INFORMATION QUALITY STRATEGY:

AN EMPIRICAL INVESTIGATION OF THE RELATIONSHIP BETWEEN INFORMATION QUALITY IMPROVEMENTS AND ORGANIZATIONAL OUTCOMES

by

John P. Slone

A Dissertation Presented in Partial Fulfillment

Of the Requirements for the Degree

Doctor of Philosophy

Capella University

October 2006

UMI Number: 3233727

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

INFORMATION QUALITY STRATEGY:

AN EMPIRICAL INVESTIGATION OF THE RELATIONSHIP BETWEEN INFORMATION QUALITY IMPROVEMENTS AND ORGANIZATIONAL OUTCOMES

by

John P. Slone

has been approved

October 2006

APPROVED:

APIWAN D. BORN, Ph.D., Faculty Mentor and Chair

H. PERRIN GARSOMBKE, Ph.D., Committee Member

YANG W. LEE, Ph.D., Committee Member

ACCEPTED AND SIGNED:

APIWAN D. BORN, Ph.D.

Kurt Linberg, Ph.D.
Dean, School of Business & Technology

Abstract

The relationship between information and decision-making is a complex one and has been the subject of extensive research spanning several decades, with roots in nineteenth and early twentieth century economic theories. More recently, researchers have suggested a relationship between the quality of information and the quality of decision-making, with a consequent relationship with organizational strategy; however, there has been very little research in which this relationship was investigated systematically. This research set forth contextual and conceptual models relating information quality to strategy and then provided an empirical investigation of the relationship between information quality and organizational outcomes, with information intensity hypothesized as a moderator of that relationship. Data for this study were collected through a Web-based survey of individuals associated with an industry consortium and were evaluated through a combination of multiple regression analysis, moderated regression analysis, and subgroup analysis. Data analysis revealed evidence that the relationship between the quality of information and organizational outcomes is systematically measurable, in that measurements of information quality can be used to predict organizational outcomes, and that this relationship is, for the most part, positive. An unexpected finding was that different regression models emerge when stakeholder roles in an information system are taken into consideration. Data analysis did not reveal support for the hypothesis that information intensity moderates the relationship between information quality and organizational outcomes.

Dedication

This dissertation is dedicated to my wife, Nan, for the unending and unconditional love and support she has given me through this doctoral journey; to my mother, Dr. Gwen Curry, for showing me that it is possible to earn a Ph.D. midway through one's career and for encouraging me at each step along the way; and to my late grandfather, Dr. S. E. Cranfill, for altering the course of our family's history by summoning the courage to leave the family farm and enter the world of academe. It is also dedicated to each and every member of my immediate and extended family for the love and encouragement they have shown me through this process, but especially to my daughter, Melissa, for setting my sights on the honor of being called "Dr. Dad."

Acknowledgements

I truly stand on the shoulders of giants as I write this. This dissertation could not have been completed without the support and assistance of many individuals, institutions, and organizations. First and foremost, I would like to thank my faculty mentor and committee chair, Dr. Apiwan Born, who was always ready and eager to help me at each step of the way, and whose feedback and suggestions helped ensure that this would become a quality piece of scholarly work. I will remain forever grateful for the advice and encouragement she provided from start to finish. I would also like to express my gratitude to the other members of my committee, Dr. H Perrin Garsombke and Dr. Yang W. Lee, each of whom has provided valuable insights and suggestions. In this regard, the eagerness and enthusiasm with which Professor Lee fulfilled her role as visiting scholar are especially acknowledged with deep gratitude and the utmost of respect.

Another individual who deserves special recognition at this time is Dr. Richard Y. Wang. Not only did he put the topic of Information Quality on the map, as evidenced by the numerous citations of his work found throughout this dissertation, he also took a personal interest in my work and my success to a degree that I never could have expected, and for which I will always owe a tremendous debt of gratitude. In addition, I would be remiss if I failed to acknowledge that the genesis of the idea for my dissertation topic came directly from a conversation that he and I had during the summer of 2005.

Two great institutions deserve recognition for their role in this work. First, of course, is Capella University, a relative newcomer on the academic scene. The willingness and creativity with which this institution has worked to break down educational barriers for working professionals such as myself are matched only by the quality of the faculty members, librarians, and support staff they have assembled. Toward this end, in addition to those who served on my committee, I would like to acknowledge Dr. John Whitlock, Dr. Martha Hollis, and Dr. Gary Robinson for the encouragement and support that they provided along the way, probably in ways they did not even realize at the time.

The second institution that deserves recognition in this work is the Massachusetts Institute of Technology, a pillar of excellence in the academic world. In addition to being the home institution for Professor Wang, as well as the alma mater for Professors Wang and Lee, MIT has opened its doors to researchers and practitioners from around the world to promote the advancement of Information Quality research and practice. Through the MIT Information Quality program and research consortium, and through its hosting of the annual International Conference on Information Quality, not to mention the community of scholars associated with those activities, MIT has both welcomed me and helped propel me down this path in ways that I never could have imagined. The members of this community who have helped me in ways big and small make up a list too long to enumerate, for fear of leaving someone out.

Two organizations also deserve special recognition for their roles in this work. First, I would like to thank my employer, the Lockheed Martin Corporation, for supporting me financially through its tuition reimbursement program, and for allowing me the opportunity to find synergy between my academic pursuits and my daily work. I would especially like to thank two members of my management through most of this process, Mr. Massimo Villinger and Mr. Ajit Kapoor, both now recently retired, for their roles in this support. Secondly, I would like to thank The Open Group for enabling me to leverage the contents of their contacts database and for their enthusiastic assistance in inviting well over 3,000 people to take part in my survey. Toward this end, the following individuals deserve special recognition: Mr. Allen Brown, Mr. Steven Nunn, Ms. Patricia Donovan, Ms. Maggie Roth, and Mr. Chris French. In addition, I would like to acknowledge the special role played by another employee of The Open Group, Dr. Chris Harding, for piquing my interest in Information Quality in the first place. And finally, a special word of thanks is due to a long-time member of The Open Group, Dr. Bill Estrem, who sagely reminded me that the most important thing about a dissertation . . . is to finish!

This acknowledgement would be incomplete without mentioning my family, friends, peer learners, and colleagues for the interest shown in my pursuit of this degree, for the continuous support and encouragement, and undoubtedly, for simply humoring me at times. To each and every one of you, I must humbly say, “Thank you!”

Table of Contents

List of Tables

List of Figures

CHAPTER 1. INTRODUCTION

Introduction to the Problem

On January 28, 1986, the Space Shuttle *Challenger* exploded seconds after lift-off, killing the seven astronauts aboard. Seventeen years later, the Space Shuttle *Columbia* broke apart during reentry, killing another seven astronauts. In July 1988, the U.S. Navy Cruiser USS *Vincennes* shot down an Iranian commercial passenger jet, killing all 290 people aboard. On September 11, 2001, nineteen hijackers slipped through airport security unnoticed and proceeded to turn four commercial passenger jets into guided missiles, resulting in the deaths of approximately 3,000 people. Other than the tragic loss of life and the historically significant consequences, what did these four events hold in common? There are at least two answers to this question: each of them clearly ran counter to the objectives of the respective organizations, and each of them had poor information quality as an underlying cause (9/11 Commission, 2004; Columbia Accident Investigation Board, 2003; Fisher & Kingma, 2001; Rogers Commission, 1986).

The Rogers Commission, which investigated the *Challenger* accident, found that the decision to launch "was based on incomplete and sometimes misleading information" (Rogers Commission, 1986, ch. 5, para. 3). Fisher and Kingma (2001), who performed an in-depth analysis of the *Vincennes* incident, concluded that "data quality was a major factor in the USS *Vincennes* decision-making process" (p. 113). The 9/11 Commission (2004) observed that "relevant information from the National Security Agency and the CIA often failed to make its way to criminal investigators" (p. 79). Finally, the Columbia

Accident Investigation Board (2003) found that "the information available about the foam impact during the mission was *adequate*" (p. 62, emphasis added), yet also noted that "if Program managers had understood the threat that the . . . foam strike posed . . . a rescue would have been conceivable" (p. 174). In none of these incidents was an information quality problem considered the only cause; however, as Fisher and Kingma concluded with respect to both the *Challenger* and *Vincennes*, "it still remains difficult to believe that proper decisions could be made with so many examples of poor data quality" (p. 114).

While these incidents are instructive in emphasizing the importance of information quality, they are by no means the norm. A far more typical example is the hospital staff that misplaced a decimal point, failed to catch the error, and then allowed a fatal overdose to be administered to a pediatric patient (Belkin, 2004), or the eyewear company that failed to recognize that data errors were the root cause of their fifteen percent lens-grinding rework rate, costing the company at least $1 million annually (Wang, Lee, Pipino, & Strong, 1998), or the managed health care organization that consistently overpaid $4 million in claims annually on behalf of patients who were no longer eligible for benefits (Katz-Haas & Lee, 2005). Even organizations that recognize the importance of information quality can find themselves expending significant resources to resolve billing discrepancies (Redman, 1995), or failing to make an adequate profit due to incomplete cost data (Campbell, Douglass, & Smith-Adams, 2004), or losing customers due to their dissatisfaction with the company's Web site (McKinney, Yoon, & Zahedi, 2002). Overall estimates of business losses resulting from poor information quality vary widely, but are consistently measured in the billions of dollars annually; the collective human costs, measured in lives lost or permanently altered, are equally staggering (Bovee, 2004; Fisher & Kingma, 2001; Melkas, 2004; Parssian, 2002; Redman, 1998).

Background of the Study

The relationship between information and decision-making is a complex one and has been the subject of extensive research spanning several decades. A branch of economic theory, known as the behavioral theory of economics, sought to explain how people in organizations make decisions in the face of "imperfect information" (Simon, 1979, p. 503). This explanation is found in such concepts as bounded rationality and game theory (Seth & Thomas, 1994; Simon; Winter, 1971). A forerunner of the behavioral theory, known as institutional economics (Commons, 1931; Cranfill, 1940; Simon), was characterized by its focus on collective action in the presence of conflict. Unlike the behavioral economic theorists, the institutional economic theorists assumed that the necessary information was available to decision makers (Seth & Thomas). Even so, institutional economic theorists had a more sophisticated view of the relationship between information and decision-making than did their predecessors, the classical economic theorists. Although the classical theorists assumed that necessary information was available to decision-makers, their focus was on the inherent value of property and commodities, with the implication that a single correct decision could be reached, given enough information (Commons; Cranfill; Ricardo, 1960).

Current theories of strategy fall into two primary schools of thought: focus on the competitive environment (Porter, 1991, 1996) and the resource-based view (Barney, 1991, 2001; Bharadwaj, 2000; Wade & Hulland, 2004; Wernerfelt, 1984). The school of thought that focuses on the competitive environment draws significantly from behavioral economics, and finds its roots in institutional economics, given the focus on collective action in the presence of conflicting objectives. The resource-based view, on the other hand, focuses its attention on resources owned or controlled by the firm, and how those resources can be exploited for competitive advantage, much in the same way that classical economics focused on property and commodities and on maximizing the economic value of each.

Given the different characteristics of these schools of thought, logic suggests that certain mixtures of information quality improvements fit more naturally within one strategy school or the other, depending on the type of advantage sought. For instance, improvements that focus on delivering more timely and relevant information can help an organization respond more quickly to changes in the competitive environment. On the other hand, improvements that focus on maximizing the completeness and accuracy of information can help a firm extract more value from its resources.

Separately from the study of strategy and economics, awareness of information quality as an issue emerged slowly during the early years of computers. Authors such as Maffei (1958) and Trueblood (1960) used the vague, intuitive term, “better information” (Maffei, p. 186; Trueblood, p. 48). Maffei recognized that a price was being “paid by deviating from a ‘best’ course of action” (p. 186) in the absence of better information, while Trueblood sought to improve "management judgment" (p. 48) through the provision of better information. By the mid-1990's, information quality had begun to emerge as a research discipline in its own right. Wang, Storey, and Firth (1995) conducted a thorough literature search on the topic, finding more than 100 research articles published between 1970 and 1995, and proposed a framework for organizing and guiding information quality research moving forward.

Statement of the Problem

In 1996, Wang and Strong published a seminal article establishing a hierarchical, multi-dimensional construct of information quality. Since that time, research in the field has made significant theoretical and practical advances. From a theoretical standpoint, several frameworks have been put forth for managing information quality in an organizational context (Ballou, Wang, Pazer, & Tayi, 1998; Kahn, Strong, & Wang, 2002; Wang, 1998). From a practical perspective, various approaches to applying knowledge about information quality management in an organizational setting have been developed (Ballou et al.; Kahn et al.; Lee, Strong, Kahn, & Wang, 2002; Levitin & Redman, 1998; Wang).

Despite these advances, there has thus far been very little understanding from either a theoretical or practical perspective of the relationship between information quality improvement activities and organizational outcomes. A review of the literature revealed few examples of research addressing information quality strategy; those that were identified were written from a variety of perspectives with little or no commonality in approach or findings. Consequently, a need was identified for research providing a common conceptual framework for information quality strategy and for research evaluating the relationship between information quality and organizational outcomes.

Purpose of the Study

The purpose of this research was to investigate the relationship between the management of information quality and organizational outcomes. This relationship includes several aspects of information quality and different types of organizational outcomes. A literature review provided the basis for the development of the research model. That model identified four specific aspects of information quality (soundness, dependability, usefulness, and usability) and two categories of organizational outcome (strategic benefit and transactional benefit). Each of these six items constituted a variable in the conceptual model. This research also investigated the moderating effect of information intensity on the relationship between information quality and organizational outcomes. Information intensity characteristics were captured as two variables, one representing the information content of the organization’s products and services, and the other representing the information intensity of the organization’s value chain.

