"Practicing Safe Spreadsheeting" – A case study examination of spreadsheet use and the challenges and risks associated with their use in a Private Sector Business operating in Halifax, Nova Scotia (Canada)

By

Scott Laing, CPA, CMA, CHRP

A Thesis Submitted to

Saint Mary's University, Halifax, Nova Scotia

In Partial Fulfillment of the Requirements for

The Degree of Masters of Business Administration

September, 2015, Halifax, Nova Scotia

Approved: David Bateman

Associate Professor

Approved: Margaret McKee

MBA Academic Director

Date: September 14, 2015

Acknowledgements

This labour of lines would not have been possible without the support of a few people, namely my friend 'JS', who generously offered up his time and that of some of his colleagues/co-workers who he solicited as participants in this case study when he was under no obligation to do so, my colleagues from the Rowe School of Business at Dalhousie University, who afforded me the time and provided me with the necessary encouragement and motivation for me to undertake this case study during what was already shaping up to be a short summer, and Professor David Bateman (at Saint Mary's University), who provided the necessary mentorship and patience in guiding me along the path of scholarly research. Lastly, my family who spent many summer's day (once the summer finally arrived) enjoying the sun and the surf (poolside) watching me, watching them from my office window above, wishing I too was poolside.

Abstract

"Practicing Safe Spreadsheeting" – A case study examination of spreadsheet use and the challenges and risks associated with their use in a Private Sector Business operating in Halifax, Nova Scotia (Canada)

By Scott Laing

Abstract: Spreadsheets have been and continue to be one of most commonly used analytical tools by many businesses today. With the increasing pressures on businesses today associated with making sense of increasingly larger data sets, it begs the question: are they the 'right' tool to be used. The purpose of this case study is to examine how spreadsheets are being used and the risks associated with their (spreadsheet) use within a private sector organization operating in Halifax, Nova Scotia, using a subject group of participants comprised of various types of spreadsheet users working in different functional areas of the organization. The data were collected through a series of one-on-one interviews with each participant, using a standard list of questions developed specifically for this research study, that captured the participants' experiences using spreadsheets within their organization. The aim of this case study is to determine the extent of use and reliance on spreadsheets by the participants and their organization, and the challenges and risks (due to potential errors) associated with their use. Additionally this case study also looks at potential techniques and tools that may be used to mitigate the risks of spreadsheet errors and the organizations reliance on spreadsheets, as well as offering an assessment of the capabilities (skill level) of the spreadsheet users/creators.

September 14, 2015

Table of Contents

List of Figures

List of Figures - *continued*

List of Supplements

Chapter 1 - Introduction

1.1 Overview

It is not hard to find spreadsheets being used somehow in most businesses today, from managing lists to preparing budgets and financial reporting to data analytics and visualization used for management decision support. In fact, it is the exception not the norm for companies to not be using spreadsheets in some capacity. As a result it is no wonder that companies are finding errors in their spreadsheet models/templates and reports. There is a substantial amount of empirical evidence to suggest that errors are far more common in spreadsheets than companies may want to admit to and that the nature and severity of these errors should be cause for concern for most businesses since it is unlikely that many are even aware of all the risks associated with using spreadsheets (Panko, 1996, 1998, 2008; Croll, 2005, 2009; Baker, Powell & Lawson, 2007).

Does this mean it is a case of "ignorance is bliss" or is it more the lack of understanding or a symptom of the insidious nature of organization's reliance on spreadsheets where they just don't know what they don't know, which may end up being something quite the opposite of blissful once it becomes known. In other words, depending on the nature and severity of the errors intrinsic in spreadsheets and the level and number of management decisions made as a result of relying on them, there is the distinct possibility that the individuals, organizations and systems that are using and relying on them may at some point face significant adverse consequences as a result.

According to Panko's paper (2008, para. 88), titled: *What We Know About Spreadsheet Errors*, there are several things that he and other researchers have observed/learned over recent years. Panko concluded, that "all in all, the research done to date in spreadsheet development presents a very disturbing picture [and] every study that has attempted to measure errors [in spreadsheet], without exception, has found them at rates that would be unacceptable in any organization [and] …most large spreadsheets will have multiple errors, and even relatively small "scratch pad" spreadsheets will have a significant probability of error". Panko believes organizations and those within it that rely on spreadsheets for decision support are in denial as to the potential risks that spreadsheet errors pose. Panko's position is that, based on this high probability of errors in spreadsheets and their associated risks, organizations need to acknowledge this and do more to combat spreadsheet errors such as thorough testing of spreadsheets (in particular during the development stage of the spreadsheets life cycle) as a means of mitigating/eliminating errors and their potentially adverse impacts. According to Panko, challenges facing organizations in executing, on this objective are "…few spreadsheet developers have *spreadsheeting* in their job descriptions at all, and very few do spreadsheet development as their main task. In addition, because spreadsheet development is so dispersed, the implementation of policies has to be left to individual department managers. While organizations might identify critical spreadsheets and impose hard disciplines on them (Panko, 1988), this would still mean that many corporate decisions would continue to be made on the basis of questionable analyses" (Panko, 2008, para. 90).

According to Pryor (2003, para.1) "results summarized by Panko [2000] indicate that about (80%) to (90%) of spreadsheets contain significant errors". This means that at most only two (2) out of every ten (10) spreadsheets will yield the correct result. There are several other well documented cases of severe spreadsheet errors that have cost organizations from thousands to millions and even billions of dollars. According to an article written by Wailgum (2007, para. 12) spreadsheet errors and spreadsheet model [template] errors can have a significant financial impact on organizations using them. There are several cases listed in the article along with the dollar amount of their impact such as: $2.4M (US), in the case of the University of Toledo (2004) due to what they called an "internal budgeting error", $9M (US), in the case of Kodak (2005) due to a formula miscalculation error that they referred to as "an internal control deficiency", $24M (CAD), in the case of TransAlta, a big Canadian power generating company (2003) because of what they called a "cut and paste" clerical error, $1.3B (US), in the case of Fannie Mae, (2003) due to what they called "an honest mistake" at the time and a $2.6 B (US) in the case of Fidelity's Magellan fund due to the omission of a negative [-] sign. These findings are alarming and should serve as a wake up call for organizations.

To put this issue of spreadsheet risk in context, the amount of people using Excel (spreadsheet users) worldwide and the potential impact of errors that they may contain is alarming. According to a recent article in an industry blog site, titled: *The Case for QARP*, "financial models (*aka* spreadsheets) are one of the most powerful, complex and widely used computational tools on earth. Microsoft recently estimated that Excel has 500+ million active users worldwide, with 50+ million using the tool for heavy duty

decision making" (Persico, 2011, para. 5). Such an extensive use of spreadsheets makes for a high potential for error. The extent and magnitude of spreadsheet errors will be discussed further in sections 1.3 and 2.4 of this paper.

1.2 Purpose of Study

The purpose of this study is to examine and synthesize the various independent empirical research that has already been conducted and published in the domain of spreadsheet use and spreadsheet errors. Then, by applying a case study research methodology to a single private sector organization operating in Halifax, Nova Scotia (Canada), to determine if their use of spreadsheets correlates to the types of findings from these studies. The focus of this research study will be in the following four (4) primary areas: 1) understanding how spreadsheets are being used, 2) the frequency and impact of errors/risks associated with spreadsheet use, 3) the skill levels of (and training provided to) those creating/using spreadsheets experienced by this single organization, and 4) how these results compare to the findings of other independent empirical studies. This case study will also offer some theoretical propositions based on the other research studies included in the literature review. It is the hope of the principal investigator that as a result of this qualitative case study that other opportunities and questions for further research in the area of spreadsheet management will be identified.

1.3 Background

Historically, spreadsheets were designed to be used exclusively for basic business accounting, record keeping and budgeting purposes primarily by accountants. Their use has grown over the years to include more depth and breadth of use beyond accounting, to data management, data analysis and information reporting to more wide spread adoption across many business disciplines/functional areas within organizations and many types of market/industry verticals. Over the years their use has continued to proliferate and expand into many other areas of organizations to the point where spreadsheets are almost ubiquitous in today's global business environment. In the case of some organizations spreadsheets have become essential or even critical to the functioning of the organization. As a result of this reliance or dependency on spreadsheets, organizations are susceptible to risks due to the many types (nature) and occurrence (frequency) of errors that may be present in their spreadsheet models, data templates, analysis and reports. In many cases they are not even aware of the types of errors present or the severity of the risks that these errors present to them.

Much empirical research has been conducted in Europe on the frequency and types of errors that occur in spreadsheets and the reasons why they occur as well as the magnitude or severity of their impact to organizations. With the advent of Business Intelligence (BI) tools that are designed specifically for data analytics and data visualization that are better equipped to handle/manipulate large (and increasing) data sets in a variety of file formats (from a variety of data sources) is it time that businesses investigate the possible adoption and use of these BI tools to replace their use and reliance on spreadsheets. Especially

since spreadsheets are not as robust or reliable for data analytics processing as BI tools are? One of the impediments to doing this may be users comfort level with spreadsheets, the ubiquitous nature of spreadsheets due to their ease of use and sharing and relative low cost, and the fact that they have become more robust and more functional than they used to be. The following are several of the areas that will be examined in this literature review and case study: organizational use and reliance on spreadsheets, the occurrence/frequency of errors in spreadsheets as observed in other independent research studies as well as in the participant's organization, the skill level of users and training provided by organizations including the participant's organization. The Systems Development Life Cycle (SDLC) methodology will be used as a framework to review the life cycle of spreadsheets and the types of controls that can be used to help identify and mitigate the impact of errors. The outcome of this case study will also include the identification of the risks involved in using spreadsheets and guidelines for *practicing safer 'spreadsheeting'* (spreadsheet use).

The organization chosen for this case study has been operating in Halifax, NS Canada, for approximately 18 years, with more than 500 employees less than 10% of when are regular users of spreadsheets. The organization has some standards and guidelines in place for spreadsheet use; however, it is unclear whether or not everyone is using them or consistently using them since there is little to no incentive or consequence or monitoring of this.

The organization uses MS Excel (spreadsheets) exclusively as their data analytics tool and they currently have both 2010 and 2013 versions available for employees to use. The organization's representative (a manager) acknowledges that the organization is reliant on spreadsheets for many purposes and considers them to be very important as they are used in many processes and for decision support in many areas throughout the organization.

The organization's representative also acknowledges awareness of the risks associated with spreadsheet use and the value/benefit of testing spreadsheets to ensure they are working properly before using them but admit that they only do this sometimes and when they do test them they invariably find errors in ones that they or others create. The organization's representative believes that one of the biggest reasons for the spreadsheet errors is lack of testing spreadsheets prior to use due to the challenge of having to provide the results or reports under tight (short) time constraints. As such they are willing to accept the risk of there being some errors in their spreadsheets if it means that they will receive the results/information faster.

Upon completion of this case study a copy of the summarized findings will be provided to each of the participants.

1.4 Statement of Problem

Spreadsheets have been and continue to be one of most commonly used data analytical

tools by many organizations today both large and small. With the increasing pressures on business professionals of all stripes to make sense of ever increasing large data sets, it begs the question: are spreadsheets the *right* tool to be used. The purpose of this study is to examine the rewards (benefits) and risks (errors) associated with extensive spreadsheet use by various types of spreadsheet users (participants) and their skill levels, within various functional areas of the participant's organization. Additionally this study will shed some light on the magnitude and impact of spreadsheet errors on organizations, and market systems such as domestic and/or international financial markets that rely on information generated by them. It is also the aim of this study to determine the extent of their (spreadsheet) use, challenges with, and risks (possibility of errors) from their use/overuse or misuse and whether the skill level of spreadsheet users/creators is a contributing factor to the frequency and types of errors found in them. The potential techniques and tools which could help improve both the skill levels of spreadsheet users/creators and the capability and reliability of their spreadsheet models/templates will also be discussed.

In summary the specific objectives of this case study research are to:

1) Understand how spreadsheets are being used, the frequency and impact of errors/risks associated with spreadsheet use and the skill levels of (and training provided to) those creating/using spreadsheets experienced by a single organization.
2) Determine if the findings from this case study are consistent with the findings of the other empirical research studies conducted in other countries. In the same

domain areas of; spreadsheets use/overuse up to and including critical reliance on spreadsheets for operational support and/or decision support, frequency and severity of spreadsheet error occurrence, documentation and testing of spreadsheets before use/reuse/sharing, whether standard practices exist in organizations for spreadsheet design, and the skill level and training opportunities provided to spreadsheet users.

3) Analyze the findings of this study in order to provide recommendations to the case study participants as to ways they may be able to reduce the risk of errors and improve the quality of the spreadsheets being used in their organization by:
    a. developing and adopting spreadsheet documentation and design standards throughout the organization
    b. having guidelines around spreadsheet use/reuse and sharing practices,
    c. improving their controls and testing techniques on spreadsheets to mitigate/remove the occurrence and severity of some/all errors.

    These changes could potentially lower the organization's risk exposure, improve the quality of their decision making (associated with reliance on spreadsheets), improve the skill level of staff as it relates to their design/use of spreadsheets, identify/suggest alternatives to using spreadsheets models/templates in an effort to reduce their reliance on spreadsheets. The combination of which could save the organization embarrassment, frustration, time and money.

Chapter 2 - Literature Review

2.1 Purpose & Scope of Literature Review:

The purpose and scope of the literature review is to better understand the independent research that has already been conducted on spreadsheet use and spreadsheet risk with respect to the type of errors found in them and the frequency of their occurrence as well as the benefits of spreadsheets. This literature review will focus predominantly on the research studies conducted over the past several years. These studies are more recent and are more focused on spreadsheet risks associated with error identification, reasons for occurrence and type of errors found in spreadsheets as well as some of the benefits of using spreadsheets. As such they should provide a better context for the case study research that has been undertaken for this report.

In order to better understand the reasons for the occurrence of the various types of errors found in spreadsheets and the severity of those errors it is important to first understand the circumstances around the use of spreadsheets. The following issues all play a role in the frequency and severity of the errors that occur in spreadsheets: the quality (integrity and completeness) of the data, how spreadsheets are designed, how or if spreadsheets are tested before being used/reused or shared with others, the skillset and training of those who create/modify the spreadsheets (*aka* human error), and how complex spreadsheets are both in terms of the their layout and design as well as the formulas and functions that are used in them and what purpose the spreadsheets are being used for.

These circumstances will be explained further throughout the remainder of the literature review, which is broken down into the following five (5) sections for ease of organization of content:

- The inception and evolution of spreadsheets
- Recent areas of independent empirical research
- Findings of recent independent research studies
- Conclusions that can be drawn from the research
- Areas that may require further research

2.2 The Inception and Evolution of Spreadsheets

Electronic or computerized spreadsheets were first introduced in the early 1960s as a mainframe tool for business accounting. So it is no wonder that accountants are still one of the heaviest, if not the heaviest, users of spreadsheets. Many years later in 1978, a better known electronic spreadsheet tool called VisiCalc (which stood for Visible Calculator) was introduced to the micro-computer user market (rather than the mainframe market) by Daniel Bricklin, who was a Harvard Business School student at the time. The name VisiCalc was appropriate since this tool was essentially an electronic version of a spreadsheet (which in hard copy form is a large sheet of paper consisting of rows and columns where data are entered in order to perform some type of calculation e.g. sum, count, etc.), which allowed the user to do this visually or in real time. Although it had only basic functionality, VisiCalc was fairly successful for several years until a new,

more functional and easier to use tool was developed and launched in 1982, called Lotus 123. Lotus 123 contained enhanced functionality including charting capabilities and became very popular with micro-computer users. Lotus 123 was designed for mass adoption, and as a result it quickly became the new spreadsheet standard and enjoyed a very successful run throughout most of the 1980s. In 1984, Lotus, started to face some competition when Microsoft (MS) released its spreadsheet tool, Excel (for Apple computer users), but it was not until 1987 that Excel really started to erode Lotus 123's dominant market share with the release of Excel for the Windows operating system, which was the operating system for all PC devices.

By the late 1980s, with spreadsheets still growing in popularity and their functionality increasing to include more advanced capabilities and features such as a graphical user interface (GUI), improved graphics, more built-in functions and data formatting tools, other spreadsheet products, such as Borland's Quattro (Pro), entered the marketplace and started to compete with Lotus 123 and MS Excel for market share. By 1995, MS Excel had eclipsed Lotus 123, and became the spreadsheet market leader.

Concurrent with this shift in spreadsheet market dominance in the mid to late 1990s, business intelligence (BI) tools also started emerging on the scene designed to handle the larger data sets that were associated with ERP (Enterprise Resource Planning) systems and Data Warehouses (large data repositories) that were also gaining popularity at that time and starting to take hold in the marketplace by the early 2000s. These various BI tools were superior to spreadsheets not just in their ability to handle larger data sets but

also in their ability to aggregate and access other data sources either statically or dynamically and in their functionality for analyzing and visualization of the data results. Although these BI tools were superior to spreadsheets this came at a higher cost than spreadsheets and usually required a considerably higher volume of training in order for users to become proficient in their use. As a result, throughout the early 2000s, MS Excel spreadsheets continued to remain the *go to* tool for many businesses when it came to analyzing and reporting their data even with the growing availability of the of new BI tools and even new spreadsheet tools like *Morphit* and *Power Pivot*, which appeared on the scene in 2009. The primary reason for this was Excel's low cost (both in general and relative to the cost of BI tools), their availability (part of the MS Office suite of software and their convenience of use.

Fast forward to the present day (2015), there are now a myriad of Business Intelligence, (*aka* Data Analytics & Data Visualization tools) readily available for organizations to choose from which are also better at handling the increasingly larger data sets (up to and including *Big Data – peta data*) than their predecessor BI tools were. Perhaps because early BI tools were at a much higher cost it may have acted as an impediment to their wider adoption. However, due to the proliferation of cloud-based versions (Software as a Service - SaaS format) of these BI tools, they are now more accessible and more affordable (due to the lower cost of this delivery model) to businesses of all sizes. In response to this enhanced functionality and robust nature of these new more affordable and easily accessible BI tools, MS Excel, has continued to evolve and release its own new functionality (e.g. pivot tables, pivot charts, regression analysis, what-if analysis, new

add-on tools such as solver and the developer toolkits), improved usability features and the ability to analyze even larger data sets.

However, given the information age that we are now in with better more affordable data analytic tools available, it begs the questions: why are spreadsheets still being used by many companies for data analysis and decision support and are spreadsheets the best or even a viable tool to be used by businesses for this purpose. If the answer to either of these questions is "no", then why are spreadsheets still being broadly used by many/most businesses and business professionals today? At what point will they no longer be able to keep up with the larger data sets and the increasing functionality of the various BI tools, which facilitate better data connectivity, data analysis, data visualization and reporting capabilities? Will spreadsheets continue to maintain their place of usefulness in business but in a diminishing capacity or utility such that over time they are only used for simple and/or adhoc analysis/reporting when there is a time constraint to provide it? These and other similar questions will be discussed in the conclusions and recommendations section of this report.

2.3 Recent Areas of Research

The predominant focus of many of the recent research studies conducted on spreadsheets are as follows: spreadsheet usage and user behavior, spreadsheet user experience and training, spreadsheet design, spreadsheet error taxonomy, spreadsheet controls and

testing/auditing, spreadsheet risks and errors (both quantitative and qualitative) and benefits of spreadsheets among others.

Over the past several years, a substantial amount of the spreadsheet research conducted in the areas of spreadsheet controls, risks, detection, severity and type of errors has been presented to and/or collated by various meta research websites such as the European spreadsheet risks Interest Group (EuSpRIG - *pronounced "yewsprig"*), which is comprised of a consortium of representatives from a variety of European nations. According to the EuSpRIG website (2015, para. 5),

"EuSpRIG offers Directors, Managers and Professionals in all disciplines the world's only independent, authoritative and comprehensive web-based information describing the current state-of-the-art in spreadsheet risk management. EuSpRIG is the largest source of information on real-world, implementable methods for introducing into organizations processes and methods to inventory (keep records of), test, fix, document, backup, archive, compare and control the legions of spreadsheets that support critical corporate infrastructure."

Additionally, according to EuSpRIG's website (2015, para. 2),

"Research has repeatedly shown that an alarming proportion of corporate spreadsheet models are not tested or controlled to the extent necessary to meet these obligations (e.g. statutory, fiduciary, reporting and compliance obligations such as those for Sarbanes-Oxley, SEC and/or other external parties including government). Uncontrolled and untested spreadsheet models pose significant business risks, including:
• Lost revenue, profits, cash, assets & tax
• Mispricing and poor decision making due to prevalent but undetected errors
• Fraud due to malicious tampering
• Systemic financial failure, due to overdependence

Furthermore, an inability to show that spreadsheet-based business information has been subject to procedures designed to ensure it is reliable, is in itself a failure of fiduciary and regulatory compliance."

Sarbanes Oxley legislation addresses this requirement as well. More details on the risks

that spreadsheets pose for organizations and systems will be discussed in the next section of this report.

Other spreadsheet meta research sites similar to but not as extensive as EuSpRIG include:

• the Spreadsheet Research (SSR) website (a repository for research on spreadsheet development, testing, use, and technology maintained by Ray Panko of the University of Hawaii),

• the Euses Consortium (a collaboration by researchers at Oregon State University, Carnegie Mellon University, Drexel University, Penn State University, University of Nebraska and Cambridge University whose goal is to develop and investigate technologies for enabling end users to shape effective software),

• Systems Modelling Ltd. (provides their useful site for a number of links relevant for information on spreadsheet design, other sites on 'good practices', spreadsheet auditing and inspection tools, mail lists, research, testing, validation and verification, and US business modeling) and the spreadsheet engineering

• The Project Research Project (undertaken by a team of researchers at the Tuck School of Business at Dartmouth). The purpose of this three-year project (2005-2007), funded by a grant from the National Institute of Standards and Technology (NIST), was to improve the design and use of spreadsheets by individuals and organizations. The rationale behind their project was that spreadsheets and the software packages that have evolved to support their use have become one of the major tools for mathematical and statistical analysis for people at all levels of sophistication.

These are a few of the sites that have been actively researching, collecting and sharing the findings of other research studies on spreadsheet use, design and testing for a number of years now. Their existence is an indication that an awareness of the risk of inappropriate spreadsheet use exists and needs to be studied further.

For the purpose of this literature review and subsequent case study, the main areas of focus will be on the *use* of spreadsheets, the types of *errors/risks* and *controls* associated with spreadsheets, the *testing methods* that can be used, *guidelines* for spreadsheet design and practices, *benefits* of spreadsheets and the *skill level* and *training/ development* of spreadsheet users.

2.4 Findings of recent independent research studies

Spreadsheet users & usage

It is generally accepted by most, if not all, of the researchers in the field of spreadsheets that spreadsheet usage is ubiquitous in business today and that there are many different user types with certain users like professional practitioners in the disciplines of accounting and finance using them more heavily than others. It is also clear that most, if not all, researchers support the premise that spreadsheets contain error(s), the extent (frequency and type) and impact (cost and severity) of which is not known but some research studies have attempted to quantify. It is unclear however, if the general consensus among researchers is that spreadsheet usage will continue to be as extensive

over the next decade as it has been in the past decade, but it is almost certain that spreadsheet use will continue into the foreseeable future.

Murphy (2007) looked at the use of spreadsheets by practitioners and the challenges that they face such as the organization's level of reliance on spreadsheets, the quality of the spreadsheet models being used by organizations, and the responsibility of maintaining and changing requirements necessary in order to support the use and reuse of a spreadsheet over its lifespan in the organization. The paper, which is based primarily on anecdotal evidence asserts that, "Commercial use of spreadsheets raises issues well beyond the quality of individual models. The overall process of managing the use of this critical resource can have a dramatic effect on the risks to which an organisation is exposed and the value it can leverage from its investments (Murphy 2007, p. 19)." Put another way, the more reliant an organization is on spreadsheet use the more exposed they are to risks associated with spreadsheet errors whose impact may not be easily measured or felt by them until the damage is already done. This may prove in the end to have a devastating or irreversible impact on the organization.

Spreadsheet testing and error types

It is generally accepted by most, if not all, spreadsheet researchers that errors in spreadsheets are a real problem and represent a significant risk to businesses, particularly those who have a heavier reliance on spreadsheets and for those in certain industries.

For clarity purposes ***Spreadsheet risk*** needs to be defined. The following definition, retrieved from www.definedterm.com, will be used for the purposes of this report: Spreadsheet risk is, "the risk of financial losses or other adverse effects resulting from errors, omissions or duplications in a spreadsheet. Losses and other adverse effects may also result from fraud, overconfidence in the spreadsheet's results, overdependence on the spreadsheet, misinterpretation of results, failure to communicate assumptions and limitations, or failure to understand the consequences of assumptions and limitations" (para. 1).

This section reviews several of the recent independent research studies and discusses their findings in an effort to identify commonalities in the elements of recent research objectives and their findings. The spreadsheet Engineering Project (SERP) conducted by Baker, Powell and Lawson (2006) at the Tuck School of Business at Dartmouth College focused on "…improving the practice of spreadsheet engineering. Four activities are envisioned: evaluation of spreadsheet models and modeling processes in use; survey of corporate training and standards; compilation of best practices; and design of a training program" (Baker et al. 2006, p. 208).

Although the focus of their research project was on the final activity, namely the design of a training program on spreadsheet engineering for spreadsheet practitioners, which they believed would lead to improvements in the design and quality of spreadsheets, their research findings also support the generally accepted belief that spreadsheet errors pose significant risks for organizations. They concluded that better and more testing done by

those creating and using the spreadsheets are necessary to increase the quality of the spreadsheet models being used or shared throughout the organization. Quality has many different meanings, ranging from the conformance to one or more specified standards, to the overall accuracy and/or completeness of the end product. For clarity and the purposes of this literature review, spreadsheet *quality* will include the aesthetics and efficiency of the design/formatting and layout of the spreadsheet (qualitative measures), the organization and accuracy of the formulas and functions contained within it (quantitative measures), as well as the integrity of the data used by the model/template.

The first activity in the research study focused on the evaluation of spreadsheet models and consisted of three main objectives; "One objective of the spreadsheet audits is to assess the quality of design, technical correctness, and suitability for use of these models and to compare our findings with those described in the literature. A second objective is to identify the purposes for which the models were built and to determine whether they have been used for those purposes. A third objective is to learn about the process by which these models were developed and the life cycles that they have subsequently experienced" (Baker et al. 2006, p. 208).

Based on the results of the survey by Baker et al. (2006, p. 6), a large percentage of the respondents (over 80%) spent 10% or less of their time testing spreadsheets they were using and just half (approx. 50%) of the respondents said they 'usually' or 'always' tested spreadsheets that they either created or used. These findings suggest that a large number of spreadsheets are not being audited or tested for errors prior to their use. As a result of this lack of testing the spreadsheets may pose inherent risks to organizations who use the

data analysis or results provided by them. Furthermore, many errors may go undetected for some time causing a potential prolonged and adverse effect on the organization should these spreadsheets continue to be used or shared.

