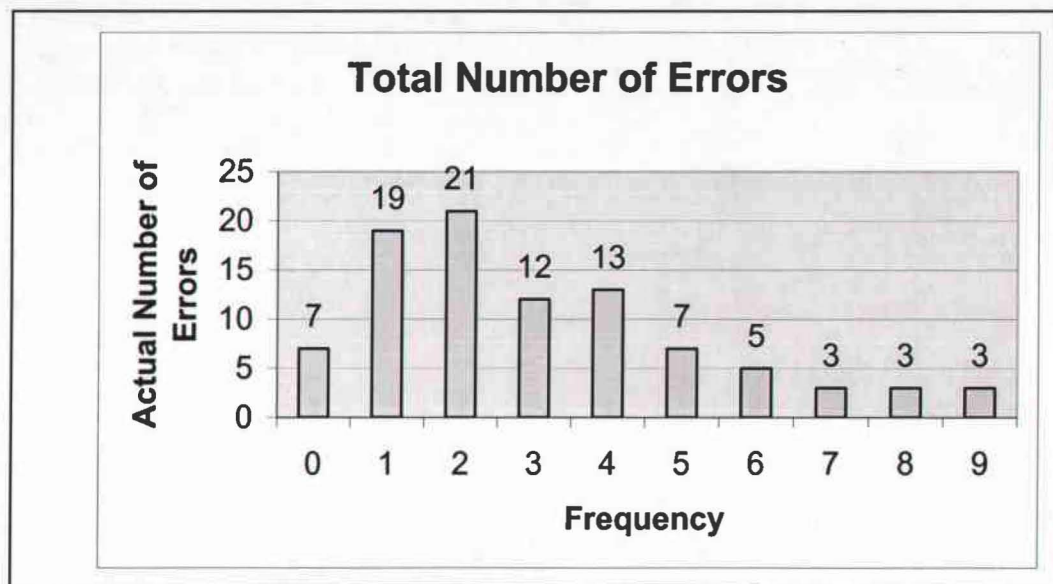
Figure 4.1:   Total actual number of errors
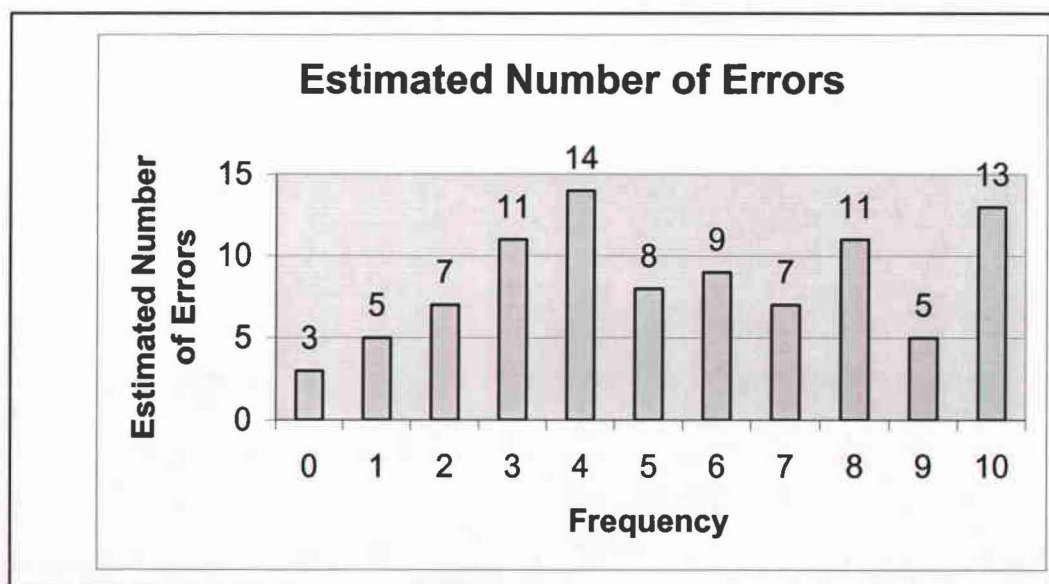
Figure 4.2:   Estimated number of errors

With regards to the scales, Table 4.3 and Table 4.4 provide the means, medians, and standard deviations for the computer self-efficacy items and the confidence items respectively. The medians in Table 4.3 for each of the computer self-efficacy items are consistent as well as the variance. On the other hand, the variance for item #1 (Qu#1) in Table 4.4 for the confidence measure shows a greater variance.

Table 4.3: Means, medians, and standard deviations for computer self-efficacy items

	Mean	Median	Standard Deviation
SE#1 -- I believe I have the ability to manipulate the way a number appears in a spreadsheet.	65.05	70.00	27.88
SE#2 -- I believe I have the ability to use and understand the cell references in a spreadsheet.	67.63	80.00	26.39
SE#3 -- I believe I have the ability to use a spreadsheet to communicate numeric information to others.	64.73	70.00	25.86
SE#4 -- I believe I have the ability to write a simple formula in a spreadsheet to perform mathematical calculations.	73.01	80.00	26.69
SE#5 -- I believe I have the ability to summarize numeric in a spreadsheet.	61.18	70.00	29.30
SE#6 -- I believe I have the ability to use a spreadsheet to share numeric information with others.	66.24	70.00	25.96
SE#7 -- I believe I have the ability to use a spreadsheet to display numbers as graphs.	89.60	80.00	26.69
SE#8 -- I believe I have the ability to use a spreadsheet to assist me in making decisions.	65.59	70.00	26.60

Table 4.4: Means, medians, and standard deviations for confidence items

	Mean	Median	Standard Deviation
Qu#1 -- What is the probability that YOU made an error in this task?	52.59	50.00	35.86
Qu#2 -- My confidence in the *accuracy* of my spreadsheet model is	60.98	57.14	24.21
Qu#3 -- I believe that the spreadsheets were error free.	44.09	42.86	27.92
Qu#4 -- The number of errors I probably made for both spreadsheet tasks.	44.73	50.00	29.33

4.2 Are Confidence and Self-efficacy Different?

4.2.1 Principal Component Analysis Investigation

A factor analysis was conducted to examine the two affective constructs, weighted computer self-efficacy (WtSE) and weighted confidence (Wt-Conf), and to determine if they are separate factors. A principal component analysis is intended to show whether psychometrically computer self-efficacy and confidence are the same variable. Factor analysis is used when researchers believe that certain latent factors exist that has causal influences on the observed variable being investigated (Hatcher, 1994, p. 10).

An initial principal component analysis was conducted to determine if the data resulted in separate factors. Based upon the data, two factors emerged from the principal component analysis using the "eigenvalue equal to one or larger" criterion. According to Hatcher (1994), the eigenvalue-one criterion, known as the Kaiser criterion, is one of the most commonly used criteria in extracting the number of components or factors from the data. Table 4.5 lists the communalities of the items examined. The communalities explain the percent of variance in the observed variables in the components or factors. These two components account for 80.3% of the variance in the data.

Table 4.5: Communalities

	Initial	Extraction
SE#1 -- I believe I have the ability to manipulate the way a number appears in a spreadsheet.	1.000	.786
SE#2 -- I believe I have the ability to use and understand the cell references in a spreadsheet.	1.000	.840
SE#3 -- I believe I have the ability to use a spreadsheet to communicate numeric information to others.	1.000	.898
SE#4 -- I believe I have the ability to write a simple	1.000	.749

70

formula in a spreadsheet to perform mathematical calculations.		
SE#5 -- I believe I have the ability to summarize numeric in a spreadsheet.	1.000	.738
SE#6 -- I believe I have the ability to use a spreadsheet to share numeric information with others.	1.000	.903
SE#7 -- I believe I have the ability to use a spreadsheet to display numbers as graphs.	1.000	.787
SE#8 -- I believe I have the ability to use a spreadsheet to assist me in making decisions.	1.000	.822
Qu#1 -- What is the probability that YOU made an error in this task?	1.000	.745
Qu#2 -- My confidence in the *accuracy* of my spreadsheet model is	1.000	.758
Qu#3 -- I believe that the spreadsheets were error free.	1.000	.845
Qu#4 -- The number of errors I probably made for both spreadsheet tasks.	1.000	.765

Note: SE#n indicates the computer self-efficacy items, while Qu#n indicates the confidence items, where n references the item number within each scale.

An oblique (correlated) run was performed. The obliquely rotated component matrix in Table 4.6 reflects the weighting each item contributed to the extracted components, which reflects the two extracted components. This suggests that the weighted computer self-efficacy (WtSE) and the weighted confidence (Wt-Conf) are two separate factors.

Table 4.6: Obliquely rotated component matrix

	Component #1	Component #2
SE#1 -- I believe I have the ability to manipulate the way a number appears in a spreadsheet.	.886	.415
SE#2 -- I believe I have the ability to use and understand the cell references in a spreadsheet.	.917	.420
SE#3 -- I believe I have the ability to use a spreadsheet to communicate numeric information to others.	.948	.464
SE#4 -- I believe I have the ability to write a simple formula in a spreadsheet to perform mathematical calculations.	.865	.444
SE#5 -- I believe I have the ability to summarize numeric in a spreadsheet.	.858	.443
SE#6 -- I believe I have the ability to use a spreadsheet to share numeric information with others.	.950	.478

71

SE#7 -- I believe I have the ability to use a spreadsheet to display numbers as graphs.	.877	.290
SE#8 -- I believe I have the ability to use a spreadsheet to assist me in making decisions.	.900	.517
Qu#1 -- What is the probability that YOU made an error in this task?	.322	.858
Qu#2 -- My confidence in the *accuracy* of my spreadsheet model is	.534	.858
Qu#3 -- I believe that the spreadsheets were error free.	.376	.917
Qu#4 -- The number of errors I probably made for both spreadsheet tasks.	.596	.845

The oblique rotation resulted in the component correlation matrix in Table 4.7 reflecting a correlation of 0.469. This latter result suggests that by performing an oblique (correlated) rotation, an interaction of these separate factors exists and accounts for 21.9% (0.469^2) of the variance. The oblique rotation shows that computer self-efficacy and confidence are psychometrically distinguishable variables, although correlated.

Table 4.7: Component correlation matrix

	Computer Self-efficacy	Confidence
Computer Self-efficacy	1.000	.469
Confidence	.469	1.000

4.3 Reliability of Scales

The Cronbach's coefficient alpha for the weighted computer self-efficacy instrument is 0.966 indicating a very good reliability. Nunnally (1978) considers a reliability of 0.70 or higher as sufficient (p. 245). Similarly, the Cronbach's coefficient alpha for the confidence instrument is 0.885 indicating good reliability.

4.4 Confidence and Computer Self-efficacy Findings

Table 4.8 presents the correlation results from the composite responses after the completion of both spreadsheet tasks in this experiment.

Table 4.8: Correlation statistics

	Wt-Conf	WtSE	Total # of Errors	Est # of Errors	Act - Est # Errors
Wt-Conf	1.0				
WtSE	.53**	1.0			
Total # Errors	-.19	-.34**	1.0		
Est # Errors	-.87**	-.59**	.22*	1.0	
Act - Est # Errors	.64**	.29**	.50**	-.74**	1.0

*   p<.05   (1-tail)
** p<.01   (1-tail)

4.4.1 Hypotheses 1a, 1b, and 1c

H_{1a} states "*confidence in the accuracy of spreadsheets, as measured by the four rating scale items, will be positively correlated to the number of errors per spreadsheet*" and is rejected. Confidence, used in this spreadsheet development research, refers to an assessment of being accurate and error free in performing a spreadsheet task and is measured as the weighted confidence of the four items (Section 3.2.2). The correlation between weighted confidence and the 'Total # of Errors' is not significant and slightly negative (-.19).

H_{1b} states "*overconfidence in the accuracy of spreadsheets, as measured by the difference between actual and estimated number of errors, will be positively correlated to the number of errors per spreadsheet.*" Based on this four-item confidence measure, overconfidence is calculated as the difference between the subject's actual number of errors and the subject's estimate of the number of errors made in the spreadsheet tasks.

An overconfidence condition exits when the actual number of errors exceeded the estimated number of errors. The results of the analyses shows that the correlation of the overconfidence condition ('Act – Est # Errors') is significantly (p < .01) and positively correlated with actual total number of errors ('Total # of Errors') (.50). The effect size of the correlation is large and in the expected direction to support the investigated H_{1b}. Cohen (1988) refers to the effect sizes as small, medium, and large for correlation coefficients of 0.1, 0.3, and 0.5 respectively.

H_{1c} predicts *"confidence, as measured by the four rating scale items, and overconfidence, as measured by the difference between actual and estimated number of errors, in the accuracy of spreadsheets will be positively correlated."* As expected, the correlation of the weighted confidence is significantly (p < .01) and positively correlated to overconfidence (.64). The effect size of the correlation results is large and in the expected direction to support the investigated H_{1c}.