These variables were used to measure stakeholder perceptions in terms of importance, current state, organizational benefits derived from information quality improvement, and stakeholder perceptions of their organization’s information intensity characteristics. Analysis of the measurement data provided an understanding of the relationship between stakeholder perceptions of information quality and perceptions of organizational outcomes. The data collection and analysis together provided empirical evidence regarding the validity of the proposed model.

Rationale

This research approach provided empirical evidence regarding the relationship between information quality and organizational outcomes. The variables selected provided a reasonable balance between parsimony and explanatory completeness. With respect to information quality, the four aspects selected had previously been demonstrated to represent sixteen dimensions of information quality, measurable using a validated data collection instrument (Kahn et al., 2002; Lee et al., 2002). With respect to organizational outcomes, the two categories of outcomes selected had previously been demonstrated to represent six dimensions of organizational benefits that result from information systems projects. These variables were also measurable using a validated data collection instrument (Mirani & Lederer, 1998). In addition, these outcomes variables were related to both of the strategic schools of thought discussed earlier in this chapter.

Research Questions

The objective of this research was to answer the following research questions:

1. What is the nature of the relationship between stakeholder perception of information quality improvement and organizational outcomes?
2. What interaction effects exist between different aspects of information quality improvement and organizational outcomes?

3. In what way does information intensity affect the relationship between information quality improvement and organizational outcomes?

Significance of the Study

Prior evidence in the research literature supports the notion that lack of attention to information quality problems leads to substantial losses, measured in both human and economic terms, either of which constitutes a negative organizational outcome. This study contributes to the body of knowledge in two ways. First, it presents a conceptual model of the relationship between information quality and organizational outcomes, including empirical evidence regarding the validity of this model. Second, it provides empirical details regarding the nature of stakeholder perceptions of the relationship between information quality and organizational outcomes.

These contributions are expected to benefit both researchers and practitioners. Researchers can benefit by applying the conceptual model in the conduct of similar research in other organizational settings, and in the conduct of research extending the model and investigating different aspects of the model in more specific strategic contexts. Practitioners can benefit by applying the results of the analysis to their own information quality management decisions with an understanding of how those decisions relate to the organization's strategic outcome.

Definition of Terms

Dependable information. Information that is delivered in a sufficiently timely fashion and in a sufficiently secure manner to conform with specifications associated with its intended use.

Information intensity (of the organization). A construct representing the information intensity of the organization's products and services, and the information intensity of the organization's value chain.

Information intensity (of the product and services). A measure of the extent to which information is relied upon in the selection, purchase, use, and maintenance of products and in the selection, purchase, and performance of services.

Information intensity (of the value chain). A measure of the extent to which information is relied upon within the value chain, including the amount of information used and the frequency with which that information is updated.

Information (or data) quality. A multi-dimensional construct that characterizes the extent to which information is fit for use for a particular purpose.

Information quality dimensions. Characteristics of information that determine aspects of its quality, including accessibility, accuracy, appropriate amount, believability, completeness, conciseness of representation, consistency of representation, ease of operation, interpretability, objectivity, relevance, reputation, security, timeliness, understandability, and value-added. Definitions of the individual dimensions are found in Wang and Strong (1996).

Information quality strategy. A pattern or plan for improving an organization's information quality for the purpose of influencing specific organizational outcomes.

Organizational benefits. Positively oriented organizational outcomes that include strategic benefits and/or transactional benefits.

Sound information. Information that is sufficiently free of error, complete, concisely represented, and consistently represented to conform to specifications associated with its intended use.

Strategic benefits. Organizational outcomes that positively affect competitive advantage, alignment, or customer relations.

Strategy. A pattern of behavior or a purposeful plan that influences an organization's goals, policies, and actions toward the ultimate goal of organizational success, where success is defined by the organization in question.

Transactional benefits. Organizational outcomes that positively affect the organization's business efficiency, communication efficiency, or system development efficiency.

Usable information. Information that meets or exceeds the expectations of the consumers of that information with respect to its appropriateness of amount, interpretability, objectivity, relevancy, or understandability.

Useful information. Information that meets or exceeds the expectations of the consumers of that information with respect to its accessibility, believability, ease of operation, reputation, or the value that it adds.

Assumptions and Limitations

This study was based upon the assumption that the participant responses accurately reflected the knowledge and perceptions of those participants with respect to information quality, organizational outcomes, and information intensity, and that measurement of those perceptions provided a reasonable representation of reality. The sampling frame consisted of a specific industry consortium's contacts database, from which a systematic sample was selected for participation. As such, the results of this study may not necessarily reflect the generalized views of broader selections of organizations or of organizations outside the industries represented in the sample.

The study was also conducted from the perspective of the post-positivist paradigm. Although it is recognized that research in the field of information quality transcends multiple research paradigms, and as such is not limited to post-positivism alone, the selected approach was considered appropriate for a study of this nature. As such, the contribution of this study to the body of knowledge is limited to that which can be provided from within the post-positivist paradigm.

Nature of the Study

This study was a quantitative analysis employing multiple regression to explore the ability to predict organizational outcomes based on information quality. Data were collected through the use of a survey instrument employing a Likert scale to measure stakeholder perceptions of information quality, organizational outcomes, and information intensity. Multiple regression was selected as it is appropriate for predicting a single dependent variable in the presence of multiple independent variables when all of the variables are of the interval type (Lattin, Carroll, & Green, 2003; Mertler & Vannatta, 2005).

Organization of the Remainder of the Study

This study is presented in five chapters. Chapter 1 introduces the study, highlighting the importance of understanding the relationship between information quality and organizational outcomes, and noting the lack of research in this area. Chapter 2 provides a review of literature related to information quality and strategy, connecting each of these fields to its foundational theories, and setting forth contextual and research models. Chapter 3 details the methodology employed in the study. Chapter 4 presents the analysis of the data, including descriptive analysis, construct analysis, and hypothesis testing. Finally, chapter 5 presents a discussion of the implications of the findings from this analysis, along with conclusions, limitations, and recommendations for further research.

CHAPTER 2. LITERATURE REVIEW

Introduction and Organization of the Chapter

This chapter presents a review of the literature related to information quality, strategy, and the relationship between information quality and strategy. Information quality is presented in terms of its theoretical roots in information and quality, and in terms of contemporary research addressing formal definitions, measurement techniques, management approaches, and contributing factors. Strategy is presented in terms of its theoretical roots in nineteenth and early twentieth century economic theory, and in terms of contemporary research. The latter examines two schools of strategic thought, namely the resource based view of the firm and the focus on the competitive environment in which the firm operates. Literature examining relationships between strategy and information systems, technology, and quality is also presented. Based on this review, the chapter establishes both the historical and contemporary underpinnings of the current research.

This chapter reveals an important gap in the research literature, in that the linkage between information quality and strategy has only been minimally examined to date, with relatively little theoretical grounding. This chapter thus sets forth a contextual framework within which information quality strategy research can be viewed, and it establishes a research framework and model for examining a set of strategic relationships between information quality aspects and organizational outcomes. By investigating this relationship, the current research has contributed to the body of knowledge by examining the nature, direction, and strength of specific connections between information quality improvement initiatives and organizational outcomes.

Information Quality

Information quality has been the subject of research for many years. This section of the chapter explores the literature documenting such research, beginning with the theoretical roots forming the foundation of information quality theory, followed by a discussion of the predominant research focused on establishing a rigorous definition of information quality. The section continues with an examination of research on managing information quality and then with an examination of research examining factors that contribute to high information quality.

Theoretical Roots of Information Quality

Prior to exploring the nature of information quality, this section of the chapter explores the theoretical roots of information quality. These theoretical roots are found in the separate disciplines of information and quality, each of which is discussed below.

The Nature of Information

Central to the concept of information quality is an understanding of the nature of information. This section explores two theories fundamental to this understanding. The first of these theories is *information theory*, developed primarily by Claude Shannon and his colleagues at Bell Labs in the 1940s (Shannon, 1948; Shannon & Weaver, 1949). A key development of information theory was the novel application of the thermodynamic concept of entropy as a representation of uncertainty. According to this theory, information is that which serves to reduce this uncertainty. An aspect of information excluded from information theory is *meaning*. Semiotics, which focuses on such meaning, is thus presented following the discussion of information theory.

Information theory. The work of twentieth century mathematician Claude Shannon is widely considered to be the most influential work in the field of information theory (Avery, 2003; Bovee, 2004; McEliece, 2002). In describing the importance of this work, McEliece observed:

> While of course Shannon was not working in a vacuum in the 1940's, his results were so breathtakingly original that even the communication specialists of the day were at a loss to understand their significance. Gradually, as Shannon's theorems were digested by the mathematical/engineering community, it became clear that he had created a brand-new science, and others began to make first-rate contributions of their own. (p. 13)

Shannon's early research, conducted at Bell Labs, was primarily focused on the handling of information by a communication channel (Avery; McEliece). As Shannon (1948) described it, "the fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point" (p. 379).

Among the key contributions of Shannon (1948) was the recognition that information acts to reduce uncertainty. Uncertainty, in this sense, is likened to the thermodynamic concept of entropy, hence the term *entropy* was introduced into the study of information theory as well. In a two-symbol system such as binary, entropy is maximized when the chance of encountering each symbol is approximately equal, as shown in Figure 1. As Bovee (2004) described it, "the potential [underlined in original] for a message to reduce uncertainty between two states in the system is maximized at that point" (p. 8). He went on to note that information "is related [underlined in original] to the reduction in uncertainty associated with the received message" (p. 8). The notion of uncertainty reduction was also discussed by Handscombe and Patterson (2004), who described the quantity of information as "the ratio of the number of possible answers before and after receiving information" (p. 28).

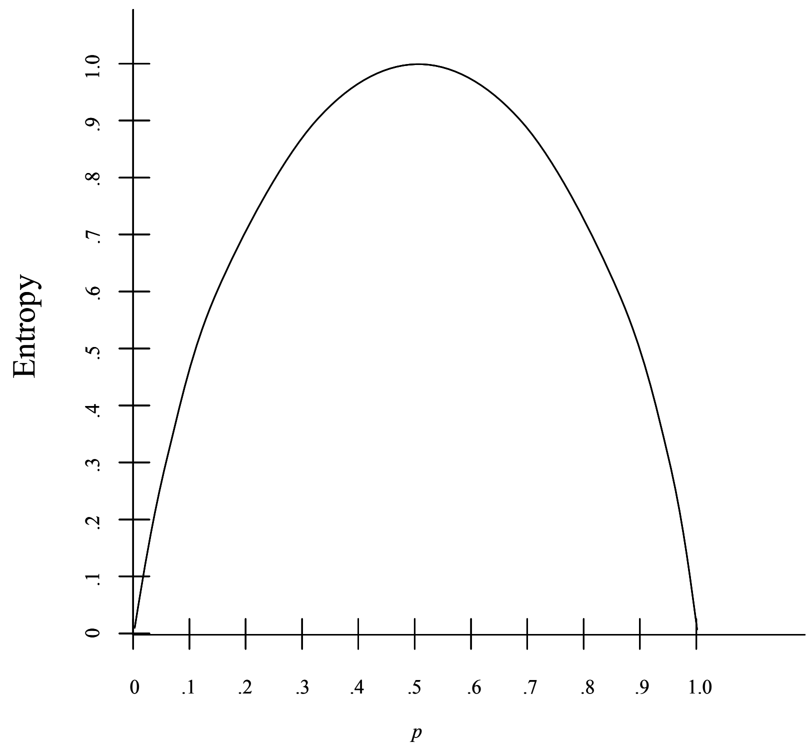

Figure 1. Maximum entropy in a two symbol system.

Note. Adapted from Shannon, C. E., A mathematical theory of communication. The Bell System Technical Journal*, 27, 379-423, 623-656.*

Shannon (1948) based his work in part on the prior research of Henry Nyquist, who had examined the information handling capacity of telegraph lines. Shannon extended Nyquist's work by considering "the case where the signal is perturbed by noise during transmissions or at one or the other of the terminals" (p. 379). Noise, in this sense, increases entropy in a channel, thereby limiting the amount of information the channel can convey.

Despite the groundbreaking significance of this work, information theory alone is insufficient for the purposes of this research. In particular, Shannon (1948) considered the content of a message to be outside the scope of information theory. As he put it, "frequently the messages have *meaning*; that is they refer to . . . certain physical or conceptual entities" (p. 379). He went on to preclude further consideration of information content in his research by stating, "these semantic aspects of communication are irrelevant to the engineering problem" (p. 379).

Semiotics. While dismissed as irrelevant in Shannon's mind, hence out of scope with respect to information theory, the semantic aspects of communication are central to the study of semiotics. Semiotics is the study of signs, where "a sign is something which stands to someone for something else in some respect or capacity" (Liu, 2000, p. 13). Humans use signs routinely to communicate with one another, thus a degree of mutual understanding of the meaning of those signs is necessary for communication to take place (Liu). Even so, the essence of meaning remains stubbornly elusive. As Merrell (1997) put it, "meaning flows along within the *semiosic* process, resisting any and all pigeon-holes. . . . If we try to specify meaning in precise terms we are playing a tail-chasing-dog game inevitably ending in frustration" (p. x).

The semiosic process referred to by Merrell (1997) is central to an understanding of semiotics. This process is depicted in Figure 2. In that figure, a sign (lower left) is an encoding that denotes an object (lower right). When the sign is decoded by an interpretant (top) it functions as an indication of the object. The notion of an interpretant deserves further explanation, in that it "can be an individual, a group or a social community with certain knowledge and obeying certain norms" (Liu, 2000, p. 16). The semiosic process then iteratively uses signs to build "structures of experience" (p. 16).