Since spreadsheets are often shared with others in the organization and typically have a longer than initially anticipated lifespan (whether or not this is recommended is debatable), the issues of maintenance and security (controls) of the spreadsheet become important in order to mitigate the risk of errors perpetuating throughout the organization throughout the life span of the spreadsheet. To this end, Vlootman and Hermans (2013, p.1) considered these factors in their research on spreadsheets and developed a "…checklist aimed at measuring the maintainability of a spreadsheet" where they created a series of questions and grouped them into several categories (e.g. structure, formatting, skills/functions used, etc.) to evaluate and assess the safety of sharing/reusing a spreadsheet.

An earlier study by Hermans, Pinzger and Deursen (2012, p.1) also found that spreadsheet understandability was of significant importance if the spreadsheet was going to be reused and or shared with others in the organization so as part of their study they proposed a set of spreadsheet understandability metrics. It was their hypothesis that in order for spreadsheet users to be able to assess the quality of spreadsheets they first needed to be able to understand the spreadsheet. In their study they identified characteristics that they believed would aid in the understandability of spreadsheets allowing others to use, edit and reuse the spreadsheets more readily. They also identified

characteristics that contributed to the misinterpretation of spreadsheets, such as "the number of ranges (cell ranges), the nesting depth (length) and the presence of conditional operations (IF tests) in formulas." It would appear from their study's findings that the simpler and more clear a spreadsheet's design and formulas are, the easier it is for users to understand its functionality and the higher its quality (lower rate of errors). Although this may not always be possible or practical, 'clarity' and 'quality' are a good place to start when considering how to design a spreadsheet, especially if it is intended for use by others who may not be as familiar with spreadsheets.

The findings of a survey conducted by Coster, Leon, Kalbers and Abraham (2011, p.10) on organizational controls over spreadsheets for financial reporting in practice found that "there are problems in all stages of a spreadsheets life cycle and suggested several important areas for future research." It also found that "companies continue to use spreadsheets for financial reporting… [and] even with such a strong incentive for companies to have strong controls [due to the Sarbanes Oxley Act of 2002], many weaknesses in controls exist". As a result one can deduce that where controls are lacking or absent the frequency of error occurrence may be higher.

Wu (2011, p.1) discusses ways that the finance function of organizations can "improve spreadsheet controls …and start managing the risks of errors in key spreadsheets by strategically selecting controls that complement existing user practice". It is Wu's position that the use of spreadsheets in the *finance* functional area of organizations is not likely to slow down anytime soon. Consequently, Wu recommends that, to mitigate some of the impact of possible errors in spreadsheets that are used (and created) by those in the

finance area, implementing simple yet effective spreadsheet controls…such as data validity checks, clear data placement and labels and display of constants, should be considered. Given that it is generally accepted that spreadsheets are used by many individuals in finance/accounting roles in organizations and given the nature and sensitivity of the financial information that these individuals are reporting, the use of spreadsheets for this purpose may pose an even higher risk to organizations.

Rittwegera and Langan (2010, p.1) discuss the controls that organizations have in place to manage spreadsheet risk and errors in the context of financial reporting. The findings of their study support the findings of Panko (1998), "that errors occur frequently in spreadsheets and that there is little or unenforced [inadequate] controls employed." However, their research findings suggest that "attitudes are changing with regard to spreadsheet risk", and that organizations are becoming more aware of the perils of spreadsheet use without proper controls and are developing policies on the development and control of spreadsheets. Additionally, in their study they also reference the different error taxonomies developed by others: such as, "Panko and Halverson (1996) who created a taxonomy with several distinctions... firstly, it distinguishes between quantitative errors which give a wrong number immediately and qualitative errors which are likely to lead to wrong numbers later. Secondly, based on Allwood's (1984) work in mathematics, the taxonomy distinguishes between mechanical errors, logical errors in creating formulas and omission errors; which is the result of leaving something out of a model." or possibly an incorrect interpretation of the numbers. The findings of their study also support the previous research findings by Panko (1998) that "spreadsheet usage is high and is used

extensively in financial reporting and spreadsheet risk is considered to be an important issue within organizations" (Rittwegera and Langan 2010, p. 11).

In a case study conducted by Lemon and Ferguson (2010, p.1) they propose that spreadsheets are used extensively in todays organizations and are here to stay and that "although spreadsheets have many benefits [to organizations] they can also represent a significant risk exposure, requiring appropriate management". Their case study discusses a practical and pragmatic approach that was recently taken by a large global organization to manage their spreadsheet risk and the authors propose that a similar approach could be scaled and customized to meet the requirements of different organizations. The approach the 'client' in the case study took to manage their risk of spreadsheet errors was to develop a spreadsheet control framework (end-user based) that they used to define the spreadsheet risks and the associated controls that should be considered for each type of risk. The old adage 'you can't manage what you don't measure' applies to this situation. Management must realize the importance of better understanding the risks of using spreadsheets in order to help them better manage these risks.

Many of the sources of spreadsheet research studies and research findings cited in this literature review have come from the EuSpRIG website, which they claim is "…the largest source of information on real-world, implementable methods for introducing into organizations processes and methods to inventory (keep records of), test, fix, document, backup, archive, compare and control the legions of spreadsheets that support critical corporate infrastructure" (2015, para. 1). According to their website, EuSpRIG's mission

is to "…bring together academics, professional bodies and industry practitioners throughout Europe to address the ever-increasing problem of spreadsheet integrity" (2015, para. 1). EuSpRIG holds frequent (annual) conferences to continue the debate on spreadsheet risks and rewards.

As previously stated, untested spreadsheet models pose significant risks to many businesses. Errors can be of an intentional or unintentional nature. According to Pryor (2004, p.1), "testing is a vital part of software development, and spreadsheets are like any other software in this respect". Although Pryor supports the systematic testing of spreadsheets and identifies several different types of testing techniques that can be used there are invariably some common problems associated with using any of them. Essentially there does not appear to be either a simple solution or a single solution to avoidance of all errors (risks) latent in spreadsheets. Techniques like testing at various levels (unit, system, regression, user acceptance) during the spreadsheet development and implementation process can be effective. Unit level testing involves testing isolated components or formulas, system level testing involves testing the final result, regression testing refers to back-testing where the new result is compared to the previous result to check for consistency and finally user acceptance testing is ensuring that the spreadsheet has everything that the user requested. One of the challenges with which method of testing to use and how to test spreadsheets is whether to automate the testing or manually test as there are pros and cons associated with either approach. A few of the studies in the foregoing compare results of automated testing tools versus manual testing for spreadsheet errors.

A paper by Kerr (2012, p.1) supports this same notion that "due to human error it is impossible to ensure a process like this (using spreadsheets to manipulate and transform data) is always error free." Kerr's paper goes on to describe "a method that has been used to improve reliability and efficiency, and reports on how it has worked in practice" (Kerr 2012, p. 1). This 'method' is reusable code (built using VBA-Visual Basic for Applications in Excel) that automates the testing of cell data contained in various Excel spreadsheets. The code is based on the organization's business rules so it can be adapted to other organizations. They also support the notion that there would be value in developing and applying a base set of 'good practices' (which should mitigate some of the contamination due to human error) and employing testing whenever spreadsheets are used.

Cost of spreadsheet errors

One of the more important questions that needs to be answered and also better understood by organizations is "what is the cost of spreadsheet errors?" and what is it relative to the cost of doing something in an effort to mitigate or possibly eliminate errors in spreadsheets. The following section will discuss the issues around the cost of known (or unknown) errors in spreadsheets and organization's reliance on them. Awareness of these costs should help justify the need for more research to be conducted in this area. Errors in spreadsheets have the potential to substantially impact both individual organizations as well as larger national and international/global systems making it significantly relevant to most if not all organizations.

In two (2) studies, Croll (2005; 2009) looked at the impact that spreadsheets had on the collapse of the banking sector in Jamaica in the late 1990s, as well as the influence they may have had on the collapse of the global financial system in 2008, and the potential risk spreadsheets pose in the contemporary financial system of the UK. What Croll found was that spreadsheets played a key role in the recent collapse of the financial system and that spreadsheets played a role in the collapse of the Jamaican financial system.

> "We have confidence in concluding that spreadsheets played a role, perhaps even a significant role, in the recent collapse of the financial system, affirming our research hypothesis. In our opinion, their primary role is centered around the fact that they were one of the principal technologies used in the Credit Derivatives marketplace [a systemic collapse of the global financial system occurred during the period 2007-2009 where credit derivatives played a significant part in the destruction of capital.]" (2009, p. 12).

Another one of the findings from this same study is consistent with the findings of many other independent empirical studies discussed in this literature review is that "…human error and other human factors, ...will remain a problem [in spreadsheets] for the future unless and until resolved" (2009, p. 12). What this illustrates is that the costs of spreadsheet errors and the inherent insidious reliance on them by organizations of all stripes in all industry verticals can range from low organizational specific impacts to extremely high impacts on a global scale and everything in between. In other words, the potential and real costs associated with spreadsheet errors are often not known (measureable) until after the problem or errors are detected and the organization/system is impacted, at which time it is often too late to fix or abate the problem. It can be likened to the difference between a 'false positive' and a 'false negative' hiring error for an organization; they are not able to easily measure the cost of the false negative (a candidate who was right for the position but who scored low on the interview or

recruitment criteria so was not hired) but they can certainly measure and *feel* the cost of hiring the false positive (a candidate who was not right for the position but who scored high on the interview or recruitment criteria so was hired). The bottom line is if it can happen to large organizations, including financial institutions and the financial systems of various countries who claim to have adequate control systems in place, then it can happen to just about any organization/system in any jurisdiction if they are relying too heavily on inadequately tested spreadsheet models. In fact, it may already be happening unbeknownst to them.

In many cases the starting position for many organizations is reliance on the results of spreadsheets even though there is substantial literature and studies that suggest they would be wise to not rely too heavily on spreadsheet results at least at their face value without performing some testing to validate the results. An empirical study conducted by Przasnyski, Leon, and Seal (2011, p.10) in the United Sates focused on the type of errors found in spreadsheets and proposed the design of a taxonomy for classifying the qualitative (design, layout) types of errors. Their findings showed that there were many qualitative (design) type errors present in the spreadsheets used by the organizations they studied and that these errors led to confusion, misinterpretation, and understandability issues and consequently affect the readiness of the spreadsheet for use by others in the organization for decision support purposes. They proposed four (4) categories of errors as follows: input data structure, semantics, extendibility and formula integrity. It was their position that "qualitative errors in spreadsheets are as serious as quantitative errors." They have developed their own proposed taxonomy for qualitative (design) errors, which they believe lead to quantitative errors during operational use of the spreadsheet. They

believe the reason these errors occur are due to design characteristics like poor identification of user input cells, incorrect cell documentation, or ambiguous cell documentation to name a few. These design characteristics may result in users entering the right data in the wrong place producing a mechanical or quantitative error. This means that everything from design errors to mechanical errors can affect spreadsheet results or the interpretation of their results.

In a research study by Mittermeir, Clermont and Hodnigg (2005, p.1) they make a distinction between two (2) main types of errors that occur in spreadsheets, ones that happen inadvertently (unintentionally) from the creation (*ab initio – from the beginning*) of the spreadsheet *aka mistakes,* and ones that happen with intent *aka faults/ fraud*. The difference between them is as follows: mistakes are easier to prevent by using various tools "…that notify the spreadsheet writer (author/editor) about potential problems whereas faults that are introduced on purpose have to be discovered by auditors without the cooperation of their originators." As a result, the faults/fraud types of errors are not only harder to detect but are also harder to protect against since not all fraud coding/syntax is known. A way to mitigate spreadsheet fraud would be to borrow from a technique used by accountants (and businesses) to mitigate fraud and that is segregation of duties where the tasks required to complete the spreadsheet in its entirety are split up and assigned to different people to complete. That way unless they are working in collusion there is less chance of fraud occurring. According to the authors, additional checks and balances like separation of the data from the spreadsheet and various inspection techniques might also help to improve the quality of the spreadsheet.

A study done by Powell, Lawson and Baker (2007, p.57) provided "…the first fully-documented evidence on the quantitative impact of errors in operational spreadsheets." They found that many of the errors observed in the 25 operational spreadsheets from 5 different organizations studied had no quantitative impact on the spreadsheet and that those errors that had an impact often affected unimportant portions of the spreadsheet but the remaining errors did "…sometimes have substantial impacts on key aspects of the spreadsheet" (Powell et al. 2007, p. 57). This study does not conclusively attest that spreadsheet errors have little or no impact only that errors were present and in some cases had a substantial or severe impact on the key sections of the spreadsheet. Although they concluded from their study that devastating errors are rare (based on the spreadsheets that they reviewed) they did caution that their conclusions should not be taken as 'proven' and are simply suggested hypothesis that would warrant future research.

In an empirical study conducted by Bishop and McDaid (2007, p.165) on spreadsheet error detection and correction, comprised of a group of end-users consisting of both professional practitioners *(aka experts)* and students (*aka novices*) and found that "the professional [end users] significantly outperformed student [end users] in correcting certain error types" and that "a strong correlation exists between the percentage of cells inspected and the number of errors corrected." They also found that overall "professionals (experts) are more efficient and effective spreadsheet debuggers than students (novices)" (Bishop et al. 2007, p.173). This may also be due to the more experienced users spending more time testing for and correcting errors.

There are a series of different studies that were done over the past several years in the area of spreadsheet error testing and the various methodologies and techniques used for error testing. One such study conducted by Aurigemma and Panko (2010, p.11) had human subjects look for "seeded" errors in spreadsheets and then their success rates were compared to the success rates of error-flagging by spreadsheet static analysis tools applied to the same spreadsheets. There are a number of ways to detect errors in spreadsheets including; testing, inspection, and static analysis tools (SAT). The study did not include auditing as a testing technique because auditing is more of a statistical sampling approach to finding nonconformities rather one that tests for all possible types of errors. The results of their findings were as follows: "Human subjects detected only 48% of the errors, while the software programs found only 0.25% (one quarter of a percent)." What this study found was that overall human error checking success far surpassed that of the software programs used in the study to test for errors albeit they still only found approximately half of the seeded errors in the spreadsheets.

A small localized study by Balson (2010, p.1) found that it was unlikely that errors could be eliminated regardless of what methodologies were enlisted to prevent or detect them so they (errors) needed to be managed in some other way and they found that "…the biggest driver of spreadsheet quality was found to be user attitudes", which may be affected by training (or lack thereof) and/or the presence of organizational guidelines around spreadsheet quality that are supported by management. "This experience suggests that user attitudes can be effectively addressed with attitudinal guidelines, provided they are strongly supported by management and reinforced with skills training and support.

Guidelines are not a substitute for training and discipline, but they provide relevance for safety techniques, and they also appear to make users more receptive to improving their skills, leading to more effective training" (Balson 2010, p. 5).

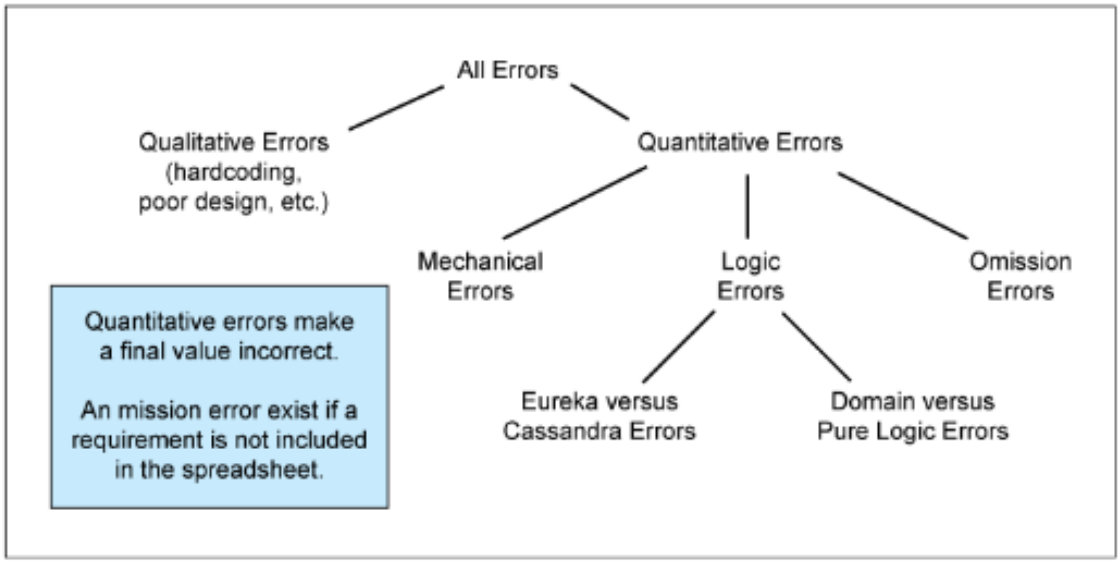
This finding is interesting given that anecdotally one might expect other reasons such as; user skill level, specific training provided, time constraints given to prepare spreadsheets, unclear or ambiguous requirements and/or assumptions made about the design and purpose of the spreadsheets, the quality of the data being used and basic human error may be more likely candidates that cause the unintentional types of errors. Although we know that attitudinal/behavioral changes take time, studies have shown that they can be altered (sometimes by modifying one in order to influence the other). With the proper level of management support, the right incentives and appropriate training and enforcement, the targeted behaviors and attitudinal shift can result in the desired outcomes of more thorough testing and improvements in design standards yielding higher quality spreadsheets.

Panko has conducted many research studies and collected a considerable amount of empirical data in the areas of spreadsheet error types and spreadsheet testing techniques. In one of his many studies, Panko (2007, p.69) stated that "…both academics and practitioners generally have ignored the rich findings produced by a century of human error research [that suggest ways to reduce errors]." Panko further states that among the key conclusions from the extensive amount of human error research that has been conducted are the following; "that thinking is bad, that spreadsheets are not the cause of spreadsheet errors, and that reducing errors is extremely difficult" (Panko 2007, p. 69). In the preceding statement that 'thinking is bad', Panko was referring to spreadsheet

designers/developers (*humans*) making errors the more they have to think about the complexity of the problem domain associated with the spreadsheets they are building. The main points raised in Panko's 2007 study are: 1) that errors are not due to spreadsheets themselves but with the thinking being done by those who design/build them (this refers to what they think they need to do versus what they are being asked to do and the different ways that they could do it among other factors), 2) that eliminating or even reducing spreadsheet error is either theoretically/practically impossible or at best is extremely difficult, and finally, 3) that replacing spreadsheets with other software packages does not eliminate errors and may not even reduce them. Having said that, certain types of errors such as those related to data integrity, may be reduced by having a software package that has better controls and rules designed to improve data quality. Although the findings in the foregoing studies may be true in terms of human (thinking) errors, they may not hold true in terms of new analytics software packages being a better option for data analysis and reporting than spreadsheets, especially given the ever growing large data sets that businesses are analyzing from a variety of different data sources.

Pankos and Halversons (1996, p.17) Taxonomy of Error Types (see figure below) was designed "…to support quantitative research studies to demonstrate that quantitative spreadsheet errors are frequent…difficult to detect and that many spreadsheet errors are significant."

Figure 1: Panko and Halverson's (1996) Taxonomy of Error Types



Much of Panko's (1996) early research findings and original classification system still seems to hold true today based on the findings of many other subsequent research studies. However, Panko believes that moving forward it is more desirable/beneficial for academics as well as practitioners to focus their research on understanding the broader types (spectrum) of errors that can occur and the various influences or triggers that cause them to occur rather than simply relying on adhering to a static taxonomy.

In Panko's (2008) paper he describes the different types of quantitative errors found in spreadsheets,

"Panko and Halverson (1996), following Allwood (1984) also found it useful to distinguish between three types of quantitative errors. Mechanical errors are simple mistakes, such as mistyping a number or pointing to the wrong cell. Logic errors involve entering the wrong formula because of a mistake in reasoning. As noted earlier, logic error rates are higher than mechanical error rates. Logic errors are also more difficult to detect and correct (Allwood, 1984). The most dangerous type of error is the omission error, in which something is left out. Omission errors appear to be extremely difficult to detect (Allwood, 1984; Bagnara, Stablum, Rizzo, Fontana, & Ruo 1987; Woods, 1984)" (Panko 2008, para. 46).

Another research study by Panko (2008) focused on revisiting his original 1996 study, on the *Taxonomy of spreadsheet errors,* to evaluate its use and usefulness over the preceding decade and to update it given the discovery of new additional errors not previously known or tested in his original study. He found that as a result of its use over the preceding ten years, and due to the discovery of more errors from other studies conducted over the same time period, that he needed to revise his earlier taxonomy by expanding on the previous classification system used to include more types of errors. Panko's revised taxonomy will be discussed further in section 2.5 of this report.

Spreadsheet Good/Best Practices debate

A study performed by Bekenn and Hooper (2009, p.1) found that "…poor layout (design, structure) choices can compromise spreadsheet quality", where quality is inferred to mean the low occurrence of errors. Their position was that any layout (design, structure) mistakes may be avoided by being able to prevent/detect them during the initial design phase by applying some simple good practice guidelines and conventions each time one is creating/modifying a spreadsheet. This concept is also supported by Kulesz (2011, p.1) who proposes a retrospective approach to identifying a list "…of good practices for spreadsheets" when it comes to designing spreadsheets. It is his hypothesis that this list of good practices would be able to be teased out of the findings of existing studies conducted by human domain experts after cross-validating (using an evaluation loop) them against the rules implemented in a semi-automated spreadsheet workbench that would also take into account the context in which the spreadsheet is being used. His

hypothesis is still theoretical in nature as the concept of the semi-automated spreadsheet workbench has not yet been developed.

On the other side of the debate, Colver (2004, p.1) suggests there is no one "...set of 'best practices', because no such set is optimal in all spreadsheet applications." In other words, what may work for one company in one industry may not work for all companies in all industries or even all companies in any one industry. Essentially, Colver is saying that designing a "one-size fits all model" is not practical or worth the effort since it would not be universally beneficial to a large enough audience. Therefore any perceived benefit associated with developing a single model that may only work in some applications/organizations, or may be short-lived, would be outweighed by the time and cost that would be required to develop the model. Applying the concept of a cost-benefit analysis to this situation would require that a spreadsheet model have more universal adaptability and/or a longer life expectancy (benefit) in order for it to be worth the effort needed to be invested in developing the model (cost). Colver's position is that there would appear to be few spreadsheet models that would satisfy this criteria. It sounds like this side of the argument is saying to waste time on something this impractical would essentially be an exercise in futility or about as useful as Don Quixote's efforts of *tilting at windmills*. Colver's perspective is also supported by Grossman (2002) in his article, *Best practices are situation-dependent*.

According to a paper by Dunn (2010, p.157), since spreadsheets are ubiquitous and prone to error, in order to control the occurrence of these many types of errors there should be a defined list of "good practices (a set of characteristics) that a spreadsheet must possess

and as bad practices another set (of characteristics) that it (spreadsheets) must avoid." This list of "good practices" and "bad practices" should in theory be easy to assemble; however, being able to say at any point in time that any one spreadsheet is in compliance with all of the good attributes and none of the bad ones may be challenging. Having said that it is, Dunn's position that "the use of automated spreadsheet development could markedly help in ensuring and demonstrating such compliance" (2010, p. 157). One tool mentioned is the Operis Analysis Kit (OAK). They further propose a suggested list of standard attributes in their paper that all spreadsheets should have some additional desirable and non-desirable attributes of spreadsheets – see Appendix B of this report)

Spreadsheet Benefits

Regardless of the findings of the many quantitative studies conducted on errors in spreadsheets there are many compensating benefits and successes associated with the continued use of spreadsheets. Several of these areas include; increased productivity at a relatively low cost, ease of using spreadsheets given the limited training needed in order for users to be able to use them, the ability to share them across the organization (between departments and/or subsidiaries) as a collaboration tool, often being the lowest common denominator in many organizations for data aggregation (combining data from diverse source information systems). In short, spreadsheets are similar to a middleware tool in terms of their data aggregation abilities and being an easy to use data analytics tool for many organizations.

Croll (2012, p.77) investigated the long term survival rates of some small but representative samples of the 30,000 largest UK limited companies who were using spreadsheet and associated decision analysis software and found "that there is a material and statistically significant increase in the long term survival rate of all of these groups of companies compared to the control." In other words the companies in the study that used spreadsheets and/or associated spreadsheet software had an improved survival rate over those that did not, showing that there was clearly a correlated benefit between a company's going concern value (longevity) and their use of spreadsheets. One could infer that one of the primary benefits of spreadsheet use is their ability to aid organizations in competitive survival by providing management with timely and relevant information to support decision making.

Skill level/Training of spreadsheet users

Several studies suggest that spreadsheet user skill level or training provided has an impact on the quality of the spreadsheet design and resulting analysis. A small study (using a pilot questionnaire) by Chadwick (2007, p. 197) proposes that because of the high frequency of errors in spreadsheets and the importance of spreadsheet use in industry and academia that "spreadsheet training courses should specifically address risk management in the development process both from a generic and a domain-specific viewpoint." Their research study specifically focused on generic issues of risk management that should be present in a training course that attempts to meet 'good-practice' within industry. The pilot questionnaire used in the study was designed to evaluate the necessary criteria for a

good spreadsheet training course that addresses risk management and good spreadsheet practices within industry. The study's findings show that the following two (2) criteria scored among the highest results in terms of what participants (comprised of a small group comprised of private trainers, academic trainers and industry practitioners) believe should be included in a good spreadsheet training course: 1) documentation within the spreadsheet itself and, 2) built-in audit functions e.g. those integral to Excel, Password mechanisms [for security]. The study's findings also show that "The pilot questionnaire [used in their study] has given an indication of some of the generic skills of spreadsheet risk management that need to be included in a good-practice training course" (2007, p. 200). Chadwick, states that, "these findings will in turn be used to establish a set of criteria for defining 'good-practice' in the training of spreadsheet risk management wherever this may occur" (2007, p. 200). Chadwick's results study are consistent with the findings of other studies which support the need for companies to adopt or make improvements in the standards and/or practices they are using for the design, documentation and sharing of spreadsheets.

Much of the research aggregated by the Euses Consortium, whose goal it is to develop and investigate technologies for enabling End Users to Shape Effective Software, focuses on the various testing techniques and tools available for testing software (in which they include spreadsheets). These techniques enable spreadsheet users to test spreadsheets that they create and/or use, and automatically detect errors (although not necessarily all possible errors), mitigating the risks of relying on spreadsheets which are generally accepted to contain at least some errors. In addition to testing for and detecting

spreadsheet errors some of these testing tools can also be used for other improvements to spreadsheets. One study conducted by Chambers and Scaffidi (2010, p.1), "…revealed several opportunities to improve spreadsheet editors, including developing different modes for spreadsheet creation, improving support for spreadsheet reuse, and helping users to find and use features." They are saying that spreadsheet editors could be improved by temporality hiding features allowing users to just focus on key features used often as a way of standardizing spreadsheet design and improving skill levels. By allowing users to develop expertise using only specific standardized features until they are more comfortable learning/using other features, or by improving support for spreadsheet reuse by providing assistance through the use of online forums, may help users overcome learning barriers and at the same time facilitate improving the quality of spreadsheets being shared within the organization. This study suggests that the Pareto principle (aka 80/20 rule) may be applicable in terms of spreadsheet features where 20% of the features are used 80% of the time.