To understand the mixed results above, the estimated number of errors exceeded the actual number of errors, resulting in an underconfidence condition as indicated in Table 4.2; this occurred in about 75% of the cases. To confirm whether the underconfidence condition was significant, a t-test was conducted (refer to Table 4.2):

$$t = -2.45 / (3.292 / 93^{1/2})$$

$$t = -7.18^{**} \quad (^{**}p < .01)$$

This indicates the sample has a tendency to be underconfident, or at least cautious or uncertain about the individual's abilities. Further elaboration of this condition is subsequently discussed.

4.4.2 Hypothesis 2

H_2 predicts *"computer self-efficacy, as measured by the eight rating scale items, will be negatively correlated to the number of errors per spreadsheet."* The correlation of the weighted computer self-efficacy measure is significantly ($p < .01$) and negatively correlated with the total number of errors (-.34). The effect size of the correlation results is medium and in the expected direction to support the investigated H_2.

4.4.3 Hypotheses 3a and 3b

H_{3a} states *"there will be a positive correlation between the computer self-efficacy, as measured by the eight rating scale items, and the confidence of subjects, as measured by the four rating scale items."* The correlation between the weighted self-efficacy and the weighted confidence (.53) is significant ($p < .01$). The effect size of the correlation results is large and in the expected direction to support the investigated H_{3a}.

H_{3b} states *"there will be a positive correlation between the computer self-efficacy, as measured by the eight rating scale items, and the overconfidence of subjects, as measured by the difference between actual and estimated number of errors."* The correlation between the weighted self-efficacy and overconfidence (.29) is significant ($p < .01$). The effect size of the correlation results is medium and in the expected direction to support the investigated H_{3b}.

4.4.4 Hypothesis 3c

H_{3c} states *"there will be an interaction between computer self-efficacy, as measured by the eight rating scale items, and confidence, as measured by the four*

rating scale items, on the total number of errors." To initially investigate this hypothesis, a correlation was performed on confidence and total number of errors as seen in Table 4.9. The correlation of the low computer self-efficacy subjects show significance in both the weighted confidence measure (.49) (p < .01) and the 'Total # of Errors' (-.29) (p < .05), but the correlation for the high self-efficacy subject result is significant for only the weighted confidence measure (.28) (p < .05). The effect size of the correlation results is medium to small in support of the investigated H_{3c}. The implication of this result is that there is a significant relationship for low computer self-efficacy subjects in their confidence and total number of errors. This is not entirely true for the high computer self-efficacy subjects. Further analysis is needed.

Table 4.9: Correlation of confidence and total number of errors by high computer self-efficacy and low computer self-efficacy subjects

	Wt-Conf	Total # of Errors
High Computer Self-efficacy	.278*	-.091
Low Computer Self-efficacy	.490**	-.288*

*   p<.05   (1-tail)
** p<.01   (1-tail)

In further investigation for a distinction between the high and low self-efficacy subjects, a chi-squared test was conducted to categorically compare high and low computer self-efficacy with overconfidence and no overconfidence. The sample was equally segregated into high self-efficacy and low self-efficacy categories with 46 subjects in the former category and 48 subjects in the latter category. The difference of subjects in each category was due to the separation of equal weighted computer self-efficacy subjects included in the same category. The other variable, overconfidence, was segregated into whether the actual number of errors exceeded the estimated number of

errors of the subjects, or whether the estimated number of errors exceeded or equaled the actual number of errors. Table 4.10 reflects the categorization for the chi-squared test. The results (χ^2 = 14.33, df = 1, p < .01) indicate that we reject the null hypothesis that there is a no difference in overconfidence between the high and low self-efficacy groups. So, there is a difference between the high and low self-efficacy groups; 63% of the low self-efficacy subjects estimated their actual number of errors exceeded their estimated number of errors compared with 27% of the high self-efficacy subjects.

Table 4.10: χ^2 results of number of high and low computer self-efficacy subjects by the differences in actual and estimated number of errors

	Actual – Estimated # of Errors (Underconfident)	Actual – Estimated # of Errors (Not underconfident)	Total
High Self-efficacy	26	19	45
Low Self-efficacy	44	4	48
Total	70	23	93

To further investigate this relationship, Figure 4.3 is the framework used in performing a multiple regression analysis to determine the relationship between the two affective constructs of computer self-efficacy and confidence and their interaction with the number of spreadsheet errors.
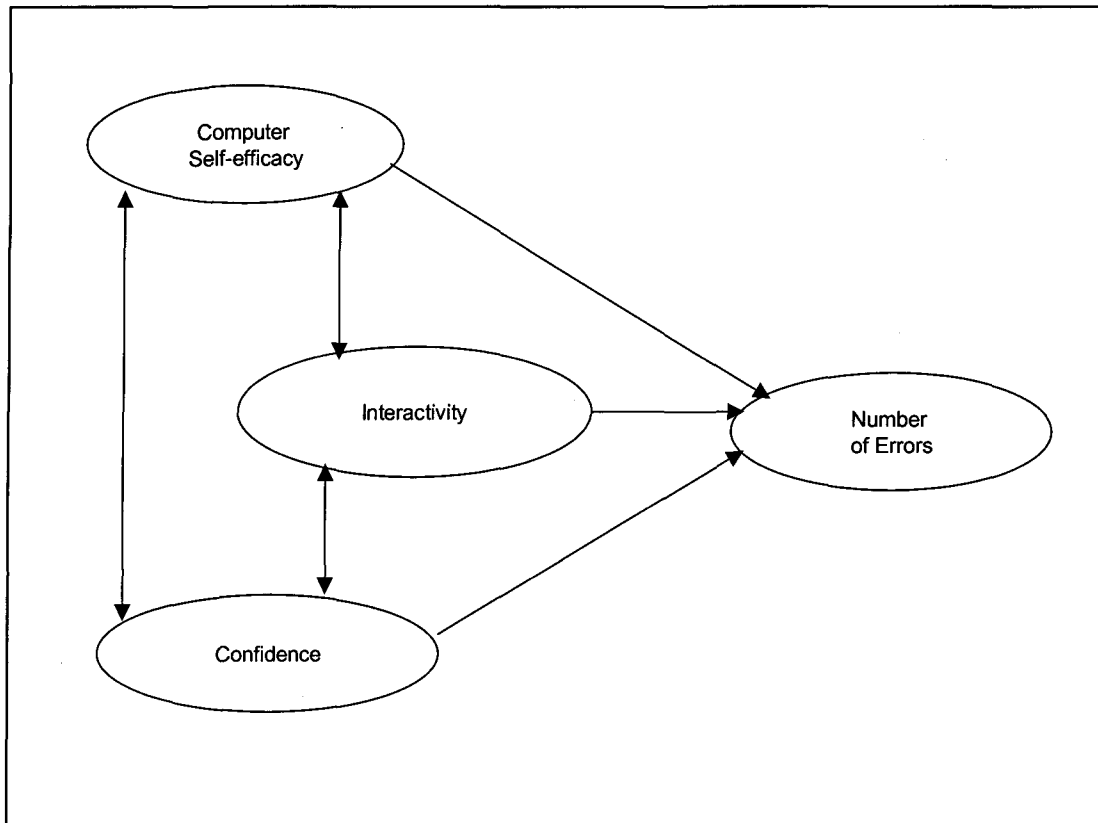
Figure 4.3: Multiple regression model

Using multiple regression, the total number of errors were then regressed on the linear combination of WtSE, Wt-Conf, and their interaction (the product of WtSE and Wt-Conf). The equation containing these three variables accounted for 17.8% of the variance in the number of errors, $F(3,92) = 6.406$, $p < .01$, adjusted $R^2 = .150$. It should be noted that the two-variable multiple regression effect of WtSE and Wt-Conf accounted for 11.8% of the variance, indicating less accountability.

B weights (unstandardized multiple regression coefficients) were then reviewed assessing the relative importance of the three variables in the prediction of the total number of errors. The WtSE and Wt-Conf measures were reverted back to the original data to amplify the effects of B, the unstandardized coefficients. The unstandardized coefficients (B) are presented in Table 4.11, which shows that the weighted confidence

and the interaction (product of weighted confidence and weighted computer self-efficacy) displayed statistically significant unstandardized coefficients.

Table 4. 11:   Unstandardized coefficients obtained in multiple regression analysis

	Unstandardized coefficients (B)	t^a
WtSE	.011	0.054
Wt-Conf	1.512	2.293**
Interaction	-.271	-2.536**

[a] For t tests that tested the significance of the unstandardized coefficients with df = 92.

** p<.01

The results of the analyses produced the following multiple regression equation with unstandardized coefficients:

$$\hat{Y}_{\#Errors} = 3.156 + 0.011(WtSE) + 1.512(Wt\text{-}Conf) + (-0.271)(\text{Interaction effect})$$

Since the Interaction effect is the product of WtSE and Wt-Conf, this is equivalent to:

$$\hat{Y}_{\#Errors} = 3.156 + 0.011(WtSE) + 1.512(Wt\text{-}Conf) + (-0.271)(WtSE)(Wt\text{-}Conf)$$

By observation, in examining Wt-Conf, when the Wt-Conf is low, for instance at -2 standard deviations:

$$\hat{Y}_{\#Errors} = 3.156 + 0.011(WtSE) + 1.512(-2) + (-0.271)(-2)(WtSE)$$

$$= 3.156 - 3.024 + 0.553(WtSE)$$

$$= 0.132 + 0.553(WtSE)$$

Similarly, in examining Wt-Conf, when Wt-Conf is high, for instance at +2 standard deviations:

$$\hat{Y}_{\#Errors} = 3.156 + 0.011(WtSE) + 1.512(2) + (-0.271)(2)(WtSE)$$

$$= 3.156 + 3.024 - 0.531(WtSE)$$

79

$$= 6.180 - 0.531(\text{WtSE})$$

Alternatively, by observation, in examining WtSE, when the WtSE is low, for instance at -2 standard deviations:

$$\hat{Y}_{\#Errors} = 3.156 + 0.011(-2) + 1.512(\text{Wt-Conf}) + (-0.271)(-2)(\text{Wt-Conf})$$

$$= 3.156 - 0.022 + 2.054(\text{Wt-Conf})$$

$$= 3.134 + 2.054(\text{Wt-Conf})$$

Similarly, in examining WtSE, when WtSE is high, for instance at +2 standard deviations:

$$\hat{Y}_{\#Errors} = 3.156 + 0.011(2) + 1.512(\text{Wt-Conf}) + (-0.271)(2)(\text{Wt-Conf})$$

$$= 3.156 + 0.022 + 0.970(\text{Wt-Conf})$$

$$= 3.178 + 0.970(\text{Wt-Conf})$$

These observations show that people with higher computer self-efficacy and higher confidence make fewer errors than those with lower computer self-efficacy and higher confidence. Also, people with higher computer self-efficacy and lower confidence make more spreadsheet errors than those with lower computer self-efficacy and lower confidence. This computer self-efficacy finding supports the self-efficacy research reviewed earlier.

This finding sheds some light that confidence ratings may have different effects on the ratings by a high self-efficacy individual as opposed to the ratings by a low self-efficacy individual as related to these two spreadsheet tasks.

4.5 Protocol Analysis Findings

Individual sessions, used specifically for the "think-aloud" sessions, were solicited of all subjects at the same time as the general sessions were offered. Each subject volunteered for this portion of the experiment either by personal session selection, or due to a scheduling conflict, willingly participated in either another group session or in one of the individual sessions. The self-selected 10 subjects were composed of 6 female and 4 male subjects. English was not the primary language for only one subject in this group. Table 4.12 provides descriptive statistics comparing the "think-aloud" subjects with the other general session subjects.

The mean and median error rates for the "think-aloud" subjects were higher than the general session subjects. The variance was also higher. However, the estimated mean number of errors was lower for the "think-aloud" subjects. The individual attention given the "think-aloud" subjects could account for some of the difference between individual and group session subjects; otherwise, a random selection process was maintained throughout the experiment.

Table 4.12: Comparison between think-aloud subjects and general session subjects

	Think-Aloud Subjects (n = 10)	General Session Subjects (n = 83)
Mean Number of Errors	3.55	3.01
Median Number of Errors	4.00	2.00
Variance of Errors	7.27	4.95
Estimated Mean Number of Errors	4.27	5.70
Minimum/Maximum Number of Actual Errors	0/9	0/9
Minimum/Maximum Number of Estimated Errors	3/8	0/10

4.5.1 Observations of Error Types

Utilizing Panko and Haverson's (1997) categorization of error types, the results of the protocol analyses of the experiment contained all categories of mechanical, logical, and omission errors. The same spreadsheet tasks were administered in the individual sessions for the protocol analysis as previously used by all subjects in their group sessions. The solution for the spreadsheet tasks was known to the experimenter in advance and each spreadsheet was corrected to obtain the known solution in order to identify and categorize the actual errors.

Table 4.13 provides descriptive statistics comparing the "think-aloud" subjects with the other general session subjects using Panko and Haverson's (1997) categorization of error types, and based upon this researcher's assessment of error type, which could be an artifact of the category the error was assigned.