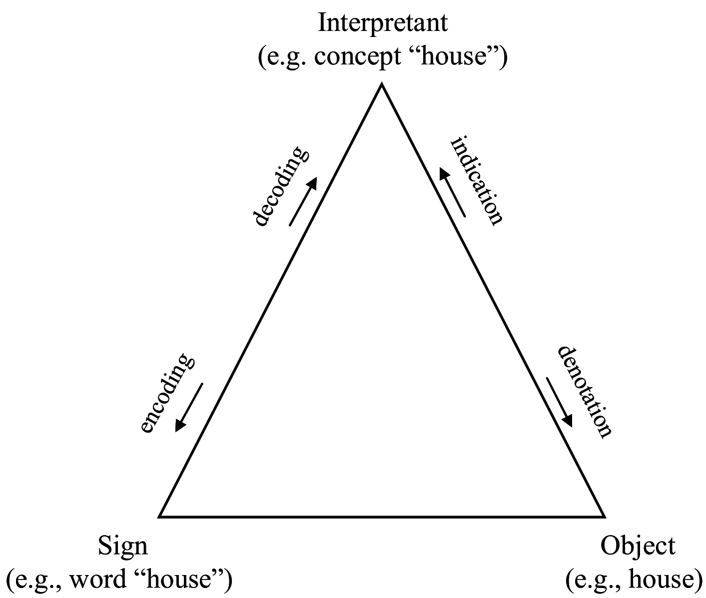

Figure 2. The semiosic process.

Note. Adapted from Liu, K. (2000).Semiotics in information system engineering, Cambridge, UK: Cambridge University Press.

As originally developed by the nineteenth century logician, Charles Sanders Peirce, semiotics consisted of three fields, "syntactics (or syntax), semantics, and pragmatics" (Liu, 2000, p. 13), which consider, respectively, "the structures, meanings, and usage of signs" (p. 26). More recently, three more fields have been added, namely the physical world, empirics, and the social world, to incorporate consideration of additional aspects of signs. Figure 3 illustrates the resulting framework. In this framework, the lower layers, in particular the empirics and syntactics layers, are most closely associated with Shannon's work, while the upper three layers are concerned with those aspects of communication that Shannon excluded. It is at these upper layers that semiotics is concerned with the usage of signs to communicate meaning and intention, and to consider how signs function within a social context – precisely those aspects of communication excluded by Shannon (1948).

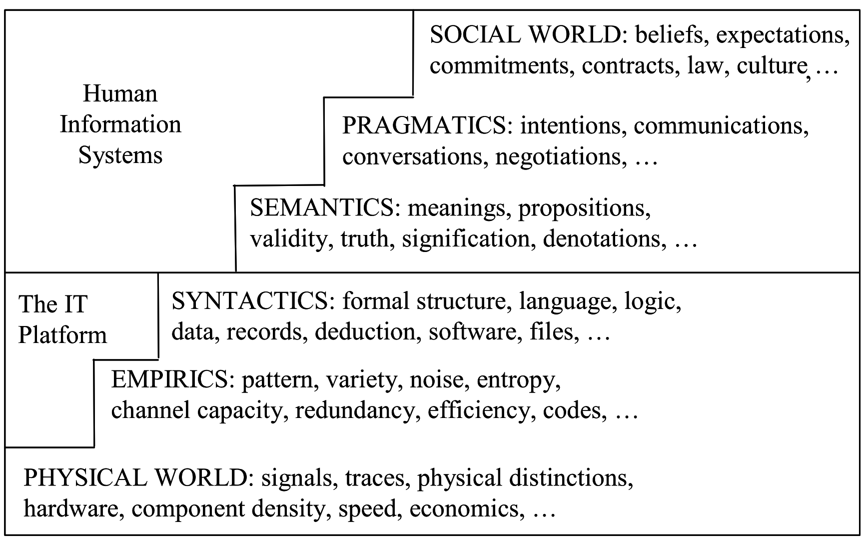

Figure 3. Semiotic framework.

Note. Adapted from Liu, K. (2000).Semiotics in information system engineering, Cambridge, UK: Cambridge University Press and Stamper, R..(1996) Signs, information, norms, and systems. In B. Holmqvist, P. B., Andersen, H. Klein, and R. Posner (Eds.), Signs of work: Semiosis and information processing in organizations *(pp. 349-398). Berlin: Walter D. Gruyter.*

Despite the heavy emphasis on philosophical issues, Liu (2000) identified a number of research and practical applications of semiotics. Among these are linguistics, education, anthropology, computer science, information systems, and organizational theory. Of particular relevance to the context of this research are the study of organizational semiotics and the application of semiotics to information systems development. Most notably, when viewed through a semiotic lens, the organization itself can be thought of as an information system. When conceived in this manner, the organization can be described in terms of three nested layers of information systems. At the outermost layer one finds the informal information system, “a sub-culture where meanings are established, intentions are understood, beliefs are formed and commitments with responsibilities are made, altered and discharged” (p. 109). Next are the formal information systems, consisting of bureaucratic forms and rules, which serve to replace meaning and intention with codified systems. Technical information systems, where information technologies are deployed to automate portions of the formal systems, makes up the innermost layer. Together these layers form what Liu refers to as the “organisational onion” (p. 109), depicted in Figure 4.

The Nature of Quality

Much like information, the concept of quality is defined in different ways by different people. Among the earliest proponents of quality was W. Edwards Deming, best known for his work in the industrial reconstruction of post-World War II Japan. In recognition of this work, the Japanese Union of Scientists and Engineers established the Deming Prize in 1951 to recognize businesses that achieved a certain level of quality

(Deming, 1982; Mahoney & Thor, 1994). Deming (1982) asserted that quality improvements inevitably lead to productivity improvements, hence improvements in competitive position. In his view, low quality wastes effort and production capacity, and causes rework, each of which brings down productivity, increases cost, and has the potential to damage the firm's reputation. He also emphasized that "the customer is the most important part of the production line" (p. 225). In particular, he noted that "the cost to replace a defective item on the assembly line is fairly easy to estimate, but the cost of a defective unit that goes out to a customer defies measure" (p. 225). Ironically, "the most intriguing feature of the [criteria for the Deming Prize] is that there is no mention of customer satisfaction" (Mahoney & Thor, 1994, p. 12).

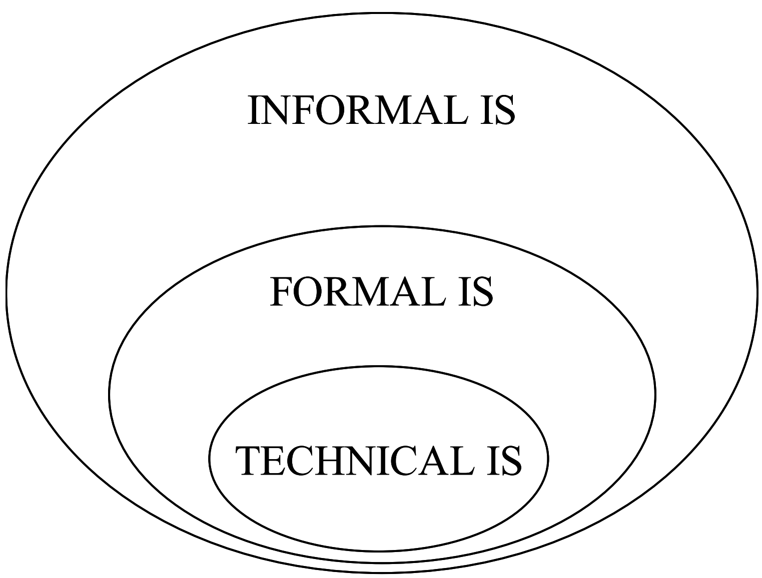

Figure 4. The organizational onion.

Note. Adapted from Liu, K. (2000). Semiotics in information system engineering*, Cambridge, UK: Cambridge University Press.*

Another important contribution to quality is the work of Juran. As did Deming, Juran (1988) emphasized the importance of the customer in defining and measuring quality, and proposed that "a simple definition of quality is 'fitness for use.'" (p. 5). He also noted, however, that "that definition must quickly be enlarged, because there are many uses and users" (p. 5). In a departure from Deming, Juran (1988) greatly expanded the definition of customers "to include *all persons who are impacted by our processes and our products*" (p. 8). He then went on to elaborate about various internal and external customers, including essentially everyone involved in processing or handling a product until it reaches its eventual end user. Juran identified what he saw as three compelling reasons for an organization to pay attention to quality: loss of sales, the costs of poor quality, and threats to society, such as those resulting from product defects. To manage quality effectively, he defined and prescribed what he termed a "trilogy" (p. 12) of quality management processes: quality planning, quality control, and quality improvement.

A third major contributor to the work on quality is Crosby (1992, 1996). Echoing Deming and Juran, Crosby (1992) emphasized the role of the customer, stating, "the only absolutely essential management characteristic [of the twenty-first century] will be to acquire the ability to run an organization that deliberately gives its customers exactly what they have been led to expect and does it with pleasant efficiency" (pp. 16-17). To fulfill this mission, he advocates that an organization work to enable the success of all its key constituents, most notably its employees, suppliers, and customers. He cautions, however, that "quality is hard to pin down, because each person thinks everyone else defines it the same way he or she does" (Crosby, 1996, p. 48).

Two other developments in quality deserve mention in addition to the contributions of these three influential individuals. First, the late U. S. Secretary of Commerce, Malcolm Baldridge, pushed for the establishment of a quality award within the U. S. as part of his strategy for enhancing the competitiveness of U. S. businesses. Shortly after his untimely death in 1987, Congress established the Malcolm Baldridge National Quality Award, which evaluates businesses on seven major criteria. The award strongly emphasizes customer satisfaction and a preventative, rather than reactive, approach to quality management (Mahoney & Thor, 1994; National Institute of Standards and Technology, 2005).

The second major development is the ISO 9000 series of international standards. The focus of these standards is on organizational capabilities in regard to quality management. Organizations doing business internationally may voluntarily seek ISO 9000 certification as a way of assuring their international customers that they have the demonstrated organizational ability to provide quality products and services (Mahoney & Thor, 1994).

Defining Information Quality

Considerable research attention has been focused on the need for a rigorous definition of information quality. This section traces the streams of research literature establishing such a definition, beginning with a look at the early attempts at defining information quality, followed by a discussion of research that led eventually to the establishment of a consistent definitional model, including a look at the model used as the basis of this research. The section concludes with a look at the issue of information quality versus data quality, and identifies the position taken by this researcher.

The Emergence of the Information Quality Concept

Awareness of information quality as an issue emerged slowly during the early years of computers, when researchers gradually developed an awareness of the need to measure data quality, and began the work of convincing others of that need. In 1958, for instance, Maffei wrote, “a theory of the cost and value of information is needed. We need to know quantitatively what price is being paid by deviating from a ‘best’ course of action and weigh this against the cost of getting better information” (Maffei, 1958, p. 186). In describing this as a cost issue, he was referring to cost in the broadest possible sense, including such notions as opportunity cost and the cost of making poor decisions based on inferior information.

Related to this, Trueblood (1960) focused on what was at the time the newly emerging field of operations research. As he put it, “the purpose [of operations research] is not to replace management judgment but to provide more and better information” (p. 48). As with Maffei (1958), the vague, intuitive notion of “better information” found its way into his writing, but without further refinement or definition. In addition to better information, he also recognized a fundamental set of changes in information needs, pointing out that operations research not only provided new forms of information, it also demanded it. Unlike Maffei, Trueblood explicitly recognized the connection between operations research and computers by noting, “in one of [the accounting profession’s] more thoughtful business magazines, these changes have been viewed as generating an entirely new 'information technology'" (p. 49).

Not long afterward, literature began to emerge in which a conceptual model of information quality was taking shape in the context of information technology. Among the key concepts were the separation of roles between different information actors as well as the identification and definition of various aspects of information quality (Feltham, 1968). Most research during this period focused on accuracy, although a few began to explore other dimensions as well (Bovee, 2004; Wang et al., 1995). Feltham observed that "relevance, timeliness, and accuracy are often listed as desirable attributes of information" (p. 684). In terms of timeliness, he wrote of delays in reporting and the collection of information changes until some specified condition or time interval occurs, and then defined the value of timeliness in terms of the cost versus benefit of creating a change in either reporting delay or reporting interval. Within this context, his work was focused on the development of a model for assessing the value of a change to an information system, by measuring the cost of the change versus the benefits of the change. As such, it stands as an early example of literature evaluating information quality management in terms of its costs and benefits.

In Search of a Workable Model

For most of the next three decades, researchers explored a variety of ways to conceptualize data quality. For example, Gallagher (1974) considered such factors as usefulness, desirability, meaningfulness, and relevance, among others, in determining the value of information systems. Halloran et al. (1978) focused on accuracy, relevance, completeness, recoverability, access security, and timeliness, and specified metrics for each of these in terms of the overall system. With respect to accuracy, Halloran et al. wrote, “an organization can keep error statistics relating to data accuracy” (p. 7). Relevance is defined as “how [the system’s] inputs, operations, and outputs fit in with the current needs of the people and the goals it supports” (p. 10). A few years later, Bailey and Pearson (1983) measured system satisfaction with measures that included accuracy, precision, currency, completeness, relevance, and other similar information quality attributes.

Ballou and Pazer (1985) similarly recognized that the issue goes beyond accuracy alone, observing that errors can be “amplified or diminished by processing” (p. 151) and noting that “it has become apparent that data quality is a relative rather than absolute term” (p. 151). They also explored what they referred to as the “accuracy-timeliness tradeoff” (p. 151), and proposed a theoretical framework and algorithm for calculating the effect of this tradeoff. As they explained it, “information regarding some situation or activity at a fixed point in time becomes better with the passage of time. However, as a consequence of the dynamic nature of many environments, the information also becomes less relevant over time” (p. 151). They made a strong case for expanding the scope of information quality beyond a mere focus on accuracy, stating that other attributes “include timeliness, consistency, completeness, relevance, and reliability” (p. 152).