In 2009, Panko embarked on a project to develop a framework for thinking about spreadsheets in research and corporate management. The following are highlights from the 1st round paper in which Panko proposes the following 6 level risk framework:

Figure 2: Panko (2009) proposed 6 level framework for researching spreadsheet issues

6. External Environment Level
5. Corporate Level
4. Group Level
3. Individual Level
2. Spreadsheet System Level
1. Individual Spreadsheet Level

This framework illustrates the increasing level of risk to the organization and system as you move up levels from the individual levels to the group level to the corporate level to the external reporting level (which includes government, regulatory and lender, shareholder reporting). A variation of Panko's framework above is proposed later in this paper (refer to section 6.1) that links the impact of risk to each of the respective levels, which can then be used to provide a risk assessment and associated risk management strategies for the issues faced at each level. The proposed framework (Figure 7) has condensed, Panko's 6 levels to 4 primary levels consisting of: Individual (levels 1-3), Group (level 4), Enterprise Wide (level 5), and External Environment (level 6). At each level there are different types of risks to consider, along with their impact to the organization. As previously noted in this paper there is a considerable cost associated with the risk of spreadsheet errors. Arguably there may be an even higher cost or risk to organizations for doing nothing.

Another study by Chambers, Erwig and Luckey (2010, p.8) researched a tool call SheetDiff which "...gives end-users the ability to see the changes made between versions with the click of a button and determine if the new version has been changed correctly or if there are any unexpected changes. This makes *SheetDiff* a very useful tool in a business setting and it holds the potential to help facilitate the reuse and sharing of spreadsheets." This may be helpful in reducing risk for organizations in terms of Panko's Six (6) level framework, because as spreadsheets are shared between the levels particularly between levels 3 – 4 – 5, the risk exposure to the organization increases. In

general, as spreadsheet sharing increases throughout the enterprise so too does the risk of errors and their impact on the organization.

Spreadsheet Data Quality/ Integrity

The research studies discussed in this section focus on the quality of the data that is used by spreadsheet users and what effect it can have on the resulting spreadsheet analysis.

Systems Modelling Limited is a site that contains a number of links related to studies and information on spreadsheets related to risks, design, training courses offered on spreadsheet use, 'best practices' and productivity as well as links to other sites for 'good practices', spreadsheet auditing and inspection tools. The following is an excerpt from one of the papers submitted to EuSrpig, prepared by O'Beirne (2008, p.171). The position in O'Beirne's paper is that "much of the focus on spreadsheet quality is naturally concerned with the formulas and their integrity. While most users and their managers are worried about the problems caused by Garbage In, Garbage Out (GIGO)." What he is saying is that, since the quality of the data is as integral to the spreadsheet results as the spreadsheet itself there should be a considerable amount of effort also extended to ensure that the integrity and quality of the data being used in spreadsheet analysis or reporting is good to begin with (clean, *aka* error free and complete). One way to improve data quality (integrity, currency and completeness) would be to improve upon (coding/programming) or create better controls in the source information system(s) so as to force data completeness or ensure referential integrity as data is entered (input), or as data records are modified or archived so that when the data is imported into or accessed by a spreadsheet it is clean (of the highest possible quality) to avoid GIGO (garbage in garbage

out) syndrome when the data was analyzed by the spreadsheet. A single version of the truth is also important to ensure that the data analyzed in the spreadsheet is the right data set (comes from the right source and has the right attributes and parameters e.g. date range, codes) so that the analysis and reports can be replicated anytime by anyone and produce the same results.

2.5 Conclusions that can be drawn from the research

For the most part, the studies reviewed produced findings that were consistent with each other, one of which was the need for the continued use of and expansion of a spreadsheet error taxonomy like the one originally designed by Panko (1996). Another general conclusion supported by many of the research studies is that many companies and participants studied relied extensively on spreadsheets even though in most cases they were aware of the many uncontrolled risks associated with doing so. This shows that there appears to be a tendency for users/companies to use what they know how to use spreadsheets and that they are either ignorant of the alternatives to spreadsheets like BI tools or are willing to accept the risks and sacrifice of accuracy for the benefits of increased productivity, convenience, relatively low cost of use and timeliness of analysis/results.

Another general conclusion from many of the quantitative research studies reviewed is that there are a variety (type) of errors both qualitative (design) and quantitative (formula) found in most, if not all, spreadsheets being used. Additionally, these studies found that many spreadsheet users (both authors and editors) were not testing or not adequately testing their spreadsheets for any/all of these errors before using or sharing them with others. This is possibly due to: 1) not being aware of the risks associated with insufficient testing, 2) time constraints in users having to perform the analysis and/or provide the results, 3) decision latency challenges that may impede completion of the work in a timely manner, and 4) possible lack of support or awareness within the organization as to the need for more testing.

Based on the research conducted by Bishop and McDaid (2007, p.173) on spreadsheet error detection behaviour by end-users, their overall results show that

"…Professionals (experts) are more efficient and effective spreadsheet debuggers than students (novices). Professional subjects outperformed student subjects in detecting and correcting errors of certain categories, namely formula errors, with a 16%-25% greater correction rate. An important finding is that a relationship exists between the percentage of critical cells inspected and the number of errors detected and corrected. In traditional software testing, predicting the reliability of software programs based on code coverage and defect density is a tried and tested method, which could possibly be applied to the spreadsheet paradigm. This study utilises a small, well-structured spreadsheet. But the question remains whether the findings can be applied to larger, poorly-structured spreadsheets. …. that experts would outperform novices in debugging regardless, but that greater variance in debugging behaviour would occur with larger, real-world spreadsheets."

Based on their findings spreadsheet size and complexity as well as user experience level appear to be critical factors in the ability of users to find and fix errors.

Many of the researchers contend that organisations need to adopt and enforce policies to control the risk of spreadsheet errors occurring. According to the research conducted by Rittweger and Langan (2010, p.72) "the biggest risk which Panko (1998) has identified is the omission error; leaving something out of a model. Thorne, (2009) confirms that there is audit software and tools to detect spreadsheet errors, however the impact is unknown regarding how effective they are in reducing spreadsheet errors. These errors can only be detected if proper SDLC's [Software Development Life Cycles] stages are adopted in organisations regarding core spreadsheets as in the case of those that fall into levels 4 - 5 of Panko's taxonomy (Panko 2009), therefore, Thorne, is of the opinion that it is impractical to have a SDLC for every spreadsheet."

The concept of using a system development life cycle (SDLC) approach to the designing and developing of all spreadsheets is not one that is generally accepted by all the researchers (Thorne, 2009; Rittweger & Langan, 2010) but many researchers in this area do believe that applying a software engineering approach regardless of methodology chosen would be beneficial when designing and developing spreadsheets (Panko, 2008; Pryor, 2004; Vlootman & Hermans, 2013; Kerr, 2012). Essentially this approach would entail aligning the types of errors with the stages of the spreadsheet life cycle that they are more likely to occur at as illustrated below by Panko (2008).

Figure 3: Panko (2008) Spreadsheet Life Cycle and Types of Errors

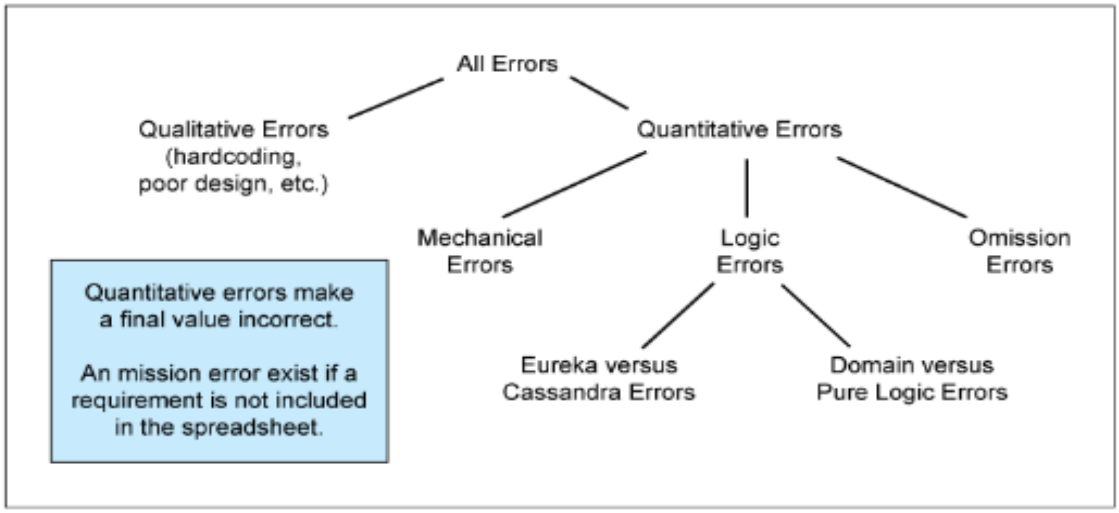
	Violations	Qualitative Errors	Mistakes	Slips and Lapses
Analysis				
Requirements Development				
Module Development				
Spreadsheet Development				
Implementation				
Operation				
Maintenance				
Termination/Replacement				

Although the above framework from Panko's 2008 study does not list all the possible types of errors that can be or have been found in spreadsheets at each stage of the spreadsheet life cycle (system [spreadsheet] development life cycle - *SDLC*) it is a good guide to build upon. This case study research report will explore further this relationship between error type and stage of SDLC and includes an expanded list of error types found in spreadsheets (including qualitative spreadsheet errors). In addition to mapping this more inclusive list of errors types (refer to the various taxonomies found in Appendix D of this report) with the stage of SLDC where they are most likely to occur it also attempts to correlate these error types with the types of internal controls that would be available to help counteract (prevent, detect and/or correct) them and offers an assessment as to the feasibility of organizations applying them to their spreadsheet models.

Over the years various error classification systems (taxonomies) have been developed and proposed by researchers, the most notable of which was Panko's (1996, 2008), original quantitative error taxonomy. This original taxonomy developed and proposed by Panko

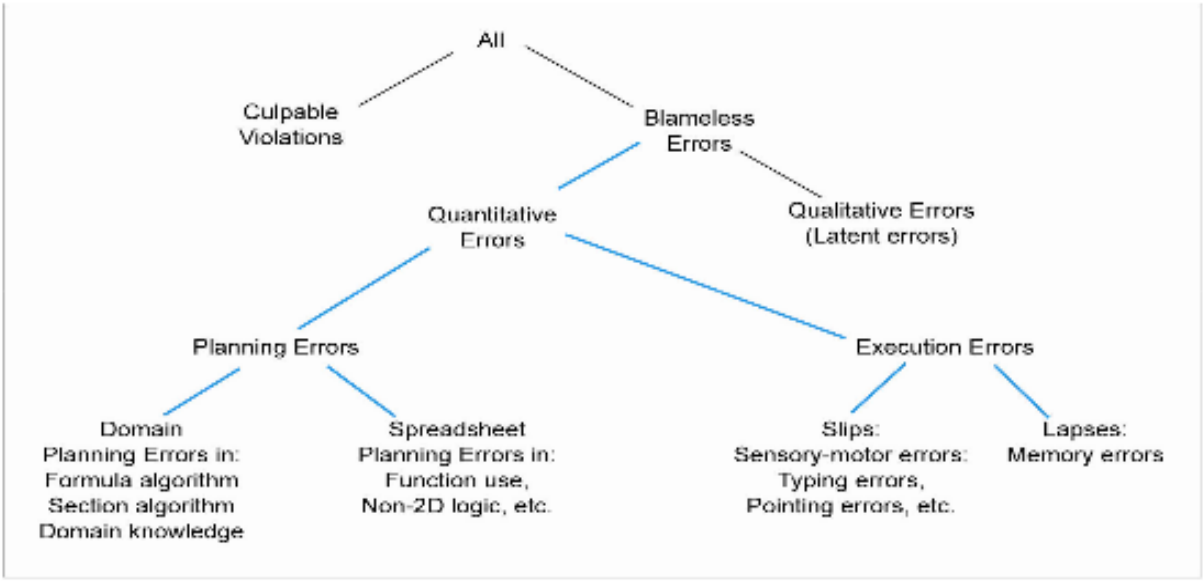
and Halverson (1996) focused on the quantitative error types and classified them in the following categories:

Figure 1: Panko and Halverson's (1996) Taxonomy of Error Types



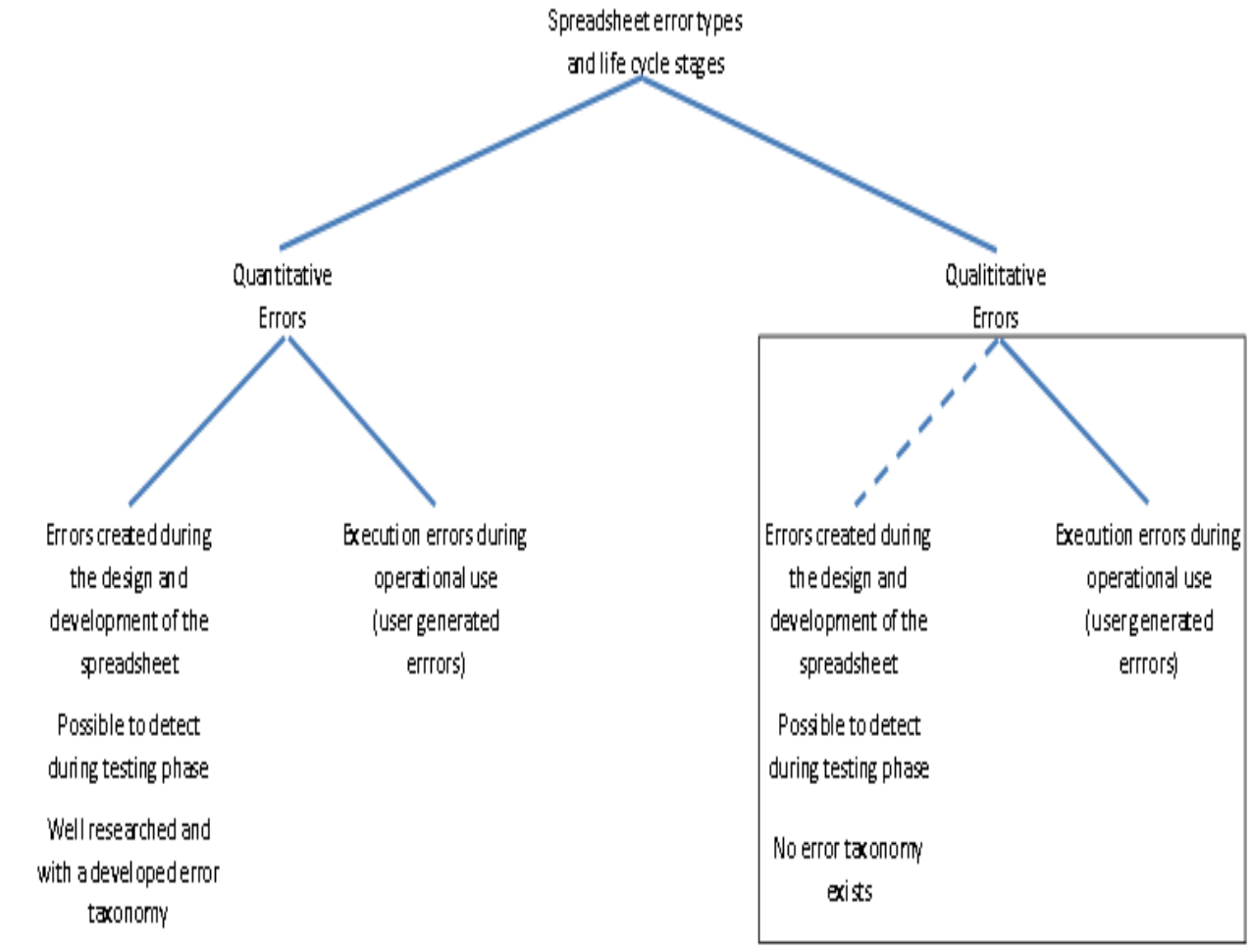
In a later revision (2008) of their original taxonomy (1996), they expanded their classification system to include more error types in the following categories:

Figure 4: Panko and Halverson's (2008) Revised Taxonomy of Error Types



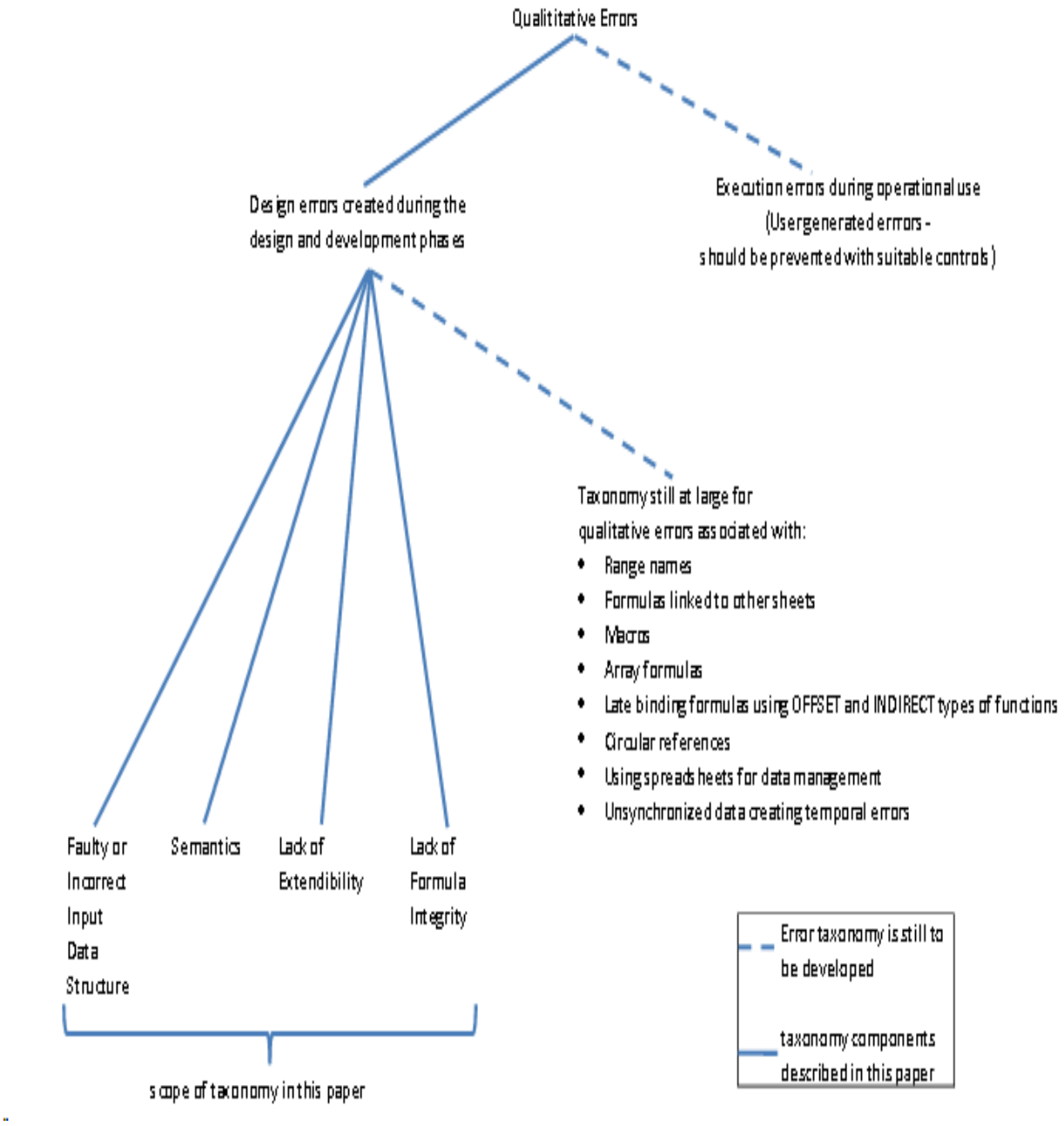
Another proposed error taxonomy that was modeled after Panko's original one was developed by Panko and Aurigemma (2010) and it included a hierarchy of both quantitative and qualitative error classifications:

Figure 5: Panko and Aurigemma (2010) Spreadsheet Error Classification - Revisited



A subsequent error classification model was developed and proposed by Przasnyski, Leon and Seal (2011) focusing only on the types of qualitative errors found in spreadsheets (see below):

Figure 6: Przasnyski, Leon and Seal (2011) Qualitative Spreadsheet Error Taxonomy



in which they proposed four (4) categories of qualitative errors as follows: input data structure, semantics, extendibility and formula integrity. These various taxonomies and other similar spreadsheet error taxonomy models from several studies (refer to Appendix

D of this report) have helped frame the categories and types of errors that have been found in spreadsheets to date and knowing this should help spreadsheet users better understand what methods/tools can be utilized to prevent them from occurring in the first place and where and how to test for others that may occur at different stages of the SDLC. Without knowing what the errors are and understanding what is causing them to occur it becomes impossible or at the very least challenging for organizations and users to develop strategies enabling them to attempt to mitigate and/or reduce the frequency of their occurrence and the severity of their impact.

There are various risk management strategies that organizations can adopt to mitigate and possibly eliminate some of the spreadsheet errors – one such strategy referenced in the paper by Coster, Leon, Kalbers and Abraham (2011, p.3) was originally proposed by PricewaterhouseCoopers, (2004) who propose that organizations should use "…a high-level five (5) step process to manage spreadsheet risk:

1. Create an inventory of spreadsheets that are in the scope of SOX regulations
2. Perform a risk assessment of financial misstatement (materiality and likelihood) by evaluating the use and complexity of the spreadsheet
3. Determine the necessary level of controls for "key" spreadsheets
4. Evaluate existing controls for each spreadsheet
5. Develop action plans for remediating control deficiencies."

This would suggest that due to the elevated importance of accountability associated with reporting requirements of regulatory bodies, compliance with Sarbanes Oxley, its

Canadian counterpart Bill 198 (*aka* CSox), and other similar regulations in other countries it is important for organizations to consider how to implement and use these different types of controls. This strategy supports both Panko's 6 level framework for spreadsheet risk (refer to Appendix D of this report) specifically in relation to moving from level 4 to level 6 of his framework, and the proposed *Figure 7* in this report (refer to section 6.1 of this report) in relation to moving between levels 3 and 4.

According to this same paper by Coster et al. (2011, p.10), they found the following,

"Our findings demonstrate that companies continue to use spreadsheets for financial reporting. However, even with such a strong incentive for companies to have strong controls, many weaknesses in controls exist. Formal policies and procedures are still lacking in most companies for most of the stages of spreadsheets. More than half, and often most, of the companies report no policy in place to describe the required qualifications for individuals who develop, modify, review, or use spreadsheets.

More formal policies and procedures that set requirements for processes and expertise for domain knowledge and spreadsheet expertise are needed, particularly in the development, review, and use stages. We note again that the weaknesses found in this study are for controls in an area that is highly regulated and visible. We would further suggest that practitioners consider and apply similar analyses to operational spreadsheets, where errors may lead to poor business decisions."

Their findings indicate that there is still a lot of work needed to be done by organizations to implement necessary spreadsheet controls to ensure both proper financial reporting as well as improve the reliability of spreadsheets used internally. This study also supports the relevance and need for Panko's 6 level spreadsheet risk framework (refer to Appendix D of this report) in particular when moving from levels 4 through level 6 as well as in the proposed *Figure 7* of this report (refer to section 6.1 of this report), when moving between levels 3 and 4. Both proposed frameworks (Figures 7 & 8) illustrate the elevation of spreadsheet risk associated with moving up the levels.

2.6 Areas that may require further research

The future of spreadsheet use is yet to be determined although it is clear from many of the researchers that it is almost certain that they expect spreadsheets to be around for the foreseeable future. However, it is less certain how they will be used in the future given the increasing awareness of the risks associated with using them in terms of the errors and potential for fraud. There are other challenges with using spreadsheets such as the increasingly larger data sets that businesses need to analyze and compliance issues related to Sarbanes-Oxley. The proliferation of other more robust data analytics tools may also have an impact on future spreadsheet use. Having said that, because spreadsheets are one of the most widely used technologies today, it seems less definite that even if spreadsheets eventually fall out of favour with many businesses that accountants, finance professionals or other heavy users of spreadsheets will stop using them anytime soon. However, spreadsheet use at some point may begin to be phased out by certain organizations or organizations in certain industries where over-reliance on spreadsheets can have more disastrous effects on their financial performance as well as that of their clients or partners/affiliates/global markets, as in the case of the studies conducted by Croll on the banking sectors. Spreadsheets may continue to be used as a tool for lower risk data analysis or estimation purposes when a "quick and dirty" answer is needed in a timely manner. Whereas, for more complicated data analysis or reporting needed to support regulatory or external reporting or senior management decision making (in particular when moving between levels 4-6 of Panko's 6 level framework for spreadsheet risk) organizations may prefer the use of more functionally powerful data analytics tools

that can handle the growing large data (big data) sets coming from their own internal systems or other external data sources.

It is not clear from the research that switching from using spreadsheets to using a Business Intelligence (BI) or other data analytics tool would solve all the problems businesses are experiencing with respect to data integrity and human errors in data model design and analysis, although it would facilitate the analysis of larger data sets and more types of data sources which spreadsheets are currently not able to manage easily. However, the underlying problems with respect to errors and data integrity errors/issues would still remain since as much of the research shows it is often human error and design flaws/assumptions that are the root cause of most spreadsheet errors.

According to Dunn (2010, p.160) "to date, EuSpRIG has concluded that search for codified best practice would be akin to search for the end of the rainbow, pointless and to be avoided. Nevertheless, there does seem to be some degree of consensus about desirable and undesirable spreadsheet characteristics. Given that EuSpRIG exists to address the ever increasing problem of spreadsheet integrity, its aims would be furthered if it is able to help those who develop spreadsheets to make the greatest use of the essential and desirable characteristics and to avoid those that are undesirable and even downright dangerous. A challenge, therefore, for practitioners and academics is; would a code of good and bad practice be useful? If so, how can compliance best be enforced [accomplished] and how can compliance or its absence best be recognised?" And finally,

who would volunteer to provide comment on such guidelines and contribute to a process which seeks an agreed code."

As to the debate on whether or not there should or could be a set of guidelines developed for good/best practices for spreadsheet design that would work for all practitioners and organization in all industries globally, there does not seem to be a clear winner yet. Therefore, an argument could be made for setting this topic area aside until such time as more future research studies can be conducted, given the polarized views held by the opposing sides of this debate.

According to Rittweger and Langan (2010, p.71), in order to gain a more in depth understanding of how organisations manage spreadsheet risk further research could be conducted that would include; "…other professionals in an organization from other functional areas of an organisation such as IT [Information Technology] and operations as well as finance professionals; to gain a more in depth understanding of how organizations manage spreadsheet risk." They propose that "…a questionnaire using a larger sample could be applied as it possesses an anonymous trait, which may result in participants disclosing sensitive information" (Rittweger and Langan 2010, p. 71). They also recommend that, "further research should consider the development of a best practice model both for the reduction in errors and to minimize the risks" (Rittweger and Langan 2010, p. 71).

Considering the afore mentioned findings, there appears to be a need for more qualitative research to be conducted in this spreadsheet domain in order to better identify and understand the areas of importance.

The next section discusses the research methodology and design used in this case study.