Table 4.13: Comparing error types of "think-aloud" subjects and general session subjects

	"Think-Aloud" Subjects	Percentage of Total	General Session Subjects	Percentage of Total
Wall Task: (subtotal)	10	33	108	38
- Mechanical Errors	1	3	13	5
- Logical Errors	7	23	40	14
- Omission Errors	2	7	55	19
MicroSlo Task:	20	67	178	62
- Mechanical Errors	0	0	15	5
- Logical Errors	14	47	120	42
- Omission Errors	6	20	43	15
Total Errors	30	100	286	100

Errors were prevalent; a possible reason for the differences in all the categories of errors between the "think-aloud" and the group session subjects is the individual sessions required for the "think-aloud" sessions.

Mechanical errors were evident; while developing the spreadsheets, many

mechanical errors were caught and corrected before the final solution was submitted. For instance, a cell calculation was entered but an equal sign did not precede the calculation; this caused an unanticipated result, which an insertion of a leading equal sign corrected. One subject's "think-aloud" monologue was "Oh! I had to add in an equal sign to make it a formula." Using Panko and Halverson's approach to measuring errors, mechanical errors were fewer than expected, as many were corrected during the spreadsheet task development, and related primarily to procedural errors (e.g., copying wrong cells or including erroneous cells in calculation).

Far more logical and omission errors also occurred. The logical errors included using wrong calculations; for instance, corporate income tax was computed based on individual salaries, rather than the net corporate income, where all expenses were deducted from all the revenues. As an example one subject's comment was "It's a negative value, so I think I did something wrong. So I'm going to delete this answer." Another subject commented that "I caught an error after rechecking the answer, and going back to redo a total for expected costs." Using Panko and Halverson's categorization, logical errors accounted for the most number of errors.

Omission errors were also found as the spreadsheet tasks were developed; for example, rental expense was omitted. "Think-aloud" comments would not illustrate this type of error, but the outcomes of the spreadsheet tasks indicate that there were many omitted errors.

By triangulation of methods, Table 4.14 provides descriptive statistics based on this researcher's assessments of using Allwood's categories of errors in assessing the "think-aloud" sessions. Many of the detected errors were corrected during the spreadsheet development effort; however, a few errors still remain.

Table 4.14: Error types of "think-aloud" subjects based on Allwood's categories

	Number of errors detected	Number of errors corrected
Direct error hypothesis	9	7
Error suspicion	2	2
Standard check	4	3

As previously detailed, the direct error hypothesis formation was initiated by a presumed detection of an error, while error suspicion was initiated when the solution appeared strange or unexpected to the subject and not directly diagnosed to a specific error. The standard check was comprised of general check independent of any specific condition.

This triangulation of errors types provided an added dimension in understanding sources of errors. Allwood's direct error hypothesis captured many of the mechanical errors of Panko and Halverson's categories of errors, as these types of errors occurred during the problem-solving activities. To this researcher, the direct error hypothesis appeared related to the mechanical errors. Many logical errors were committed, but the subjects were unaware of many of its occurrences. While performing a direct error hypothesis and a standard check of a detected error, additional errors were encountered during the correction of an original detected error. Similarly, errors of omission occurred that subjects were also not aware of their existence.

The data reflects results similar to Allwood's and Panko's research findings. As an observations by this researcher, many of the subjects in the "think-aloud" sessions implicitly assumed that the electronic tool would detect many of their errors, which in some cases the tool did detect the error causing the subject to reassess their error in problem solving.

4.5.2 System Design and Analyses Observations

One characteristic displayed by most subjects was the exclusion of any detailed systems design and analysis effort while solving the problems. The subjects started the development of the spreadsheet almost immediately after reading the problem statement. Almost no consideration is given whether all the needed information was provided in the problem statement. Sternberg (2003) distinguished experts from novices by how they spent their time solving problems. Experts spend proportionately more time determining how to represent a problem, while novices spend proportionately more time implementing their solution (Sternberg, 2003). This effort of how to represent a problem is referred to in IT parlance as system design and analyses.

This approach of either no or only an incidental system design and analyses effort is contrary to the SDLC approach taken in software engineering development. Software design and analysis is done early in the development life cycle. IT software developers implicitly assume that their resultant product will have a life cycle of a minimum of several years. On the other hand, the spreadsheet developer might perceive their efforts to be a one-time effort, but field studies have also shown that spreadsheets remain in organizations for many years.

One subject entered all the input data provided in a separate area and subsequently referred to these data; another subject jotted notes on a piece of paper while working on the solution. Perhaps in spreadsheet development these activities can be perceived to be a precursor to the system design and analysis effort.

Hendry and Green (1994) posited that spreadsheets were good at incremental growth type of requests, which solved immediate problems very quickly but portends future problems in debugging and comprehension. This characteristic is exemplary of ad hoc and exploratory one-time efforts. Hendry and Green (1994) also offered

transformation and presentation as other categories, which were less suitable for spreadsheet development.

The data clearly reflects that system design and analysis is lacking. If this characteristic is carried forward into the business mileau, the results from the spreadsheet used for decision-making would become suspect.

4.5.3 Lack of Testing Observations

Another characteristic displayed consistently by the subjects was the lack of testing performed by the spreadsheet developers as compared to the testing performed in the software engineering arena. As exemplified in "think-aloud" protocol analysis by Tesch (1990), the following coding scheme was employed:

1.   Check solution for reasonableness,

2.   Verify if all problem statement requirements met, or

3.   Use alternate input data to assess if model produces similar

    results and still operational.

One subject concluded the development effort once the output format looked presentable and reasonable (coding scheme #1 above). The appearance aspect is consistent with the findings of Reithel et al. (1996), where a large neatly formatted output provided the impression of being correct while a less structured and small formatted output suggested additional effort.

After the initial spreadsheet development effort, another subject verbally repeated the problem statement in an attempt to assure that all aspects of the spreadsheet tasks were addressed and included in the final solution (coding scheme #2 above).

No subject in the "think-aloud" sessions provided an alternative input data

(coding scheme #3 above) to produce similar and reasonable results to verify the spreadsheet just developed. This situation could be due to the lack of training offered these subjects or a lack of experience. Since most subjects in the BUS 311 course were in their first year of their academic curricula, they might not have had the opportunity to take a system design and analysis course. The spreadsheet experience outside of the classroom for the "think-aloud" subjects averaged 4 hours per week and averaging 0.33 years for the past 2 years.

The data reflect that testing of spreadsheets is not extensive. This observation also adds suspicion to the results of spreadsheets used in decision-making.

4.5.4 Confidence, Overconfidence, and Computer Self-efficacy Observations

An interpretive reading of the "think-aloud" data entails the experimenter to develop or document a version of what the data represent, or what is inferred from the data (Mason, 2002). The observed characteristics of a lack of testing and a lack of problem design and analysis effort in the problem solving activity suggests that training may be an integral component needed for spreadsheet development; however, this would still not preclude errors from occurring, rather minimize the effects of the spreadsheet on which the resultant decisions were based.

The mean computer self-efficacy for the "think-aloud" subjects was 620, or slightly above the median of 600. The mean overconfidence for the "think-aloud" subjects was -0.73 with a median of −1.00 for these subjects. Compared to the entire group, the "think-aloud" subjects were similar to the other subjects.

Chapter 5. Discussion

This research is the first attempt to integrate the self-efficacy, confidence, and spreadsheet error performance research streams. The addition of self-efficacy theory and confidence theory to spreadsheet research provides an added dimension in our understanding of error performance in spreadsheet development. The results of this research show the applicability of computer self-efficacy to spreadsheet error performance, and to a lesser extent, the applicability of confidence to spreadsheet error performance.

5.1 Contribution

The main finding from this research was that there was a negative correlation between computer self-efficacy and the number of errors (hypothesis H_2). Consistent with research in other areas, self-efficacy is a good predictor of computer error performance. People with higher computer self-efficacy make fewer spreadsheet errors than those with lower computer self-efficacy.

When confidence was added to the picture, confidence in the accuracy of spreadsheet development by itself was not significantly correlated to the number of errors (hypothesis H_{1a} was rejected). However, the interaction of confidence with computer self-efficacy adds to the predictability of spreadsheet error performance (hypothesis H_{3c}). People with higher computer self-efficacy and higher confidence make less spreadsheet errors than those with lower computer self-efficacy and higher confidence. Also, people with higher computer self-efficacy and lower confidence make more spreadsheet errors than those with lower computer self-efficacy and lower confidence.

Self-efficacy has been shown to be a good predictor of performance in many

88

fields. This research confirms the applicability of computer self-efficacy to computer and IT performance. Although extensively researched, confidence and overconfidence have not been historically related to performance. By themselves, confidence and overconfidence did not predict performance in this research. However, confidence did supplement the predictability of computer self-efficacy when added; so its usefulness is proven.

5.2 Confidence, Overconfidence and Computer Self-efficacy

Another concern is the apparent similarity between the constructs of computer self-efficacy and confidence. Computer self-efficacy, as hypothesized in H_{3a} and H_{3b}, is significantly positively correlated to both confidence and overconfidence; the effect sizes of the correlations are large (0.53) and medium (0.29) respectively.

As previously stated, there was a negative correlation between computer self-efficacy and the number of errors (hypothesis H_2). But, confidence in the accuracy of spreadsheet development was not significantly correlated to the number of errors (hypothesis H_{1a}). This divergence questions the similarity of the computer self-efficacy and confidence constructs.

A principal component analysis was performed on the experimental data. The analysis showed that computer self-efficacy and confidence are distinguishable factors. Principal component analysis is based on the correlations between the data items sharing common variance-covariance characteristics.

The regression analysis found that 17.8% of the shared variance was accounted for by confidence, computer self-efficacy, and their interaction on the number of errors. Without the interaction effect, only 11.8% of the variance was accounted for.

5.3 Demographic Findings

In investigating the demographic data, there were no significant correlations in the sample when gender, English as a second language, and grade point average were correlated with computer self-efficacy and confidence.

5.4 Protocol Analysis Findings

In a companion research effort, the "think-aloud" protocol analysis methodology was utilized with a sample of 10 subjects; only observations are reported. The data clearly showed that no system design and analysis activities occurred before the spreadsheet development.

Another salient observation was a lack of testing in most of the spreadsheets. Only one subject did testing during the development of his spreadsheet, while another did some testing after the spreadsheet development was completed.

Future research utilizing more experienced subjects may reveal differences in planning and testing in spreadsheet development. However, other studies have also shown that even experienced spreadsheet developers do not take the necessary planning and testing precautions used by the formal software programmers (refer to section 2.1.4).

5.5 Limitations

A limitation in this research is the problematic non-independence of error rate and overconfidence. From the literature reviewed by this researcher, human errors are

pervasive in most cognitive activities. Similarly, overconfidence appears to be a prevalent condition in most human activities. Since both conditions can occur, this researcher expected both occurrences, yet the data did not support this condition.

Based on earlier research, it was expected that most subjects would be very confident in their spreadsheet development and most subjects would be overconfident in their assessment of the accuracy of their spreadsheet tasks. We expected that their actual number of errors would be higher than their estimated number of errors. In this experiment, the expected overconfidence was not seen; rather subjects were underconfident causing the rejection of H_{1a}. This particular artifact may be due to the sample selection. The subjects selected were upper division undergraduates in their first year in the business school. Although these subjects had previous exposure(s) to spreadsheet development and to basic accounting and finance courses, the level of their experience using these tools places them towards the novice end of the spectrum on a novice-expert continuum.

An observation from the "think-aloud" protocol analysis found that subjects expended a minimum of planning effort, with almost all their time devoted to implementing a solution to the tasks assigned. Experts spend relatively more time determining how to represent a problem than novices; experts expend less time actually implementing the strategy for solution than novices (Sternberg, 2003). Expertise level knowledge requires extensive exposure. Ericsson (2003) posits that at least ten years of practice in the domain is needed to attain the performance level of experts. The mean age of the subjects in the experiment was 22.6 years with 17 subjects (out of the 93), or 18% of the subjects, averaging 1.8 years of development experience and averaging just over a day a week of work where their spreadsheet skills are used. The other 82% of the subjects had only classroom exposure to spreadsheet tasks. This suggests that

91

experience was very limited. Further research using a more experienced sample set would determine whether the underconfidence condition or the cautious and conservative nature is a factor, which possibly would yield a different concluding result.

Another possible explanation for this underconfidence appears to be related to the relative low levels of the integration of domain knowledge observed in the subjects. The integration of academic work in spreadsheets and its application to functionally related areas is weak. Several "think-aloud" subjects found that using the spreadsheet tool for accounting purposes was novel; this also suggests a limited experience level. Subjects may consider academic courses as discrete and independent undertakings. Ericsson (2003) suggests successful practice requires identifying specific goals to change the performance levels. The low level of integration was evident especially in the pro-forma income statement spreadsheet task (one of the spreadsheet task used in this research experiment). The pro-forma income statement draws from previous course materials in accounting (income statement, revenue, and expenses), information technology (spreadsheet creation and manipulation), and finance (sales, material costing, and salaries). Although the subjects were in their first year of their business curricula, their ability to draw from different courses appeared to be challenged.