Towards a Consistent Model

It was not until the mid-1990s that information quality research began to coalesce around a common framework. In particular, Wang et al. (1995) proposed a framework derived from ISO 9000 for use in categorizing data quality research. Wang et al. systematically categorized research on the topic up to that point in time. In addition to the literature cited above, they identified dozens of other articles. Among these, they found numerous examples using different combinations of dimensions, as well as a variety of approaches to the research. Of the dimensions they observed, the ones most commonly occurring were accuracy, timeliness, completeness, and consistency. Some dimensions occurring less frequently included credibility and traceability. They noted that one researcher had identified more than 20 dimensions.

Among their findings, Wang et al. (1995) noted that “previous research has focused primarily on the accuracy requirements” (p. 629). They also noted, “since data quality is a multi-faceted concept that includes not only accuracy, . . . more research is needed on the other dimensions as well” (p. 630), and they called for research on “an overall data quality metric” (p. 637).

Building on the suggestion of Paradice and Fuerst (1991), Wang et al. (1995) drew the analogy between the manufacture of products and the processing of data. That is, information systems were considered analogous to manufacturing systems, with the difference being that data are used as the raw material, and processed data, sometimes referred to as information, are the output. In this model, data stores are comparable to inventory. The ISO 9000 concept of “Specification and Design” (Wang et al., p. 624) translates into the need to specify different quality aspects of data, such as acceptance and rejection criteria, consistent with management policy, and subject to management processes. Adopting a customer perspective similar to the one advocated by Juran (1988), Wang et al. noted, the "use of the term 'data product' emphasizes the fact that the data output has value that is transferred to customers, whether internal or external to the organization" (p. 624).

This perspective subsequently became one of the driving forces behind Wang and Strong (1996), which was written "to develop a framework that captures the aspects of data quality that are important to data consumers" (p. 5). Wang and Strong wrote that "although firms are improving data quality with practical approaches and tools, their improvement efforts tend to focus narrowly on accuracy" (p. 5). They then went on to report on the result of a major study involving a two-stage survey. Beginning with a very broadly based set of nearly 200 data quality attributes, they used factor analysis to narrow the set to a much more parsimonious set of 20 dimensions. Based on the second-stage survey, they reduced this set even further to 15 dimensions, along with four categories for grouping those dimensions: intrinsic, contextual, representational, and access. They summarized their findings as follows:

> Intrinsic DQ denotes that data have quality in their own right. Contextual quality highlights the requirement that data must be considered within the context of the task at hand. Representational DQ and accessibility DQ emphasize the importance of the role of systems. These findings are consistent with our understanding that high-quality data should be intrinsically good, contextually appropriate for the task, clearly represented, and accessible to the data consumer. (p. 22)

Figure 5 depicts Wang and Strong's model of data quality as a multi-dimensional construct. Although the exact number of dimensions considered and the arrangement of the dimensions varies somewhat from researcher to researcher, the essence of this model now has broad support among the information quality research community.

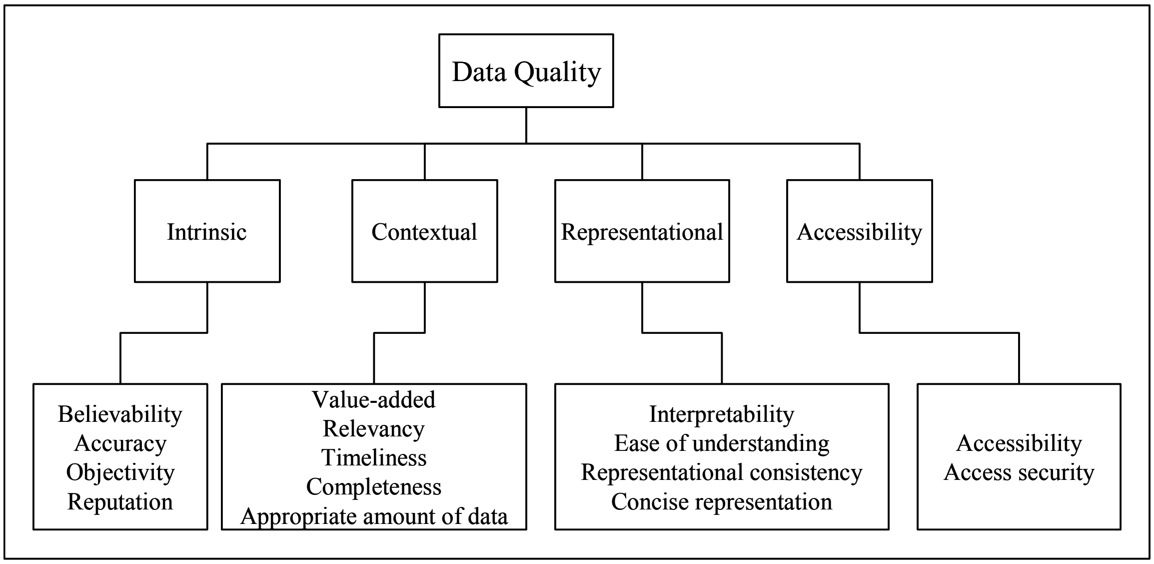

Figure 5. Data quality as a multi-dimensional construct.

Note. Adapted from Wang, R. Y., and Strong, D. M. (1996).Beyond accuracy: What quality means to data consumers. Journal of Management Information Systems*, 12(4), 5-34.*

Strong, Lee, and Wang (1997) used this model to frame their research into data quality problems and their solutions at three organizations. Among their findings were that there were patterns of data quality problems that cross from one hierarchical grouping to another. For instance, a believability problem with a particular database can lead to perceptions of low added value, thus crossing from intrinsic to contextual. Similarly, problems with inconsistent data representation can be perceived as accessibility problems. In general, Strong et al. found “two different approaches to problem resolution: changing the systems or changing the production process” (p. 107). As a result of these findings, the authors advocated strongly for expanding one’s view of data quality problems, along with one's approach to solving those problems, beyond the limited perspective of the intrinsic quality dimensions.

The Product and Service Performance Model for Information Quality

Kahn et al. (2002) acknowledged that the prevailing conceptual models treated information as a product, yet noted that it "can also be conceptualized as a service" (p. 186). A service, unlike a product, "is perishable, for you cannot keep it; it is produced and consumed simultaneously" (p. 186). In addition to recognizing the service aspects of information quality, they drew upon general quality literature to identify additional ways to characterize quality, two of which they adopted for their purposes: "conformance to specifications" (p. 185) and "meeting or exceeding customer expectations" (p. 185). By combining these two characterizations with the product and service aspects of information quality, they developed a significant extension of the Wang and Strong (1996) model, termed the "product and service performance model for information quality (PSP/IQ)" (Kahn et al., p. 184).

The PSP/IQ model is represented as a two-by-two grid, as shown in Figure 6. Product quality and service quality are represented in the figure as rows, and specification versus expectations make up the columns. As shown, the various information quality dimensions from the Wang and Strong (1996) model map onto this two-by-two grid, and each of the quadrants has been assigned a short, descriptive name. On the product side, the product-conformance quadrant is referred to as "sound information" (Kahn et al., 2002, p. 189) and the product-expectations quadrant represents "useful information" (p. 189). On the service side, the service-conformance quadrant represents "dependable information" (p. 189), with "usable information" (p. 189) making up the service-expectation quadrant.

	Conforms to Specifications	Meets or Exceeds Expectations
Product	Soundness • Free of error • Concise representation • Completeness • Consistent representation	Usefulness • Appropriate amount • Relevancy • Understandability • *Interpretability* • *Objectivity*
Service	Dependability • Timeliness • Security	Usability • Believability • Accessibility • Ease of operation • Reputation • Value-added

Figure 6. The PSP/IQ model. Italics in "usefulness" quadrant indicate dimensions that only marginally map into this quadrant.

Note. Adapted from Kahn, B. K., Strong, D. M., and Wang, R. Y. (2002). Information quality benchmarks: Product and service performance. Communications of the ACM*, 45(4), 184-192.*

Information Quality Versus Data Quality

Usage of the terms *information quality* and *data quality* is highly inconsistent from one researcher to another. Many researchers consider the terms to be synonymous and treat them as such; others do not. While exploring this issue of terminology, Kahn, Pierce, and Melkas (2004) conducted a survey of session titles for papers presented at the International Conference on Information Quality during the eight year period from 1996 to 2003. They reported that they expected to see the term *data quality* predominate during the earlier years, followed by a shift to a predominant use of the term *information quality*; however, they noted that their data did not support that expectation. Instead, they found no discernable pattern distinguishing the use of the terms and thus decided to use them synonymously.

Bovee (2004) conducted a thorough exploration of the terms *data* and *information* in hopes of resolving this dilemma before moving on to define the compound terms *data quality* and *information quality*. Instead of finding resolution, he found numerous instances in which, if a distinction was to be made, one term was defined by its relationship to the other, leaving neither term well-defined. After many pages of well-reasoned, well-documented consideration, he decided to "bypass the circularity found between these two constructs" (p. 32), choosing instead to use the terms synonymously. Given these findings in the literature, the terms will likewise be treated as synonyms in this research unless specifically noted otherwise.

Managing Information Quality

This section examines the research literature addressing the topic of information quality management. The section begins with a look at why information quality is considered sufficiently unique to warrant its own approach to management, as opposed to the extant approaches to managing quality in general. Next, the section provides an in-depth look into research regarding the measurement of information quality, including both subjective and objective measurement approaches, as an understanding of measurement is essential to any management approach. The section then continues with a look at several of the major approaches to managing information quality, namely Total Data Quality Management (TDQM), data production maps, and benchmarking.

Why Information Quality is Unique

Many of the frameworks and approaches proposed for managing information quality are rooted in an analogy between physical product manufacturing and information product manufacturing (Ballou et al., 1998; Paradice & Fuerst, 1991; Scannapieco, Pernici, & Pierce, 2005; Shankaranarayanan, Wang, & Ziad, 2000; Wang, Lee, et al., 1998; Wang, Storey, et al., 1995; Wang & Strong, 1996). It is reasonable to ask, therefore, why previously available approaches developed for managing the quality of physical products (Crosby, 1992, 1996; Deming, 1982; Juran, 1988; Mahoney & Thor, 1994) were deemed unsuitable or insufficient for managing the quality of information products. As Ballou et al. put it, the differences "arise from the nature of the raw material" (p. 463). More generally, these differences have to do with the nature of information itself, differences in the nature of information quality as contrasted with physical product quality, with the difficulty of measuring information, and with the contexts in which information is used.

As for the nature of information itself, a key difference between information and physical products is in the fact that data can be consumed repeatedly, indeed indefinitely, without being depleted (Ballou et al., 1998; Paradice & Fuerst, 1991; Shankaranarayanan et al., 2000; Wang, 1998). In this sense, information is "more analogous to a tool crib than to inventory" (Ballou et al., p. 463). However, even this analogy falls short, given that tools are not incorporated into final products. In pointing out that a single piece of raw data can be captured once and then used in multiple information products, Shankaranarayanan et al. asserted that it is “imperative that a good representation accurately tracks the details of what triggered the capture of this data along with how, who, and where” (p. 4). Data may also be collected continuously and stored indefinitely without knowledge of whether they will ever be incorporated into an information product. Unlike physical manufacturing, the collection and storage of an additional piece of raw information material on the chance that it might be used in the future results in relatively little additional expense to the organization.

Differences in the nature of quality between information and physical products can be explained in part by considering specific dimensions of information quality that lack physical counterparts. For instance, as Wang (1998) observed, “one could say that a raw material arrived just in time, but one would not ascribe an intrinsic property of timeliness to the raw material” (p. 59). Similarly, “dimensions such as believability simply do not have a counterpart in product manufacturing” (p. 59). These differences also manifest themselves in the aggregate, in that for information products, the quality of the individual data items that make up an information product are as important to the consumer as the quality of the overall product (Shankaranarayanan et al., 2000).

Another difference between information quality and product quality has to do with the difficulties associated with measuring information, given that information has no physical properties to measure (Redman, 1995). With respect to the accuracy dimension, which Wang and Strong (1996) identified as one of the intrinsic dimensions, accuracy cannot be measured intrinsically; its measurement must always reference something else, such as the real world situation that the data represent (Redman, 1995; Wand & Wang, 1996).

In terms of the usage context, Redman (1995) pointed out a subtle, but important distinction between information quality and physical product quality, namely that "most useful data are novel or unique" (p. 23). As a hypothetical example, he considered the absurdity of including genus and species fields in employee records. With every entry identifying the employee as *homo sapiens*, the data would be highly accurate, but uninteresting. Instead, it is the uniqueness of the values that makes them interesting. "This stands in contrast with most manufacturing processes where one strives for uniformity, and standard measures can be applied" (p. 23). To handle this uniqueness while maintaining quality control, Pierce (2005) suggested the use of automatic range checking or an assortment of feedback mechanisms, such as customer-driven, staff-driven, or management-driven feedback, or a combination thereof.

Measuring Information Quality

The ability to measure anything is essential to one's ability to manage it. The same is true for information quality. The following sections of this chapter present an in-depth discussion of information quality measurement, beginning with a brief historical look at the issue, followed by discussions of subjective measurements, objective measurements, and combined approaches.

Early measurement approaches. As early as the late 1950s, Maffei (1958) noted the difficulty of measuring information quality, stating, "more often than not, the information available is not only not what we would like to have but is, in addition, unknowably inaccurate" (p. 171). He also mused, "when is it important to know when to collect information carefully and precisely and when is it not so important?" (p. 171). Trueblood (1960), put it in more accounting-specific terms when he wrote that "there are today no generally accepted criteria for the design of an integrated information system for a firm – deciding what information is needed, how frequently the information is needed, [and] how accurate it needs to be" (p. 50).