Chapter 3 - Research Methodology & Design

3.1 Research Objectives and justification for use of case study approach

As previously stated the research objectives of this case study will focus on the following primary areas: 1) understanding how spreadsheets are being used, 2) the frequency and impact of errors/risks associated with spreadsheet use and 3) the skill levels of (and training provided to) those creating/using spreadsheets experienced by this single organization and 4) how that compares to the findings of other independent empirical studies. The reason for using a qualitative case study research approach in this situation is due to the fact that an understanding of how spreadsheets are used and the risks (due to potential errors) that they pose for organizations and systems remains unresolved and not fully understood. This makes it a contemporary issue that deals with the "how" and the "why" organizations are still struggling with the challenges associated with their usage and reliance on spreadsheets. Several of these spreadsheet challenges include design standards, formula creation, lack of documentation, extensive sharing and reuse, omission and human errors, data integrity and skill level of creators to name a few.

Another consideration for using a case study research approach is when researchers also have little or no control over the "how" or "why" questions associated with the situation being studied [rather than the "what" used for more quantitative studies]. Yin (2014, p. 14) defines a case study as "an empirical inquiry that investigates a contemporary phenomenon (the "case") in depth and within its real-world context, especially when the boundaries between phenomenon and context may not be clearly evident." In terms of features "a case study inquiry copes with the technical distinctive situation in which there

will be many more variables of interest than data points and as one result relies on multiple sources of evidence, with data needing to converge in a triangulating fashion, and as another result benefits from the prior development of theoretical propositions to guide data collection and analysis."

The qualitative case study research process begins with a thorough literature review followed by the formulation and creation of research objectives and then the careful design of interview questions and procedures for the collection, analysis and interpretation of the data findings, concluding with the report write up. It may also lead to the identification and exploration of additional areas of research.

3.2 Research Methodology

The research design for this study is one of a case study research methodology using a single case study holistic model applied to a private sector organization operating in Halifax, Nova Scotia, Canada. The case study design included choosing a convenience sample of approximately 5-10 participants from different functional areas within the chosen organization. Next, a series of approximately 60 interview questions were developed using several questions from other previous independent studies along with a few new ones. The questions were then classified and grouped into the following categories; 1) User/Usage, 2) Risk/Controls & Errors, 3) Training, 4) Documentation & Sharing and 5) Design (qualitative) Errors that represented the various areas of focus of

the case study. After a convenience sample of management and non-management participants from a cross section of the organization's functional areas was selected, consent was obtained and individual one-on-one interviews were conducted with each participant. Each interview was approximately 25-40 minutes in duration.

The data were collected by the principal investigator through a series of one-on-one interviews with participants, where each participant was asked a standard set of interview questions designed specifically for this case study (see Appendix A for a complete listing of all the interview questions and tombstone data used in this study). The question types and corresponding answer types included in the interview varied from multiple choice, multiple answer, tabular to short answer. As each individual interview was conducted participant's responses were recorded by the principal investigator.

Limitations of this Study

There are limitations associated with the use of any case study research methodology such as methodological rigor, researcher subjectivity and external validity. According to Yin (2014, p. 17) there are five main areas of concern when doing case study research: 1) Rigor, in particular the need for the researcher to ensure that they are rigorously following systematic procedures and not allowing questionable evidence to influence their findings, 2) confusion between case study research and the use of *case studies* as a pedagogical approach to teaching, the latter of which are almost always altered in order to aid in focusing on a specific teaching point or outcome, 3) the inability to generalize the results

from a single case study to a larger population limits the usefulness of the results, 4) unmanageable level of effort, associated with case studies frequently taking a long time to complete and producing a largely unusable report and 5) unclear comparative advantage because it does not offer the same assurance of effectiveness as more experimental methods of research do.

The most notable of which according to Yin (2009, p.14) is "the absence of systematic procedures for case study research is traditionally the greatest concern due to a relative absence of methodological guidelines."

These limitations aside, over the past several years, case study research has established itself as a useful methodology for qualitative research studies, and it is for this reason and the additional ones previously listed that support the use of the case study approach in this situation.

The chosen research methodology for this case study is one of a "single case study - holistic design" (Yin 2014, p. 50). The subject group of participants will be small, therefore there are some limitations as to how the findings can be interpreted and how the results can be used. The findings of the case study are not intended to be used to make broad generalizations like the findings of a larger empirical research study could be used to do. Neither can the results be used to extrapolate on what could be expected from the findings of a larger population group. Additionally due to the small sample size of participants it would also not be possible with any level of certainty to affirm that the

sample group used is representational of the larger population group or that the findings would be the same even if a quantitative research methodology were to be used. However, it is the objective of the researcher that the findings of this case study may validate many of the other empirical research findings that have been conducted in the area of spreadsheet use and occurrence of errors and explore some of the issues in greater detail. As well it is also the objective of the researcher that this case study and its literature review will identify areas for additional/new research.

This single case study is a qualitative research study that is expected to yield similar findings as those of other empirical research studies conducted on organizations usage/reliance on spreadsheets, presence of errors and frequency of reuse/sharing of spreadsheets given the proliferation of spreadsheet use by most organizations. The reason for the selection of the organization that was chosen for this case study is due to their size and the nature of their business. By using a large multi-layered service-based organization it was the expectation of the researcher that there would be a variety of spreadsheets being used by many, if not all, of the levels of management and non-management staff for a variety of different purposes throughout the organizations many business units.

Chapter 4 - Data Collection & Findings

4.1 Data Collection

The following is an overview/background of the organization chosen for this case study; they have been operating in Halifax, NS, Canada for approximately 18 years. They have more than 500 employees, less than 10% of whom are regular users of spreadsheets. The organization has some standards and guidelines in place for spreadsheet use, however it is unclear whether or not everyone is using them or consistently using them since there is little to no incentive, consequence or associated monitoring. They have their own internal audit department, but they are not directly involved in auditing spreadsheet conformity to any organizational standards. They also have their own in-house IT support team, but they too are not involved in spreadsheet oversight or controls.

The organization uses MS Excel spreadsheets exclusively as one of their data analytic tools and they currently have both 2010 and 2013 version available for employees to use. For most of the organizations positions, they typically only require a general working knowledge (basic skill level) of the MS Office application suite consisting of Word, Excel and PowerPoint. The exception is their accounting/finance area where they require more than just a general working knowledge of Excel from those employees or new hires. Although they often include a provision for working knowledge of MS Office Suite in the skill section of many of their job postings and job descriptions they do not require or check for this when hiring new employees. The expectation or assumption of those responsible for hiring for the organization is that most candidates will have acquired

spreadsheet skills through their previous work experience, rather than through training or other post-secondary academic programs.

Training is provided by the organization at no cost to all employees on an annual basis, primarily through CBT/online training that consists of a large catalogue of courses which employees are free to choose from. MS Office suite is among them and is offered at various levels from basic to advanced. Although it is expected that all employees will use at least some of the training provided annually, there is no requirement that they use all that is offered to them or that the training they do take be in any specific area such as Excel. The organization does not anticipate there being any changes to either their hiring qualifications or training requirements as it relates to Excel skills in the foreseeable future. They acknowledge that they are reliant on spreadsheets for many purposes within the organization. They consider them very important to their organizations as they are used in many processes and for decision support in many areas throughout the organization. They also acknowledge awareness of the risks associated with spreadsheet use and the value/benefit of testing spreadsheets to ensure they are working properly before using them, but admit that they only do this occasionally. When they do test them, they invariably find errors in spreadsheets that they or others create. They believe that one of the biggest reasons for the spreadsheet errors or lack of testing spreadsheets prior to use is due to the challenge of having to provide the results or reports under tight time constraints. They were willing to accept the risk of there being some errors in the spreadsheets they use if it means that management would receive the results/information faster. This rationale is referred to as "decision latency" (Ross, 2007, para. 5) which is

the time it takes for data to be converted to information that the organization can then use for decision-making. The longer it takes to assemble and obtain the data (capture latency) and then analyze the data (analysis latency) the longer it takes to distribute the results and make a decision with it (decision latency). If this process is too long in aggregate it may impede the organization's ability to make the necessary decisions or make them in a timely enough manner as they may be time sensitive in nature.

Preventative controls, like training, typically tend to be less expensive relative to other forms of controls, like detective (auditing) or corrective (after the error or mistake has already had an impact). The types of controls and the risks associated with insufficient testing will be discussed later in the report in the conclusions and recommendations sections.

Data was collected through one-on-one interviews with each participant, during which the principal investigator, recorded each of the participant's responses to each question ensuring that each of the participants names and position titles were kept anonymous. The type of questions used were comprised of a variety of the following: multiple choice, multiple answer, tabular and a few short answer questions. The interview findings were tabulated and analyzed on both an individual question basis – as well as summarized and categorized by the nature of each question and similarity in participants responses to them in order to identify commonality/themes. A list of interview questions and their responses classified by category have been provided in the appendices to this report for reference purposes (Appendix A and B respectively).

4.2 Interview Findings

Once all the interviews were completed, the responses provided for each question were tabulated by question, and the frequency of response for each question was calculated. Where there was commonality in the answers given for the short answer questions or in the responses provided for the questions that had an 'other' option, they were grouped by general theme of response. The types and frequency of responses were then analyzed and visualized to illustrate the pattern and frequency of responses. In the following section a more thorough analysis of the findings of this case study and the results of the findings will be discussed and compared to the findings of other independent empirical research studies. Some minor themes will be drawn from the data that were collected and the analyzed and compared to the expected findings based on other research studies. Since this is a single case study with a small sample size of participants the findings cannot be considered conclusive. Finally, in the conclusions and recommendations sections of this report a new form of classification system will be proposed that synthesizes the findings and error classifications proposed by other empirical research studies with those found in this case study and the controls that may be appropriate for each error type and the feasibility of using them. As well in the recommendations section of this report, some possible areas for further consideration in terms of additional areas of research will be discussed. A summary of all of the interview question responses is provided in Appendix B of this report.

The next section of the report provides an analysis of the interview data and a summary of the results.

Chapter 5 - Analyzing Case Study Findings

5.1 Results

According to Yin (2014, p. 136) there are four (4) general strategies to use when analyzing case study evidence (findings). The one chosen for this case study is based on the reliance on the *theoretical propositions* that led to the case study in the first place, which are "…reflected a set of research questions, reviews of the literature, and new hypothesis or propositions." This case study is being used as a means of developing a theoretical proposition for a single consolidated classification system that integrates the findings/results from several published taxonomies for quantitative and qualitative spreadsheet error types, with the stages of the spreadsheet life cycle and internal controls used for spreadsheets and the operational feasibility of applying these controls.

There were several interesting findings from the data of this small sample group of participants. Based on the analysis of the interview question responses some of the observations include; 1) eighty percent (80%) of participants said that spreadsheets were either important or critically important to their job, and 2) two-thirds (67%) of participants also said that they were spending approximately fifty percent (50%) or more of their time each week working with at least ten (10) spreadsheets either on a daily or weekly basis that were authored/created by them or shared with them by others. All participants said that Microsoft Excel was the primary business tool used by the organization for a variety of purposes from maintaining lists to tracking, and analyzing data to and reporting information.

From a risk controls perspective (a measure of all 3 types of internal controls – preventative, detective and corrective), the observations were that, although there was an internal audit function in the organization, the participants did not know if they spent any time auditing spreadsheets specifically. At least half of the participants were unaware of whether or not there were any standards or policies in place within the organization for spreadsheets. All but one (1) participant admitted to finding some errors in their own spreadsheets and all participants admitted to finding errors in spreadsheets created by others. What was interesting about this particular finding is that in all cases the responses for the amount of errors the participants admitted finding in spreadsheets created by others was the same or higher than the amount that they had admitted to finding in their own spreadsheets. Eighty percent (80%) of participants cited the following reasons as the ones that they thought were the most common causes of spreadsheet errors; rushed/short timeline to get things done, not testing their spreadsheets before using/sharing them, not understanding the initial requirements requested, and inadequate spreadsheet training by those creating/editing spreadsheets. Also eighty percent (80%) of participants responded that the level of risk spreadsheets posed to their organization was of medium severity, where severity was defined as the nature of the adverse impact to the organization in monetary or other equivalent terms. They also all thought that the organization was either somewhat aware or fully aware of the risks that spreadsheet use posed to the organization. This finding is interesting because the response to the question of whether there were any standards or policies in place within the organization for spreadsheets use was that at least half of the participants were unaware of whether or not there were any. When you add this to the fact that all participants found errors in

spreadsheets created by others and in almost all cases found errors in their own spreadsheets and that testing of spreadsheets was either not done at all or only done some of the time by two-thirds of participants, this should be cause for concern for the organization.

Although it is clear that testing is being done and errors are being found, it may not be sufficient to mitigate the potential risks facing the organization associated with spreadsheet errors. Of the methods available for testing spreadsheets for errors, the ones used most commonly by the participants were using a calculator to test specific cell formula results and applying judgment or common sense when reviewing the reported results to ensure they met a reasonableness test. This could consist of a mental calculation or spot checking of the amount for accuracy, e.g., a measurement to determine the validity of an action or process. The most common response given by participants when asked how much of their time was spent checking for errors was less than twenty percent (20%). When asked about the severity of the impact of the errors that were found, the most frequent response was minor to moderate. The gauge that was used by participants to assess severity was either a dollar amount (if known) or other equivalent measure of impact on the organization.

When evaluating the skill level and training provided, participants most commonly responded that the organization offered various online courses in spreadsheet training (at various levels from basic to advanced) whereas some participants responded that they were not aware of any spreadsheet training courses offered by the organization. Having

said that though, during a follow-up interview with one of the senior human resources professionals responsible for recruitment in the organization, they stated that most positions did not explicitly state in their job qualifications that a working knowledge of spreadsheets was a requirement. However, according to them, it was also the expectation of the organization that spreadsheet skills would have been acquired by most applicants during their prior work experience, rather than through any post-secondary program, so they would have this skill already before joining the organization. All participants indicated that they had received some form of spreadsheet training course prior to working for the organization. All participants also indicated that the organization provided approximately a week of training annually to all employees in an online format where they were free to choose from a large catalogue of diverse topic areas. The organization also compensated employees with paid time off to take the training so for the most part participants were utilizing all or almost all of the training time provided to them each year as it was supported and strongly encouraged by organization that everyone participate in the training offered. Training is generally considered to be more of a preventative type of control, the cost of which for organizations may be lower than the cost of detective or corrective types of controls and/or the impact associated with spreadsheet errors that occurred due to the skill level of their creators/modifiers.

The findings of this particular series of interview questions suggest that the organization has a requirement for spreadsheet skills for many of its positions although in most cases they do not explicitly state it as a skill qualification requirement. However, they do in cases when it is position specific, such as when hiring for an accounting or clerical

position. In the case of all the participants interviewed, it is clear that they did have spreadsheet training prior to joining the organization through some type of formal or informal training course (not taken while with their current employer) which supports the organization's needs associated with their heavy reliance on spreadsheets. This is further supported by the value the organization puts on spreadsheet skills due to their heavy reliance on spreadsheet usage and that it does offer various spreadsheet training courses as part of their online course catalogue. It is also clear that many of the participants rated themselves as being at either an intermediate or expert level in terms of their spreadsheet proficiency. What is not clear from the findings is whether enough training has been provided to spreadsheet users in the specific area of spreadsheet testing and auditing techniques, which has been shown by other studies to aid in reducing the frequency and/or severity of error occurrence.

From the documentation and sharing section of the interview questions which are also considered to be more preventative types of controls, the findings indicate that all participants admit that spreadsheets are shared frequently (daily) and widely (throughout all functional areas of the organization). In most cases the entire spreadsheet is shared, either without any protection (e.g. locked cell values or password protection on either the workbook or the worksheets) or with just password protection on the spreadsheet file which would have to be given to the recipient in order for them to be able to open and edit it. As for the documentation that may accompany spreadsheets that are shared, participant responses indicated inconsistency and fell between not providing any documentation to usually providing documentation that consisted of either in-cell or in-

worksheet documentation. All participant responses indicated that they spent less than ten percent (10%) of their time with one third (33%) of participants spending no time at all) documenting their spreadsheets.

What the findings of this section of the interview questions suggest is that spreadsheet documentation is not considered an important aspect of the work of spreadsheet users, or spreadsheet users are having to skip this step in order to adhere to the tight timelines they have been given to provide the analysis/results. This may be an excuse used to mask the real root cause of some of these errors, which is the lack of importance being given to this necessary and important step of spreadsheet development by the organization. Other studies have shown that spreadsheet documentation can contribute to reducing the types of errors associated with interpretation/understanding of the spreadsheets purpose by its users. More of which will be discussed in the next section of this report.

The findings from the final section of the interview questions (spreadsheet design) which is also more of a preventative type of control indicate that participants created more spreadsheets from templates than they created from scratch with eighty-percent (80%) of participants indicating that they spent less than twenty-percent (20%) of their time creating spreadsheets from scratch. Fifty-percent of the participants said they worked independently when creating their spreadsheets, which means that if they did not document them it may be challenging for others to understand what the spreadsheet is supposed to do or how they are supposed to use it. Those that are created from scratch typically lack uniform standards. This can pose risks to the organization due to confusion

and/or misinterpretation caused by inconsistencies in layouts and formats used by the creators of spreadsheets when they share them with others in the organization. This practice may also pose other risks that can lead to a number of different types of errors if the creator or recipient makes formulae errors because of improper layout of the spreadsheet and/or incorrect/inconsistent locations of data sources or cell references. In terms of the templates that are being used in the organization it is not clear if they have all been adequately tested to ensure that they are error free. Even if they are once they are shared and edited by others there is no guarantee that they will remain error free due to lack of protection. Therefore, frequent and focused testing, as well as organization standards and practices for spreadsheet use, are important in mitigating the frequency and severity of errors.

The next section of this report compares and discusses the findings of this case study with the findings of other independent research studies. Namely, it looks at the results of the survey conducted by the Spreadsheet Engineering Research Project (SERP) in the Tuck School of Business at Dartmouth (refer to Appendix C of this report).

5.2 Analysis of Findings

One of the most well know research studies measuring similar attributes of spreadsheets and spreadsheet users as the one used in this case study was the one conducted by the Spreadsheet Engineering Research Project (SERP) in the Tuck School of Business at

Dartmouth between 2005 – 2006. Their study, which focused on spreadsheet use, error types, testing, risk management, sharing, documentation and training of spreadsheet users, had almost 1600 responses to a 67-question online survey. It included participants from seven different institutions, which provided them with a picture of spreadsheet designers and their practices. According to Tuck's researchers, their study "…may well be the largest number of responses to a comprehensive survey of its kind" (Baker, Powell and Lawson 2006, para. 2).

Similar to the Tuck survey, Excel was the predominant software tool used by the participants in this case study. Over eighty percent (80%) of Tuck participants responded that spreadsheets were either very important or critically important in their job, which is consistent with the results of this case study where eighty percent (80%) of participants also responded that spreadsheets were either very important or critically important in their job. Many Tuck participants (over 90%) also rated themselves as experienced or experts in terms of their spreadsheet proficiency, and eighty percent (80%) worked independently when creating spreadsheets, which is also consistent with the results of this case study research.

The vast majority (over 80%) of participants in the Tuck survey indicated that they shared their spreadsheets with others in the organization and that only about fifty percent (50%) tested their spreadsheets on a regular basis. These findings are again consistent with the findings of this case study, where eighty percent (80%) of participants also said they shared their spreadsheet but only thirty three percent (33%) said they tested them on a regular basis. The most common method used by participants in the Tuck survey for

testing spreadsheets was also the same common method used by the participants in this case study. It was a "common sense" approach to testing the spreadsheet for errors. This is not to say that it is the best approach only that it is a consistent approach used by participants in both studies. Based on the Tuck study, this appears to be indicative of organizations either not having standards / policies / protocols in place for spreadsheets or spreadsheet testing or that users may or may not necessarily be following them.

One of the findings of this study was the frequency of error observation by participants. The number of errors each participant found in spreadsheets created by themselves was less than fifteen (15) whereas the number of errors that each participant found in spreadsheets created by others was higher, between 16 and 50. On the whole this may seem low; however, when compared to the time most participants spent testing for errors (less than 20%) it makes more sense that the lower error detection rate is probably due to less time being spent testing for errors. This finding is supported by the findings of other independent research studies that found errors were observed in almost all, if not all, of the spreadsheets used in their respective studies. Essentially then, one might extrapolate from this case study that if the participant's organization were to increase its spreadsheet testing frequency and testing criteria (making it both broader and more in depth) they may find and correct more errors. Also, although the severity (monetary or equivalent) of the actual impact of each error detected by participants was not known (just that they were assessed by participants as being low to moderate in severity) the fact remains that they still occurred and that they and others undetected errors pose potential risks for the organization.

Another finding of this case study is that, much like the Tuck survey results, a high percentage of participants in the case study (100%) also believed that they and others in their organization are at least somewhat aware, if not fully aware, of the risks involved in using spreadsheets. Given the high percentage of risk awareness observed in the Tuck survey it is surprising that the correlation between awareness and testing effort was not more positive. Only about fifty percent (50%) of Tuck participants tested their spreadsheets on a regular basis which although low relative to the percentage who were aware of the risks of using spreadsheets, is still higher than the results of this case study which had only thirty three percent (33%) of participants testing on a regular basis.

Approximately forty percent (38.30%) of participants in the Tuck study responded that spreadsheets posed a medium risk to their organization. In this case study the response was double that, with eight percent (80%) thinking they posed a medium risk.
The reason for this study's higher percentage may be due to their slightly higher reliance on them. Approximately eighty percent (80%) responded that they were very or critically important to the organization, as opposed to seventy percent (70%) in the Tuck survey. The difference may also be due to the purpose that the participant's organizations were using them for.

The Tuck survey found that approximately two-thirds (66.40%) of participants responded that their organization had no standards for spreadsheet use; however, one third (33%) of the case study participants thought their organization did not have any standards for

spreadsheet use either. I think what this says is that many organizations don't have any standards and for those that do spreadsheet users may not always be aware of them.

In both surveys the findings were similar in terms of the frequency of sharing spreadsheets, which was either daily or weekly (by more than 50% of participants in both studies). When you combine the frequency of sharing spreadsheets with the surveys findings that show a lack of protection used in spreadsheets, which for both surveys was between fifty percent (50%) for the case study and over sixty percent (60%) for the Tuck study it can increase the risk level for organizations. When protection was used for spreadsheets, in both cases password protection was the most common form used. This all suggests that protection is not used often enough when sharing spreadsheets, which can increase the risk to organizations as the spreadsheets are shared between levels. Risk increases even more so when spreadsheets are being shared up the levels specifically between levels 3-4-5-6 (per Panko's, proposed 6 level framework for spreadsheet research found in Appendix D of this report).

Another finding that was consistent between the two studies is the amount of time participants spent documenting their spreadsheet, which is used as an aid in understanding how it works and making it easier for others to use. In both studies, the amount of time that almost all participants (97.5%-100%) spent documenting spreadsheets was less than 10%. This suggests that not enough emphasis is being placed on the importance of documentation to help mitigate the occurrence of qualitative as well as quantitative errors.

The final area of commonality between the two surveys is in the area of spreadsheet training. In the case of both surveys, over fifty percent of the participant's organizations offered some form of spreadsheet training to them, ranging from basic to intermediate to more advanced topic areas. This is a positive sign that organizations value this skill in their employees and are supporting their skill development.

The findings of this case study also suggest that participants believe that the following are the main reasons that spreadsheet errors occur; 1) short time lines to complete the analysis/report, 2) not testing the spreadsheet before using/reusing or sharing it, 3) not understanding the requirements requested for building/modifying the spreadsheet and 4) inadequate spreadsheet training. As previously noted, even given the fact that the participant's organization in this case study and those in more than half of the Tuck survey respondents cases are providing spreadsheet training to participants it appears as if still more training is needed.

The next section will synthesize and distill the findings of this case study and other independent research studies and propose two new/modified frameworks. The first of which considers assessing spreadsheet risk at different levels and the other one uses the system development life cycle stage that the error(s) are more likely to occur in as a way of applying the appropriate type(s) of controls to mitigate/eliminate them from occurring and the feasibility of applying those controls.

Chapter 6 - Case Study Research Report

6.1 Conclusions

As referenced in this report the historical evolution of spreadsheets and their use has been well documented and researched. The future role of spreadsheets is somewhat less certain. It is clear that they have a future in the sense that they are not likely to go away anytime soon. What seems uncertain at this time is what that future state will look like and what role spreadsheets will play in it. The future global business landscape will no doubt be a more competitive and regulated environment. Businesses will likely have to analyze even larger data sets than are available today, and have to choose from even more data analytics tools than are available today. The need for better, easier to use and more reliable data analytic tools will be more important to their survival than they are today.

Will spreadsheets be relegated to a lesser role in organizations and systems than they are today? Certain industries or professions will likely continue to use them. If they are to be used in future as they are being used today, then we must learn from the lessons of these various research studies in order to prevent making the same mistakes going forward.

The literature review for this study identified several issues with a few common themes namely; that spreadsheets are ubiquitous, that spreadsheet errors are equally ubiquitous, and that there are a variety of different taxonomies that can be used for classification of these errors (both qualitative and quantitative). Other issues identified were: the notion that there should be standards for spreadsheet design and development (in order to mitigate or help eliminate some errors), that more testing and auditing techniques are

what is needed to improve the quality of spreadsheets, and that user training is necessary to improve the skillset of those that create/modify spreadsheets.

In spite of all the research that has been conducted and compiled in the spreadsheet domain there still seems to be several unresolved issues, namely, what error classification system (taxonomies) should be used, what types of errors are more problematic or pose a higher risk to organizations and systems, and what is the real cost of spreadsheet errors to organizations and systems. Other unresolved issues include; which testing techniques should be used and whether any of them they can ever eliminate all spreadsheet errors, and of course the debate as to whether or not there could ever be an agreed upon single framework for best (or even good) practices for the design and development of spreadsheets. On one front there seems to be little or no debate, and that is the belief that spreadsheets will be around for some time to come. Microsoft is surely hoping that this is the case. So, we had better get used to them and figure out a way to better manage them and the risks they pose.

It is with these issues in mind that this report proposes two (2) frameworks. The first proposed framework is a classification system that links the level of spreadsheet use with an assessment of the risk impact at each level. The second proposed framework is a hybrid classification system comprised of various types of errors (both qualitative and quantitative) that aggregates the types of known errors (from various studies) with the stage of the system development life cycle (SDLC) that they are more likely to occur in. Lastly, it combines this with the type(s) of control(s) that may be applicable in dealing

with the types of risks the errors pose at each stage, along with the feasibility of the organization using these controls, to manage those risks.

Figure 7: Spreadsheet Use Level & Risk Impact Assessment - Proposed

Level	Risk Impact Assessment
4. External Environment Level	Risk Level = high; due to regulatory / governmental / shareholder / creditor reporting requirements SOX/CSOX, Government, Banks
3. Corporate Level	Risk Level = moderate to high; due to broader use/sharing/reuse, internal use but may affect longer term decision making,
2. Group Level	Risk Level = low to moderate; due to being shared, cost$ and impact of errors increases as it is shared/reused, supports decision making
1. Individual Level	Risk Level = low; due to limited internal audience / not shared, limited decision making use, limited cost$ to signing authority of user

Note: This framework is based on Panko's (2009) 6 level framework for spreadsheet risk (*see Figure 2 below)* and has been condensed to represent the primary 4 levels of impact and associated risk that spreadsheets (and their potential for errors) pose: e.g. Individual (levels 1-3), Group (level 4), Enterprise wide (level 5), External environment (level 6).