A third possible explanation for the underconfidence may have been the incentive(s) for the subjects to participate in the experiment. Extra credit over and above normal classroom credit was offered. If this experiment had been conducted earlier in the semester, a greater incentive might have been evidenced for the subject's to attain a higher course grade; or at least higher experiment participation level, as seen in the pilot experiment, might have occurred. Since this experiment was offered after the first midterm exam, the timing of this experiment might have an influence, as some subjects would have better determined whether extra credit would influence their class standings

and be beneficial for them. One potential subject, who elected not to participate in the experiment, commented that she had enough alternate extra credits, as another experiment was offered concurrent with this one. Further research on non-participants from the selected sample may assist in determining whether incentive is a factor or another characteristic of the non-participants, which may affect the outcome.

In addition to the limited experience level of the subjects selected, a companion limitation to this research is the subject selection from the classroom environment rather than a more naturalistic environment, such as the workplace milieu.

5.6   Research Implications

Prior to the advent of cognitive psychology, factors affecting the environment were considered the primary cause of errors. With the shift to today's focus on cognitive processes, cognitive bias became a major area of study. In the 1980s, the question of bias in confidence judgment appeared settled with people being overconfident on all but the easiest questions (Klayman et al., 1999).

The confidence judgment question was reopened in the 1990s. Opponents argued that people are imperfect but unbiased judges of confidence; they argued that the choices of questions were biased. This debate continues today with differences in overconfidence occurring between domains and between individuals with differences caused by information content, information processing, and their interactions (Klayman et al., 1999).

Confidence in spreadsheet development as previously researched assessed an individual's estimated error rate in comparison to the group's average actual error rate.

Subjects were found to be overconfident. This is contrary to the finding in this research. This research, however, measures the individual's estimated number of errors in comparison to his or her own performance. This measure is new with this research effort. Further research is needed to study differences in confidence measures.

Another area for future research involves the non-participants of the BUS311 sample set. Obtaining computer self-efficacy and confidence data on the non-participants would provide data as to whether there is another factor differentiating participants and non-participants. Potential subjects did not participate in this experiment for various reasons. One factor for non-participants is the extra credit offered as not being a sufficient motivator. Perhaps their computer self-efficacy and confidence are high, or they were performing sufficiently well in the course that the extra credit offering would not be beneficial. Or, these non-participants may be overconfident. This research's experimental design leaves certain questions unanswered, specifically whether the non-participants were overconfident or underconfident.

The underconfidence seen in the data was contrary to the earlier findings that this researcher reviewed. This underconfidence could be due to the inexperience level of the selected subjects, or the minimal integration of academic coursework. A suggestion that future experiment subjects be selected later in their academic career would be a consideration. As previously mentioned, Ericsson (2003) posits that at least ten years of practice in the domain is needed to attain the performance level of experts. But, errors are still made by even experts. A field experiment soliciting more experienced subjects might be more applicable to the core question of the existence of overconfidence in spreadsheet errors.

5.7 Practical Implications

Errors are inherent in all human cognitive activities. Recognition that errors do occur and that efforts are needed to reduce errors are important. Training and instruction are integral to addressing this condition of error containment. Skills in spreadsheet development, functional knowledge, and their interaction are essential key ingredients to reducing the number of errors. Testing should be included. A lack of testing is especially acute with the non-professional end-user. Further research into these areas would provide the essential components needed for error containment. Spreadsheet error detection becomes more and more important due to the requirements imposed by the Sarbanes-Oxley bill passed in 2002, where significant financial and penal penalties are at stake.

Another area needing attention is a revision to the academic curricula to include an integration of different components of the business curricula for the students to better understand the interaction and integration effects of their academic topics. These interrelationships actually exist in the business world. This is analogous to the case study approach that has been successfully employed by most business schools. Since the spreadsheet development and analysis is more detailed in nature than the typical case study, a greater content detail is a major distinction required for integrating business topics at the operational and managerial levels. This will expose the students to the potential cascading effects of errors typical in spreadsheet errors.

References

Agarwal, R., Sambamurthy, V., & Stair, R. M. (2000). Research report: the evolving relationship between general and specific computer self-efficacy – an empirical assessment. *Information Systems Research, 11*(4), 418–430.

Allwood, C. M. (1984). Error detection processes in statistical problem-solving. *Cognitive Science, 8*, 413–437.

Arkes, H. R., Christensen, C., Lai, C., & Blumer, C. (1987). Two methods of reducing overconfidence. *Organizational Behavior and Human Decsion Processes, 39*, 133–144.

Ayton, P., & McClelland, A. G. R. (1997). How real is overconfidence? *Journal of Behavioral Decision Making, 10*(3), 279–285.

Banaji, M. R., Bazerman, M. H., & Chugh, D. (2003). How (un)ethical are you? *Harvard Business* Review, *81*(12), 56–64.

Bandura, A. (1977a). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84*, 191–215.

Bandura, A. (1977b). *Social learning theory.* Englewood Cliffs, NJ: Prentice-Hall.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory.* Englewood Cliffs, NJ: Prentice-Hall.

Bandura, A. (1997). *Self-efficacy: The exercise of control.* New York: W.H. Freeman.

Barber, B., & Odean, T. (1999). The courage of misguided convictions. *Financial Analyst Journal, 55*, 41–55.

Barr-Hillel, M. (1973). On the subjective probability of compound events. *Organizational Behavior and Human Performance, 9*, 396–406.

Brown, I., Jr., & Inouye, D. K. (1978). Learned helplessness through modeling: the role of perceived similarity in competence. *Journal of Personality & Social Psychology, 36*(8), 900–908.

Brown, P. S., & Gould, J. D. (1987). An experimental study of people creating spreadsheets. *ACM Transactions on Office Information Systems, 5*(3), 258–272.

Butler, R. J. (2000). Is this spreadsheet a tax evader? How H.M. Customs & Excise tax test spreadsheet applications. Paper presented at the meeting of the *Proceedings of the 33rd Hawai'i International Conference on System Sciences*, Kīhei, HI.

Cale, E. G., Jr. (1994). Quality issues for end user developed software. *Journal of Systems Management, 45*(1), 36–39.

Chan, Y. E., & Storey, V. C. (1996). The use of spreadsheets in organizations:

Determinants and consequences. *Information & Management, 31,* 119–134.

Chase, W. G., & Simon, H. A. (1973). The mind's eye in chess. In W. G. Chase (Ed.), *Visual information processing* (pp. 215–281). New York: Academic Press.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences (2nd ed).* Hillsdale, NJ: Lawrence Erlbaum Associates.

Compeau, D. R., & Higgins, C. A. (1995). Application of social cognitive theory to training for computer skills. *Information Systems Research, 6*(2), 118–143.

Compeau, D. R., & Higgins, C. A. (1999). Social cognitive theory and individual reactions to computing technology: A longitudinal study. *MIS Quarterly, 23*(2), 145–158.

Cragg, P. G., & King, M. (1993). Spreadsheet modeling abuse: An opportunity for OR? *Journal of Operational Research Society, 44*(8), 743–752.

Davis, N., & Ikin, C. (1987). Auditing spreadsheets. *Australian Accountant,* 54–56.

De Groot, A. D. (1965). *Thought and choice in chess.* The Hague, The Netherlands: Mouton & Co.

DiClemente, C. C., Prochaska, J. O., Fairjurst, S. K., Velicer, W. F., Velasques, M. M., & Rossi, J. S. (1991). The process of smoking cessation: An analysis of pre-contemplation, contemplation, and preparation stages of change. *Journal of Consulting and Clinical Psychology, 59,* 295–304.

Ericsson, K. A. (2003). The acquisition of expert performance as problem solving: Construction and modification of mediating mechanisms through deliberate practice. In J. E. Davidson & R. J. Sternberg (Eds.), *The psychology of problem solving* (pp. 31–83), Cambridge, United Kingdom: Cambridge University Press.

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data.* Cambridge, MA: MIT Press.

Fagan, M. E. (1976). Design and code inspections to reduce errors in program development. *IBM Systems Journal, 15*(3), 182–211.

Floyd, B., Walls, J., & Marr, K. (1995). Managing spreadsheet model development. *Journal of Systems Management, 46*(1), 38–43, 68.

Galletta, D. F., Abraham, D., El Louadi, M., Lekse, W., Pollalis, Y. A., & Sampler, J. L. (1993). An empirical study of spreadsheet error-finding performance. *Accounting Management Information Technology, 3*(2), 79–95.

Galletta, D. F., Hartzel, K. S., Johnson, S. E., Joseph, J. L., & Rustagi, S. (1997). Spreadsheet presentation and error detection. *Journal of Management Information Systems, 13*(3), 45–63.

Gigerenzer, G., Hoffrage, U., & Kleinbolting, H. (1991). Probabilistic mental models: A

Brunswickian theory of confidence. *Psychological Review, 98,* 506–526.

Goo, J. (2002). The effect of negative feedback on confidence calibration and error reduction in spreadsheet development. *Dissertation Abstracts International, 63*(06), 2032.

Gould, D., & Weiss, M. (1981). Effects of model similarity and model self-talk on self-efficacy in muscular endurance. *Journal of Sports Psychology, 3,* 17–29.

Graham, S., & Weiner, B. (1996). Theories and principles of motivation. In D. C. Berliner & R. C. Calfee (Eds.), *Handbook of educational psychology* (pp. 63–84), New York: Simon & Schuster MacMillan.

Hall, M. J. J. (1996). A risk and control oriented study of the practices of spreadsheet application developers. Paper presented at the meeting of the *Proceedings of the 29th Hawai'i International Conference on System Sciences,* Kīhei, HI.

Hatcher, L. (1994). *A step-by-step approach to using the SAS system for factor analysis and structural equation modeling.* Cary, NC: SAS Institute.

Hays, R. D., & Ellickson, P. L. (1990). How generalized are adolescents' beliefs about pro-drug pressures and resistence self-efficacy? *Journal of Applied Social Psychology, 20,* 321–340.

Hendry, D. G., & Green, T. R. G. (1994). Creating, comprehending, and explaining spreadsheets: A cognitive interpretation of what discretionary users think of the spreadsheet model. *International Journal of Human-Computer Studies, 40*(6), 1033–1065.

Horowitz, A. (2004). Spreadsheet overload? *Computerworld,* May 24, 2004, 46–47.

Johnson, R. D., & Maracas, G. M. (2000). Research report: The role of behavioral modeling in computer skills acquisition — Toward refinement of the model. *Information Systems Research, 11*(4), 402–417.

Jones, N., & Rogers, P. J. (2003). Preoccupation, food, and failure: An investigation of cognitive performance deficits in dieters. *International Journal of Eating Disorders, 33*(2), 185–192. Retrieved April 19, 2005, from the Psychological and Behavioral Science Collection databases.

Keillor, G. (1987). *Leaving home.* New York: Viking Penguin.

Keren, G. (1992). Improving decisions and judgments: The desirable versus the feasible. In G. Wright & F. Bolger (Eds.), *Expertise and decision support* (pp. 25–46), New York: Plenum Press.

Klayman, J., Soll, J. B., Gonzalez-Vallejo, C., & Barlas, S. (1999). Overconfidence: It depends on how, what, and whom you ask. *Organizational Behavior and Human Decision Processes, 49*(3), 216 – 247.

Koriat, A., Lichtenstein, S., & Fischoff, B. (1980). Reasons for confidence. *Journal of Experimental Psychology: Human Learning and Memory, 6*, 107–117.

Kruck, S. E., Maher, J. J., & Barkhi, R. (2003). Framework for cognitive skill acquisition and spreadsheet training. *Journal of End User Computing, 15*(1), 20–37.

Lacher, J. (1997). The power of spreadsheets [Electronic version]. *Journal of Accountancy, 183*(5), 66–71.

Lichtenstein, S., & Fishoff, B. (1980). Training for calibration. *Organizational Behavior and Human Performance, 26*, 149–171.

Looney, C. A., Valacich, J. S., & Akbulut, A. Y. (2004). Online investment self-efficacy: development and initial test of an instrument to assess perceived online investing abilities. Paper presented at the meeting of the *Proceedings of the 37th Hawai'i International Conference on System Sciences,* Kona, HI.

Marakas, G. M., Yi, M. Y., & Johnson, R. D. (1998). The multilevel and multifaceted character of computer self-efficacy: Toward a clarification of the construct and an integrated framework for research. *Information Systems Research, 9*(2), 126–162.

Mason, J. (2002). *Qualitative researching (2nd edition).* Thousand Oaks, CA: Sage.

Mayer, R. E. (1992). *Thinking, problem solving, cognition (2nd edition).* New York: W.H. Freeman.

Mayer, R. E. (2003). *Learning and instruction.* Upper Saddle River, NJ: Merrill Prentice Hall.