Ballou and Pazer (1985) recognized that information quality "is a relative rather than absolute term" (p. 151). In support of this work, they proposed a model for evaluating the "magnitude of errors" (p. 152). They also identified four dimensions that they considered relevant: accuracy, timeliness, completeness, and consistency, and proposed measurements for each in terms of its differential relative to some reference point.

Taking a somewhat different approach, Agmon and Ahituv (1987) applied the concepts of quality control theory as used in industrial engineering to the issue of data reliability in information systems. In doing so, they subdivided the concept of data reliability into three components: internal reliability, relative reliability, and absolute reliability. As they used the terms, internal reliability is most closely associated with what they call "commonly accepted" (p. 34) data usage and characteristics, such as only allowing positive values for quantities in an inventory control system. Relative reliability is as measured against user requirements, such as requiring that every vendor have the name field specified. Absolute reliability is as measured against reality and verified by observation.

Against this backdrop, a considerable number of researchers remained focused on accuracy and little else. Paradice and Fuerst (1991), for example, who opened their article by stating, “the importance of data quality in management information systems (MIS) increases daily” (p. 48), focused their efforts on a proposed formula for computing a “stored error rate” (p. 51). They defined this as a combination of the ratio of a particular data element classified as being in error and the percent of time it is classified as being correct, weighted by the probability, as measured by random samples, that any given element will be in error.

Despite the limited focus on accuracy, Paradice and Fuerst (1991) provided a valuable contribution to the field of data quality management. For instance, while noting that nearly all of the prior literature they reviewed relied upon the use of internal control processes, such as audits, rather than a quantifiable mechanism such as a calculated error rate, they observed a lack of literature applying the quality control methods of manufacturing to information processing. In this context, they proposed the metaphor of data as raw material being consumed by a data manufacturing system to produce information. They noted, however, that “data, unlike most raw materials, is not consumed when processed and therefore may be reused repeatedly” (p. 50).

Subjective measurements. In 2002, Lee et al. observed that “despite a decade of research and practice, only piece-meal, ad hoc techniques are available for measuring, analyzing, and improving IQ in organizations” (p. 133). In response to this situation they developed a measurement instrument, known as the Information Quality Assessment

(IQA), which measures stakeholder perceptions of each dimension in the Wang and Strong (1996) model.

This instrument, which employs 69 items to measure the various information quality dimensions, has been used as the basis of several studies requiring information quality measurement (Kahn et al., 2002; Pipino, Lee, & Wang, 2002; Pipino, Wang, Kopsco, & Rybolt, 2005) as well as for studies that extend this measurement concept further, such as the PSP/IQ model (Kahn et al., 2002). The PSP/IQ model aggregates the results of the 69 items and 16 dimensions measured by the IQA to produce a measure of information quality consisting of only four numbers. By using the IQA to measure the dimensions, the quadrant measurements are derived by calculating the mean scores for the dimensions associated with each quadrant (Kahn et al., 2002; Lee et al., 2002).

Objective measurements. Despite the quantitative nature of the measurements in the previous section, these measurements are subjective, based on human perceptions and subject to the vagaries of human interpretation of the state of information quality and the meaning of the questions asked. This section shifts the focus to objective measurements, beginning with a look at formal definitions, followed by an introduction to the difficulties associated with measuring information quality objectively, and proceeding to a discussion of proposed metrics and measurement scales.

Wand and Wang (1996) used an ontological perspective to develop rigorous definitions of the dimensions. Drawing on communication theory and information economics, they adopted the fundamental notion that "the information system is to provide a representation of an application domain (also termed the real-world system) as perceived by the user" (p. 88). Based on this notion, they developed a formal definition of an information system and its ideal state as a correct representation of the real-world system. Information quality problems thus manifest themselves as one of four types of deficiency: incomplete representation, ambiguous representation, meaningless states, and incorrect states. Each of these is defined precisely and formally in terms of a mapping from the real-world system to the information system and back. Information quality dimensions (or their negative counterparts) are defined in terms of represented states and deficiencies. For example, an information system is inaccurate if it "represents a real-world state different from the one that should have been represented" (p. 93). Similarly, inconsistency is a state in which "the representation mapping is one to many" (p. 94).

Although such formal definitions have been developed, the ability to operationalize measurements for some information quality dimensions continues to be problematic for researchers. Accuracy, in particular, is especially troubling. As Redman (2005) put it, "there is nothing akin to length, viscosity, impurities in parts per million, impedance, or other physical dimensions" (p. 23). He went on to note that "all measurements of data accuracy must, of necessity, make reference to human knowledge, other data, or the real world" (p. 23).

Redman (2005) proposed a four-component framework for measuring accuracy. First, one must consider the point of measurement, which can be the point at which data are transmitted from a data supplier, as they enter a database, as delivered to an end user, as the user perceives them, or across the entire chain. Second, one must decide which data to include in an accuracy measurement. For instance, one might include all the data in a database or perhaps only specified key attributes. Third, the measurement device or mechanism must be considered. Accuracy can be measured in a number of ways, such as by inspection, by tracking its flow through an information chain, by comparison against the real world, or by comparison against a set of permissible answers in accordance with specified business rules. Finally, one must determine the level of analysis, such as the field level versus the record level.

Redman (2005) also proposed a set of simple accuracy metrics useful at either the field or record level. For fields, the proposed metric is the ratio of the number of fields judged correct to the number of fields tested. For records, the ratio is specified as the number of records judged correct in all their fields to the total number of records tested. By definition, these metrics have limited applicability. A more generic set of metrics is proposed by Pipino et al. (2002), who specified three basic functional forms: simple ratio, minimum or maximum operator, and weighted average. Regardless which of these forms is used, the metric should be normalized to a value between zero and one, with one representing the ideal.

The simple ratio is very similar to prior constructs for accuracy, such as Oman and Ayers' (1988) measure of percent correct or Paradice and Fuerst's (1991) stored error rate. The Pipino et al. (2002) approach differs slightly in that it proposes arranging the construct in such a way that improvement always results in a higher number. For example, rather than measuring the ratio of errors, they proposed measuring the error-free rate. To measure this dimension, one simply divides the number of units in error by the number of units tested. In addition to freedom-from-error, the simple ratio is deemed useful for measuring completeness, consistency, conciseness, relevancy, and ease of manipulation (Pipino et al.).

The minimum or maximum operator is slightly more complex. The minimum operator takes the lowest value from a set, and is deemed useful for dimensions such as believability or appropriate amount of data. For example, if one has three different measures of believability representing a source rating, an experience-based rating, and a definition-based rating, the minimum operator assigns the lowest of the three to the believability dimension. As Pipino et al. (2002) put it, "assume the believability of the data source is rated as 0.6; believability against a common standard is 0.8; and believability based on experience is 0.7. The overall believability rating is then 0.6 (the lowest number)" (p. 214). By contrast, the maximum operator takes the highest value from a set and assigns this value to the dimension. This operator is appropriate for the accessibility and timeliness dimensions. In particular for the latter, a suggested formula is the maximum of "0 and 1 minus the ratio of currency to volatility" (Pipino et al., p. 215), where currency is the age of the data and volatility is the "length of time the data remains valid" (p. 215). The net effect of this formula is a number very close to one until shortly before the data cease to be valid, at which time the value rapidly moves to zero. Other dimensions for which they consider these operators appropriate include the appropriate amount of data, measured using minimum, and accessibility, measured using maximum (Pipino et al.).

For their final measurement form, Pipino et al. (2002) proposed the use of weighted average as an alternative for wherever the minimum operator might otherwise be used. As with the other metrics, the relative weightings should be normalized, resulting in a score between zero and one. They further suggested that the determination of which is more appropriate (i.e., minimum/maximum or weighted average) is a matter of the degree of an organization's understanding of the relative importance of the factors considered. This metric is essentially a multivariate alternative to the minimum operator.

A final objective measurement concept to be discussed is the need to consider the scale type, such as ratio, interval, ordinal, or nominal. Pipino et al. (2005) noted, "typically the definition of these dimensions and their associated metrics are based on an intuitive understanding or industrial experience" (p. 37). They went on to caution that "lack of attention to scale type . . . can lead to improper interpretation and application of measurement results" (p. 38), especially when combining dimensions to obtain a single metric. To address this problem, Pipino et al. developed precise, formal definitions for completeness, correctness, system currency, storage time, and volatility, and they demonstrated that each of these dimensions can be measured with ratio scale.

Table 1 provides a summary comparison of the various dimensions and their measurement, as described in this and the preceding sections. The first column identifies the dimensions included in either the Wang and Strong (1996) model or the PSP/IQ model (Kahn et al., 2002; Lee et al., 2002). The second column indicates the hierarchical category assigned by Wang and Strong (1996), the third column shows the associated PSP/IQ quadrant, and the fourth column lists the objective metric type proposed by (Pipino et al., 2002).

Combined measurements. The sections above consider various ways of measuring information quality, either subjectively or objectively, from an assortment of perspectives. Each of these approaches falls short of Wang and Strong's (1996) call for "an overall data quality metric" (p. 637). One recommendation that comes close is described in Pipino et al. (2002). In addition to the three generic forms, Pipino et al. proposed the use of a simple two-by-two grid, with high-low subjective assessments on one axis and objective assessments on the other. The resulting mapping onto one of the four quadrants can then be used as an overall gauge of information quality. While this approach does allow the assignment of a single metric, the differences in scale limit what one can do with that metric (Pipino et al., 2002; Pipino et al., 2005).

Total Data Quality Management

According to Wang (1998), TDQM is an adaptation of Deming's Total Quality Management (TQM). In a manner reminiscent of Deming's cycle, TDQM has its own four-stage quality cycle, in which the stages are *define*, *measure*, *analyze*, and *improve*. TDQM draws on the parallels between physical product manufacturing and information product (IP) manufacturing, and thus treats information as a product. This analogy emphasizes "the fact that the information output from an information manufacturing system has value that is transferable to the consumer" (p. 60). Given this manufacturing analogy, four stakeholder roles have been identified: information suppliers, who "create or collect data" (p. 60); information manufacturers, who "design, develop, or maintain the data and systems infrastructure" (p. 60); information consumers, the recipients and users

Table 1.
Information Quality Dimensions, Categorization, and Measurement

Dimension	Wang and Strong (1996) Category	PSP/IQ Quadrant	Pipino et al. (2002) Metric Type
Access security	Accessibility	Dependable	Min/max or weighted
Accessibility	Accessibility	Useable	
Appropriate amount	Contextual	Useful	Min/max or weighted
Believability	Intrinsic	Useable	Min/max or weighted
Completeness	Contextual	Sound	Simple ratio
Conciseness	Representational	Sound	
Consistency	Representational	Sound	Simple ratio
Ease of operation		Useable	Simple ratio (ease of manipulation)
Ease of understanding	Representational	Useful	
Free-of-error (accuracy)	Intrinsic	Sound	Simple ratio
Interpretability	Representational	Useful	
Objectivity	Intrinsic	Useful	
Relevancy	Contextual	Useful	Simple ratio
Reputation	Intrinsic	Useable	
Timeliness	Contextual	Dependable	Min/max or weighted
Value-added	Contextual	Useable	

of the IP; and an IP manager, who is "responsible for managing the entire IP production process throughout the IP life cycle" (p. 60).

The definition stage of the TDQM cycle involves several components. First, characteristics of the IP are defined at two levels: the consumer's functional requirements, and the basic IP elements and their interrelationships. The definition stage also includes defining requirements from the perspective of each of four roles identified above. Finally, the information manufacturing system itself is defined in terms of its inputs, outputs, processing and storage points, and points at which quality improvement steps can be inserted (Wang, 1998).

Metrics are collected during the measurement stage. These metrics can be basic quality metrics, such as error rates or referential integrity rates, or they can measure more complex quality concepts such as conformance to business rules. Metrics may also be collected to track production-related issues, such as updates made by various departments or unauthorized access attempts over a certain period of time (Wang, 1998).

During the analysis stage, root causes of errors are investigated. It is also appropriate at this stage to evaluate the metrics themselves, asking, for example, whether they address the right set of requirements or considering the degree to which they represent the issue at hand. Finally, following analysis, improvement takes place. This can involve the correction of errors or it can involve process-related steps aimed at reducing the introduction of new errors (Wang, 1998).

Wang (1998) recommends a four step methodology for implementing and following the four stages of TDQM. First, one should "clearly articulate the IP in business terms" (p. 61). Second, establish an IP team, including a senior executive to be TDQM champion, one or more IP engineers, who must be familiar with TDQM concepts, and representatives from each of the four roles. Third, teach information quality skills to the relevant constituencies, and fourth, "institutionalize continuous improvement" (p. 61) of the organization's information products.

Data Production Maps

As with TDQM, data production maps build upon the analogy between physical product manufacturing and that of information products. Davidson, Lee, and Wang (2004) defined an information product as "a collection of data element instances that meet . . . specified requirements" (p. 225). The data production map approach sets forth a method of visualizing the movement of data through the production process, much as physical products move through an assembly line, and it provides tools for analyzing and tracing quality aspects of the IP at different stages of production, including in the information product supply chain. Data production maps are based on the notion that an IP has intrinsic value to consumers, and that there is cost associated with both the production and improvement of an IP. Based on this consumer-focused value perspective, the approach helps to identify the trade-offs associated with decisions about the information production process (Ballou et al., 1998; Davidson et al.; Scannapieco et al., 2005; Shankaranarayanan, 2005; Shankaranarayanan & Cai, 2006; Shankaranarayanan et al., 2000).