The risk assessment at each level is assessed as either Low → High and is based on cost or impact to the organization. The Level influences the potential for risk and the severity of its impact, and is based on the types of errors that may occur, the impact of those errors to the organization including the cost of those errors, the cost of not testing for errors (*aka* cost of doing nothing), cost / benefit of testing for errors, the importance of using a controls framework for spreadsheets, and the sensitivity of the information being reported.

Figure 2: Panko (2009) proposed 6 level framework for researching spreadsheet issues

6. External Environment Level
5. Corporate Level
4. Group Level
3. Individual Level
2. Spreadsheet System Level
1. Individual Spreadsheet Level

Next, a second framework is proposed. This one for the classification of error types based on the stage of the systems development life cycle (SDLC) where they are more likely to occur given the activities that occur at each stage and the nature of the error. The framework then provides various types of controls that may be used by organizations to mitigate or eliminate errors and identify them in relation to the stage of the SDLC that they may occur in. It also offers an assessment of the feasibility of the organization being able to implement the recommended controls.

The proposed framework (see Figure 8 below) acts as a guide for both the identification of the types of errors that can occur, as well as where they can occur and how to best manage their potential risk. If management is aware of the stage that they are in with their spreadsheets and the level of risk that the represent then they can use this proposed framework as a tool. It can help them improve their understanding of what causes errors to occur, where they occur and their risk level in order to better control (mitigate and manage) their impact on the organization and to improve the quality of the information their spreadsheets provide.

Figure 8: Spreadsheet Error type and Risk Management Strategy - Proposed

System Development Life Cycle (SDLC) /Error Type		Internal Controls & Feasibility Assessment of their use			
Spreadsheet (SS) SDLC Stage	**Error type (Quantitative & Qualitative)**	**Preventative**	**Detective**	**Corrective**	**Feasibility Assessment**
Planning (* added)	**Planning Errors** **Objective:** to understand the SS's intended purpose and radius and frequency of its use including its intended audience internal vs. external Incomplete parameters / requirement given – due to not understanding them or not writing them down - Changing, ambiguous or unknown requirements at the beginning	Have clear understanding of intended purpose, whether the spreadsheet will be shared/reused and who the intended audience will be and have all parties sign off on the objectives before commencing work. Stakeholders should sign-off on this before design/development begins		These will be the most likely type of controls for this stage. Review of planned and sign-off requirements before subsequent stages begin	Challenges would arise for *adhoc* requests that are of a unique nature and as such may be ambiguous or where the true purpose is not easily conveyed due to sensitivity of the information being requested. Sign-off or approvals before commencing next steps may need a further control to mitigate this. Its more of a concern/risk when spreadsheets are being shared throughout the enterprise or when results are intended for external parties – must follow stricter guidelines and in some cases regulatory (SOX, CSOX) or other reporting requirements.

Figure 8: - continued

System Development Life Cycle (SDLC) /Error Type		Internal Controls & Feasibility Assessment of their use			
Spreadsheet (SS) SDLC Stage	**Error type**	**Preventative**	**Detective**	**Corrective**	**Feasibility Assessment**
Analysis & Design (requirements development)	**Design Errors (Qualitative)** **Objective:** to standardize the layout and design of the spreadsheet, use of templates, models, tools and guidelines for design that are good → best practices. As well as policies/guidelines covering the type and format of documenting spreadsheets Layout, documentation, format, model limitations (due to being created in older versions of MS Excel that need to be updated to newer versions or based on release of new functionality) Uncertainty of requirements, skill level of creator may affect quality of the design and the end spreadsheet model/template result Inconsistency in design between creators based on experience/skill level may make prove to be problematic for users	Design Controls - Use of templates - Use of standard guidelines or policies around formatting and layouts - Provide user training on standards Documentation controls - use of standards guidelines or policies around type and locations of documentation - updating of standards as new and improved layouts/designs features are released	Manual review	Enforce guidelines/policy on standards standards Ensure that Master templates are updated	This would appear easy enough to implement and enforce – challenges may be getting more experienced workers to agree on and adopt standards and master templates/models

Figure 8: - continued

System Development Life Cycle (SDLC) /Error Type		Internal Controls & Feasibility Assessment of their use			
Spreadsheet (SS) SDLC Stage	**Error type**	**Preventative**	**Detective**	**Corrective**	**Feasibility Assessment**
Spreadsheet / Module Development & Testing	**Formula Errors – Quantitative** **Objective:** to standardize the use of certain tools e.g. macros, pivot tables/pivot charts that have been used and vetted already, and that are tools that users are familiar with. **risks** - Use of built-in functions vs. creation of formulas when needed - Skill level of creator/modifier- - Complexity of calculations - omissions related to formula syntax or entire formulas or incomplete data **Data Integrity (Quality and Completeness) Errors – including data extraction – and volatility - frequency of refreshing and responsibility for performing this process**	Use of software development approach such as Systems Development Life Cycle (SDLC) Input controls Design controls Training - Provide user training basic, intermediate & advanced spreadsheet tools	Internal Excel testing tools – such as tracing dependents, precedents and errors and watch windows to monitor changes External testing tools Auditing approaches using tools outside of Excel External testing tools - *testing sample, or key cells for errors* Kulesz D, Ostberg J-P., 2013 Practical Challenges with Spreadsheet Auditing Tools Output checking – using various tools / calculator verify amounts	Change controls – as changes are made to the design have them approved and documented Version controls – have a master version once all changes are final and make it the only one available – code it in the documentation quality system and make it available on a public drive for those who should have access to it – need to password lock macros, cells worksheets within the workbook or the entire workbook itself (before it can even be opened)	Adopting a software development methodology requires in-house expertise in this area to drive this process and may require a shift in organizational culture to one of attention to detail and focus on following standards. Spreadsheet support function (IT support for users) needed as complexity goes up Forcing the use of specific tools such as macros and pivot tables and pivot charts requires that all users be adequately trained and supported in these tools.

Figure 8: - continued

Internal Controls & Feasibility Assessment of their use		**Internal Controls & Feasibility Assessment of their use**			
Spreadsheet (SS) SDLC Stage	**Error type**	**Preventative**	**Detective**	**Corrective**	**Feasibility Assessment**
Implementation & user acceptance testing (UAT)	Data Integrity (Quality and Completeness) Errors – including data extraction – and volatility - frequency of refreshing and responsibility for performing this process	Input controls - Data Security Controls – to ensure that the right data is accessed/used and that it is non-volatile so that data refreshes are being managed Design controls – using templates or standard functions/tools	Testing controls – testing approach & protocols - cell formula testing, excel audit tools, others…	Change controls – have version controls on the spreadsheet, sign-off for any user changes	This stage takes time dependent on frequency of use of the spreadsheet – adhoc vs. a repeat use spreadsheet e.g. a permanent corporate asset
Operation – reuse and sharing	Data Integrity (Quality and Completeness) Errors – including data extraction – and volatility - frequency of refreshing and responsibility for performing this process	Input controls - Documentation provided in the spreadsheet or a companion document Access Controls - to ensure that only those who are supposed to use it can – file storage location, passwords on files and cell protection/data validation - Data Security Controls – who has access to the data – data volatility and who is responsible for refreshing the data source and how often	-	Change controls - Version controls – password lock cells, password entire lock worksheets, macros, or the entire workbook	This stage takes time dependent on frequency of repeat use and sharing of the spreadsheet
	Errors due to Mistakes – *including slips & lapses*	Documentation and appropriate training and support			
	Omission Errors – leaving something out of the model	Testing tools	Testing tools		
	Intentional Errors – Fraud/Faults	Auditing tools	Auditing tools	New controls	

Figure 8: - continued

Internal Controls & Feasibility Assessment of their use		Internal Controls & Feasibility Assessment of their use			
Spreadsheet (SS) SDLC Stage	**Error type**	**Preventative**	**Detective**	**Corrective**	**Feasibility Assessment**
Maintenance – modifications	- number of formulas/functions used - type of formula/functions used	Access Controls - passwords on spreadsheets and various worksheets as well as on the cell formulas - Design Controls documentation - Access Controls – only those with proper clearance can modify/edit spreadsheets others can only view spreadsheet		Change controls – sign-off before changes are performed, adherence to spreadsheet standards, guidelines when making changes	Different process with some more onerous than others depending on the spreadsheet level (refer to Figure 7)
Termination (retirement) /Replacement	When new versions of MS Excel or when new/ improved functions / tools are released	Access Controls – to ensure that access to the old file(s) are revoked or that old retired spreadsheet is removed from public drive(s) and archived Documentation of old/retired spreadsheet and documentation and change process for replacement spreadsheet Process repeats itself when new spreadsheets replace old ones. Continue training to ensure users skills are updated		Monitoring of version – need version control system in place – *aka* a QA system	Develop a process for archiving old files once one of the following happens: 1) new version of MS Excel is installed 2) new Excel functionality is made available 3) new standards are adopted by the Organization 4) new templates are made available old ones are no longer used – need version control system in place Start the process all over again with the new file

This proposed framework is a hybrid framework that integrates content from several published taxonomies for both quantitative and qualitative spreadsheet error types, with the stages of the spreadsheet life cycle and internal controls used for spreadsheets and assessment of the operational feasibility of applying the controls.

Panko's (2008) spreadsheet lifecycle framework (Figure 3) provides part of the basis for the second framework proposed in this study (Figure 8). Several of the proposed error

taxonomies from other studies (Panko & Halverson, 1996, 2008; Przasnyski, Leon, & Seal, 2011) also provide the basis for the types of errors listed in the proposed framework (Figure 8). In an effort to aggregate all of the various frameworks / taxonomies into one consolidated *useable* format, not all of the various error classifications or error types have been included in this initial version of the proposed framework. It is the intent of the researcher to expand on the content of this proposed framework. The internal controls framework, Stockton, is primarily a risk aversion one and uses three (3) types of controls detective, corrective and preventative as a means of bringing risks and errors down to an acceptable level. This controls framework has been used as a guide for the types of controls that can be used for the proposed framework (Figure 8). Lastly, Pankos (2009) proposed 6 level framework for researching spreadsheet issues was used to assess the feasibility of applying the suggested controls for each of the error types included in the proposed framework (Figure 8).

The case study conducted by Lemon and Ferguson (2010) supports the preceding proposed *Figure 8* of this report. In their case study, Lemon and Ferguson discuss a practical and pragmatic approach that was recently taken by a large global organization to manage their spreadsheet risk. They propose that a similar approach could be scaled and customized to meet the requirements of different organizations. The approach that the 'client' in their case study took to manage their risk of spreadsheet errors was to develop a spreadsheet control framework that they used to define the spreadsheet risks and the associated controls that should be considered for each type of risk based on the end-user.

Some of the underlying assumptions made when developing the proposed frameworks above (Figures 7 & 8) were: 1) listing each of the error types would be listed only once under the various stages of the systems development life cycle even though they may occur in more than one stage, 2) the type of internal control suggested to mitigate and/or eliminate the errors identified may not work in all organizations or situations and are simply the recommended type of control to apply, 3) that the errors listed under each control may not be the only ones that fit into each category of control as they are just the errors that were used for the purpose of this initial proposed version of the framework, 4) that the controls that fall under each category of internal control may not be inclusive of all the possible types of controls for each category and 5) that the feasibility of implementing each of the internal controls is based on the risk assessment for each individual level of the abridged version of Panko's six (6) level framework proposed in this report (Figure 2), which may not work in all cases in all organizations.

Both of these proposed frameworks would require further research and adoption/use in practice for a period of time in order to determine their validity and usefulness. The next section proposes some recommendations and areas of opportunity for future research of both a qualitative and quantitative nature.

6.2 Recommendations

Before proceeding with recommendation it is important to remind the reader that this case study, and was conducted on one organization with a small sample group of participants and as such there are limitations on how the findings can be used (Yin, 2014). The intent of the case study was to investigate how a single organization was using spreadsheets and the challenges that they faced in doing so and how those findings compared to the findings of other independent research studies. Additionally, this case study was also conducted with the intent to offer theoretical propositions, based on the premise that content from the various frameworks proposed in other research studies in the spreadsheet domain can be integrated and synthesized into a couple one aggregate framework. This hybrid framework identifies these various types of known spreadsheet errors from taxonomies proposed in other studies, allocates them to the stage of the Systems Development Life Cycle (SDLC) where they are more likely to occur, and links them with the types of controls (internal) that may be able to mitigate or eliminate their occurrence along with an assessment of the feasibility of the organizations ability to implement said controls.

Specifically the case study focused on four (4) primary areas: 1) understanding how spreadsheets are being used, 2) the frequency and impact of errors/risks associated with spreadsheet use, 3) the skill levels of (and training provided to) those creating/using spreadsheets experienced by this single organization and 4) how that compares to the findings of other independent empirical studies included in the literature review that accompanies this case study.

After reviewing and analyzing the findings of this case study, one of the key recommendations is that more qualitative research should be conducted in order to better understand the importance of the issues in this domain that can not properly be measured through quantitative research methods. For instance other case studies using one or more of the alternative case study research design methodologies, such as single-cases *embedded*, multiple-case *holistic* or multiple-case *embedded* could be applied to future case studies to validate the findings of this case study and others like it.

It is clear that more quantitative research also needs to be conducted with a larger sample from different populations in order to empirically test and evaluate the underlying assumptions for the way that spreadsheets are being used, the challenges associated with their use, and ways to improve their reliability and validity as a data analytics tool for business. One way to do this is would be to develop a new survey questionnaire using the main themes of the interview questions that were used in this case study (refer to Appendix A) as a guide, and incorporating them with questions from other similar surveys. By adapting those questions and combining them with some additional ones designed to capture more of the attributes of spreadsheet use and spreadsheet users it would result in a more focused survey questionnaire. Then by distributing this to a larger population of spreadsheet users the data collected from a survey of this nature would provide more useful results that would offer more insight in areas for improving spreadsheet design, development and use. The combination of more questions targeted at specific attributes of spreadsheets and their user distributed to a lager population of

spreadsheet users could provide more useful and conclusive results about the occurrence and impact of risks, ways to mitigate risk and any training gaps for spreadsheet users.

Another area that represents an opportunity for further investigation is whether or not a set of criteria could be developed for when to use spreadsheets. This would be an opportunity to test assumptions about when it may be safer (less risky) and more beneficial for organizations to use an alternative to spreadsheets, such as a business analytics or business intelligence tools that may provide higher quality results such as for regulatory reporting and more reliable information to organizations supporting their long term, strategic planning, capital investment decisions.

Based on the specific findings of this case study, there is an opportunity for further research to be conducted to study the assumptions and integration of content combined in the two (2) proposed frameworks in the conclusion (Section 6.1) of this report to validate or disprove their theoretical proposition.

6.3 Areas for further research

Several areas that may warrant further consideration or further investigation are as follows:

a. whether or not 'best or *good* practices' in design standards (universal or industry specific) would help eliminate/reduce the level and/or types and frequency of errors
b. understanding of the root cause of errors *e.g.* are they due more because of the behavioral attributes of users or the types of human error (judgment, skill, formula) or some other type of error like design (qualitative) or perhaps related to the data quality (integrity and completeness) of the data that is being analyzed rather than the spreadsheet being used to do the analysis
c. identification of the most common types of errors both qualitative and quantitative
d. evaluating various spreadsheet error testing techniques that could be used, *e.g.,* is there a checklist that can be developed and applied to spreadsheet design (templates/models) and/or the process of testing for errors in spreadsheets, *e.g.,* ones that could prevent errors from occurring in the first place or ones that could detect them and challenges with their adoption and feasibility of use by organizations

The concept described in *a) above* is also supported by Kulesz (2011, p. 1) who proposes "an expert-based, retrospective approach to the identification of good practices for spreadsheets" when it comes to developing policies for preventing spreadsheet errors.

In conclusion, the following take-aways can be distilled from the case study and preceding literature review; it has been proven that spreadsheets contain errors, that there are risks associated with using spreadsheets whether they are acknowledged and/or recognized or not, that insufficient testing is being done on spreadsheets prior to them being used or shared, that more spreadsheet training is required, and that it is not likely that spreadsheets are going away anytime soon. Additionally, management has a fiduciary duty to recognize and manage (mitigate) spreadsheet risks more effectively. Organizations are not just responsible to their internal stakeholders; they are also responsible to their external stakeholders. They must provide external information (sometimes of a financial nature) to various regulatory and government bodies, as well as to shareholders that complies with certain standards on which various systems, such as the financial markets, rely. Without the appropriate spreadsheet controls in place to ensure that the financial information and other information being shared is of the highest quality, organizations domestic and global, systems and other stakeholders, including the general public are all at risk.

Appendices

Appendix A – Interview Questions

MRP INTERVIEW QUESTIONS – SL **JULY 13, 2015**

SUMMARY	# QUESTIONS	QUES NUMBERS	SHEET LINK
Tombstone (General)		**n/a**	Additional Questions
CATEGORY	**# QUESTIONS**	**QUES NUMBERS**	**SHEET LINK**
User/Usage Profile	**12**	**1-12**	Usage Questions
Risks, Controls & Errors	**18**	**13-30**	Risks, Controls & Errors
Training	**8**	**31-38**	Training
Documentation & Sharing	**7**	**39-45**	Documentation & Sharing
Design (Qualitative)	**8**	**46-53**	Design (creation) Qualitative
Total Questions	**53**		

Tombstone (General)	Years with the current employer	
Tombstone (General)	Gender (M/F)	
Tombstone (General)	Age	
Tombstone (General)	Functional area	
Tombstone (General)	Highest level of education	
User/Usage Profile	1. Indicated which one of these best describes you: **choose all that apply**	
	a.	Spreadsheet Creator (author)
	b.	Spreadsheet Modifier (editor)
	c.	Spreadsheet Viewer (receiver)
User/Usage Profile	2. Years using spreadsheets	
	a.	0-2
	b.	3-5
	c.	6-10
	d.	> 10

User/Usage Profile 3. Rate your spreadsheet skills using the following scale

a. Novice
b. Intermediate
c. Expert

User/Usage Profile 4. Level of importance spreadsheets have in your job.

a. Unimportant
b. Moderately important
c. Very important
d. Critically important

User/Usage Profile 5. Approximate percent of time spent with spreadsheets in your job.

a. 0-25%
b. 26-50%
c. 51-75%
d. 76-100%

User/Usage Profile 6. Spreadsheet software (and version) you are using.
Choose all that apply

a. excel 2010
b. excel 2013
c. excel for Mac (2011)
d. excel (older/other) - which one _____?
e. other spreadsheet - which one _____?

User/Usage Profile 7. Main purposes of spreadsheets you use.
Choose all that apply

a. Maintaining lists (e.g. names and addresses)
b. Tracking data (e.g. budgets, sales, inventories)
c. Analyzing data (e.g. financial, operational)
d. Determining trends and making projections
e. Evaluating alternatives
f. Other _____

User/Usage Profile 8. How often each of the following specific spreadsheets tools are used by you:

	Tool/Function	**Never**	**Rarely**	**Occasionally**	**Frequently**
a.	Goal Seek Tool				
b.	LOOKUP Functions				

	c.	Pivot Tables
	d.	Conditional formatting
	e.	IF Function
	f.	Formula Auditing Tools
	g.	Chart Wizard
	h.	Function Wizard
	i.	Solver
	j.	Financial Functions (e.g. NPV, IRR, PMT)
	k.	Find/Replace
	l.	Macros
	m.	Data Filter Tool
	n.	Data Sort Tool

User/Usage Profile

9. Number of different spreadsheets you normally use per week.

a.	0-1
b.	2-5
c.	6-10
d.	more than 10

User/Usage Profile

10. Number of hours per week of your time normally spent using spreadsheets.

a.	0 -1
b.	1-3
c.	3-5
d.	5-10
e.	10-20
f.	> 20

User/Usage Profile

11. Frequency of usage (reuse or sharing) of a typical spreadsheet after first use. **Choose all that apply**

a.	daily
b.	once or twice a per week
c.	monthly
d.	quarterly
e.	annually
f.	less than once a year

User/Usage Profile

12. Please check the types of software that you use in your job. **Choose all that apply**

a. Microsoft Excel
b. Microsoft Access
c. SAP and/or SAP Business Objects (e.g. Lumira)
d. Oracle BI or Oracle database
e. IBM database and/or Cognos
f. Microsoft Dynamics
g Tableau
h. Crystal Reports
i. Other _____

Risks, Controls & Errors

13. Do you have an internal audit function?

a. Yes
b. No
c. Don't know

Risks, Controls & Errors

14. If there is an internal audit function are they involved in spreadsheet audits/testing?

a. Seldom
b. Usually
c. Always
d. Don't know

Risks, Controls & Errors

15. Organization has standards or polices for spreadsheets.

a. No standards
b. No written standards, only informal guidelines
c. Basic written standards
d. Detailed written guidelines and protocols
e. Don't know

Risks, Controls & Errors

16. How many errors have you found in your own spreadsheets?

a. 0
b. 0-15
c. 15-50
d. > 50

Risks, Controls & Errors

17. How many errors have you found in spreadsheets prepared by others?

a. 0
b. 0-15
c. 15-50
d. >50

Risks, Controls & Errors

18. What is the main reason you believe spreadsheet errors occur. Choose all that apply

a. not paying attention
b. rushed / short timeline to get it done
c. not testing it before using/sharing it
d. not understanding the requirements requested
e. poor design
f. inadequate spreadsheet training
g. other - explain

Risks, Controls & Errors

19. Spreadsheet Standards and/or Polices are followed in your organization.

a. Seldom
b. Usually
c. Always
d. Don't know

Risks, Controls & Errors

20. Impediments to following the Spreadsheet standards provided by your organization. **Choose all that apply**

a. No impediments
b. Too stringent
c. Lack of spreadsheet knowledge, training, or support
d. No incentives
e. No enforcement
f. Others do not follow the standards
g. don't understand the standards
h. Not applicable

Risks, Controls & Errors

21. Importance of spreadsheets to your organization as a whole.

a. Unimportant
b. Moderately important
c. Very important
d. Critical

Risks, Controls & Errors

22. Level of risk (impact of errors) spreadsheets pose in your company.

a. High risk
b. Medium risk
c. Low risk
d. No risk

Risks, Controls & Errors

23. Awareness of your organization of the risk of spreadsheets (incl. sharing)

a. Full awareness
b. Some awareness
c. No awareness

Risks, Controls & Errors

24. Specific problems encountered with the creation or use of spreadsheets.

Explain _____

Risks, Controls & Errors

25. Practices that have been particularly helpful to you or your organization in improving the quality of spreadsheets or the manner in which they are used.

Explain _____

Risks, Controls & Errors

26. Frequency of testing of spreadsheet models that you or others create.

a. Never
b. Sometimes
c. Usually
d. Always

Risks, Controls & Errors

27. Which of the following methods are used to test spreadsheets Choose all that apply

a. Test extreme case
b. Use a calculator to check selected cells
c. Display all formulas
d. Examine formulas individually
e. use Go To – Special
f. Test performance for plausibility
g. Error Checking option
h. Formula Auditing Toolbar
i. Use common sense
j. External reviews
k. Other tools:_____

Risks, Controls & Errors

28. What is the Percentage of time devoted to spreadsheet testing by you or your organization?

a. 0%
b. 1-10%
c. 11-20%
d. 21-30%
e. 31-40%
f. 41-50%
g. > 50% and other

Risks, Controls & Errors

29. What has been the severity of the impact of spreadsheet errors detected by you? **Choose all that apply**

a. Major (High)
b. Moderate
c. Minor (Low)
d. No impact

Risks, Controls & Errors

30. What primary type of data source has been used by you for spreadsheet analysis?

a. data that has been manually entered in a spreadsheet/other file
b. data that has been downloaded from a system (export/import)
c. data source that has been accessed directly or dynamically
d. other, explain _____

Training

31. Types of training in spreadsheets made available by your organization. **Choose all that apply**

a. None
b. In-house training (face to face)
c. Training by external party
d. One basic session is available
e. Several sessions, including advanced topics, are available.
f. Spreadsheet specialist dedicated to assisting designers and users.
g. online training
h. Other _____

Training

32. Topics covered in the training program offered to you. **Choose all that apply**

a. Basic spreadsheet techniques (for example, copy and past, simple formulas)
b. Advanced spreadsheet techniques (e.g. use of built-in functions, conditional formatting
c. Data analysis (sorting, filter, pivot tables)
d. Use of specialized add-ins and other tools
e. Macros
f. Other _____

Training

33. Number of days of training offered to you each year.

a. None
b. 1 or 2 days
c. 3 to 5 days
d. More than 5 days

Training

34. Number of days of training you use each year.

a. None
b. 1 or 2 days
c. 3 to 5 days
d. More than 5 days

Training

35. The biggest impediments to your further participation in company-sponsored training. **Choose all that apply**

a. Not enough time
b. High cost
c. Poor quality of training
d. Lack of personal interest
e. Lack of support from management
f. Not applicable

Training

36. Incentives offered to you for organization-sponsored training. **Choose all that apply**

a. None
b. Organization pays cost of training
c. Organization provides paid time off
d. Training is a prerequisite for promotion
e. Not applicable

Training

37. Probability of participating in training, if made available in your organization.

a. Probably not
b. Perhaps
c. Definitely
d. Not applicable

Training

38. Type(s) of training have you had using spreadsheets. **Choose all that apply**

a. None
b. Formal classroom instruction
c. Occasional informal training sessions
d. Books and manuals
e. Demonstrations from colleagues
f. online, e-learning

Documentation & Sharing

39. Are spreadsheets shared?

a. Yes
b. No
c. Don't know

Documentation & Sharing

40. Ways you share your spreadsheets. Choose all that apply

a. rarely share any part of spreadsheet
b. provide a summary of results
c. provide parts of the spreadsheet
d. I share the entire model

Documentation & Sharing

41. Frequency of sharing this kind of information with others.
   a. daily
   b. weekly
   c. monthly
   d. quarterly
   e. annually
   f. less than once a year

Documentation & Sharing

42. Type of protection normally used for these spreadsheet models when shared.
   a. None
   b. Password protection
   c. Cell protection
   d. Data validation
   e. Other

Documentation & Sharing

43. Documentation of spreadsheets (either in the spreadsheets or separate document)
   a. Never
   b. Sometimes
   c. Usually
   d. Always

Documentation & Sharing

44. Techniques used to document spreadsheets. choose all that apply
   a. Text in spreadsheet
   b. Cell Comments
   c. Documentation sheet in workbook
   d. Separate document
   e. None of the above

Documentation & Sharing

45. Percentage of work time devoted to spreadsheet documentation.
   a. 0%
   b. 1-10%
   c. 11-20%
   d. 21-30%
   e. 31-40%
   f. 41-50%
   g. > 50%

Design (Qualitative)

46. Create spreadsheets *from scratch*
   a. Always
   b. Sometimes
   c. Never

Design (Qualitative)

47. Create spreadsheets *from template*
   a. Always
   b. Sometimes
   c. Never

Design (Qualitative)

48. Percentage of work time devoted to spreadsheet *creation.*
    a. 1-10%
    b. 11-20%
    c. 21-30%
    d. 31-40%
    e. 41-50%
    f. > 50%

Design (Qualitative)

49. Size of models/templates normally created.
    a. under 100 KB
    b. 101 KB to 1MB
    c. over 1 MB

Design (Qualitative)

50. Size of models/templates normally created (number of sheets).
    a. 1-2 worksheets/tabs
    b. 3-5 worksheets/tabs
    c. 6-10 worksheets/tabs
    d. over 10 worksheets/tabs

Design (Qualitative)

51. Size of models/templates normally created (number of cells).
    a. 1-75 cells
    b. 75-150 cells
    c. 150-250 cells
    d. > 250 cells

Design (Qualitative)

52. Best description of your work in creating spreadsheets.
    a. Work independently
    b. Seek advice from another person(s)
    c. Work with a peer group
    d. Work with a project team

Design (Qualitative)

53. Other people normally use the spreadsheets you create.
    a. No, my spreadsheets are for my personal use.
    b. My spreadsheets are shared with one or two others
    c. My spreadsheets are used by a number of people.
    d. My spreadsheets often become permanent assets in my organization.