Merriam-Webster's Collegiate Dictionary — 11th Edition (2003). Springfield, MA: Merriam-Webster.

Nahl, D. (1995). Affective elaborations in Boolean search instructions for novices: Effects on comprehension, self-confidence, and error type. *Proceedings of the 58th ASIS Annual Meeting, 32,* 69–75.

Nahl, D. (1996). Affective monitoring of Internet learners: Perceived self-efficacy and success. *Proceedings of the 98th ASIS Annual Meeting, 33,* 100–109.

Nahl, D., & Harada, V. H. (1996). Composing Boolean search statements: Self-confidence, concept analysis, search logic, and errors. *School Library Media Quarterly, 24*(4), 199–206.

Nardi, B. A., & Miller, J. R. (1991). Twinkling lights and nested loops: Distributed problem solving and spreadsheet development. *International Journal of Man-Machine Studies, 34*(2), 161–184.

National Aeronautics & Space Adminstration (1992). Recommended approaches to software development: Revision 3, Software Engineering Laboratory Series, SEL–81–305. Retrieved September 30, 2004 from

http://sel.gsfc.nasa.gov/website/documents/online-doc81-305new.pdf

Norman, D. A. (1981). Categorization of action slips. *Psychological Review, 88*(1), 1–15.

Nunnally, J. C. (1978). *Psychometric theory* (2nd edition). New York: McGraw-Hill.

Onglatco, M., Yuen, E. C., Leong, C. C., & Lee, G. (1993). Managerial self-efficacy and managerial success in Singapore. *International Journal of Management, 10,* 14–21.

Pajares, F. (1997). Current directions in self-efficacy research. In M. Maehr & P. R. Pintrich (Eds.), *Advances in motivation and achievement* (pp. 1–49), *10,* Greenwich, CT: JAI Press.

Panko, R. R. (1996). Hitting the wall: Errors in developing and debugging a 'simple' spreadsheet model. Paper presented at the meeting of the *Proceedings of the 29th Hawai'i International Conference on System Sciences,* Kīhei, HI.

Panko, R. R. (1999). Applying code inspection to spreadsheet testing. *Journal of Management Information Systems, 16*(2), 159–176.

Panko, R. R. (2000). *Journal of End User Computing, 10*(2), Spring, 1998, 15–21. Retrieved February 24, 2005, from http://panko.cba.hawaii.edu/ssr/

Panko, R. R. (2005a). *Human Error Website.* Retrieved February 22, 2005, from http://panko.cba.hawaii.edu/HumanErr/

Panko, R. R. (2005b). *Spreadsheet Research (SSR) Website.* Retrieved February 24, 2005, from http://panko.cba.hawaii.edu/ssr/

Panko, R. R., & Featherman, M. (2000). *Overconfidence in spreadsheet development.* Unpublished manuscript, University of Hawai'i at Mānoa.

Panko, R. R., & Halverson, R., Jr. (1996). Spreadsheets on trial: A survey of research on spreadsheet risks. Paper presented at the meeting of the *Proceedings of the 29th Hawai'i International Conference on System Sciences* (pp. 324–325), Kīhei, HI.

Panko, R. R., & Halverson, R., Jr. (1997). Are two heads better than one? (At reducing errors in spreadsheet modeling?). *Office Systems Research Journal, 15*(1), 21–32.

Panko, R. R., & Sprague, R. (1998). Hitting the wall: Errors in developing and code inspecting 'simple' spreadsheet model. *Decision Support Systems, 22*(4), 337–353.

Pender, L. (1998). Lotus 1-2-3. *Computer Reseller News, 816,* 71–72.

Petersen, D. (1996). *Human error reduction and safety management (3rd ed.).* New York: Van Nostrand Reinhold.

Pintrich, P. R., & Schunk, D. H. (1996). *Motivation in education: Theory, research, and application.* Englewood Cliff, NJ: Merrill.

Plous, S. (1993). *The psychology of decision making*. Philadelphia: Temple University Press.

Pulford, B., & Colman, A. (1996). Overconfidence, base-rates and outcome positivity/negativity of predicted events. *British Journal of Psychology, 87*(3), 431–445.

Rasmussen, J. (1986). *Information processing and human-machine interaction: An approach to cognitive engineering*. New York: North-Holland.

Rasmussen, J. (1990). The role of perception in organizing behavior. *Ergonomics, 33*(10/11), 1185–1199.

Reason, J. T. (1987). Generic error-modeling system (GEMS): A cognitive framework for locating common human error forms. In J. Rasmussen, K. Duncan, & J. LePlat (Eds.), *New technology and human error* (pp. 63–83), New York: John Wiley & Sons.

Reason, J. T. (1990). *Human error*. New York: Cambridge University Press.

Reithel, B. J., Nichols, D. L., & Robinson, R. K. (1996). An experimental investigation of the effects of size, format, and errors on spreadsheet reliability perception. *Journal of Computer Information Systems, 36*(3), 54–64.

Ronen, B., Palley, M. A., & Lucas, H. C. Jr. (1989). Spreadsheet analysis and design [Electronic version]. *Communications of the ACM, 32*(1), 84-93.

Rumar, K. (1990). The basic driver error: Late detection. *Ergonomics, 33*(10/11), 1281–1290.

Schneider, J. A., O'Leary, A., & Agras, W. S. (1987). The role of perceived self-efficacy in recovery from bulimia: A preliminary examination. *Behaviour Research and Therapy, 25*, 429–432.

Schulthesis, R., & Sumner, M. (1994). The relationship of application risks to application controls: A study of microcomputer-based spreadsheet applications. *Journal of End User Computing, 6*(2), 11–18.

Schunk, D. H. (1983). Reward contingencies and the development of children's skills and self-efficacy. *Journal of Educational Psychology, 75*, 511–518.

Schunk, D. H. (2001). Social cognitive theory and self-regulated learning. In B. J. Zimmerman & D. H. Schunk (Eds.), *Self-regulated learning and academic achievement: Theoretical perspectives* (pp. 125–151), Hillsdale, NJ: Lawrence Erlbaum Associates.

Schunk, D. H., & Zimmerman, B. J. (1997). Social origins of self-regulatory competence. *Educational Psychologist, 32*, 195–208.

Senders, J. W., & Moray, N. P. (1991). *Human error: Cause, prediction, and reduction*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Shanteau, J. (1992). The psychology of experts: An alternative view. In G. Wright & F, Bolger (Eds.), *Expertise and decision support* (pp. 11–23). New York: Plenum Press.

Speir, C., & Brown, C. V. (1996). Perceived risks and management actions: Differences in end-user application development across functional groups. Paper presented at the meeting of the *Proceedings of the 29th Ninth Hawai'i International Conference on System Science*, Kīhei, HI.

Stajkovic, A. D., & Luthans, F. (1998). Self-efficacy and work-related performance: Meta-analysis. *Psychological Bulletin, 124*(2), 240–261.

Sternberg, R. J. (2003). *Cognitive psychology (3rd ed)*. Belmont, CA: Wadsworth/Thomson.

Stone, D. N., & Black, R. L. (1989). Using microcomputers: Building structured spreadsheets. *Journal of Accountancy, 168*(4), 131–142.

Sweller. J., & Chandler, P. (1994). Why some material is difficult to learn. *Cognition and Instruction, 12*(3), 185–233.

Tesch, R. (1990). *Qualitative research: Analysis types and software tools.* New York: The Falmer Press.

Troutt, M. D. (2005). Spreadsheet usage: Some new challenges and opportunities [Electronic version]. *Journal of Organizational and End User Computing, 17*(1), i–v.

Tversky, A., & Kahneman, D. (1982a). Belief in the law of small numbers. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgement under uncertainty: Heuristics and biases,* (pp. 23–31). Cambridge, MA: University Press.

Tversky, A., & Kahneman, D. (1982b). Judgments of and by representativeness. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgement under uncertainty: Heuristics and biases,* (pp. 84–98). Cambridge, MA: University Press.

Tversky, A., & Kahneman, D. (1982c). Judgment under uncertainty: Heuristics and biases. In , D. Kahneman, P. Slovic, & A. Tversky (Eds), *Judgment under uncertainty: Heuristics and biases* (pp. 3–20). Cambridge, MA: University Press.

U. S. Department of Labor, Bureau of Labor Statistics (2001). Retrieved September 30, 2004, from http://www.bls.gov/news.release/ciuaw.tn.htm

Vealey, R. S. (1986). Conceptualization of sport-confidence and competitive orientation: Preliminary investigation and instrument development. *Journal of Sport Psychology, 8*(3), 221–246.

Vealey, R. S. (1988). Sport-confidence and competitive orientation: An addendum on scoring procedures and fender differences. *Journal of Sport & Exercise Psychology, 10,* 471–478.

Vealey, R. S., Hayashi, S. W., Garner-Holman, M., & Giacobbi, P. (1998). Sources of

sport-confidence: Conceptualization and instrument development. *Journal of Sport & Exercise Psychology, 20,* 54–80.

Vicente, K. J., & De Groot, A. D. (1990). The memory recall paradigm: Straightening out the historical record. *American Psychologist, 45*(2), 285–287.

Weber, N., & Brewer, N. (2004). Confidence–accuracy calibration in absolute and relative face recognition judgments. *Journal of Experimental Psychology/Applied, 10*(3), 156–172. Retrieved March 4, 2005, from the EBSCO Host Research databases.

Yi, M. Y., & Davis, F. D. (2003). Developing and validating an observational learning model of computer software training and skill acquisition. *Information Systems Research, 14*(2), 146–169. Retrieved April 19, 2005, from the Business Source Premier databases.

Yi, M. Y., & Im, K. S. (2004). Predicting computer task performance: Personal goals and self-efficacy. *Journal of Organizational and End User Computing,16*(2), 20–37. Retrieved April 19, 2005, from the ABI-INFORM Global databases.

Zeldin, A. L., & Pajares, F. (2000). Against the odds: Self-efficacy beliefs of women in mathematical, scientific, and technological careers. *American Educational Research Journal, 37*(1), 215–246.

Zimmerman, B. J. (2000). Attaining self-regulation. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 13–39). San Diego, CA.: Academic Press.

Zimmerman, B. J. (2001). Theories of self-regulated learning and academic achievement: An overview and analysis. In B. J. Zimmerman & D. H. Schunk (Eds.), *Self-regulated learning and academic achievement: Theoretical perspectives* (pp. 1–37). Hillsdale, NJ: Lawrence Erlbaum Associates.

Appendix A: Instructions

There are 4 sections to this experiment. Please confirm that you have 4 separate sections and a diskette.

You are to perform this experiment in the specified **SEQUENCE.** This order is critical.

1. General questionnaire
2. MicroSlo Task
3. Wall Task
4. Final questionnaire

During the experiment, if you are confused on any of written questions or requested information, please make a note and provide an itemized documentation in the last section under Comments.

Please feel free to take a break at your convenience. A break between each of the spreadsheets (Sections 2 and 3) would be a good time. Please help yourself to some candy at the sign-in desk.

Remember that your answers are confidential and you are NOT graded, so please do your best as your input will provide invaluable research data. You will receive extra credit for your participation, so **please follow instructions**.

After this experiment, please do not discuss the contents of this experiment with your classmates or peers. A debriefing will be held after all students have completed this experiment. Your cooperation is appreciated.

Thank you for your participation.

Appendix B: Questionnaire

Step 1

Description of the Experiment,

Agreement to Participate

Please read the following description of the experiment. If you have any questions about the experiment after reading the description, please ask.

> The purpose of this experiment is to help us understand the types of errors that students make when they build spreadsheet models.
>
> We ask that you do your best, but we do not expect everybody to build a perfect spreadsheet, and you will get full credit if you do your best.
>
> You will build two spreadsheet models. After building each spreadsheet model, you will complete a questionnaire.
>
> You do not have to format the models. Merely add basic text, numbers, and formulas.

After you have read the description, please go on to the next two pages and fill out the two copies of the agreement to participate in the experiment.

If you have questions, please ask the experimenter for clarification. Please rip off and keep the second copy. Let the experimenter know you are finished.

AGREEMENT TO PARTICIPATE IN AN

Experiment in Spreadsheet Development
Dr. Raymond R. Panko & Steven T. Takaki
College of Business Administration / University of Hawai'i
2404 Maile Way / Honolulu, HI 96822
(808) 956-5049

Project Description

You are asked to participate in a research experiment. In this experiment, you will develop two spreadsheets working alone. After each, you will fill out questionnaires asking you to describe your experience.

There are no hidden agendas. Although there is no direct benefit to you, we hope this project will help the academic community better understand the problems that people have developing spreadsheet models. As you are aware, spreadsheeting is one of the most widespread computer applications in business, so there is a strong need to understand spreadsheeting.

Your responses will be confidential in all reporting. The researchers believe that there is little or no risk to participating in this research project.

You may refuse to do the experiment. If you start filling the experiment, you may stop at any time. No reprisals or other negative actions will be taken if you refuse to participate or discontinue participation. Extra course credit will be given for student participation; extra credit alternative will be available to each student.