The unit of analysis for data production maps, referred to as a data unit, can represent an array of constructs, such as "a number, a record, a file, a spreadsheet, or a report" (Ballou et al., 1998, p. 463). Originally, data production maps included five building blocks: data vendor blocks, processing blocks, data storage blocks, quality blocks, and customer blocks. Data vendor blocks represent the point at which data units enter the system, and can represent any source, internal or external to the organization. Processing and storage blocks represent information system processes and storage, whether automated or manual. Quality blocks represent points at which steps are taken for the express purpose of improving the quality of data units. The "effect of a quality block could be modeled by specifying the fraction of apparently defective units entering and the fraction leaving" (p. 466). Finally, customer blocks represent the points at which end users receive the finished IP (Ballou et al.).

In addition to the building blocks, Ballou et al. (1998) described the use of an "information manufacturing analysis matrix" (p. 472), in which each row represents a data unit and each column represents a building block. The intersecting cells are used to record values for a set of prescribed data cost and quality parameters. Once recorded in this manner, the matrix forms the basis for determining net costs and values, which are of interest to information producers and customers, respectively.

Shankaranarayanan et al. (2000) extended and enhanced the Ballou et al. (1998) model in several important ways, and named the resulting extended version of the concept IP-MAP. First, they formalized a set of value constraints from the customer's perspective, enabling those requirements to drive the entire design. They also specified constraints on data units entering the system, permitting quality controls to be placed on incoming data prior to their use. In addition, they added metadata specifications, providing a way to capture and track information about the various blocks and about data quality throughout the process. Finally, they added specifications for three additional block types: decision blocks, organizational boundary blocks, and information system boundary blocks.

Another important enhancement was proposed by Scannapieco et al. (2005), who leveraged the extension mechanism in the Unified Modeling Language (UML) specification to develop a UML profile based on IP-MAP. The UML extension mechanism permits the development of such profiles, which are specified as sets of constraints, tag definitions, tag values, and stereotype class definitions. UML analysis "is based on the notion of a model element, defined as an abstraction drawn from the system being modeled" (p. 117). Examples of model elements include classes and relationships. The IP-MAP profile consists of three models: a data analysis model, a quality analysis model, and a quality design model. The data analysis model includes three UML stereotype classes, representing information products, raw data, and component data, each of which is a direct analog of the corresponding concepts in IP-MAP. The quality analysis model includes stereotype classes representing quality requirements, permitting the specification of any set of information quality dimensions. Finally, the quality design model captures the dynamic characteristics of an IP-MAP.

A methodology called IP-UML uses this UML profile to produce a set of "design patterns for quality improvement" (Scannapieco et al., 2005, p. 123). The methodology consists of three phases: data analysis, quality analysis, and quality improvement. Together, these three phases produce class diagrams and activity diagrams that document the information production process, and a set of design patterns for improving information quality (Scannapieco et al.).

Benchmarking

Lee et al. (2002) developed a methodology useful for identifying aspects of information quality needing attention. Their methodology expands on the use of PSP/IQ to establish benchmarks and includes two forms of gap analysis. The first is referred to as a “Benchmarking Gap Analysis” (p. 140). In conducting this analysis, an organization obtains information representing “tough competitors, industry leaders, or other sources of best practice” (p. 140). This information, along with data on one’s own organization, is then plotted on a two-dimensional grid, with the horizontal axis representing the cumulative percentage of respondents, sorted from lowest to highest scores, and with the vertical axis representing the scores themselves. By plotting the reference scores and the organization’s scores as separate lines, one can then visually assess the differences. Lee et al. recommended consideration of three indicators: the size of the gap area between the two lines, the location of the gap with respect to the vertical axis, and differences in the size of the gap at various places along the horizontal axis. This analysis should be performed for the aggregate scores and for each of the quadrants separately.

The second gap analysis is the “Role Gap Analysis” (Lee et al., 2002, p. 141). Although not a benchmark in the traditional sense, it is useful as an intra-organizational benchmark. For this analysis, the scores from each of the different stakeholder groups (i.e., data providers, custodians, and consumers) are plotted separately. Instead of producing a line representing sorted responses, as in the previous analysis, the responses are aggregated, and single points are plotted to represent the mean scores within each stakeholder group. Once plotted, the following indicators are considered: the size of the gap, the vertical location of the gap, and the direction of the gap. The size of the gap provides an indication of how close or far apart the perceptions of each group are. For instance, a large gap indicates a significant difference of opinion. The location of the gap indicates the nature of improvements needed: A gap at the low end suggests that radical changes may be in order, whereas a gap at the high end suggests incremental improvements. Finally, the direction of the gap is an important indicator in the sense that if, for example, information system professionals hold a higher opinion of the data quality than do the information consumers, there is probably a lack of awareness on the part of the information systems staff (Lee et al.).

Contributors to High Information Quality

Research literature suggests that there are two primary contributors to high information quality: information systems themselves, and organizational factors. This section presents key findings and concepts from literature examining either of these two factors. First, the contribution of information systems is examined, focusing on a line of research that identifies information quality as an outcome of successful information systems deployment, that is, as a dependent variable. Second, an array of information quality literature is presented, painting a composite picture of the role of organizational factors and their contribution to information quality.

The Contribution of Information Systems

In 1992, DeLone and McLean published a landmark article in which they explored the notion of information system (IS) success. In seeking an explanation for IS success as a dependent variable, they developed a taxonomy of IS success consisting of six dimensions: system quality, information quality, use, user satisfaction, individual impact, and organizational impact. Relating this taxonomy to information theory (Shannon, 1948; Shannon & Weaver, 1949), they proposed that system quality correlates to Shannon and Weaver's technical level, whereas information quality correlates to the semantic level. These two dimensions contribute to both use and user satisfaction, which in turn contribute to individual impacts, ultimately impacting the organization as a whole. This taxonomy is illustrated in Figure 7. As DeLone and McLean described it, the arrangement of these dimensions is intended "to suggest an *inter*dependent success construct while maintaining the serial, temporal dimension of information flow and impact" (p. 83).

In a follow-up study, DeLone and McLean (2003) investigated the body of available research based on the 1992 model. During their investigation they noted that a number of researchers had expressed difficulty in applying the model. They attributed this difficulty, in part, to a lack of clarity with respect to the independent and dependent variables. They used a process model to clarify the original intent, stating, "this process model has just three components: the creation of a system, the use of the system, and the consequences of this system" (p. 16). They also proposed a modified taxonomy, adding service quality as a dimension, and combining individual and organizational impacts into a single dimension, net benefits. The revised taxonomy is shown in Figure 8.

Among the research they reviewed were several articles that evaluated the role of information quality, which was shown "to be strongly associated with system use and net benefits" (p. 21). There were also several studies focusing on system quality. In this context, it should be noted that, according to DeLone and McLean (2003), "system quality was measured in terms of ease-of-use, functionality, reliability, flexibility, *data quality*, portability, integration, and importance" (p. 13, emphasis added).

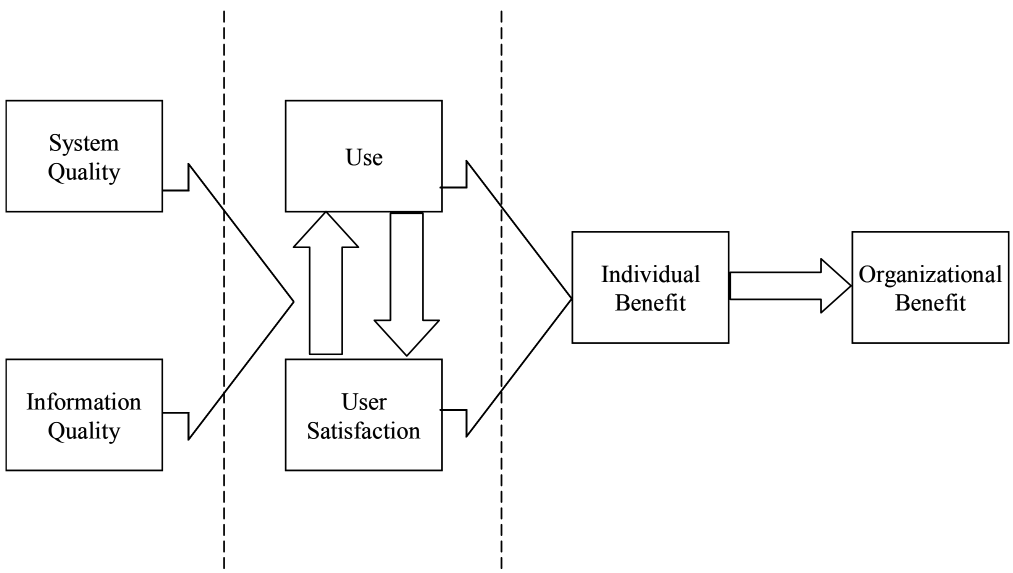

Figure 7. DeLone and McLean model of information systems success.

Note. Adapted from DeLone, W. H. and McLean, E. R. (1992). Information systems success: The quest for the dependent variables. Information Systems Research*, 3(1), 60-95.*

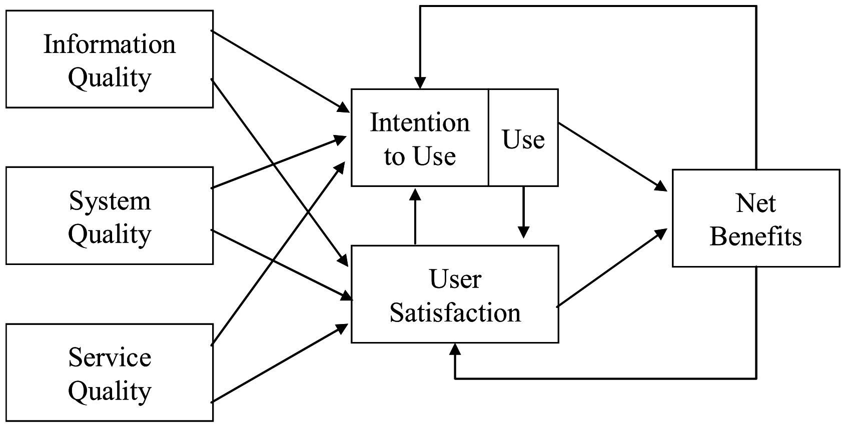

Figure 8. Modified DeLone and McLean IS success model.

Note. Adapted from DeLone, W. H. and McLean, E. R. (2003). The DeLone and McLean model of information systems success: A ten-year update. Journal of Management Information Systems, *19(4), 9-30.*

Organizational Contributors to Information Quality

This section describes and evaluates research studies employing one or more of the information quality management approaches above, and concludes with an evaluation of factors associated with success or failure. Most of the studies were reported as successes; others, while not identified as failures, were deemed as such, given their focus on information quality problems without any indication of successful resolution.

Successful improvement activities. Scannapieco et al. (2005) reported on a case study of two Italian government agencies that utilized a methodology based on data production maps (Ballou et al., 1998; Shankaranarayanan et al., 2000). The goal of the agencies was to "improve the quality of addresses stored in the public administration databases" (Scannapieco et al., 2005, p. 125). Among the problems identified was that update responsibility was highly decentralized, even for a single location. For example, in a small town one might find the city council responsible for part of the address and the post office responsible for another. An especially difficult problem with poor address data was found in Venice. As the authors put it, "the notion of street (in Italian, '*strada*') does not exist in Venice; instead there is the notion of '*calle*'" (p. 127).

The solutions to these problems focused on process issues. At the national level, they reengineered several inter-administration processes, paying particular attention to improving communication among the administrations. To handle the problems in Venice, they added a manual quality check process to be implemented "by all administrations storing locating data related to Venice" (Scannapieco et al., 2005, p. 127). To facilitate timely communications, they implemented an automatic publish-and-subscribe notification system.

Bertoletti, Missier, Scannapieco, Aimetti, and Batini (2005) reported on a project within the Italian e-Government Initiative called "Services to Business" (p. 151). This project treated information as both product and service. The focus of the project was on improving the organizational service of the agencies involved by improving the quality of the information they collect, maintain, and provide, and by improving the sharing of that information among agencies.

The solution involved a major redesign of the way master data about businesses is handled by government agencies. Rather than have each agency collect and maintain the same basic information (e.g., name of the business, primary address, key officials, etc.), a central database was created and systems were redesigned to enable a completely revised process. After redesign, the agency that first interacts with a business becomes a "one-stop shop" (Bertoletti et al., 2005, p. 152) through which that business subsequently interacts with the government; other agencies share data behind the scenes through back-end service interfaces as necessary.

The motivation behind the case was the promise of improving the quality of service while reducing the fees charged to businesses. By this measure alone, the project was deemed a success in that the initial deployment among three agencies resulted in the complete elimination of misaligned data (originally estimated at 20% to 25%) and a 40% reduction in administrative costs for the identified transactions. Organizationally, the project was assigned to a single agency, which then coordinated with each of the other agencies involved (Bertoletti et al., 2005).

Kerr and Norris (2004) described a participative case study in which the New Zealand Ministry of Health employed the Total Data Quality Management (TDQM) approach (Wang, 1998) in its effort to improve the quality of clinical data reported by the ministry. The ministry developed a data quality assessment framework that addressed the entire flow of data, including data reported to the ministry as well as that reported by the ministry. The framework served as a data quality measurement tool, as a benchmark for assessing the effectiveness of future improvement initiatives, and as a template for use by both internal and external users.

Using a variety of qualitative techniques, including interviews, open-ended questionnaires, and focus groups, Kerr and Norris (2004) determined stakeholder needs and established priorities for the framework. Next, they conducted a pilot test on "three very different health data collections" (p. 221). After a few modifications, the framework was put into production. In a follow-up assessment, the framework was deemed "an invaluable tool that helps developers to produce robust and valid clinical databases" (p. 222). It was also reported to be highly valuable to hospital personnel responsible for submitting clinical data to the government. Plans call for using the lessons learned through this project "to form a prioritized list of data quality improvement initiatives across the Ministry" (p. 224).