Appendix B – Interview Results

MRP INTERVIEW QUESTIONS – SL			JULY 13, 2015
SUMMARY	**# QUESTIONS**	**QUESTION #'s**	**SHEET LINK**
Tombstone (General)		**n/a**	Additional Questions
CATEGORY	**# QUESTIONS**	**QUESTION #'s**	**SHEET LINK**
User/Usage Profile	**12**	**1-12**	Usage Questions
Risks, Controls & Errors	**18**	**13-30**	Risks, Controls & Errors
Training	**8**	**31-38**	Training
Documentation & Sharing	**7**	**39-45**	Documentation & Sharing
Design (Qualitative)	**8**	**46-53**	Design (creation) Qualitative
Total Questions	**53**		

Questions	Responses	Numeric Responses	% Responses
Tombstone Data			
Years with the current employer		Average =	10.4
Gender (M/F)		Males = 67%	Females = 33%
Age		Ranged from	26 - 52
Functional area		67%from delivery areas	33% from support areas
Highest level of education attained		Ranged from a	Diploma to a Masters Degree
Question #	**Responses**	**Numeric Responses**	**% Responses**
User/Usage Profile			
1	a.	6	38%
	b.	6	38%
	c.	4	25%
		16	**100%**
2	a.	0	0%
	b.	1	17%
	c.	1	17%
	d.	4	67%
		6	**100%**
3	a.	0	0%
	b.	5	83%
	c.	1	17%
		6	**100%**

4	a.	0	0%
	b.	1	17%
	c.	2	33%
	d.	3	50%
		6	**100%**
5	a.	2	33%
	b.	1	17%
	c.	2	33%
	d.	1	17%
		6	**100%**
6	a.	5	42%
	b.	3	25%
	c.	0	0%
	d.	3	25%
	e.	1	8%
		12	**100%**
7	a.	6	24%
	b.	6	24%
	c.	6	24%
	d.	4	16%
	e.	2	8%
	f.	1	4%
		25	**100%**
** Question # 8 located at the end*			
9	a.	0	0%
	b.	1	17%
	c.	1	17%
	d.	4	67%
		6	**100%**
10	a.	0	0%
	b.	0	0%
	c.	1	17%
	d.	1	17%
	e.	2	33%
	f.	2	33%
		6	**100%**
11	a.	4	40%
	b.	3	30%
	c.	1	10%
	d.	1	10%
	e.	1	10%
	f.	0	0%
		10	**100%**

12	a.	6	29%
	b.	0	0%
	c.	4	19%
	d.	5	24%
	e.	3	14%
	f.	0	0%
	g	0	0%
	h.	2	10%
	i.	1	5%
		21	**100%**
Risks, Controls & Errors			
13	a.	4	67%
	b.	2	33%
	c.	0	0%
		6	**100%**
14	a.	1	17%
	b.	1	17%
	c.	0	0%
	d.	4	67%
		6	**100%**
15	a.	2	33%
	b.	1	17%
	c.	1	17%
	d.	1	17%
	e.	1	17%
		6	**100%**
16	a.	1	17%
	b.	5	83%
	c.	0	0%
	d.	0	0%
		6	**100%**
17	a.	0	0%
	b.	4	67%
	c.	2	33%
	d.	0	0%
		6	**100%**
18	a.	3	14%
	b.	4	19%
	c.	5	24%
	d.	4	19%
	e.	1	5%
	f.	4	19%
	g.	0	0%
		21	**100%**

19	a.	0	0%
	b.	2	33%
	c.	1	17%
	d.	3	50%
		6	**100%**
20	a.	0	0%
	b.	0	0%
	c.	0	0%
	d.	1	14%
	e.	1	14%
	f.	0	0%
	g.	1	14%
	h.	4	57%
		7	**100%**
21	a.	0	0%
	b.	1	20%
	c.	1	20%
	d.	3	60%
		5	**100%**
22	a.	0	0%
	b.	4	80%
	c.	0	0%
	d.	1	20%
		5	**100%**
23	a.	3	50%
	b.	3	50%
	c.	0	0%
		6	**100%**
24	cut & paste errors, cell reference errors, leak of private information, reconciling the data, incorrect formula, , , , ,		
25	use of pivot tables, use of macros, encouraged to check own work, , , , ,		
26	a.	1	17%
	b.	3	50%
	c.	0	0%
	d.	2	33%
		6	**100%**

27	a.	1	5%
	b.	3	16%
	c.	2	11%
	d.	2	11%
	e.	1	5%
	f.	2	11%
	g.	1	5%
	h.	1	5%
	i.	3	16%
	j.	1	5%
	k.	2	11%
		19	**100%**
28	a.	1	17%
	b.	2	33%
	c.	2	33%
	d.	1	17%
	e.	0	0%
	f.	0	0%
	g.	0	0%
	h.	0	0%
		6	**100%**
29	a.	0	0%
	b.	3	33%
	c.	4	44%
	d.	2	22%
		9	**100%**
30	a.	1	17%
	b.	3	50%
	c.	2	33%
	d.	0	0%
		6	**100%**
Training			
31	a.	2	29%
	b.	0	0%
	c.	0	0%
	d.	0	0%
	e.	1	14%
	f.	0	0%
	g.	4	57%
	h.	0	0%
		7	**100%**

32	a.	2	20%
	b.	2	20%
	c.	1	10%
	d.	1	10%
	e.	0	0%
	f.	4	40%
		10	**100%**
33	a.	0	0%
	b.	0	0%
	c.	4	67%
	d.	2	33%
		6	**100%**
34	a.	0	0%
	b.	1	17%
	c.	3	50%
	d.	2	33%
		6	**100%**
35	a.	5	56%
	b.	0	0%
	c.	0	0%
	d.	1	11%
	e.	2	22%
	f.	1	11%
		9	**100%**
36	a.	0	0%
	b.	5	33%
	c.	5	33%
	d.	4	27%
	e.	1	7%
		15	**100%**
37	a.	0	0%
	b.	0	0%
	c.	4	67%
	d.	2	33%
		6	**100%**
38	a.	0	0%
	b.	4	20%
	c.	4	20%
	d.	4	20%
	e.	4	20%
	f.	4	20%
		20	**100%**

39	a.	6	100%
	b.	0	0%
	c.	0	0%
		6	**100%**
40	a.	1	7%
	b.	4	29%
	c.	4	29%
	d.	5	36%
		14	**100%**
41	a.	5	83%
	b.	1	17%
	c.	0	0%
	d.	0	0%
	e.	0	0%
	f.	0	0%
		6	**100%**
42	a.	3	50%
	b.	2	33%
	c.	0	0%
	d.	0	0%
	e.	1	17%
		6	**100%**
43	a.	2	33%
	b.	2	33%
	c.	2	33%
	d.	0	0%
		6	**100%**
44	a.	2	18%
	b.	3	27%
	c.	2	18%
	d.	1	9%
	e.	3	27%
		11	**100%**
45	a.	2	33%
	b.	4	67%
	c.	0	0%
	d.	0	0%
	e.	0	0%
	f.	0	0%
	g.	0	0%
		6	**100%**

Design (qualitative) errors			
46	a.	0	0%
	b.	6	100%
	c.	0	0%
		6	**100%**
47	a.	0	0%
	b.	5	83%
	c.	1	17%
		6	**100%**
48	a.	0	0%
	b.	3	50%
	c.	2	33%
	d.	0	0%
	e.	0	0%
	f.	1	17%
	g.	0	0%
-		**6**	**100%**
49	a.	1	17%
	b.	4	67%
	c.	1	17%
		6	**100%**
50	a.	2	33%
	b.	4	67%
	c.	0	0%
	d.	0	0%
		6	**100%**
51	a.	2	33%
	b.	0	0%
	c.	1	17%
	d.	3	50%
		6	**100%**
52	a.	3	50%
	b.	1	17%
	c.	1	17%
	d.	1	17%
		6	**100%**
53	a.	1	17%
	b.	2	33%
	c.	2	33%
	d.	1	17%
		6	**100%**

*8	Tool/Function	Never	Rarely
a.	Goal Seek Tool	2	3
b.	LOOKUP Functions	1	0
c.	Pivot Tables/Charts	0	0
d.	Conditional formatting	0	1
e.	IF Function	1	0
f.	Formula Auditing Tools	3	2
g.	Chart Wizard	2	0
h.	Function Wizard	1	0
i.	Solver	3	0
j.	Financial Functions (e.g. NPV, IRR, PMT)	5	1
k.	Find/Replace	0	0
l.	Macros	2	0
m.	Data Filter Tool	0	0
n.	Data Sort Tool	0	0
	Frequency of use	20	7
***8**	**Tool/Function**	Occasionally	Frequently
a.	Goal Seek Tool	1	0
b.	LOOKUP Functions	0	3
c.	Pivot Tables/Charts	1	2
d.	Conditional formatting	1	2
e.	IF Function	1	2
f.	Formula Auditing Tools	1	0
g.	Chart Wizard	3	1
h.	Function Wizard	2	3
i.	Solver	2	1
j.	Financial Functions (e.g. NPV, IRR, PMT)	0	0
k.	Find/Replace	3	1
l.	Macros	3	0
m.	Data Filter Tool	0	3
n.	Data Sort Tool	0	3
	Frequency of use	18	21

*8	Tool/Function	Daily	Total
a.	Goal Seek Tool	0	6
b.	LOOKUP Functions	2	6
c.	Pivot Tables/Charts	3	6
d.	Conditional formatting	2	6
e.	IF Function	2	6
f.	Formula Auditing Tools	0	6
g.	Chart Wizard	0	6
h.	Function Wizard	0	6
i.	Solver	0	6
j.	Financial Functions (e.g. NPV, IRR, PMT)	0	6
k.	Find/Replace	2	6
l.	Macros	1	6
m.	Data Filter Tool	3	6
n.	Data Sort Tool	3	6
	Frequency of use	18	84

Appendix C – Other Research Study findings

Spreadsheet Engineering Research Project (SERP) Tuck School of Business at Dartmouth College
http://faculty.tuck.dartmouth.edu/images/uploads/faculty/serp/serp_results.pdf

SURVEY ON SPREADSHEET USAGE							
		ALL SURVEYs - March '06					
		(based on 1597 responses in seven surveys received by March 10, 2006)					
Spreadsheet Usage							
			#			%	
1. Please check the types of software you use in your job.							
	a.	Microsoft Excel	1586			99.30%	
	b.	Quattro Pro	24			1.50%	
	b.	Lotus 1-2-3	39			2.40%	
	c.	Microsoft Access	514			32.20%	
	d.	Visual Basic for Applications (VBA)	399			25.00%	
	e.	Oracle database	138			8.60%	
	f.	IBM database	29			1.80%	
	g.	Other	191			12.00%	
2. Level of importance spreadsheets have in your job.							
	a.	Unimportant	23			1.40%	
	b.	Moderately important	253			15.90%	
	c.	Very important	536			33.60%	
	d.	Critical	781			49.00%	
3. Please classify your experience with spreadsheets.							
	a.	Little or no experience	11			0.70%	
	b.	Some experience; still a beginner	101			6.40%	
	c.	Extensive experience; some expertise	853			53.60%	
	d.	Very experienced; high expertise.	625			39.30%	
4. Type(s) of training have you had using spreadsheets.							
	a.	None	281			17.60%	
	b.	Formal classroom instruction	602			37.70%	
	c.	Occasional informal training sessions	467			29.20%	
	d.	Books and manuals	856			53.60%	
	e.	Demonstrations from colleagues	835			52.30%	

5. When working with spreadsheets, you typically work:							
	a.	By yourself	1289			81.10%	
	b.	In a team of 2 or 3	259			16.30%	
	c.	In a larger team (4 or more)	42			2.60%	
6. Approximate percent of time spent with spreadsheets in your job.							
	a.	0-25%	712			44.70%	
	b.	26-50%	484			30.40%	
	c.	51-75%	284			17.80%	
	d.	76-100%	114			7.20%	
7. Main purposes of spreadsheets you use .							
	a.	Maintaining lists (e.g. names and addresses)	399			25.00%	
	b.	Tracking data (e.g. budgets, sales, inventories)	753			47.20%	
	c.	Analyzing data (e.g. financial, operational)	1399			87.60%	
	d.	Determining trends and making projections	875			54.80%	
	e.	Evaluating alternatives	907			56.80%	
	f.	Other	194			12.10%	
8. Techniques used in your spreadsheets.							
	a.	Statistical analysis	963			60.30%	
	b.	Optimization (e.g. Solver, What's Best)	748			46.80%	
	c.	Simulation (e.g. Crystal Ball, @Risk)	489			30.60%	
	d.	None of the above	413			25.90%	
9. How often each of the following specific spreadsheets tools are used:							
			Never	Rarely	Occasional	Frequent	Every Day
	a.	Goal Seek Tool	526	321	409	264	29
			34.00%	20.70%	26.40%	17.00%	1.90%
	b.	LOOKUP Functions	271	243	426	434	196
			17.30%	15.50%	27.10%	27.60%	12.50%
	c.	Pivot Tables	394	366	366	306	131
			25.20%	23.40%	23.40%	19.60%	8.40%
	d.	Conditional formatting	314	249	419	466	110
			20.20%	16.00%	26.90%	29.90%	7.10%
	e.	IF Function	146	137	284	595	401
			9.30%	8.80%	18.20%	38.10%	25.70%
	f.	Formula Auditing Tools	393	294	317	340	206

			25.40%	19.00%	20.50%	21.90%	13.30%
	g.	Chart Wizard	162	200	383	618	199
			10.40%	12.80%	24.50%	39.60%	12.70%
	h.	Function Wizard	217	242	417	504	85
			14.80%	16.50%	28.50%	34.40%	5.80%
	i.	Solver	491	356	339	284	76
			31.80%	23.00%	21.90%	18.40%	4.90%
	j.	Financial Functions (e.g. NPV, IRR, PMT)	250	302	369	458	188
			16.00%	19.30%	23.50%	29.20%	12.00%
	k.	Find/Replace	158	254	413	488	244
			10.10%	16.30%	26.50%	31.30%	15.70%
	l.	Macros	300	418	343	289	210
			19.20%	26.80%	22.00%	18.50%	13.50%
	m.	Data Table Tool	423	407	371	276	72
			27.30%	26.30%	24.00%	17.80%	4.60%
	n.	Data Sort Tool	110	158	404	653	226
			7.10%	10.20%	26.00%	42.10%	14.60%
10. Number of different spreadsheets you normally use per week.							
	a.	0-1	93			5.80%	
	b.	2-5	640			40.20%	
	c.	6-10	408			25.60%	
	d.	more than 10	450			28.30%	
11. Those who report to you use spreadsheets to develop recommendations.							
	a.	Yes	895			56.50%	
	b.	No	161			10.20%	
	c.	Not applicable	492			31.00%	
	d.	Don't know	37			2.30%	
12. Creator of spreadsheets in your work.							
	a.	Yes	1467			92.70%	
	b.	No (if no go to questions 22)	116			7.30%	
Spreadsheet Creation							

13. Create spreadsheets from scratch							
	a.	Always	539			36.30%	
	b.	Sometimes	922			62.10%	
	c.	Never	23			1.50%	
14. Percentage of work time devoted to spreadsheet creation.							
	a.	0%	7			0.50%	
	b.	1-10%	707			47.60%	
	c.	11-20%	385			25.90%	
	d.	21-30%	173			11.70%	
	e.	31-40%	76			5.10%	
	f.	41-50%	72			4.90%	
	g.	More than 50%	64			4.30%	
15. Division of spreadsheet models into separate, integrated modules.							
	a.	Never	62			4.20%	
	b.	Sometimes	483			32.70%	
	c.	Usually	629			42.60%	
	d.	Always	301			20.40%	
16. Size of models normally created.							
	a.	under 100 cells	127			8.60%	
	b.	101 to 1000 cells	624			42.40%	
	c.	1001 to 10,000 cells	471			32.00%	
	d.	10,001 to 100,000 cells	184			12.50%	
	e.	over 100,000 cells	66			4.50%	
17. How often you separate all data inputs form the formulas in your spreadsheet.							
	a.	Never	77			5.20%	
	b.	Sometimes	457			31.10%	
	c.	Usually	608			41.40%	
	d.	Always	327			22.30%	
18. Typical first step in creating a spreadsheet.							
	a.	Borrow a design from another spreadsheet	335			22.80%	
	b.	Sketch the spreadsheet on paper	256			17.40%	
	c.	Write the fundamental relationships using algebra	85			5.80%	
	d.	Enter the data and formulas directly into a computer	717			48.70%	
	e.	Other	78			5.30%	

19. Frequency of usage of systems development methodologies (e.g. SDLC, RAD)							
	a.	Always	15			1.00%	
	b.	Sometimes	112			7.60%	
	c.	Never	1343			91.40%	
20. Best description of your work in creating spreadsheets.							
	a.	Work independently	1140			77.30%	
	b.	Seek advice from another person(s)	116			7.90%	
	c.	Work with a peer group	85			5.80%	
	d.	Work with a project team	133			9.00%	
21. Other people normally use the spreadsheets you create.							
	a.	No, my spreadsheets are for my personal use.	169			11.50%	
	b.	My spreadsheets are shared with one or two others	619			42.00%	
	c.	My spreadsheets are used by a number of people.	456			30.90%	
	d.	My spreadsheets often become permanent assets.	231			15.70%	
Spreadsheet Testing							
22. Testing of spreadsheet models that you or others create.							
	a.	Never, (if never, go to questions 25)	271			17.10%	
	b.	Sometimes	505			31.90%	
	c.	Usually	422			26.70%	
	d.	Always	383			24.20%	
23. Which of the following methods used to test spreadsheets.							
	a.	Test extreme case	733			45.90%	
	b.	Use a calculator to check selected cells	613			38.40%	
	c.	Display all formulas	290			18.20%	
	d.	Examine formulas individually	729			45.60%	
	e.	use Go To - Special	100			6.30%	
	f.	Test performance for plausibility	693			43.40%	
	g.	Error Checking option	163			10.20%	
	h.	Formula Auditing Toolbar	447			28.00%	
	i.	Use common sense	1076			67.40%	
	j.	Other tools:	121			7.60%	

24. Percentage of time (approximate) devoted to spreadsheet testing.							
	a.	0%	56			4.20%	
	b.	1-10%	1051			78.30%	
	c.	11-20%	156			11.60%	
	d.	21-30%	53			3.90%	
	e.	31-40%	11			0.80%	
	f.	41-50%	10			0.70%	
	g.	more than 50%	5			0.40%	
Spreadsheet Documentation							
25. Documentation of spreadsheets (within spreadsheets or in separate document)							
	a.	Never (If never, go to question 28)	278			17.70%	
	b.	Sometimes	780			49.50%	
	c.	Usually	404			25.70%	
	d.	Always	113			7.20%	
26. Techniques used to document spreadsheets.							
	a.	Text in spreadsheet	1019			63.80%	
	b.	Cell Comments	955			59.80%	
	c.	Documentation sheet in workbook	463			29.00%	
	d.	Separate document	291			18.20%	
	e.	None of the above	29			1.80%	
27. Percentage of work time devoted to spreadsheet documentation.							
	a.	0%	1172			88.10%	
	b.	1-10%	125			9.40%	
	c.	11-20%	21			1.60%	
	d.	21-30%	6			0.50%	
	e.	31-40%	5			0.40%	
	f.	41-50%	2			0.20%	
	g.	More than 50%	0			0.00%	
Spreadsheet Implementation/Use							
28. Hours per week of your time normally spent in using a typical spreadsheet.							
	a.	0-1	278			17.70%	
	b.	03-Jan	567			36.20%	
	c.	05-Mar	307			19.60%	
	d.	10-May	245			15.60%	
	e.	20-Oct	115			7.30%	
	f.	over 20	55			3.50%	

29. Number of other users for a typical spreadsheet you use.							
	a.	None	211			13.50%	
	b.	1 other person	295			18.80%	
	c.	2-5 other people	782			49.90%	
	d.	6-10 other people	138			8.80%	
	e.	more than 10 other people	140			8.90%	
30. Frequency of usage of a typical spreadsheet after first use.							
	a.	daily	220			14.10%	
	b.	once or twice a per week	724			46.40%	
	c.	monthly	401			25.70%	
	d.	quarterly	122			7.80%	
	e.	annually	32			2.10%	
	f.	less than once a year	60			3.80%	
Spreadsheet Sharing							
31. Ways in which you share your spreadsheets.							
	a.	I rarely share any part of a spreadsheet	157			9.80%	
	b.	I provide a summary of results	608			38.10%	
	c.	I provide parts of the spreadsheet	428			26.80%	
	d.	I share the entire model	1080			67.60%	
32. Frequency of sharing this kind of information with others.							
	a.	daily	295			19.10%	
	b.	weekly	577			37.30%	
	c.	monthly	447			28.90%	
	d.	quarterly	126			8.10%	
	e.	annually	31			2.00%	
	f.	less than once a year	71			4.60%	
33. Type of protection normally used for these spreadsheet models when shared.							
	a.	None	998			62.50%	
	b.	Password protection	382			23.90%	
	c.	Cell protection	395			24.70%	
	d.	Data validation	208			13.00%	
	e.	Other	81			5.10%	
34. Method used to ensure version control when models are shared with others.							
	a.	No control	490			30.70%	
	b.	Save the date	615			38.50%	
	c.	Save with version number	674			42.20%	
	d.	Save with user name	171			10.70%	
	e.	Other	97			6.10%	

Spreadsheet Modification							
35. Average lifetime of major spreadsheet models you use, including refinements.							
	a.	One week	61			3.90%	
	b.	Few weeks or months	624			40.10%	
	c.	A year or two	574			36.80%	
	d.	More than two years	299			19.20%	
36. Person modifying or refining these models over time.							
	a.	The original developer	1172			73.40%	
	b.	A new developer	337			21.10%	
	c.	Users	537			33.60%	
Spreadsheet Archiving							
37. Method used to back up a spreadsheet after saving it.							
	a.	Not applicable; no back-up	217			13.60%	
	b.	Back-up to a diskette or a separate drive	451			28.20%	
	c.	Back-up to a main server	993			62.20%	
	d.	Other	93			5.80%	
38. Information recorded when archived spreadsheets are catalogued.							
	a.	I do not catalog	950			59.50%	
	b.	Creator	213			13.30%	
	c.	Version	322			20.20%	
	d.	Title	454			28.40%	
	e.	Date	455			28.50%	
	f.	Department	86			5.40%	
39. Archived spreadsheets serve as reference base for subsequent creators /users.							
	a.	Seldom, if ever	623			40.40%	
	b.	Occasionally	591			38.30%	
	c.	Frequently	195			12.60%	
	d.	Don't know	134			8.70%	
40. Frequency of using archived spreadsheets.							
	a.	Selfdom, if ever	684			44.60%	
	b.	Occasionally	705			46.00%	
	c.	Frequently	145			9.50%	
41. Specific problems encountered with the creation or use of spreadsheets.							
	Note:	This is an open-ended question not included in this summary					

42. Practices particularly helpful to you in improving the quality/use of spreadsheets							
	Note:	This is an open-ended question not included in this summary					
Training							
43. Types of training in spreadsheets made available by your organization.							
	a.	None	660			41.30%	
	b.	In-house training	616			38.60%	
	c.	Training by external party	324			20.30%	
	d.	One basic session is available	69			4.30%	
	e.	Several sessions, incl. advanced topics, are available	227			14.20%	
	f.	Spreadsheet specialist who assists designers/users	81			5.10%	
	g.	Other	82			5.10%	
44. Topics covered in the training program offered to you.							
	a.	Basic spreadsheet techniques (for example,	659			41.30%	
		copy and past, simple formulas)					
	b.	Advanced spreadsheet techniques (e.g. use of	613			38.40%	
		built-in functions, conditional formatting					
	c.	Data analysis (sorting, filter, pivot tables)	472			29.60%	
	d.	Use of specialized add-ins and other tools	287			18.00%	
	e.	Macros	225			14.10%	
	f.	Other	109			6.80%	
45. Number of days of training offered to you each year.							
	a.	None	736			52.10%	
	b.	1 or 2 days	365			25.80%	
	c.	3 to 5 days	158			11.20%	
	d.	More than 5 days	155			11.00%	
46. Number of days of training you use each year.							
	a.	None	1044			73.00%	
	b.	1 or 2 days	248			17.30%	
	c.	3 to 5 days	67			4.70%	
	d.	More than 5 days	71			5.00%	
47. The biggest impediments to your participation in company-sponsored training.							
	a.	Not enough time	564			35.30%	
	b.	High cost	137			8.60%	
	c.	Poor quality of training	143			9.00%	
	d.	Lack of personal interest	136			8.50%	
	e.	Lack of support from management	121			7.60%	
	f.	Not applicable	627			39.30%	

48. Incentives offered to you for organization-sponsored training.							
	a.	None	633			39.60%	
	b.	Organization pays cost of training	407			25.50%	
	c.	Organization provides paid time off	135			8.50%	
	d.	Training is a prerequisite for promotion	23			1.40%	
	e.	Not applicable	386			24.20%	
49. Probability of participating in training, if made available in your organization.							
	a.	Probably not	286			20.60%	
	b.	Perhaps	430			30.90%	
	c.	Definitely	276			19.80%	
	d.	Not applicable	399			28.70%	
Standards and Policies							
50. Organization has standards or polices for spreadsheets.							
	a.	No standards	1023			66.40%	
	b.	No written standards, only informal guidelines	362			23.50%	
	c.	Basic written standards	103			6.70%	
	d.	Detailed written guidelines and protocols	53			3.40%	
51. Standards and polices are followed in your organization.							
	a.	Seldom	179			16.10%	
	b.	Usually	320			28.80%	
	c.	Always	67			6.00%	
	d.	Don't know	546			49.10%	
52. Impediments to following the standards offered by your organization.							
	a.	No impediments	339			21.20%	
	b.	Too stringent	24			1.50%	
	c.	Lack of spreadsheet knowledge	126			7.90%	
	d.	No incentives	103			6.40%	
	e.	No enforcement	183			11.50%	
	f.	Others do not follow the standards	95			5.90%	
	g.	I don't understand the standards	30			1.90%	
	h.	Not applicable	705			44.10%	
Risk Management							
53. Importance of spreadsheets to your organization as a whole.							
	a.	Unimportant	51			3.30%	
	b.	Moderately important	406			26.30%	
	c.	Very important	589			38.20%	
	d.	Critical	495			32.10%	