Certification

I certify that I have read and that I understand the foregoing, that I have been given satisfactory answers to my inquiries concerning project procedures and other matters and that I have been advised that I am free to withdraw my consent and to discontinue participation in the project or activity at any time without prejudice.

I herewith give my consent to participate in this project with the understanding that such consent does not waive any of my legal rights, nor does it release the principal investigator or the institution or any employee or agent thereof from liability for negligence.

If you cannot obtain satisfactory answers to your questions or have comments or complaints about your treatment in this study, contact: Committee on Human Studies, University of Hawai'i, 2540 Maile Way, Honolulu, Hawai'i 96822. Phone: 956-5007

_____   Date: _____
Signature of participant

(Rev. 10/03)

Please return this copy.

106

AGREEMENT TO PARTICIPATE IN AN

Experiment in Spreadsheet Development
Dr. Raymond R. Panko & Steven T. Takaki
College of Business Administration / University of Hawai'i
2404 Maile Way / Honolulu, HI 96822
(808) 956-5049

Project Description

You are asked to participate in a research experiment. In this experiment, you will develop two spreadsheets working alone. After each, you will fill out questionnaires asking you to describe your experience.

There are no hidden agendas. Although there is no direct benefit to you, we hope this project will help the academic community better understand the problems that people have developing spreadsheet models. As you are aware, spreadsheeting is one of the most widespread computer applications in business, so there is a strong need to understand spreadsheeting.

Your responses will be confidential in all reporting. The researchers believe that there is little or no risk to participating in this research project.

You may refuse to do the experiment. If you start filling the experiment, you may stop at any time. No reprisals or other negative actions will be taken if you refuse to participate or discontinue participation. Extra course credit will be given for student participation; extra credit alternative will be available to each student.

Certification

I certify that I have read and that I understand the foregoing, that I have been given satisfactory answers to my inquiries concerning project procedures and other matters and that I have been advised that I am free to withdraw my consent and to discontinue participation in the project or activity at any time without prejudice.

I herewith give my consent to participate in this project with the understanding that such consent does not waive any of my legal rights, nor does it release the principal investigator or the institution or any employee or agent thereof from liability for negligence.

If you cannot obtain satisfactory answers to your questions or have comments or complaints about your treatment in this study, contact: Committee on Human Studies, University of Hawai'i, 2540 Maile Way, Honolulu, Hawai'i 96822. Phone: 956-5007

_____   Date: _____
Signature of participant

(Rev. 10/03)

Please remove and keep this copy for your own records.

Step 2

Preliminary Questionnaire for Spreadsheet Development

Experiment

If you have any questions, please ask them at any time.

1. Sign-in workstation number assigned to you: _____

3.   Which class(es) are you in? (check all that apply)
        ( ) BUS311
        ( ) BUS310
        ( ) Other – Specify _____

4.   Please list any accounting courses you *have taken* and are *currently* taking. Do NOT count the two courses you had to take for entry into the College of Business Administration.

5.   Please list any computer courses you have had and are currently taking. Do NOT list the course you had to take for entry into the College of Business Administration (ICS101, etc.) or BUS 311.

6. *At work*, have you entered data into spreadsheet models *written by others*?

( )   No

( )   Yes, for ____ years (you may use fractions), averaging ___ hours per week. Use fractions if desirable.

7. *At work*, have you *created* spreadsheet models, *checked the models of others for errors*, or *trained* people to use spreadsheet programs?

( )   No

( )   Yes, for ____ years (you may use fractions), averaging ___ hours per week. Use fractions if desirable.

THANK YOU. PLEASE CONTINUE ON THE NEXT PAGE

7. Nationality: ( ) United States ( ) Other. Please specify: _____

8. Is English your first language? ( ) Yes ( ) No

9. Age: _____ years

10. Sex: ( ) Male ( ) Female

11. Major(s):

_____

12. Cumulative GPA in College (to be kept confidential): _____

THANK YOU.

Please remember there should be NO TALKING

during the experiment.

Appendix C: The MicroSlo Task

Form A

MicroSlo
Task Instruction Sheet

Sign-in workstation number assigned to you: _____

Directions

Your task is to build a spreadsheet from a word problem.

Do not worry about formatting. However please try to be as accurate as you can be.

Please do not ask for help on the content of the problem or how to design your spreadsheet.

You may, however, ask for help with your computer or the mechanics of your spreadsheet program, Excel

Description of the Experiment, Questionnaire Sheet

Before completing the next page, <u>please read the problem statement below</u>. Then, answer the questions on the next page and **DO NOT** attempt to do the spreadsheet problem now.

> Your task is to build a two-year pro forma income statement for a company, MicroSlo.
>
> The company sells microwave slow cookers, for use in restaurants. The owner will draw a salary of $80,000 per year. There is also a manager of operations, who will draw a salary of $60,000 per year. The corporate income tax rate is expected to be 25% in each of the two years. Each MicroSlo cooker will require $40 in materials costs and $25 in labor costs in the first year. These numbers are expected to change to $35 and $29 in the second year. Unit sales price is expected to be $200 in the first year and to grow by 10% in the second year. There will be three sales people. Their salary is expected to average $30,000 per person in the first year and $31,000 in the second. Factory rent will be $3000 per month. The company expects to sell 3000 MicroSlo cookers in the first year. In the second, it expects to sell 3200.

Please continue to the next page.

Directions

Please answer the following questions as accurately as you can. The questions are focused on your personal feelings about your personal computer abilities. As such, there is no right or wrong answer to any particular question. The first part of each question asks you about whether or not you feel you have the ability to perform a particular function with a computer. If you answer YES to the <u>first part</u> of any question, the <u>second</u> asks you to indicate how confident you are with your ability to perform that particular function. Try not to second-guess yourself. Just answer each question based on your personal ability assessment rather than some comparison to another person. Also, try not to skip any questions or leave the answer blank.

As an example, please consider the following SAMPLE item for a YES response: (each YES response requires <u>two</u> answers)

	Not at All Confident				Moderately Confident				Totally Confident
	∧				∧				∧
I believe I have the (YES)	10	20	30	40	(50)	60	70	80	90 100
ability to save a file. NO									

As an example, please consider the following SAMPLE item for a NO response:

	Not at All Confident				Moderately Confident				Totally Confident
	∧				∧				∧
I believe I have the YES	10	20	30	40	50	60	70	80	90 100
ability to create a file.(NO)									

Please continue to the next page.

Please read each question carefully and provide your answers based on your personal feelings only.

1. I believe I have the ability to manipulate the way a number appears in a spreadsheet.	YES NO	10	20	30	40	50	60	70	80	90	100
2. I believe I have the ability to use and understand the cell references in a spreadsheet.	YES NO	10	20	30	40	50	60	70	80	90	100
3. I believe I have the ability to use a spreadsheet to communicate numeric information to others.	YES NO	10	20	30	40	50	60	70	80	90	100
4. I believe I have the ability to write a simple formula in a spreadsheet to perform mathematical calculations.	YES NO	10	20	30	40	50	60	70	80	90	100
5. believe I have the ability to summarize numeric information using a spreadsheet.	YES NO	10	20	30	40	50	60	70	80	90	100
6. I believe I have the ability to use a spreadsheet to share numeric information with others.	YES NO	10	20	30	40	50	60	70	80	90	100
7. I believe I have the ability to use a spreadsheet to display numbers as graphs.	YES NO	10	20	30	40	50	60	70	80	90	100
8. I believe I have the ability to use a spreadsheet to assist me in making decisions.	YES NO	10	20	30	40	50	60	70	80	90	100

Please continue to the next page.

Do answer this Before You Begin

1. Before you begin, based on the problem statement you previously read, what do you think is the probability that YOU will make an error in this task? Ask for help if you are confused by this question. You will not receive extra credit unless you answer this question.

( ) 0.1% (1/1000) ( ) 3%   ( ) 25%   ( ) 90% or higher
( ) 0.5% (1/200)   ( ) 5%   ( ) 50%
( ) 1% (1/100)   ( ) 10%   ( ) 75%

2. If you *did tasks like this many times*, what percentage of spreadsheets do you think will contain an error? For instance, if you think you would make an error in about 10% of such spreadsheets, check 10%. Ask for help if you are confused by this question. You will receive no extra credit unless you answer this question.

( ) 0.1% (1/1000) ( ) 3%   ( ) 25%   ( ) 90% or higher
( ) 0.5% (1/200)   ( ) 5%   ( ) 50%
( ) 1% (1/100)   ( ) 10%   ( ) 75%

3. What percentage of spreadsheets developed by OTHER STUDENTS do you think will contain an error? Note that we are talking about other students, rather than about you. Ask for help if you are confused by this question. You will receive no extra credit unless you answer this question.

( ) 0.1% (1/1000) ( ) 3%   ( ) 25%   ( ) 90% or higher
( ) 0.5% (1/200)   ( ) 5%   ( ) 50%
( ) 1% (1/100)   ( ) 10%   ( ) 75%

4. On a per-cell basis, what do you think is the average probability that YOU made an error *in any given cell*. For instance, if you think that you probably made an error in 3% of all cells, select 3%. Ask for help if you are confused by this question. You will receive no extra credit unless you answer this question.

( ) 0.1% (1/1000) ( ) 3%   ( ) 25%   ( ) 90% or higher
( ) 0.5% (1/200)   ( ) 5%   ( ) 50%
( ) 1% (1/100)   ( ) 10%   ( ) 75%

5. My confidence in the *accuracy* of my spreadsheet model is:

Very low   1 2 3 4 5 6 7   Very high

6. I think that my spreadsheet is error free:

Low certainty   1   2   3   4   5   6   7   High certainty

Please note your start time.

Start time: _____.

Please work as carefully as possible. Please do not rush.

Please note your ending time.

Ending time: _____

When you are <u>finished</u>, please do the following things:
1. **Be sure you have entered you finish time above.**
2. **Please save your spreadsheet on your A: drive. The file name should be MSlow.xls. Ask for help if you need to.**
3. **Print the spreadsheet MSlow.xls.**
4. **Please continue on the next page.**

114

Please Answer these Questions

1. What do you think is the probability that YOU made an error in this task? For instance, if you think the probability that you made an error in this task was 10%, check 10%. Ask for help if you are confused by this question. You will receive no extra credit unless you answer this question.

( ) 0.1% (1/1000) ( ) 3%   ( ) 25%   ( ) 90% or higher
( ) 0.5% (1/200)   ( ) 5%   ( ) 50%
( ) 1% (1/100)   ( ) 10%   ( ) 75%

2. If you *did tasks like this many times*, what percentage of spreadsheets do you think will contain an error? For instance, if you think you would make an error in about 10% of such spreadsheets, check 10%. Ask for help if you are confused by this question. You will receive no extra credit unless you answer this question.

( ) 0.1% (1/1000) ( ) 3%   ( ) 25%   ( ) 90% or higher
( ) 0.5% (1/200)   ( ) 5%   ( ) 50%
( ) 1% (1/100)   ( ) 10%   ( ) 75%

3. What percentage of spreadsheets developed by OTHER STUDENTS do you think will contain an error? Note that we are talking about other students, rather than about you. Ask for help if you are confused by this question. You will receive no extra credit unless you answer this question.

( ) 0.1% (1/1000) ( ) 3%   ( ) 25%   ( ) 90% or higher
( ) 0.5% (1/200)   ( ) 5%   ( ) 50%
( ) 1% (1/100)   ( ) 10%   ( ) 75%

4. On a per-cell basis, what do you think is the average probability that YOU made an error *in any given cell*. For instance, if you think that you probably made an error in 3% of all cells, select 3%. Ask for help if you are confused by this question. You will receive no extra credit unless you answer this question.