Helfert and Herrmann (2005) described a case at a large financial services company that adopted the product metaphor for improving the quality of data in a data warehouse. The transfer of data into the data warehouse was centrally controlled via a set of processes that provided "a few basic data-quality checks" (p. 139), such as checking for duplicate keys or missing values. The data warehouse included metadata regarding the transfer process, but none about data quality. At the start of the project, users had very low confidence in the data and complained frequently, not only about the poor quality, but also about the inconsistent quality and the inability to distinguish good data from bad.

The project proceeded in four stages, beginning with requirements gathering and problem definition, followed by strategy development, solution planning, and implementation. The identified problems fell into two groups: data models, with emphasis on interpretability and usability; and data values, primarily inconsistent values. The data modeling problems were moved to a separate metadata management project, and were not considered further in the study. During the strategy development stage, the project team narrowed the focus to a select set of quality criteria due to time and budget constraints, and developed a conceptual architecture with both technical and organizational components (Helfert & Herrmann, 2005).

Their solution filled in details needed to implement the conceptual architecture. Technical components consisted mostly of quality rules, expressed in Structured Query Language, and integrated into the metadata management system. Organizational components included recommendations on organizational structure, role responsibility specifications, and process definitions. Organizationally, they considered a centralized data quality officer role, but chose to create two positions, one focusing on technical aspects of data quality, and one focusing on business aspects. They also defined responsibilities for project managers and data suppliers, and wrote a set of processes to provide "closed feedback" (Helfert & Herrmann, 2005, p. 145).

Helfert and Herrmann (2005) reported that "overall, the initiation of data-quality management can be characterized as a success" (p. 145). Although no quantitative analysis was performed, they observed a "significant reduction" (p. 145) in complaints and an increase in user acceptance of the data warehouse. The processes were well accepted, a fact "attributed to the continuous involvement of business users and technical staff in the data-quality project" (p. 145). They reported four significant findings. First, the team leadership believed that having a clear data quality strategy was important to the project's success. Second, the ability to produce quick results was essential in overcoming initial management resistance and reluctance to support. Third, the definition and establishment of process changes were difficult, but succeeded because of stakeholder involvement and buy-in. Finally, involving technical and business people from the start and throughout the project was considered key to success (Helfert & Herrmann).

Katz-Haas and Lee (2005) reported on the results of a participative case study and action research at a large managed health care organization that had been created as the result of a merger. This study treated information as a resource, and concluded that the root problem was organizational rather than technical. The problem that brought the issue to management's attention was the waste of approximately $4 million annually in overpaid claims, primarily on behalf of former members no longer eligible for benefits. Analysis revealed that there were "data in the warehouse showing 40,000 – 60,000 members per month as active when in fact their policies had been canceled" (p. 169).

One factor that made this case particularly difficult to solve was that the organization's incentive structure was working against solving it. In particular, each organizational unit was responsible for maximizing its own bottom line without regard to how that unit's "actions unintentionally sub-optimized the bottom line of the company as a whole" (Katz-Haas & Lee, 2005, p. 176). Recognizing that cultural change would be needed to solve the problem, organizational resistance was considered a serious risk. The risk was mitigated by succinctly articulating the problem in easy-to-understand terms. "When the team crystallized the initial problems as millions in overpaid claims, the team gained more legitimacy and, thus, garnered support from participating business areas for investigating root causes of this problem" (p. 177).

Among the root causes determined during the study were process complexity, system issues, human-computer interface problems, manual processes, open feedback loops, and conflicting cycle times among the involved departments. With respect to process complexity, the authors wrote, "the case-cancellation process, a relatively 'small' process, involved fifteen departments, eighteen subprocesses, sixty third-level processes, seven systems, three of the five company segments/businesses, and over 300 employees" (Katz-Haas & Lee, 2005, p. 172).

Activities lacking success. Wang et al. (1998) reported on several cases in which information quality problems were identified, but not resolved. Despite the lack of reportable success, these cases are instructive, especially when considered together with the successful examples discussed above. One case reported by Wang et al. involved an international investment bank that had a variety of information needs focusing on customer accounts. Among these were the need to link a newly established account to all existing accounts for the same customer, the ability to enable immediate trading on new accounts, and the need to know in real time the total balance across all accounts for a single customer. They also needed to be able to immediately close all accounts for a given customer upon notification by the government in the event certain laws had been broken by the accountholder.

Among the problems identified at the bank was that various departments needed real-time information regarding changes in total account balances; however, the bank's centralized customer account database was updated nightly with balance information. Moreover, non-balance updates to the database were ad hoc. This left each department to its own devices in terms of assuring the accuracy of non-balance data and in getting account balances more current than the previous night. As a consequence, several departments developed their own databases which were inconsistent with one another. A related problem was the difficulty in giving customers investment advice appropriate for their level of risk; failure to do so led to a high level of customer dissatisfaction and created indemnification exposures (Wang et al., 1998).

Wang et al. (1998) also discussed an eyewear company that had retail outlets throughout the United States, all supported by four laboratories. Together, they fulfilled approximately 25,000 eyeglass orders per week. This company failed to recognize the needs of internal information consumers, and as a consequence, was experiencing a fifteen percent error rate in the production of glasses, requiring rework costing the company on the order of $1 million annually, plus loss of customers due to dissatisfaction. A root cause identified by Wang et al. was the failure "to treat lens specifications as an information product with the grinder as consumer" (p. 99).

Another example discussed by Wang et al. (1998) was that of a data company that would be considered a success story if one were to limit the focus to the company's commercial information products. This company was in the business of collecting data gathered from hundreds of millions of retail transactions per week and selling detailed buying behavior reports to the retail industry. The company had well-defined production processes for these products, even including the use of neural networks to estimate missing values with a high degree of accuracy. The problem encountered by this company was its failure to apply similar rigor with respect to internal information. In particular, despite its dominant position in the marketplace, the company was making very slim profits. The root problem was identified as a failure to manage the quality of information regarding the cost of producing its commercial products. This failure led to an inappropriate pricing model, hence the slim margins.

Another instructive example is found in Thornsbury, Davis, and Minton (2003). As with the cases above, this article identifies a set of information quality problems, yet does not report success in resolving them. The authors of this study proposed treating information as a resource in order to solve a quality problem with information provided by government agencies to the Florida citrus industry. The authors noted that even in this narrowly defined industry in one state, official data come from more than 50 publications of 13 different governmental agencies, creating inconsistencies and confusion for users. For example, "information concerning lemons and limes may be reported separately, the two may be combined, or they may be included with specialty fruit; however, in many publications they are reported as tropical fruits and not included with citrus" (p. 557).

Such problems notwithstanding, the information is considered highly valuable, both to the citrus industry and to researchers. In arguing that it be treated as a valuable resource, Thornsbury et al. (2003) identified four categories of end user value and linked the value of information to its role in improved decision making. Despite the value to the end users, they noted that "since data is nonrival in consumption yet may be highly valuable if exclusive, firms have a disincentive to invest unless access can be limited" (p. 551). They proposed a solution consisting of both technical and organizational components, but noted that "there are, of course, a number of political and technical issues to be resolved" (p. 565)..

Identifying the Success Factors. The research studies discussed above highlight several factors that appear related to the success or failure of an information quality initiative. Given that different authors emphasized different aspects of their studies, it is not clear whether any of these factors, by themselves, are sufficient to assure success or to doom an initiative to failure; however, it is clear that by focusing on at least some of these factors, an organization can increase its likelihood of success.

The first factor that stands out is the clear articulation, in business terms, of the strategy, objectives, or desired outcome of the activity. Helfert and Herrmann (2005) identified the clarity of their strategy as a success factor, and Katz-Haas and Lee (2005) pointed out the value of stating the problem in terms of millions in overpayments. For Bertoletti et al. (2005), the message was articulated in terms of one-stop shopping, improved service, and reduced costs. By contrast, Wang et al.'s (1998) eyewear company missed the opportunity to point out $1 million in avoidable rework, and although Thornsbury et al. (2003) clearly stated how lemons and limes defy consistent categorization, they did not articulate the value of making improvements.

Another significant factor is the need to understand stakeholder requirements. In both Helfert and Herrmann (2005) and Kerr and Norris (2004), the projects started by focusing on the needs of their constituencies. This stands in contrast to the eyewear company, which failed to take the lens grinders' needs into account, the data company, which failed to consider those who priced their products, and the investment bank, which failed to consider either their internal departments' need for accurate and current information or their external customers' needs with respect to risk management (Wang et al., 1998). Thornsbury et al. (2003) did little other than note that "a number of political and technical issues" (p. 565) were left unanswered.

Several authors pointed out the importance of paying attention to processes. Scannapieco et al. (2005) based their solution on process reengineering. For Bertoletti et al. (2005), the primary objective was to revise processes for government-to-business interaction. In Helfert and Herrmann (2005), formal process definition was a major deliverable. By contrast, Wang et al.'s (1998) financial and eyewear companies each overlooked the role of processes.

Four of the studies noted the value of starting small and attaining success quickly. After defining stakeholder requirements, Helfert and Herrmann (2005) removed one set of issues to a different project and purposefully narrowed the scope of what remained. Kerr and Norris (2004) started with a pilot test before expanding to the entire ministry. Bertoletti et al. (2005) limited their initial focus to three government agencies. Katz-Haas and Lee (2005) focused their efforts on a single process. By contrast, Thornsbury et al. (2003) appear to have targeted the entire Florida citrus industry, including at least 13 government agencies at both the state and federal levels, as well as countless agricultural producers, processors, distributors, and researchers.

The final, and perhaps most important, success factor is the clear assignment of responsibility for information quality. Only two studies, Helfert and Herrmann (2005) and Bertoletti et al. (2005), explicitly identified this as a factor; however, it was implicit in all the successful examples, in that there was a responsible entity in each study. On the other hand, for each study that lacked success, there was a corresponding lack of an identifiable responsible party (Thornsbury et al., 2003; Wang et al., 1998).

Table 2 provides a summary of these success factors. A checkmark in a table cell indicates that the identified factor was shown in either a positive or negative light as being a contributor to either the success or failure of a particular study.

Table 2.
Summary of Information Quality Success Factors

	Clearly articulated objectives	Understand stakeholder needs	Focus on processes	Start small	Identify responsible organizations
(Bertoletti et al., 2005)	√		√	√	√
(Helfert & Herrmann, 2005)	√	√	√	√	√
(Katz-Haas & Lee, 2005)	√			√	
(Kerr & Norris, 2004)		√		√	
(Scannapieco et al., 2005)			√		
(Thornsbury et al., 2003)	√	√		√	√
(Wang et al., 1998)	√	√	√		√

Summary of Factors Contributing to Information Quality

As noted in the introduction to this section, both information systems and organizational factors contribute to information quality. The information systems success literature identifies information quality as part of a complex dependent variable, and it identifies studies demonstrating an empirical connection between information quality and other success indicators. Separately, information quality literature often focuses on the organizational factors that contribute to information quality. A side-by-side comparison of this literature reveals a number of common themes.

Strategy

This section of the chapter shifts attention to the research literature addressing strategy. The section begins with an exploration of some of the major schools of thought in strategy research, including the linkage between those schools of thought and early economic theory. The section continues with a discussion of literature examining the concept of competitive advantage and the role that management plays in creating and sustaining competitive advantage. The section then moves closer to the core of this research by presenting literature exploring the relationship between information systems, information technology, and strategy. The section concludes by focusing on literature that specifically addresses the relationship between information quality and strategy from two angles: strategies for improving information quality, and information quality as a contributor to organizational strategy.

Schools of Thought in Strategy Research

General business strategy has an extensive body of literature that is at times deeply rooted in, and at other times parallel to, other theories such as economics, organizational theory, and organizational behavior (Seth & Thomas, 1994). Porter (1991) described the focus of this literature quite simply: “The reason why firms succeed or fail is perhaps the central question in strategy” (p. 95). As Seth and Thomas (1994) put it, “a strategy is the pattern or plan that integrates an organization's major goals, policies and action sequences into a cohesive whole" (pp. 166-167).

Two major schools of thought in the business strategy literature provide perspective relevant to this research. One focuses on the competitive environment in which the firm operates (Porter, 1991, 1996), while the other focuses on the firm itself and the resources the firm owns and controls (Barney, 1991, 2001). Whether these perspectives are compatible is a matter for debate (Barney, 1991, 2001; Porter, 1991; Priem & Butler, 2001; Wade & Hulland, 2004). The resource-based view considers strengths and weaknesses (Barney, 2001; Wernerfelt, 1984), whereas the environmental view considers opportunities and threats (Barney, 2001). Based on this distinction, Barney (2001) posits that the case can be made for treating the resource-based view and the environmental view as distinct strategy theories. Seth and Thomas (1994) consider the two perspectives to be "complementary" (p. 178) and suggested that "a firm's competitive position needs to be studied with respect to both dimensions" (p. 178). Porter (1991) rejected these characterizations, suggesting instead that "resources are not valuable in and of themselves, but because they allow firms to perform activities that create advantages in particular markets" (p. 108). He went on to suggest that the resource-based view should be characterized as one of several streams of literature that help explain the nature of competitive advantage in a dynamic environment, rather than as a separate and distinct theory. The issue is blurred even further by Wade and Hulland (2004), who, in discussing the resource based view, "define resources as assets and capabilities that are available and useful in detecting and responding to market opportunities or threats" (p. 109).