54. Level of risk spreadsheets pose in your organization.							
	a.	High risk	252			16.60%	
	b.	Medium risk	580			38.30%	
	c.	Low risk	553			36.50%	
	d.	No risk	130			8.60%	
55. Awareness of your organization of the risk of spreadsheets							
	a.	Full awareness	294			19.50%	
	b.	Some awareness	819			54.20%	
	c.	No awareness	397			26.30%	
56. Awareness of spreadsheet risk in your organization since SOX legislation							
	a.	Yes	196			12.90%	
	b.	No	541			35.60%	
	c.	Don't know	783			51.50%	
57. Strategies in place in your organization to mitigate the risk from spreadsheets.							
	a.	Yes	284			18.70%	
	b.	No	601			39.60%	
	c.	Don't know	632			41.70%	
58. Person in organization responsible for managing the risks from spreadsheets.							
	a.	The developer	297			19.90%	
	b.	The user	231			15.40%	
	c.	The manager	162			10.80%	
	d.	Don't know	711			47.50%	
	e.	Other	95			6.40%	
59. Spreadsheet audit packages used in your organization.							
	a.	Yes	44			2.90%	
	b.	No	1211			80.00%	
	c.	Don't know	259			17.10%	
Personal Information							
60. Your gender							
	a.	Male	1293			83.30%	
	b.	Female	260			16.70%	

61. Your age							
	a.	20-30	213			13.70%	
	b.	31-40	601			38.50%	
	c.	41-50	408			26.20%	
	d.	51-60	230			14.70%	
	e.	Over 60	108			6.90%	
62. Your highest level of education							
	a.	High School	57			3.70%	
	b.	Undergraduate	177			11.40%	
	c.	Masters	1153			74.10%	
	d.	Ph.D.	169			10.90%	
63. Your position in your organization							
	a.	Non-manager	378			23.70%	
	b.	Supervisor or manager	502			31.40%	
	c.	Executive	516			32.30%	
	d.	Other	196			12.30%	
64. Your organization would best be categorized as -							
	a.	Government	30			2.00%	
	b.	Manufacturing	291			19.10%	
	c.	Service (e.g. banking, retail, consulting)	709			46.60%	
	d.	Agriculture and natural resources	69			4.50%	
	e.	Education	121			8.00%	
	f.	Health/medicine	46			3.00%	
	g.	Other Non-Profit	34			2.20%	
	h.	Other	221			14.50%	
65. Number of employees in your organization							
	a.	10-Jan	234			15.20%	
	b.	11-50	177			11.50%	
	c.	51-100	95			6.20%	
	d.	101-500	201			13.00%	
	e.	501-1000	108			7.00%	
	f.	Over 1000	727			47.10%	

66. Functional area of your job.							
	a.	Sales	54			3.60%	
	b.	Marketing	164			10.90%	
	c.	Operations/Manufacturing	142			9.50%	
	d.	Distribution	12			0.80%	
	e.	Engineering	135			9.00%	
	f.	Research	162			10.80%	
	g.	Finance	454			30.20%	
	h.	Human Resources	20			1.30%	
	i.	Other	358			23.90%	
67. Number of people reporting directly to you.							
	a.	None	646			41.70%	
	b.	02-Jan	347			22.40%	
	c.	05-Mar	263			17.00%	
	d.	10-Jun	180			11.60%	
	e.	11-50	86			5.50%	
	f.	More than 50	29			1.90%	

Appendix D – Figures (Taxonomies, Tables, Charts)

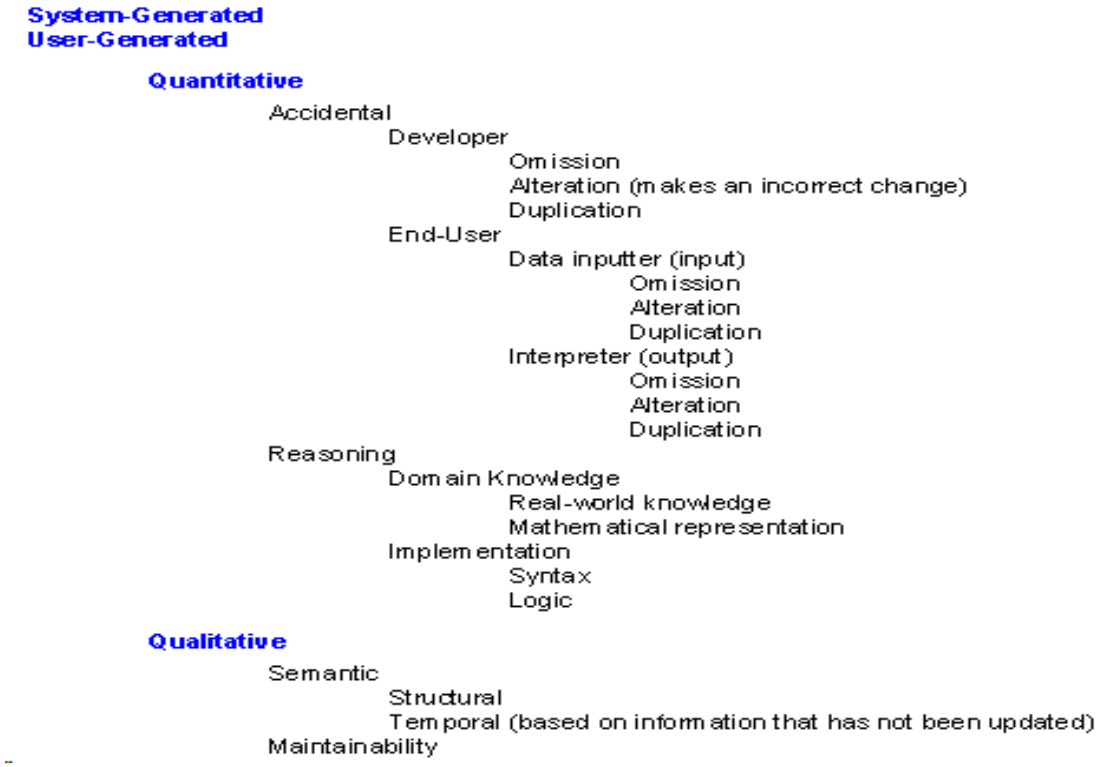
Taxonomies

Figure 9:Przasnyski, Leon and Seal (2011) Taxonomy for defining qualitative errors

Category	Type of Error	Record the error as 1 ("exists") if:	Explanatory Example
Input Data Structure	Hard-coding/jamming values into formulas	At least one value is hard-coded into a formula somewhere in the model	1-1
	Duplication of Input Values	• There is more than one place to enter the same value in the model • An intermediate calculated value must be entered directly as an input in another part of the model	1-2
	Input Cells Not Clearly Identified	One or more input cells anywhere in the spreadsheet are not clearly identified	1-3
Semantics	Missing Cell Documentation	At least one cell where a number is displayed is not labeled	2-1
	Incorrect Cell Documentation	At least one cell has a label that is clearly incorrect (e.g., refers to total profit when it is total cost; or states wrong units)	2-1
	Ambiguous Cell Documentation	At least one cell **that has not been identified** as clearly incorrect has an ambiguous label associated with it (e.g., no units are specified, general labels such as "costs" are used)	2-1
	Poor Layout for Readability	The spreadsheet is generally difficult to read/follow for any of the following reasons: • Does not read left to right and top to bottom • Related formulas are not in physical proximity of each other • Excessive blank spaces are used to create long arcs of precedence • The logic of the business problem is difficult to follow	2-2
Extendibility	Poor Layout for Model Extension	• Inserting a row or column will hurt the readability of the model or • Extending the model will require multiple insertions and manipulation of the model	3-1
	Poor Layout for Copy/Paste	Layout of the logic of the model does not allow for easy copying to a new column/row	3-2
	Poor Absolute/Relative Cell References for Copy/Paste	At least one cell included in a layout that allows for the cell logic to be easily copied to another row or column contains a formula with incorrect use or lack of use of the $ notation	3-3
Formula Integrity	Spurious Formulas	A confusing or spurious entry in a formula exists	4-1
	Lack of Explicit Formulas	An inputted number (jammed or entered in an input cell) is the result of a calculation not carried out in the spreadsheet (even if some of the numbers used to calculate the input value were generated in the spreadsheet)	4-2

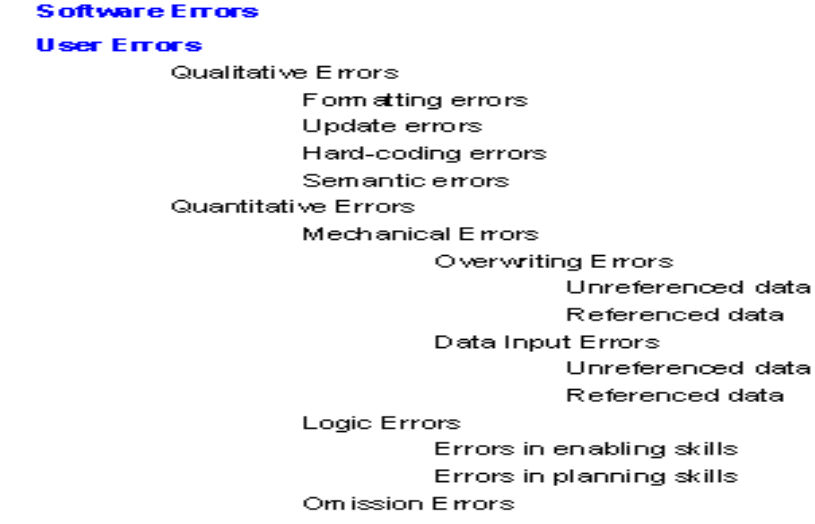
Retrieved from http://arxiv.org/ftp/arxiv/papers/1111/1111.6909.pdf

Figure 10: The Rajalingham, Chadwick, Knight and Edwards (2000) Taxonomy



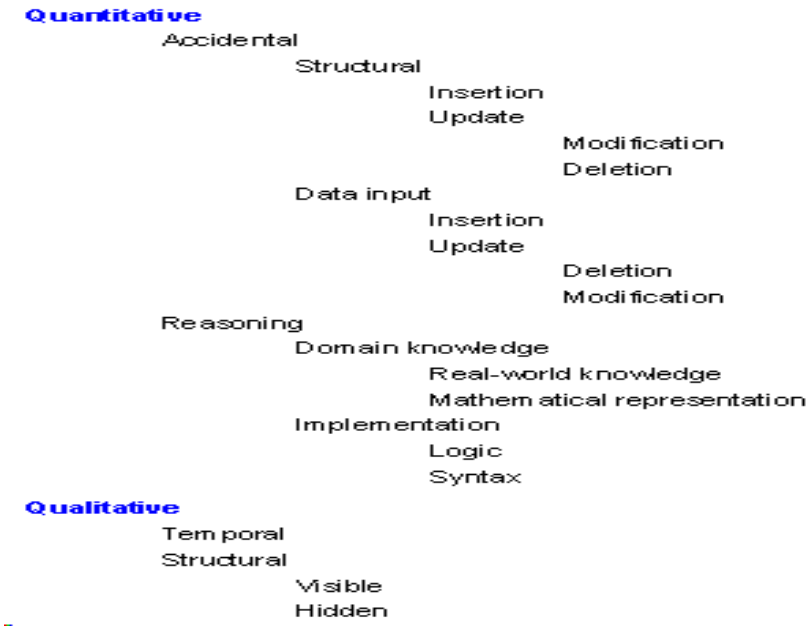
Retrieved from http://arxiv.org/ftp/arxiv/papers/0809/0809.3613.pdf

Figure 11: The Rajalingham's (2005) "Bushy" Taxonomy



Retrieved from http://arxiv.org/ftp/arxiv/papers/0809/0809.3613.pdf

Figure 12: The Rajalingham's (2005) "Binary" Taxonomy



Retrieved from http://arxiv.org/ftp/arxiv/papers/0809/0809.3613.pdf

Figure 13: Howe and Simkin (2008) Taxonomy

Type of Error	Seeded Errors	Percentage Found	Description
Data Entry Errors	5	72%	Out of range values, negative values, one value entered as a label
Clerical and Non-Material Errors	10	66%	Spelling errors, incorrect dates, etc.
Rules Violations	3	60%	Cell entries which violate a stated company policy for an ineligible employee
Formula Errors	25	54%	Inaccurate range references, embedded constants, illogical formulas
Total Errors	43	67%	

Results of their study are included in the above table
Retrieved from http://arxiv.org/ftp/arxiv/papers/0809/0809.3613.pdf

Figure 14: Powell, Lawson and Baker (2008) Taxonomy

Error Type	Description
Logic	Formula is used incorrectly, leading to an incorrect result.
Reference	A formula contains one or more incorrect references to other cells.
Hard-Coding	One or more numbers appear in formulas, and the practice is sufficiently dangerous.
Copy/Paste	A formula is wrong do to an incorrect cut and paste.
Data Input	An incorrect data input is used.
Omission	A formula is wrong because one of its input cells is blank.

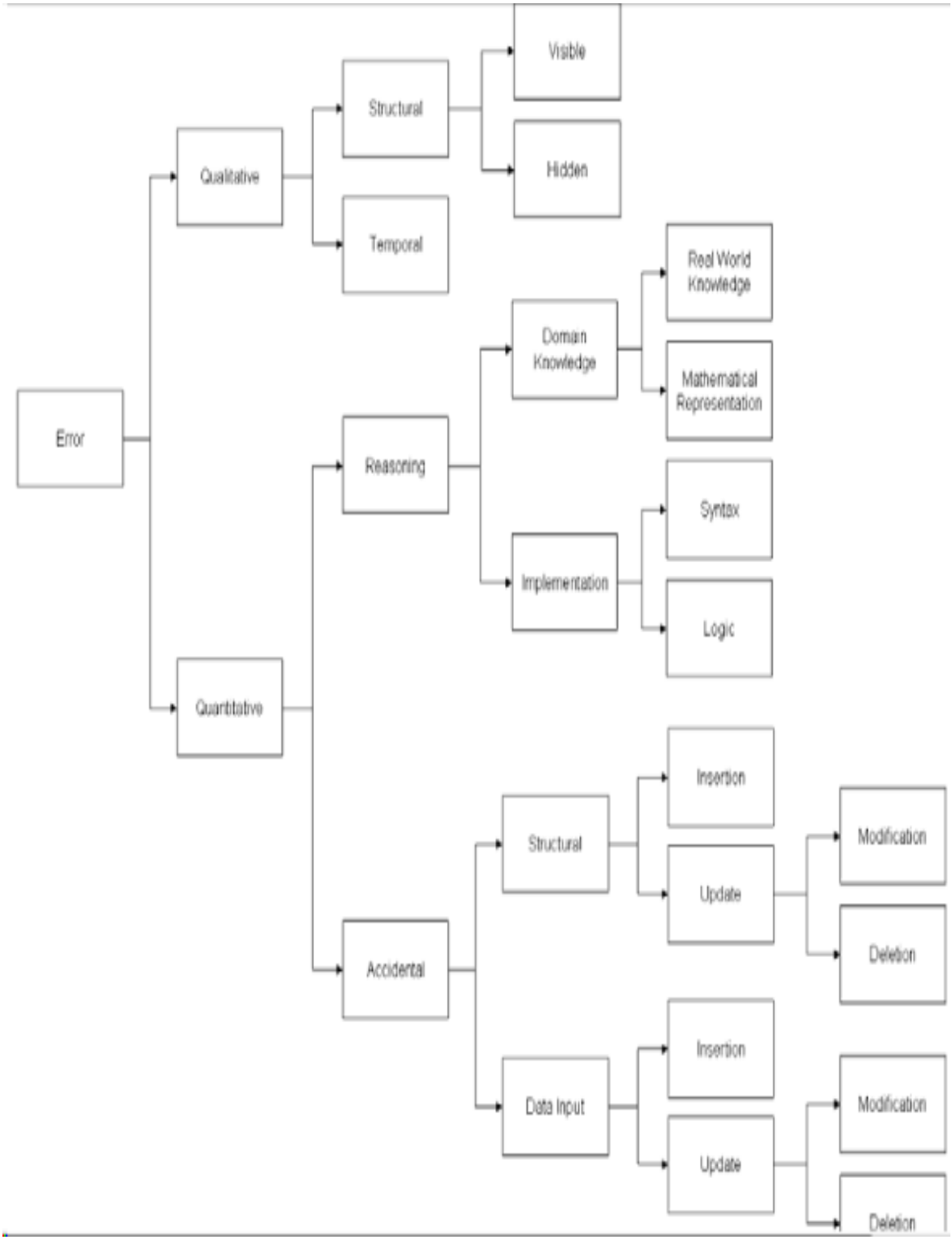
Retrieved from http://arxiv.org/ftp/arxiv/papers/0809/0809.3613.pdf

Figure 15: Madahar, Cleary and Ball (2008) Taxonomy of Spreadsheets

Dimension	Description
Dependency	How fundamentally the organization depends on the spreadsheet. Values can be operational, tactical, or strategic
Magnitude	The severity of consequences for potential errors
Time/Urgency	Deadlines that have to be met using the spreadsheet

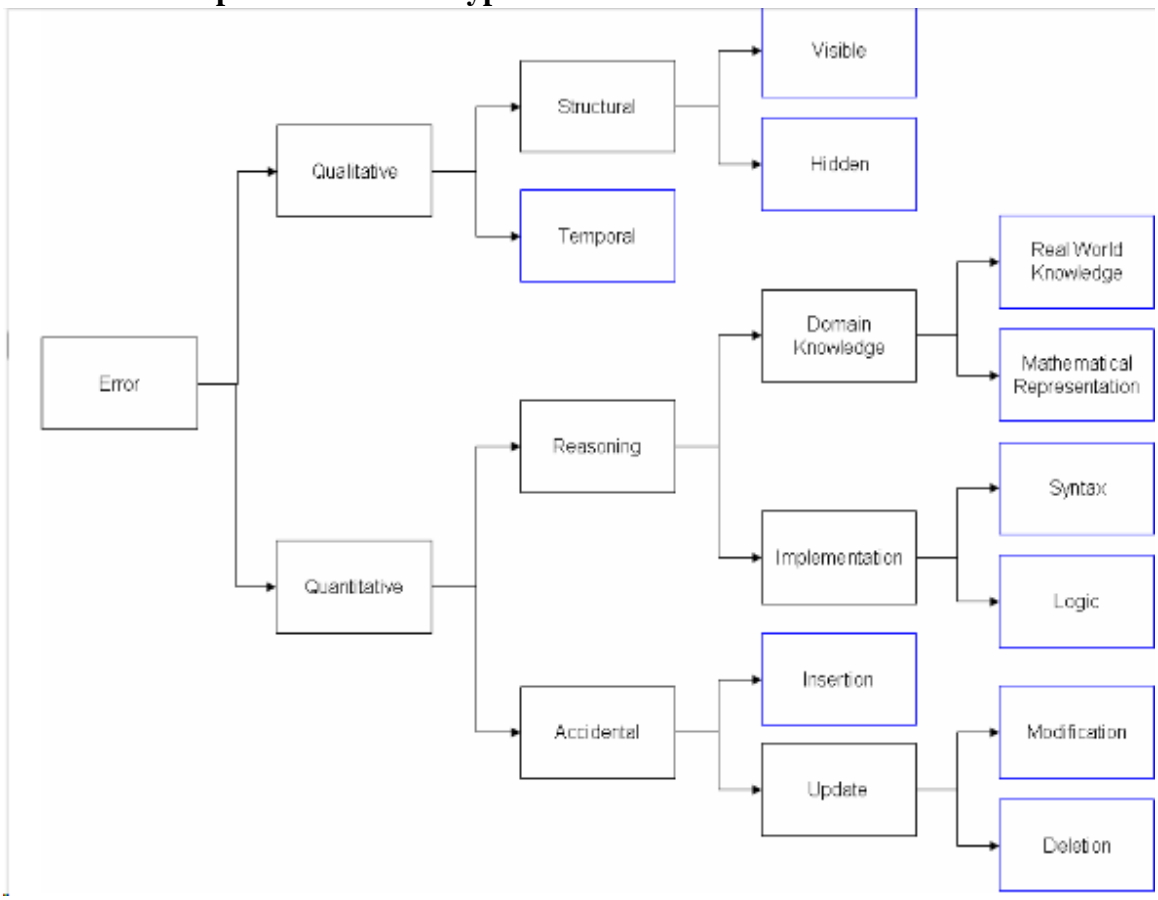
Retrieved from http://arxiv.org/ftp/arxiv/papers/0809/0809.3613.pdf

Figure 16: Rajalingham's (2005) revised classification of spreadsheet error-types



Retrieved from http://arxiv.org/ftp/arxiv/papers/0803/0803.0167.pdf

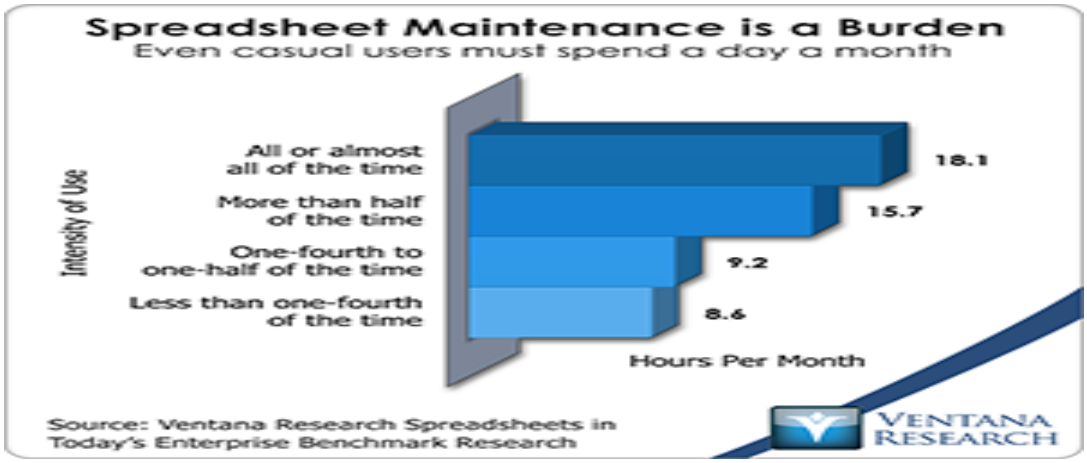
Figure 17: A revision of Rajalingham's (2005) revised classification of spreadsheet error-types



Retrieved from http://arxiv.org/ftp/arxiv/papers/0803/0803.0167.pdf

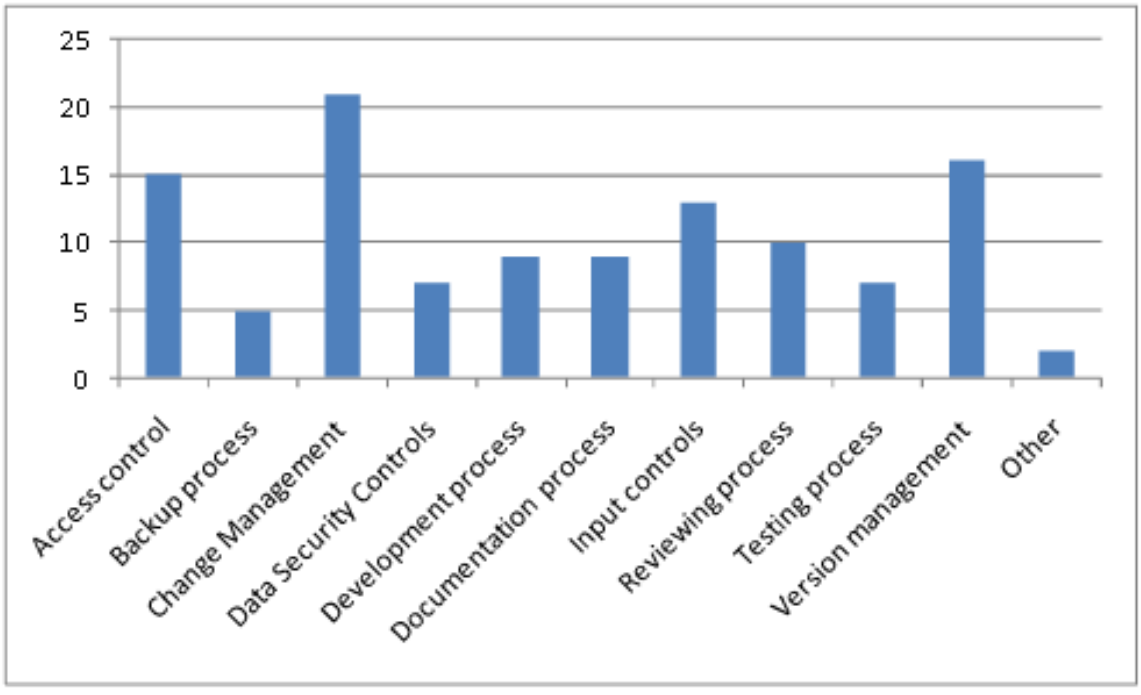
1. The error-types included in the above investigation were:
    - Insertion errors – as this represents the group that includes omissions, duplications, and typos.
    - Modification errors – as this represents the group that includes overwriting values or incorrect modifying a formula.
    - Deletion errors – as this represents the group that includes erasing values or formula.
    - Logic errors – as this represents the group that includes using absolute and relative references, or inserting a row of into range that is summed such that the sum does not include the new value.
    - Temporal errors – as this represents the group that includes values or formulae that are accurate only for a given period.
    - Structural Hidden errors – as this represents the group that includes errors that require an examination of formulae such as hard-coded values in formulae arguments.
    - Structural Visible errors – as this represents the group that includes errors that do not require an examination of formulae.