( ) 0.1% (1/1000) ( ) 3%   ( ) 25%   ( ) 90% or higher
( ) 0.5% (1/200)   ( ) 5%   ( ) 50%
( ) 1% (1/100)   ( ) 10%   ( ) 75%

5. For this task, my knowledge of the spreadsheet program mechanics was

( ) Adequate for the task
( ) Barely adequate for the task
( ) Inadequate for the task

6. For this task, my knowledge of accounting was

   ( ) Adequate for the task
   ( ) Barely adequate for the task
   ( ) Inadequate for the task

7. My overall satisfaction working on this task was.

   Very low   1 2 3 4 5 6 7 Very high

8. I felt tense and uncomfortable during this task.

   Not at all true   1 2 3 4 5 6 7 Very true

9. There was sufficient time to work on the task.

   Not at all true   1 2 3 4 5 6 7 Very true

10. I was interested in this task.

    Not at all true   1 2 3 4 5 6 7 Very true

11. I had a difficult time using the spreadsheet program itself.

    Not at all true   1 2 3 4 5 6 7 Very true

12. I had a difficult time with the accounting knowledge needed for the problem.

    Not at all true   1 2 3 4 5 6 7 Very true

13. It took me a good while to decide on issues.

    Not at all true   1 2 3 4 5 6 7 Very true

14. I had trouble coming to a firm decision; I often changed my mind.

    Not at all true   1 2 3 4 5 6 7 Very true

15. I had a general direction but not a specific target.

    Not at all true   1 2 3 4 5 6 7 Very true

16. I rate the spreadsheet task I worked on as

    Very easy   1 2 3 4 5 6 7 Very hard

17. My confidence in the *accuracy* of my spreadsheet model is

    Very low   1 2 3 4 5 6 7 Very high

18. I believe that my spreadsheet is error free.

    Low certainty   1 2 3 4 5 6 7 High certainty

19. Subjects who made errors on this spreadsheet were probably being sloppy.

    Not at all true   1 2 3 4 5 6 7 Very true

20. Subjects who made errors on this spreadsheet were probably poorly trained in accounting.

    Not at all true   1 2 3 4 5 6 7 Very true

21. Subjects who made errors on this spreadsheet were probably rushing.

    Not at all true   1 2 3 4 5 6 7   Very true

22. I would be embarassed to learn I made a mistake on this task.

    Not at all true   1 2 3 4 5 6 7   Very true

23. Subjects who make errors are *no more sloppy* than subjects who did not make errors.

    Not at all true   1 2 3 4 5 6 7   Very true

24. Subjects who made errors did not take this task seriously.

    Not at all true   1 2 3 4 5 6 7   Very true

25. I probably did better than the average subject in this experiment.

    Not at all true   1 2 3 4 5 6 7   Very true

26. I'm probably a better driver than most subjects in this experiment.

    Not at all true   1 2 3 4 5 6 7   Very true

27. I'm probably more careful than most subjects in this experiment.

    Not at all true   1 2 3 4 5 6 7   Very true

28. I'm probably a better planner than most subjects in this experiment.

    Not at all true   1 2 3 4 5 6 7   Very true

29. I'm probably better at accounting than most subjects in this experiment.

    Not at all true   1 2 3 4 5 6 7   Very true

30. What difference(s) do you think there is (are) between people who made an error in this task and those that did not?

When you finish, you may take a 10-minute break, if you wish.

Do not talk with other subjects about this task during the break.

Appendix D: The Wall Task

Form A

Wall
Task Instruction Sheet

Sign-in workstation number assigned to you: _____

Directions

Your task is to build a spreadsheet from a word problem.

Do not worry about formatting. However please try to be as accurate as you can be.

Please do not ask for help on the content of the problem or how to design your spreadsheet.

You may, however, ask for help with your computer or the mechanics of your spreadsheet program, Excel

Description of the Experiment,
Questionnaire Sheet

Before completing the next page, <u>please read the problem statement below</u>. Then, answer the questions on the next page and **DO NOT** attempt to do the spreadsheet problem now.

> Your own a company that builds walls. Your spreadsheet should allow you to create a bid to build a wall for a customer. The customer will have to choose between brick or lava rock.
> Both walls will be built by crews of two. Crews will work three eight-hour days to build either type of wall. The wall will be 20 feet long, 6 feet tall, and 2 feet thick. Wages will be $10 per hour per person. You will have to add 20% to wages to cover fringe benefits. Lava rock will cost $3 per cubic foot. Brick will cost $2 per cubic foot. Your bid must add a profit margin of 30% to your expected cost.

Please continue to the next page.

Directions

Please answer the following questions as accurately as you can. The questions are focused on your personal feelings about your personal computer abilities. As such, there is no right or wrong answer to any particular question. The first part of each question asks you about whether or not you feel you have the ability to perform a particular function with a computer. If you answer YES to the first part of any question, the second asks you to indicate how confident you are with your ability to perform that particular function. Try not to second-guess yourself. Just answer each question based on your personal ability assessment rather than some comparison to another person. Also, try not to skip any questions or leave the answer blank.

As an example, please consider the following SAMPLE item for a YES response: (each YES response requires two answers)

	Not at All Confident				Moderately Confident				Totally Confident

		‾∧				‾∧					‾∧
I believe I have the	YES	10	20	30	40	50	60	70	80	90	100
ability to save a file.	NO										

As an example, please consider the following SAMPLE item for a NO response:

	Not at All Confident				Moderately Confident				Totally Confident

		‾∧				‾∧					‾∧
I believe I have the	YES	10	20	30	40	50	60	70	80	90	100
ability to create a file.	NO										

Please continue to the next page.

Please Answer these Questions

Please read each question carefully and provide your answers based on your personal feelings only.

1. I believe I have the ability to manipulate the way a number appears in a spreadsheet.	YES NO	10	20	30	40	50	60	70	80	90	100
2. I believe I have the ability to use and understand the cell references in a spreadsheet.	YES NO	10	20	30	40	50	60	70	80	90	100
3. I believe I have the ability to use a spreadsheet to communicate numeric information to others.	YES NO	10	20	30	40	50	60	70	80	90	100
4. I believe I have the ability to write a simple formula in a spreadsheet to perform mathematical calculations.	YES NO	10	20	30	40	50	60	70	80	90	100
5. believe I have the ability to summarize numeric information using a spreadsheet.	YES NO	10	20	30	40	50	60	70	80	90	100
6. I believe I have the ability to use a spreadsheet to share numeric information with others.	YES NO	10	20	30	40	50	60	70	80	90	100
7. I believe I have the ability to use a spreadsheet to display numbers as graphs.	YES NO	10	20	30	40	50	60	70	80	90	100
8. I believe I have the ability to use a spreadsheet to assist me in making decisions.	YES NO	10	20	30	40	50	60	70	80	90	100

Do answer this Before You Begin

1. Before you begin, based on the problem statement you have previously read, what do you think is the probability that YOU will make an error in this task? Ask for help if you are confused by this question. You will not receive extra credit unless you answer this question.

( ) 0.1% (1/1000)   ( ) 3%   ( ) 25%   ( ) 90% or higher
( ) 0.5% (1/200)   ( ) 5%   ( ) 50%
( ) 1% (1/100)   ( ) 10%   ( ) 75%

2. If you *did tasks like this many times*, what percentage of spreadsheets do you think will contain an error? For instance, if you think you would make an error in about 10% of such spreadsheets, check 10%. Ask for help if you are confused by this question. You will receive no extra credit unless you answer this question.

( ) 0.1% (1/1000)   ( ) 3%   ( ) 25%   ( ) 90% or higher
( ) 0.5% (1/200)   ( ) 5%   ( ) 50%
( ) 1% (1/100)   ( ) 10%   ( ) 75%

3. What percentage of spreadsheets developed by OTHER STUDENTS do you think will contain an error? Note that we are talking about other students, rather than about you. Ask for help if you are confused by this question. You will receive no extra credit unless you answer this question.

( ) 0.1% (1/1000)   ( ) 3%   ( ) 25%   ( ) 90% or higher
( ) 0.5% (1/200)   ( ) 5%   ( ) 50%
( ) 1% (1/100)   ( ) 10%   ( ) 75%

4. On a per-cell basis, what do you think is the average probability that YOU made an error *in any given cell*. For instance, if you think that you probably made an error in 3% of all cells, select 3%. Ask for help if you are confused by this question. You will receive no extra credit unless you answer this question.

( ) 0.1% (1/1000)   ( ) 3%   ( ) 25%   ( ) 90% or higher
( ) 0.5% (1/200)   ( ) 5%   ( ) 50%
( ) 1% (1/100)   ( ) 10%   ( ) 75%

5. My confidence in the *accuracy* of my spreadsheet model is:

Very low   1 2 3 4 5 6 7 Very high

6. I believe that my spreadsheet is error free:

Low certainty 1 2 3 4 5 6 7 High certainty

Please note your start time.

Start time: _____.

Start Spreadsheet Task Development

(this problem is repeated for your convenience)

Please type your name in Cell A1

Your own a company that builds walls. Your spreadsheet should allow you to create a bid to build a wall for a customer. The customer will have to choose between brick or lava rock.

Both walls will be built by crews of two. Crews will work three eight-hour days to build either type of wall. The wall will be 20 feet long, 6 feet tall, and 2 feet thick. Wages will be $10 per hour per person. You will have to add 20% to wages to cover fringe benefits. Lava rock will cost $3 per cubic foot. Brick will cost $2 per cubic foot. Your bid must add a profit margin of 30% to your expected cost.

Please work as carefully as possible. Please do not rush.

Please note your ending time.

Ending time: _____

When you are __finished__, please do the following things:
1. **Be sure you have entered you finish time above.**
2. **Please save your spreadsheet on your A: drive. The file name should be Wall.xls. Ask for help if you need to.**
3. **Print the file Wall.xls**
4. **Please continue on the next page.**

Please Answer these Questions

1. What do you think is the probability that YOU made an error in this task? For instance, if you think the probability that you made an error in this task was 10%, check 10%. Ask for help if you are confused by this question. You will receive no extra credit unless you answer this question.

   ( ) 0.1% (1/1000) ( ) 3%   ( ) 25%   ( ) 90% or higher
   ( ) 0.5% (1/200)   ( ) 5%   ( ) 50%
   ( ) 1% (1/100)   ( ) 10%   ( ) 75%

2. If you *did tasks like this many times*, what percentage of spreadsheets do you think will contain an error? For instance, if you think you would make an error in about 10% of such spreadsheets, check 10%. Ask for help if you are confused by this question. You will receive no extra credit unless you answer this question.

   ( ) 0.1% (1/1000) ( ) 3%   ( ) 25%   ( ) 90% or higher
   ( ) 0.5% (1/200)   ( ) 5%   ( ) 50%
   ( ) 1% (1/100)   ( ) 10%   ( ) 75%

3. What percentage of spreadsheets developed by OTHER STUDENTS do you think will contain an error? Note that we are talking about other students, rather than about you. Ask for help if you are confused by this question. You will receive no extra credit unless you answer this question.

   ( ) 0.1% (1/1000) ( ) 3%   ( ) 25%   ( ) 90% or higher
   ( ) 0.5% (1/200)   ( ) 5%   ( ) 50%
   ( ) 1% (1/100)   ( ) 10%   ( ) 75%

4. On a per-cell basis, what do you think is the average probability that YOU made an error *in any given cell*. For instance, if you think that you probably made an error in 3% of all cells, select 3%. Ask for help if you are confused by this question. You will receive no extra credit unless you answer this question.

   ( ) 0.1% (1/1000) ( ) 3%   ( ) 25%   ( ) 90% or higher
   ( ) 0.5% (1/200)   ( ) 5%   ( ) 50%
   ( ) 1% (1/100)   ( ) 10%   ( ) 75%

5. For this task, my knowledge of the spreadsheet program mechanics was

   ( ) Adequate for the task
   ( ) Barely adequate for the task
   ( ) Inadequate for the task

6. For this task, my knowledge of accounting was

   ( ) Adequate for the task
   ( ) Barely adequate for the task
   ( ) Inadequate for the task

7. My overall satisfaction working on this task was.

   Very low   1   2   3   4   5   6   7   Very high

8. I felt tense and uncomfortable during this task.

   Not at all true   1   2   3   4   5   6   7   Very true

9. There was sufficient time to work on the task.

   Not at all true   1   2   3   4   5   6   7   Very true

10. I was interested in this task.

    Not at all true   1   2   3   4   5   6   7   Very true

11. I had a difficult time using the spreadsheet program itself.

    Not at all true   1   2   3   4   5   6   7   Very true

12. I had a difficult time with the accounting knowledge needed for the problem.

    Not at all true   1   2   3   4   5   6   7   Very true

13. It took me a good while to decide on issues.

    Not at all true   1   2   3   4   5   6   7   Very true

14. I had trouble coming to a firm decision; I often changed my mind.

    Not at all true   1   2   3   4   5   6   7   Very true

15. I had a general direction but not a specific target.

    Not at all true   1   2   3   4   5   6   7   Very true

16. I rate the spreadsheet task I worked on as

    Very easy   1   2   3   4   5   6   7   Very hard

17. My confidence in the *accuracy* of my spreadsheet model is

    Low certainty   1   2   3   4   5   6   7   High certainty

18. I believe that my spreadsheet is error free.

    Low certainty   1   2   3   4   5   6   7   High certainty

19. Subjects who made errors on this spreadsheet were probably being sloppy.

    Not at all true   1   2   3   4   5   6   7   Very true

20. Subjects who made errors on this spreadsheet were probably poorly trained in accounting.

    Not at all true   1   2   3   4   5   6   7   Very true

124

21. Subjects who made errors on this spreadsheet were probably rushing.

    Not at all true   1 2 3 4 5 6 7   Very true

22. I would be embarassed to learn I made a mistake on this task.

    Not at all true   1 2 3 4 5 6 7   Very true

23. Subjects who make errors are *no more sloppy* than subjects who did not make errors.

    Not at all true   1 2 3 4 5 6 7   Very true

24. Subjects who made errors did not take this task seriously.

    Not at all true   1 2 3 4 5 6 7   Very true

25. I probably did better than the average subject in this experiment.

    Not at all true   1 2 3 4 5 6 7   Very true

26. I'm probably a better driver than most subjects in this experiment.

    Not at all true   1 2 3 4 5 6 7   Very true

27. I'm probably more careful than most subjects in this experiment.

    Not at all true   1 2 3 4 5 6 7   Very true

28. I'm probably a better planner than most subjects in this experiment.

    Not at all true   1 2 3 4 5 6 7   Very true

29. I'm probably better at accounting than most subjects in this experiment.

    Not at all true   1 2 3 4 5 6 7   Very true

30. What difference(s) do you think there is(are) between people who made an error in this task and those that did not?

When you finish, you may take a 10-minute break, if you wish.

Do not talk with other subjects about this task during the break.

Appendix E: Final Questionnaire Sheet

Sign-in workstation number assigned to you: _____

Directions

Now that you have completed development of the 2 spreadsheets, please complete the final questionnaire based on your experiences and feelings of this development effort.

Please answer the following questions as accurately as you can. The questions are focused on your personal feelings about your personal computer abilities. As such, there is no right or wrong answer to any particular question. The first part of each question asks you about whether or not you feel you have the ability to perform a particular function with a computer. If you answer YES to the first part of any question, the second asks you to indicate how confident you are with your ability to perform that particular function. Try not to second-guess yourself. Just answer each question based on your personal ability assessment rather than some comparison to another person. Also, try not to skip any questions or leave the answer blank.

As an example, please consider the following SAMPLE item for a YES response: (each YES response requires two answers)

	Not at All Confident			Moderately Confident			Totally Confident

I believe I have the (YES) 10 20 30 40 (50) 60 70 80 90 100

ability to save a file. NO

As an example, please consider the following SAMPLE item for a NO response:

	Not at All Confident			Moderately Confident			Totally Confident

I believe I have the YES 10 20 30 40 50 60 70 80 90 100

ability to create a file. (NO)

126

Please Answer these Questions

Please read each question carefully and provide your answers based on your personal feelings only.

1. I believe I have the ability to manipulate the way a number appears in a spreadsheet.	YES NO	10	20	30	40	50	60	70	80	90	100
2. I believe I have the ability to use and understand the cell references in a spreadsheet.	YES NO	10	20	30	40	50	60	70	80	90	100
3. I believe I have the ability to use a spreadsheet to communicate numeric information to others.	YES NO	10	20	30	40	50	60	70	80	90	100
4. I believe I have the ability to write a simple formula in a spread-sheet to perform mathematical calculations.	YES NO	10	20	30	40	50	60	70	80	90	100
5. believe I have the ability to summarize numeric information using a spreadsheet.	YES, NO	10	20	30	40	50	60	70	80	90	100
6. I believe I have the ability to use a spreadsheet to share numeric information with others.	YES NO	10	20	30	40	50	60	70	80	90	100
7. I believe I have the ability to use a spreadsheet to display numbers as graphs.	YES NO	10	20	30	40	50	60	70	80	90	100
8. I believe I have the ability to use a spreadsheet to assist me in making decisions.	YES NO	10	20	30	40	50	60	70	80	90	100

Please Answer these Questions

1. What do you think is the probability that YOU made an error in <u>at least</u> one of the two tasks? For instance, if you think the probability that you made an error in <u>at least</u> one of the two tasks was 10%, check 10%. Ask for help if you are confused by this question. You will receive no extra credit unless you answer this question.

    ( ) 0.1% (1/1000)   ( ) 3%   ( ) 25%   ( ) 90% or higher
    ( ) 0.5% (1/200)   ( ) 5%   ( ) 50%
    ( ) 1% (1/100)   ( ) 10%   ( ) 75%

2. If you *did tasks like this many times*, what percentage of spreadsheets do you think <u>either</u> task will contain an error? For instance, if you think you would make an error in about 10% of such spreadsheets, check 10%. Ask for help if you are confused by this question. You will receive no extra credit unless you answer this question.

    ( ) 0.1% (1/1000)   ( ) 3%   ( ) 25%   ( ) 90% or higher
    ( ) 0.5% (1/200)   ( ) 5%   ( ) 50%
    ( ) 1% (1/100)   ( ) 10%   ( ) 75%

3. What percentage of the 2 spreadsheets developed by OTHER STUDENTS do you think <u>either</u> task will contain an error? Note that we are talking about other students, rather than about you. Ask for help if you are confused by this question. You will receive no extra credit unless you answer this question.

    ( ) 0.1% (1/1000)   ( ) 3%   ( ) 25%   ( ) 90% or higher
    ( ) 0.5% (1/200)   ( ) 5%   ( ) 50%
    ( ) 1% (1/100)   ( ) 10%   ( ) 75%

4. On a per-cell basis, what do you think is the average probability that YOU made an error in <u>either</u> task *in any given cell.* For instance, if you think that you probably made an error in 3% of all cells, select 3%. Ask for help if you are confused by this question. You will receive no extra credit unless you answer this question.

    ( ) 0.1% (1/1000)   ( ) 3%   ( ) 25%   ( ) 90% or higher
    ( ) 0.5% (1/200)   ( ) 5%   ( ) 50%
    ( ) 1% (1/100)   ( ) 10%   ( ) 75%

5. There was sufficient time to work on the tasks.

    Not at all true   1 2 3 4 5 6 7 Very true

6. The tasks requested were realistic and presented a real life example of speadsheet usage (except for the fictitious title).

   Not at all true   1 2 3 4 5 6 7   Very true

7. My confidence in the *accuracy* of my spreadsheet models is:

   Very low   1 2 3 4 5 6 7   Very high

8. I believe that my spreadsheets were error free:

   Low Certainty   1 2 3 4 5 6 7   High Certainty

9. The number of errors I probably made for <u>both</u> spreadsheet tasks were:

   0 1 2 3 4 5 6 7 8 9 10+

10. I believe I have the ability to manipulate the way a number appears in a spreadsheet:

    Worse than others   1 2 3 4 5 6 7   Better than others

11. I believe I have the ability to use and understand the cell references in a spreadsheet:

    Worse than others   1 2 3 4 5 6 7   Better than others

12. I believe I have the ability to use a spreadsheet to communicate numeric information to others:

    Worse than others   1 2 3 4 5 6 7   Better than others

13. I believe I have the ability to write a simple formula in a spreadsheet to perform mathematical calculations:

    Worse than others   1 2 3 4 5 6 7   Better than others

14. I believe I have the ability to summarize numeric in a spreadsheet:

    Worse than others   1 2 3 4 5 6 7   Better than others

15. I believe I have the ability to use a spreadsheet to share numeric information with others:

    Worse than others   1 2 3 4 5 6 7   Better than others

16. I believe I have the ability to use a spreadsheet to display numbers as graphs:

Worse than others 1 2 3 4 5 6 7 Better than others

17. I believe I have the ability to use a spreadsheet to assist me in making decisions:

Worse than others 1 2 3 4 5 6 7 Better than others

18. Did you discuss the contents of this experiment with your classmates or peers during the experiment or anytime during the last month? You will still get credit for your participation, but we would like to know if discussion(s) did occur.

( ) Yes

( ) No

19. After completing either or both spreadsheets, did you do any testing?

( ) Yes

( ) No

If YES, please indicate the amount of time spent in testing:

( ) Less than 1 minute

( ) More than 1 minute, but less than 2 minutes

( ) More than 2 minutes, but less than 5 minutes

( ) More than 5 minutes

20. Were there parts of either of the two spreadsheet tasks that you did not know how to do? If so, please comment on the nature of what you did not know.

_____

_____

COMMENTS: (Please feel free to provide feedback on items or questions on the questionnaires where you were unclear or you had some confusion).

When you finish, please give this packet to the experimenter. Don't forget to include the 2 spreadsheets you saved on <u>diskette and the printouts</u>.

Thank you for your time and participation in this experiment. Do not talk with other subjects about this task.

Appendix F: Experimenter's Notes (Introduction Script)

You are participating today in a study to better understand human characteristics in spreadsheet development. You will be asked demographic data initially and then some questions about how you feel regarding your capabilities. Then, you will be asked to develop 2 spreadsheets. Each of you has Instructions provided. These instructions outline the sequence in which you are to perform all 4 sections of this experiment. <u>Please follow the sequence precisely</u> as this is important. Once a section is completed, do not refer back to a previous section.

This research is important for three reasons. It will benefit the ITM in the CBA, the student population, and the UH research program. The professors at UH CBA strive to perform high quality, world-class research, so that we can increase the level of federal funding in the form of research grants. The UH overall receives about $324M this year in research grants; from UH's Research Office (Hawai'i Business magazine). In the past, these grants have been used to improve the computer facilities at the CBA and UH campus as well as pay for the increased Internet access speeds. Therefore, it is very important for us to do good research here today.

We have the opportunity today for a win-win situation. The CBA can do good research, and we might be able to receive future funding for new and upgraded computer facilities in the future.

What this means to you is that your full attention and participation is requested for the next several hours. Please read all materials presented, read each research question carefully and answer the research questions with your true responses. Your evaluative survey responses will be discussed in a group format in a future class. Remember, there are <u>no right nor wrong answers</u> you supply. If you get fatigued, simply rest for a minute or two, or take a break, then push forward until completion. Completed work is important.

You are asked for the identification number in the beginning of each survey. This is used only to match the survey responses. Nobody is trying to pry into your personal information. The names on the sign-in registration will be provided to your instructor for extra credit.

Any questions before we begin.

Thank you in advance for your support and participation of this experiment.

Appendix G: Experimenter's Notes on "Think-aloud"

This is the protocol analysis (or, "think-aloud") portion of the experiment. What this experiment entails is verbalizing your actions and problem-solving skills during the exercises. If you feel uncomfortable or unsure of this procedure, please let me know <u>NOW</u> as this is a <u>critical portion</u> of the experiment. Also, please confirm that you have 4 <u>separate</u> sections and a diskette.

You are to perform this experiment in the specified **SEQUENCE**. This <u>order is critical</u>.

1. Preliminary exercise in multiplying 2 numbers and "think-aloud" to become familiar with the procedures and talking into a microphone.
2. General questionnaire
3. MicroSlo Task *
4. Wall Task *
5. Final questionnaire

During this experiment, if you are confused on any of the written questions or requested information, please make a note and provide an itemized documentation in the last section under Comments.

Please feel free to on take a break at your convenience. A break between each of the spreadsheets (Sections 2 and 3) would be a good time. Please help yourself to some candy at the sign-in desk.

Remember that your answers are confidential and you are NOT graded, so please <u>do your best</u> as your input will provide invaluable research data. You will receive <u>extra credit</u> (or 3 homework assignment waived) for your participation, so **please follow instructions**.

After this experiment, please do not discuss the contents of this experiment with your classmates or peers. A debriefing will be held after all students have completed this experiment. Your cooperation is appreciated.

Thank you for your participation.

Equipment Preparations (unique to the audio capture portion)

1. Supplies – test ahead of experiment
   - Digital recorder – not set to voice-activation (to enable coordination with the clock).
   - Microphone
   - Extra batteries
   - Extra earphones to adjust volume level

2. Warming up exercise (extracted from Ericsson and Simon, 1993).

'In this experiment, we are interested in what you think about when you find answers to solving the spreadsheet exercises that I am going to ask you in the experiment. In order to do this, I am going to ask you to THINK ALOUD as you work on the problem given (there will be 2 problems). What I mean by "think aloud" is that I want you to tell me EVERYTHING you are thinking from the time you first see the problem statement until you give the final solution. I would like you to talk aloud CONSTANTLY from the time I present you the experiment papers until you have give your final answers to the problems. I don't want you to plan out (or figure out) what you say <u>or</u> try to explain to me what you are saying. Just act as if you are alone in the room speaking to yourself. It is most important that you keep talking. If you are silent for any long period of time, I will ask you to talk aloud. Do you understand what I want you to do?

A. Good. Now we will begin with some practice problems. First, I want you to multiply these two numbers in your head and tell me what you are thinking as you get an

answer.

"What is the result of multiplying 24 x 36?"

B.   Good.   Now I want to see how much you can remember the question until you gave the answer. We are interested in what you actually can REMEMBER rather than what you think you must have thought. If possible, I would like you to tell about your memories in the sequence in which they occurred while working on the question. Please tell me if you are uncertain about any of your memories. I don't want you to work on solving the problem again, but just report all that you can remember thinking about when answering the question. Now tell what you remember.

C.   Good.   Now I will give you two more practice problems before we proceed with the main experiment. I want you to do the same thing for each of these problems. I want you to think-aloud as before as you think about the question, and after you have answered it, I will ask you to report all that you can remember about your thinking. Any questions? Here is your next problem.

"How many windows are there in your parent's house?"

Now tell me all that you can remember about your thinking.

Good.   Now here is another practice problem. Please think-aloud as you try to answer it. There is no need to keep count, I will keep track for you.

"Name 20 animals."

Now tell me all that you can remember about your thinking.