Figure 18: **Kugel (2013) Spreadsheet Maintenance – *Ventana Research***



Retrieved from http://robertkugel.ventanaresearch.com/2013/03/01/spreadsheet-denial-is-a-big-issue/

Figure 19: **Coster, Leon, Kalbers and Abraham (2011) Top 3 Areas of Difficulty for Implementing Controls**



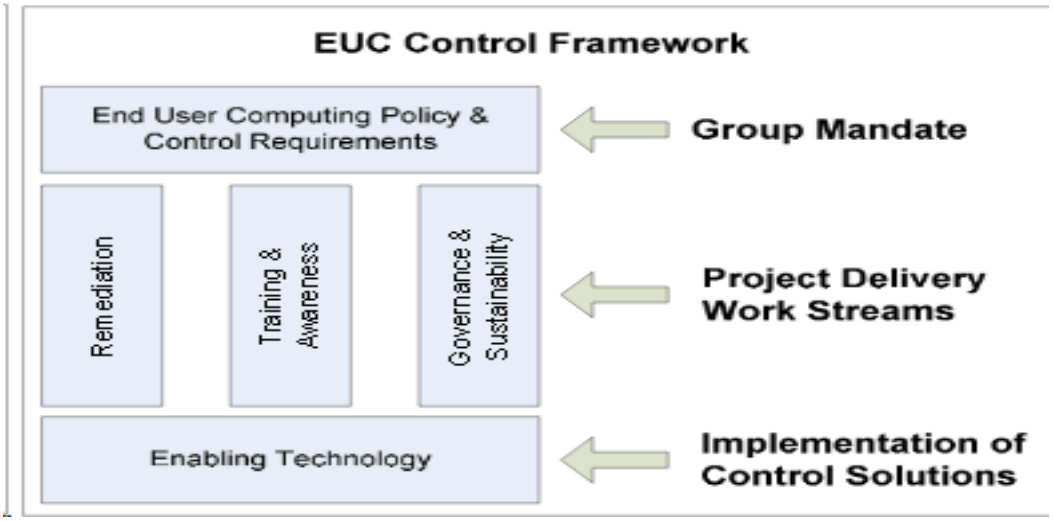
Retrieved from http://arxiv.org/pdf/1111.6887v1.pdf

Figure 20: Coster et al. (2011) Internal Controls Organizations Considered for Implementation

Internal Controls or Tools	Percent of Companies that Currently Implement Tool	Percent of Companies that Plan to Implement Tool in Future
Files secured in drives & server folders with limited access	76.9%	11.5%
Logically structured directories/folders for business units, cycles, and type of spreadsheets	65.4%	11.5%
Formal review process	57.7%	23.1%
Input controls that ensure data integrity	57.7%	15.4%
Password required to update spreadsheet	57.7%	15.4%
Cell protection (required)	50.0%	15.4%
More than one person responsible for data and maintenance	46.2%	3.8%
Independent review groups	23.1%	19.2%
Excel Track Changes (required)	19.2%	26.9%
spreadsheet computing policy stating design standards	11.5%	15.4%
Mandated training for developers	7.7%	3.8%
Third party auditing software	7.7%	15.4%
spreadsheet data consolidated into databases managed by IT	7.7%	11.5%
Third party tools for access, version, change, and archive support	3.8%	11.5%
spreadsheet converted into server-based application	0.0%	11.5%
No Stated Plans	-	26.9%

Retrieved from http://arxiv.org/pdf/1111.6887v1.pdf , http://arxiv.org/pdf/1111.6887.pdf

Figure 21: Lemon and Ferguson (2010) End User Computing Controls Framework



Retrieved from: http://arxiv.org/ftp/arxiv/papers/1009/1009.1404.pdf

Figure 22: Lemon and Ferguson (2010) Design Standards

"Design standards:

A selection of design standards were established and mandated for the most significant categories of EUC application. Standards were written for Excel spreadsheets and Access databases, although, as we've seen in most organisations, Excel spreadsheets were the more prevalent. The design standards covered the following principles:
- Improved documentation [Payette, 2006] – achieved by requiring the completion of standard documentation templates in all applications and including appropriate commentary to explain complex calculations and VBA code;
- Transparency of information – achieved by making data and calculations visible and clearly understood;
- Clear labelling – achieved by ensuring key data inputs, calculations, outputs, assumptions and units of measure are all adequately labelled;
- Separation of inputs, calculations and outputs – achieved through a combination of structural separation and visual formatting and labelling; and
- Critical cell locking – achieved by locking all cells with critical formulas and static data and activating worksheet protection.

Figure 23: Internal Controls

"Internal controls can be detective, corrective, or preventive by nature.

1. Detective controls are designed to detect errors or irregularities that may have occurred.
2. Corrective controls are designed to correct errors or irregularities that have been detected.
3. Preventive controls, on the other hand, are designed to keep errors or irregularities from occurring in the first place. "

Examples of Spreadsheet controls are:

- Change Control
  - Maintain a process for requesting changes to a spreadsheet, making changes, testing and obtaining formal sign-off from an independent individual that the change is functioning appropriately
- Version Control
  - Ensure only current and approved versions of spreadsheets are being used by creating naming conventions, directory structures and access control
- Input Control
  - Ensure that data is input completely and accurately and that it is current and secure
- Documentation
  - Ensure that it is up-to-date and communicates the business objective and specific functions of the spreadsheet

Retrieved from http://intraweb.stockton.edu/eyos/internal_audit/content/docs/icnote2.pdf

Figure 24: The SDLC traditionally divides the Project up into several solution-centric phases:

- Situation Analysis
- Business Needs Assessment
- Requirements Definition
- Solution Design
- Solution Prototyping
- Solution Construction
- Solution Testing
- Solution Deployment

Retrieved from http://vogtland.ws/markedwardvogt/?p=806

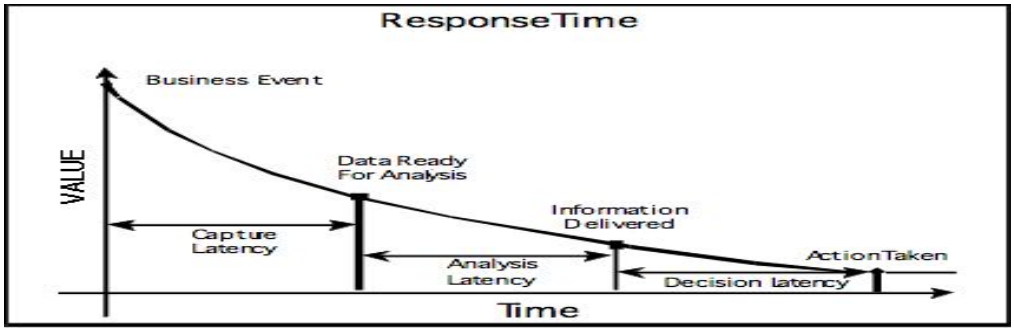
Figure 25: Panko (2007) Types of Testing

Panko (2007) identifies multiple types of testing:

- Test during development and separately through each phase.
- Requirements testing: Many errors are introduced before coding ever begins.
- Unit testing: After a developer has finished a module and implemented their own self check, the module must be subjected to unit testing.
- Integration testing: After modules are tested, they are integrated into larger units. Usually several stages of integration are needed, each with its own techniques for unit testing. According to Fagan (1976, 1986) testing by one individual will only catch 50% to 60% of errors. Team inspection can raise the detection rate to 80%.
- Agile development methods: It is assumed that a traditional software development life cycle "SDLC" model is employed. Spreadsheet development maybe done in non-traditional ways, especially agile methods.
- Eyeballing: One testing technique is looking over the spreadsheet for reasonableness or having a colleague check a spreadsheet. There is no evidence that eyeballing reduces error rates.
- Error scanning software: Excel 2003 has a built in error checking tool under the tools menu. This is simple but limited. Error checking software products such as SpACE, Comply XL, Cluster Seven and Acuate which can tend to assist in locating errors within complex spreadsheets. However, they would never be able to detect a quantitative error such as an omission.
- Auditing: In auditing, an auditor does not examine everything; they ask questions whose answers may indicate problems. Auditing will only perform spot checks and the goal of auditing is to detect indications of problems and not to reduce errors.

Retrieved from http://arxiv.org/ftp/arxiv/papers/1009/1009.2775.pdf

Figure 26: Ross (2007) Decision Latency



Retrieved from http://www.brcommunity.com/b373.php

Supplements

INFORMED CONSENT FORM

Research Case Study titled "Practicing Safe Spreadsheeting"

Investigator Name: Scott Laing

Department of Accounting
Saint Mary's University, 923 Robie Street,
Halifax, NS B3H 3C3
Phone # 902-494-1819

Email address: scott.laing@smu.ca

Supervisor Name: Prof. David Bateman
Department of Accounting

Saint Mary's University, 923 Robie Street,

Halifax, NS B3H 3C3
Phone # 902-420-5623; Fax # 902-420-5011
Email address: david.bateman@smu.ca

The purpose and aim of the study is to investigate the use of spreadsheets by a private sector organization with the intent to determine the frequency of their use, the occurrence of errors found in them and the severity of those errors. Upon completion of this study the results will be compared to the findings of other previous empirical research studies conducted by other researchers in other countries to determine if the findings of this study are consistent with those of the other independent studies. Additionally the findings from this research study may also be used to further future research studies. Recommendations and/or guidelines for improvements to spreadsheet management practices will be provided to all participants.

The location of the case study research will be on site at NTTDatas', Cogswell Tower offices, in Halifax, NS.
The data will be collected via a one-on-one interview between me (the principal investigator), and you (the participant). Each interview will be approximately 30 - 45 minutes in duration and will consist of a series of interview questions developed specifically for this case study. The data may be recorded both in written/electronic as well as audio form. All data responses provided by each participant will be kept both confidential and anonymous for the purposes of the final report findings. We want to make it clear that you as the participant are not required to answer any questions that you do not wish to answer, although we ask that you try to complete the interview as thoroughly and honestly as possible and you can withdraw from the study at any time by simply providing written notice to me.

Some sample interview questions include the following:

- What is the approximate percent of time spent with spreadsheets in your job? 0% - 100%
- What is the level of importance spreadsheets have in your job? Low – high
- What is the number of different spreadsheets that you normally use per week? 0 – 10+

The interviews will take place during the month of June, 2015 at a date/time convenient for you. There are no known risks for any participant in this research case study, although there is the possibility of some discovery of benefits for the participant and their

employer as a result of this case study. The potential direct/indirect benefits of the research to a) the participant include a brief summative report of the findings of the case study along with some recommendations based on the findings b) the field of science include the validation of some of the previous independent empirical research findings, and c) to society identification of possible new areas of research.

The data will be collected and analyzed by the primary Investigator and the Supervisor to prepare a summative report of the findings of the research study (a copy of which will be provided to the participant). The results of the research findings may be published in an academic or industry journal, professional magazine or other publication. The participants name and the name of their employer will be kept anonymous. The data will be kept secure on a local drive (and external USB flash drive) that are password protected and after a period of no more than **4 years** the data and any drives used to store the data will be physically destroyed.

Once all the data is collected and analyzed for this study, the information and findings may be shared with others as appropriate (*e.g. the research community through seminars, conferences, presentations, journal / industry magazine articles*).

If at any time the participant wants to find out more information about the study they can contact the Supervisor

Professor David Bateman
Saint Mary's University
Phone: 902-420-5623
Email: david.bateman@smu.ca

Certification:

The Saint Mary's University Research Ethics Board has reviewed this research. If you have any questions or concerns about ethical matters or would like to discuss your rights as a research participant, you may contact the Chair of the Research Ethics Board at ethics@smu.ca or 420-5728.

Signature of Agreement:

Research Case Study titled "Practicing Safe SpreadSheeting"

REB file #15-289

I understand what this study is about, appreciate the risks and benefits, and that by consenting I agree to take part in this research study and do not waive any rights to legal recourse in the event of research-related harm.

I understand that my participation is voluntary and that I can end my participation at any time without penalty.

I have had adequate time to think about the research study and have had the opportunity to ask questions.

Participant

Signature :_____Name (Printed) :_____

Date :_____ (Day/Month/Year)

Principal Investigator

Signature: _____ Name (Printed) :_____Scott Laing_____

Date:__10-June-2015_____ (Day/Month/Year)

Please keep one copy of this form for your own records.
(Participants must be provided with a copy of the signed Informed Consent Form.)

Bibliography

Allen, S., (2012). On the Interpretation of Spreadsheets within their Environment, *European Spreadsheet Risks Interest Group. Proceedings of the EuSpRIG 2012 Conference*, pp. 1-10.
Retrieved from http://arxiv.org/pdf/1209.3518v1.pdf

Ansari, S., & Block, R. (2008). Spreadsheet "Worst Practices". *CFO*.
Retrieved from http://ww2.cfo.com/technology/2008/05/spreadsheet-worst-practices/

Aurigemma, S., & Panko, R. R. (2010). The Detection of Human Spreadsheet Errors by Humans versus Inspection (Auditing) Software. *European Spreadsheet Risks Interest Group. Proceedings of the EuSpRIG 2010 Conference*, pp. 1-14.
Retrieved from http://arxiv.org/ftp/arxiv/papers/1009/1009.2785.pdf

Baker, K., Powell, S., Lawson, B., (2006). Survey & Results. The Spreadsheet Research Engineering Project. Retrieved from http://faculty.tuck.dartmouth.edu/serp/, http://faculty.tuck.dartmouth.edu/images/uploads/faculty/serp/serp_results.pdf

Baker, K., Powell, S., Lawson, B., Foster-Johnson, L., (2006). Comparison of Characteristics and Practices Among Spreadsheet Users With Different Levels of Experience. *European Spreadsheet Risks Interest Group. Proceedings of the EuSpRIG 2006 Conference*, pp. 205 - 220.
Retrieved from http://arxiv.org/ftp/arxiv/papers/0803/0803.0168.pdf

Balson, D., (2010). Changing User Attitudes to Reduce Spreadsheet Risk. *European Spreadsheet Risks Interest Group. Proceedings of the EuSpRIG 2010 Conference,* pp. 1-5. Retrieved from http://arxiv.org/ftp/arxiv/papers/1009/1009.5701.pdf

Baskarada, S., (Spring, 2012). How Spreadsheet Applications Affect Information Quality. *Journal Of Computer Information Systems*, Vol. 51(2), pp. 77-84. Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2101601

Bekenn, B., Hooper, R., (2009). Some Spreadsheet Poka-Yoke. *European Spreadsheet Risks Interest Group. Proceedings of the EuSpRIG 2009 Conference,* pp 1-12. Retrieved from http://arxiv.org/ftp/arxiv/papers/0908/0908.0930.pdf

Bekenn, B., Hooper R., (2011). Workbook Structure Analysis - "Coping with the Imperfect", *European Spreadsheet Risks Interest Group. Proceedings of the EuSpRIG 2011 Conference,* pp 1-12. Retrieved from http://arxiv.org/pdf/1111.6858v1.pdf

Bishop, B., McDaid, K. (2007). An Empirical Study of End-User Behaviour in Spreadsheet Error Detection and Correction. *European Spreadsheet Risks Interest Group. Proceedings of the EuSpRIG 2007 Conference,* pp. 165-176. Retrieved from http://arxiv.org/ftp/arxiv/papers/0802/0802.3479.pdf

Chadwick, D., (2007). Establishing a Minimum Generic Skill Set for Risk Management Teaching in a Spreadsheet Training Course. *European Spreadsheet Risks Interest Group. Proceedings of the EuSpRIG 2007 Conference,* pp. 197-208. Retrieved from http://arxiv.org/ftp/arxiv/papers/0802/0802.3481.pdf

Chambers, C., Scaffidi, C., (2010). Struggling to Excel: A Field Study of Challenges Faced by Spreadsheet Users. 2010 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC),September, *pp. 1-8.* Retrieved from http://web.engr.oregonstate.edu/%7Ecscaffid/papers/eu_20101024_sstrouble.pdf

Chambers, C., Erwig, M., Luckey, M., (2010). SheetDiff: A Tool for Identifying Changes in Spreadsheets. 2010 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC), September, *pp 1-8.* Retrieved from http://web.engr.oregonstate.edu/%7Eerwig/papers/SheetDiff_VLHCC10.pdf

Colver, D., (2004). Spreadsheet Good Practice: is there any such thing? *European Spreadsheet Risks Interest Group. Proceedings of the EuSpRIG 2004 Conference,* pp. 1-9. Retrieved from http://arxiv.org/ftp/arxiv/papers/1001/1001.3967.pdf

Colver, D., (2011). Drivers of the Cost of Spreadsheet Audit, *European Spreadsheet Risks Interest Group. Proceedings of the EuSpRIG 2011 Conference,* pp. 1-8. Retrieved from http://arxiv.org/pdf/1111.5002v1.pdf

Coster, N., Leon, L., Kalbers, L., Abraham, D., (2011). Controls over Spreadsheets for Financial Reporting in Practice. *European Spreadsheet Risks Interest Group. Proceedings of the EuSpRIG 2011 Conference,* pp. 1-13.
Retrieved from http://arxiv.org/pdf/1111.6887v1.pdf

Croll, G.J., (2005). The Importance and Criticality of Spreadsheets in the City of London. *European Spreadsheet Risks Interest Group. Proceedings of the EuSpRIG 2005 Conference,* pp. 82-92.
Retrieved from http://arxiv.org/ftp/arxiv/papers/0709/0709.4063.pdf

Croll, G.J., (2009). Spreadsheets and the Financial Collapse. *European Spreadsheet Risks Interest Group. Proceedings of the EuSpRIG 2009 Conference,* pp, 145-161.
Retrieved from http://arxiv.org/pdf/0908.4420v1.pdf

Croll, G.J., (2011). An Insight into Spreadsheet User Behaviour through an Analysis of EuSpRIG Website Statistics, *European Spreadsheet Risks Interest Group. Proceedings of the EuSpRIG 2011 Conference,* pp, 129-139.
Retrieved from http://arxiv.org/pdf/1111.4692v1.pdf

Croll, G.J. (2012). Spreadsheets and Long Term Corporate Survival, *European Spreadsheet Risks Interest Group. Proceedings of the EuSpRIG 2012 Conference,* pp. 77-96.
Retrieved from http://arxiv.org/pdf/1207.1921v1.pdf

Drew, J., (April 13, 2015). Beyond spreadsheets: Technology round table, part 1, *Journal of Accountancy*. Retrieved from http://www.journalofaccountancy.com/issues/2015/apr/accounting-technology-round-table-part-1.html

Dunn, A., (2010). Spreadsheets - the Good, the Bad and the Downright Ugly. *European Spreadsheet Risks Interest Group. Proceedings of the EuSpRIG 2010 Conference,* pp. 157-164.
Retrieved from http://arxiv.org/ftp/arxiv/papers/1009/1009.5705.pdf

Grossman, T.A., Ozluk, O., (2010). Spreadsheets Grow Up: Three Spreadsheet Engineering Methodologies for Large Financial Planning Models, *European Spreadsheet Risks Interest Group. Proceedings of the EuSpRIG 2010 Conference,* pp. 1-15.
Retrieved from http://arxiv.org/ftp/arxiv/papers/1008/1008.4174.pdf

Grossman, T.A., Mehrotra, V., Sander, J., (2011). Towards Evaluating the Quality of a Spreadsheet: The Case of the Analytical Spreadsheet Model, *European Spreadsheet Risks Interest Group. Proceedings of the EuSpRIG 2011 Conference,* pp, 1-10.
Retrieved from http://arxiv.org/pdf/1111.6907v1.pdf

Guthrie, G., McCrory, S. , (2011). Beyond The Desktop Spreadsheet, *European Spreadsheet Risks Interest Group. Proceedings of the EuSpRIG 2011 Conference,* pp. 1-11. Retrieved from http://arxiv.org/pdf/1111.6870v1.pdf

Hermans, F., Pinzger, M., van Deursen, A., (2012). Measuring Spreadsheet Formula Understandability. *European Spreadsheet Risks Interest Group. Proceedings of the EuSpRIG 2012 Conference,* pp. 77-96.
Retrieved from: http://arxiv.org/pdf/1209.3517v1.pdf

Izza, M., (June 17, 2014). 20 principles for better spreadsheets. *ICAEW.*
Retrieved from http://www.ion.icaew.com/MoorgatePlace/post/20-principles-for-better-spreadsheets

Kernochan, W., (July 10, 2014). Business Intelligence 101:A Brief History, *Enterpriseappstoday.com,* Retrieved from http://www.enterpriseappstoday.com/business-intelligence/business-intelligence-101-a-brief-history.html

Kerr, C.A., (2012). Hands-Off Spreadsheets, *European Spreadsheet Risks Interest Group Proceedings of the EuSpRIG 2012 Conference,* pp. 1-10.
Retrieved from http://arxiv.org/pdf/1210.5262v1.pdf

Kugel, R., (March 1, 2013). Spreadsheet Denial is a Big Issue, *Ventena Research.*
Retrieved from http://robertkugel.ventanaresearch.com/2013/03/01/spreadsheet-denial-is-a-big-issue/

Kulesz, D., (2011). From Good Practices to Effective Policies for Preventing Errors in Spreadsheets. *European Spreadsheet Risks Interest Group. Proceedings of the EuSpRIG 2011 Conference,* pp. 1-11. Retrieved from http://arxiv.org/pdf/1111.6878v1.pdf

Kulesz, D., Ostberg, J-P., (2013). Practical Challenges with Spreadsheet Auditing Tools, *European Spreadsheet Risks Interest Group. Proceedings of the EuSpRIG 2013 Conference,* pp. 1-13.

Retrieved from http://arxiv.org/ftp/arxiv/papers/1401/1401.7583.pdf

Lemon, T., Ferguson, E., (2010). A Practical Approach to Managing Spreadsheet Risk in a Global Business, *European Spreadsheet Risks Interest Group. Proceedings of the EuSpRIG 2010 Conference,* pp. 1-8.

Retrieved from http://arxiv.org/ftp/arxiv/papers/1009/1009.1404.pdf

Mittermeir, R.T., Clermont, M., Hodnigg, K., (2005). Protecting Spreadsheets against Fraud. *European Spreadsheet Risks Interest Group. Proceedings of the EuSpRIG 2005 Conference,* pp. 69-80.

Retrieved from http://arxiv.org/ftp/arxiv/papers/0801/0801.4268.pdf

Murphy, S., (2007). Spreadsheet Hell, *European Spreadsheet Risks Interest Group. Proceedings of the EuSpRIG 2007 Conference,* pp. 15-20.

Retrieved from http://arxiv.org/ftp/arxiv/papers/0801/0801.3118.pdf

O'Beirne, P., (2008). Information and Data Quality in Spreadsheets. *European Spreadsheet Risks Interest Group. Proceedings of the EuSpRIG 2008 Conference,* pp. 171-185.

Retrieved from http://arxiv.org/pdf/0809.3609v1.pdf

Panko, R., (1998, 2008). What We Know About Spreadsheet Errors, *Journal of End User Computing's Special issue on Scaling Up End User Development Volume 10, No 2. Spring 1998, pp. 15-21 - revised 2008 para. 88-90.*
Retrieved from http://panko.shidler.hawaii.edu/SSR/Mypapers/whatknow.htm

Panko, R. (2007) Thinking is Bad: Implications of Human Error Research for Spreadsheet Research and Practice. *European Spreadsheet Risks Interest Group. Proceedings of the EuSpRIG 2007 Conference,* pp. 69-80.
Retrieved from http://arxiv.org/ftp/arxiv/papers/0801/0801.3114.pdf

Panko, R., (1996; 2008). Revisiting the Panko-Halverson Taxonomy of Spreadsheet Errors. *European Spreadsheet Risks Interest Group. Proceedings of the EuSpRIG 2008 Conference,* pp. 199-220.
Retrieved from http://arxiv.org/ftp/arxiv/papers/0809/0809.3613.pdf

Panko, R., (2009). A Framework for Spreadsheet Research and Practice. Retrieved from http://panko.shidler.hawaii.edu/SSR/Framework%20Project/Framework%20Levels.pdf

Persico, J., (2011, August 9), The Case for QARP. Online Blog. Retrieved from http://finextra.com/blogs/fullblog.aspx?blogid=5641 on August 2, 2015.

Powell, S., Lawson, B., Baker, K. (2007). Impact of Errors in Operational Spreadsheets. *European Spreadsheet Risks Interest Group. Proceedings of the EuSpRIG 2007 Conference,* pp. 57-68.
Retrieved from http://arxiv.org/ftp/arxiv/papers/0801/0801.0715.pdf

Power, D.J., (August 30, 2004). A Brief History of Spreadsheets, *DSSResources.COM*, Retrieved from http://dssresources.com/history/sshistory.html.

Purser, M., Chadwick, D. (2006). Does An Awareness of Differing Types of Spreadsheet Errors Aid End-Users in Identifying Spreadsheet Errors? *European Spreadsheet Risks Interest Group. Proceedings of the EuSpRIG 2006 Conference,* pp. 185-204. Retrieved from http://arxiv.org/ftp/arxiv/papers/0803/0803.0167.pdf

Pryor L., (2003). Correctness is Not Enough, *European Spreadsheet Risks Interest Group. Proceedings of the EuSpRIG 2003 Conference,* pp. 117-122. Retrieved from http://arxiv.org/pdf/0808.2045.pdf

Pryor L., (2004). When, why and how to test spreadsheets. *European Spreadsheet Risks Interest Group. Proceedings of the EuSpRIG 2004 Conference,* pp. 145-151. Retrieved from http://arxiv.org/ftp/arxiv/papers/0807/0807.3187.pdf

Przasnyski, Z., Leon, L., Seal, K.C., (2011). In Search of a Taxonomy for Classifying Qualitative Spreadsheet Errors. *European Spreadsheet Risks Interest Group. Proceedings of the EuSpRIG 2011 Conference,* pp. 1-12. Retrieved from http://arxiv.org/pdf/1111.6909v1.pdf

Rittweger, B.G., Langan, E., (2010). Spreadsheet Risk Management in Organisations, *European Spreadsheet Risks Interest Group. Proceedings of the EuSpRIG 2010 Conference,* pp. 61-72. Retrieved from http://arxiv.org/ftp/arxiv/papers/1009/1009.2775.pdf

Ross, R., (November 2007). The Latency of Decisions ~ New Ideas on the ROI of Business Rules, *Business Rules Journal, Vol. 8, No. 11.*

Retrieved from http://www.brcommunity.com/b373.php

Vlootman, H., Hermans, F., (2013). A Maintainability Checklist for Spreadsheets, *European Spreadsheet Risks Interest Group. Proceedings of the EuSpRIG 2013 Conference,* pp. 1-10.

Retrieved from: http://arxiv.org/ftp/arxiv/papers/1401/1401.7814.pdf

Wailgum, T., (August 17, 2007). Eight of the Worst Spreadsheet Blunders. *CIO.com*

Retrieved from http://www.cio.com/article/2438188/enterprise-software/eight-of-the-worst-spreadsheet-blunders.html on August 10, 2015.

Wu, N., (2011). Leveraging User Profile and Behavior to Design Practical Spreadsheet Controls for the Finance Function, *European Spreadsheet Risks Interest Group. Proceedings of the EuSpRIG 2011 Conference,* pp. 1-8.

Retrieved from http://arxiv.org/pdf/1111.5007v1.pdf

Yin, R., (2014). Case Study Research: Design and Methods, Fifth Edition. *SAGE Publications Inc.*

Yin, R., (2009). Case Study Research: Design and Methods, Fourth Edition. *SAGE Publications Inc.*

Other Useful Site Links

https://www.definedterm.com/spreadsheet_risk

https://www.definedterm.com/spreadsheet_risk

http://deloitteblog.co.za/2012/05/23/avoid-tax-calculation-errors-and-duplication-by-applying-risk-intelligence-to-your-spreadsheets/

http://www.eusprig.org/conference-abstracts.htm

http://www.eusprig.org/about.htm

http://www.eusprig.org/history.htm

http://www.eusprig.org/stories.htm

http://eusesconsortium.org

http://faculty.tuck.dartmouth.edu/serp/

http://panko.shidler.hawaii.edu/SSR/index.htm

http://faculty.tuck.dartmouth.edu/serp/links/

http://faculty.tuck.dartmouth.edu/serp/project-description

http://intraweb.stockton.edu/eyos/internal_audit/content/docs/icnote2.pdf

http://mba.tuck.dartmouth.edu/spreadsheet/product_pubs.html

http://psychology.jrank.org/pages/52/Attitude-Behavior.html

http://intraweb.stockton.edu/eyos/internal_audit/content/docs/icnote2.pdf

http://www.businessdictionary.com/definition/reasonable-test.html

http://www.sysmod.com/sslinks.htm

http://www.phrases.org.uk/meanings/tilting-at-windmills.html