Each of these perspectives, and even the debate on how the perspectives relate to one another, traces back to early economic theory. Barney (2001) stated that the resource-based view is rooted in the economic theory of Ricardo in the nineteenth century. The central role of a firm's resources, which Barney defined as "all assets, capabilities, organizational procedures, firm attributes, information, knowledge, etc. controlled by a firm" (p. 101), is in fact directly traceable to Ricardo's view of the fundamental unit of economics as the set of "material commodities owned by individuals" (Cranfill, 1940, p. 73). This theoretical perspective was expanded upon by Penrose, who conceived of the firm as "a collection of productive resources" (as cited by Seth & Thomas, 1994, p. 176), thus forming the foundation of what is now known as the "resource-based view of the firm" (Seth & Thomas, p. 177). These resources include both tangible and intangible assets.

By contrast, Porter (1991) considered the firm to be "a collection of activities . . . [performed to] create advantages in particular markets" (p. 108). This perspective traces its roots to the early twentieth century work of Commons, who defined an institution "as collective action in control, liberation, and expansion of individual action" (Commons, 1931, p. 649). Commons developed the concept of institutional economics (Commons; Cranfill, 1940), tracing its origin to the philosophy of David Hume, who "found the unity of [economics, jurisprudence, and ethics] in the principle of scarcity and the resulting conflict of interest" (Commons, p. 650).

Commons considered the transaction, rather than the commodity, to be the fundamental economic unit (Commons, 1931; Cranfill, 1940). In his view, the transaction is distinct from the classical focus of the exchange of commodities, in that the transaction involves the transfer of legal rights to something in the future, rather than the transfer of physical possession in the present. As he articulated the difference between the two views,

> these individual actions are really *trans*-actions instead of either individual behavior or the 'exchange' of commodities. It is this shift from commodities and individuals to transactions and working rules of collective action that marks the transition from the classical and hedonic schools to the institutional schools of economic thinking. (pp. 651-652)

Commons posited that the transaction contains all the essential elements of economic analysis. He named these elements "conflict, dependence, and order" (Cranfill, p. 71). He also identified three basic types of transaction: bargaining, managerial, and rationing (Commons; Cranfill).

Commons did not perceive his work to be a radical departure from prior economic theory, merely a shift in perspective. Cranfill (1940) described this perspective shift by writing, "he deals with dynamics not with statics; . . . with human beings not with commodities; with social institutions not with 'natural right' individuals; . . . with conflict of interest not with harmony of interests" (p. 63). He perceived his theory to be aligned with other psychological economic theories, but emphasized that economics should also be considered in its relationship with other social sciences (Cranfill).

Writing several decades later, Seth and Thomas (1994) used the term "neoclassical" (p. 169) to describe such early economic theory, and asserted that it was based on an assumption of "economic Darwinism" (p. 169), in which "the environment rewards with survival those firms which select strategies which happen to be optimal" (p. 169). They then went on to suggest, however, that a moderated form of economic Darwinism, that is, one that recognizes the ability of the firm to make "proactive decisions to optimally utilize its unique specialized resources" (p. 170), is relevant to modern strategy theory. One of the significant issues they identified with neoclassical theory was its assumption regarding the ability of people to make rational economic decisions. Instead, they suggested that people act with "bounded rationality" (p. 174) in the presence of uncertainty. While accepting that people behave in a purposeful and intelligent manner, they emphasized the importance of recognizing that limited information creates uncertainty, hence limits the ability to make optimal decisions. In this sense they seem to agree with Commons, writing, "the bounded rationality condition is a foundation of the transactions cost framework" (p. 175).

Competitive Advantage and the Role of Management

To understand the relationship between these strategy perspectives and the goal of affecting business outcomes, one must consider the nature of competitive advantage and of management's role in influencing competitive advantage. Porter (1991) posited that competitive advantage has three central elements: price, differentiation, and scope. With respect to resources, competitive advantage arises from the control of resources that are valuable, rare, and appropriable (Wade & Hulland, 2004). Porter added that resources should be "more valuable within the firm than outside" (p. 108).

Even more important than creating competitive advantage is the ability to sustain such advantage (Seth & Thomas, 1994). Wade and Hulland (2004) noted that “although firms possess many resources, only a few of these have the potential to lead the firm to a position of sustained competitive advantage” (p. 114). They emphasized in this regard that first-mover advantage alone is insufficient to sustain an advantage. More specifically, researchers suggest that resources must be rare, inimitable, difficult to substitute, and relatively immobile (Barney, 2001; Seth & Thomas; Wade & Hulland).

Regardless which strategy perspective is considered, the role of management in influencing organizational outcomes leading to and sustaining competitive advantage boils down to making choices. Porter (1991) put it very succinctly, stating, “the essence of strategy is choice” (p. 101). Cranfill (1940), citing Commons, stated, “managerial capacity hinges mainly upon the ability to choose the strategic factor at the ‘right time, in the right place, in the right form, and in the right quantity’” (p. 74). As Barney (2001) put it, “it is almost as though once a firm becomes aware of the valuable, rare, costly to imitate, and nonsubstitutable resources it controls, the actions the firm should take to exploit these resources will be self-evident” (p. 53).

Information Systems, Information Technology, and Strategy

In essence, the strategy literature suggests that management’s role in influencing organizational outcomes is to make choices leading to competitive advantage. If management focuses on the firm and its resources, it should make choices that lead to control of a set of relatively rare, hard to imitate, and not easily substituted resources. If management focuses on the external environment, choices should affect the firm’s ability to lower its prices, differentiate its products, or position its scope in a unique way. These need not be mutually exclusive.

Porter and Millar (1985) discussed the role of information systems and technology in the context of such choices, writing, "every value activity has both a physical and an information-processing component . . . [and] every value activity creates and uses information of some kind" (p. 152). In fact, in discussing how information affects competitiveness, the factors they identified were indistinguishable from management choices for the environmental perspective in general: reducing cost, enhancing differentiation, and changing competitive scope. Barney (1991), writing from the resource based perspective, emphasized the need for rarity and inimitability, and discounted the role of cost reduction, noting that "sustained advantage related to IS comes not from machinery, but from the interaction between information systems and decision-making processes" (p. 114). A decade later, Porter (2001) recognized that the Internet had created downward pressure on profitability throughout industry, and emphasized the need to focus on differentiation and strategic positioning. Shortly thereafter, Carr (2003) reiterated this point, saying, "what makes a resource truly strategic – what gives it the capacity to be the basis for a competitive advantage – is not ubiquity but scarcity" (p. 42).

In a separate line of research focused on investigating the relationship between information systems investments and organizational outcomes, Weill (1992) recognized that, given the variety of management objectives involved in defining success, organizational outcomes should be measured along multiple dimensions. Echoing similar sentiments, Mirani and Lederer (1998) asserted that “it would be fallacious to assume that a single, best theory of the organizational benefits of IS projects even exists” (p. 805). Based on this line of thinking, Weill proposed separating organizational outcomes into three categories: strategic, informational, and transactional. Mirani and Lederer adopted Weill’s model as the basis of their own research and focused on developing an instrument for measuring the organizational benefits of IS projects. They defined each of the three categories as follows: “Strategic IT changes an organization’s product or the way in which the organization competes. Informational IT provides the information and communication infrastructure of the organization. Transactional IT supports operational management and helps cut costs” (p. 808). They then proceeded to expand each of these dimensions into a comprehensive model as shown in Figure 9. The instrument they developed was subsequently described by DeLone and McLean (2003) as “an important contribution to IS success measurement” (p. 19).

Against this backdrop, Melville, Kraemer, and Gurbaxani (2004) developed an integrative framework of information technology business value. Their work was grounded in the resource-based view, but was influenced by other theories as well. In seeking a way to link information technology and competitive advantage, they found that sustained advantage is always associated with management skills. Echoing both Commons’ (1931) and Seth and Thomas’ (1994) comments on bounded rationality, they observed a limitation of the resource based view, namely “that it assumes that resources are always applied in their best uses, saying little about how that is done” (p. 291). To more fully explain the relationship between management choices, information technology resources, and competitive advantage, they proposed a framework with three primary lenses: the focal firm, the competitive environment, and the macro environment.

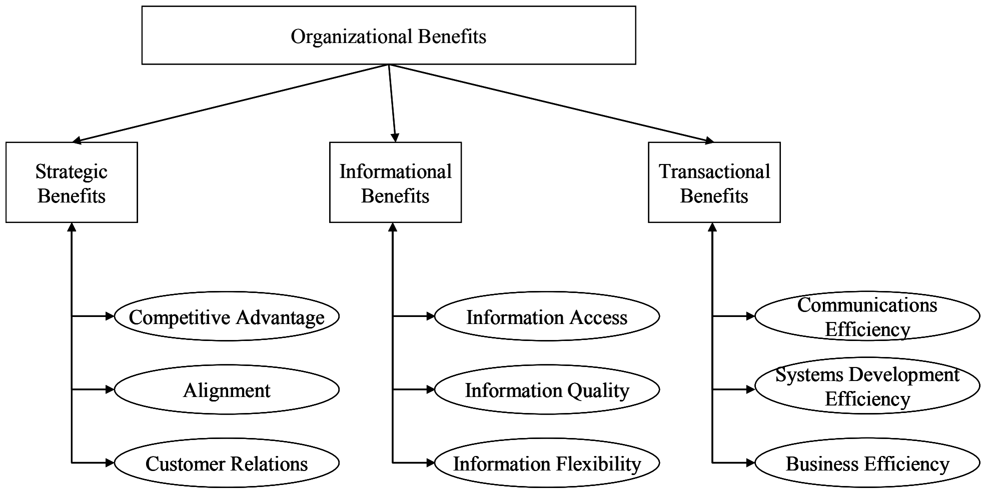

Figure 9. Organizational benefits of IS projects.

Note. Adapted from Weill, P. (1992). The relationship between investment in information technology and firm performance: A study of the valve manufacturing sector. Information Systems Research*, 3(4), 307-333 and Mirani, R. and Lederer, A. L. (1998). An instrument for assessing the organizational benefits of IS projects.* Decisions Sciences*, 29(4), 803-838.*

Information Quality and Strategy

In 1995 Redman asserted that "errors in data can cost a company millions of dollars, alienate customers, and make implementing new strategies difficult or impossible" (p. 99). Over the next decade, relatively little progress was made toward developing strategies for overcoming this problem or for relating information quality to organizational outcomes. In fact, Kerr and Norris (2004) stated, "little has been published on what constitutes a data quality strategy let alone an evaluation of a structured and tested scheme. Recently however, this type of strategy has become increasingly important as a core requirement for many businesses" (p. 225). This section presents a review of information quality strategy literature. Such literature is sparse, yet spans a period of about ten years and addresses a variety of perspectives.

Strategies for Managing Information Quality

Redman (1995) proposed a simple three-step strategy for improving data quality. First, one must identify the problem. Potential indicators of data quality problems include the expenditure of resources on inspecting and reworking data, the existence of multiple, redundant databases, difficulty obtaining the information needed to achieve a specific business purpose, and a general frustration with the organization's information technology service provider. The second step is to treat information as an asset. Toward this end, one should first develop an inventory of data owned, used, or managed by the organization. With that inventory in hand, the roles of data suppliers and customers should be identified, responsibilities should be aligned accordingly, and the organization should commit to investing "resources to improve the quality of the asset" (p. 103). The third step is the adoption of advanced quality management approaches for proactively addressing data quality problems. Rather than focusing first on cleaning up existing data, Redman recommends instead that companies focus first on preventing the introduction of errors, followed by a clean-up if necessary. To create a visual image of this third step sequence, he points out that "a database is like a lake. To ensure a clean lake, one must first eliminate the sources of pollution." (p. 106).

The specific techniques recommended by researchers vary widely. Redman (1995) advocated treating information as an asset, and suggested developing an inventory as part of one's strategy. Pierce (2004) and Campbell et al. (2004) advocated treating information as a product. In the context of this product perspective, Pierce recommended focusing on the production process, and Campbell et al. recommended a focus on measuring the quality of that product. As for techniques to improve information quality, Redman emphasized the need for analysis. Redman and Pierce each emphasized the need to assign appropriate organizational responsibilities. Redman also suggested building business relationships as part of one's improvement strategy. Kerr and Norris (2004) proposed a significantly different emphasis from the others, namely one that focuses on governance, education and training, and fostering cultural change.

Writing from a research-oriented perspective, Pierce (2004) identified four strategic information quality principles: understand the customer's needs, manage the technical and organizational processes that produce the product, manage the entire life cycle of information products, and assign an information product manager. She suggested applying five manufacturing principles adapted from total quality management (TQM): articulate a vision in business terms; establish central responsibility; educate suppliers, manufacturers, and consumers; teach the skills needed to define, measure, analyze, and improve; and institutionalize continuous improvement.

From this foundation, Pierce (2004) proposed applying marketing techniques to the development of an information quality strategy, following a six-step process. First, develop a mission statement. Second, identify customers and determine which dimensions are important. As a third step, she recommends performing conjoint analysis to obtain a multiple regression equation with which to establish the relative importance of the dimensions. Following the conjoint analysis, one can identify clusters of customers with similar requirements by using cluster analysis as if performing a market analysis. With this information in hand, the fifth step is to prioritize the organization's objectives, and the sixth is to define the strategy as a set of elements. Once developed, the strategy can be implemented and monitored. The process should be documented, the quality measured, and the remaining gaps identified. As a final step, a combination of information quality tools, TQM principles, and statistical process control techniques can be applied as appropriate.

Information Quality as a Contributor to Organizational Strategy

Redman (1995) noted that "poor quality data can cause immediate economic harm and have more indirect, subtle effects" (p. 99). Examples of the indirect effects include the erosion of trust by customers and suppliers, impediments to the ability to effectively implement a business strategy, and interruptions in the flow of just-in-time manufacturing systems. Redman (1998) expanded on these thoughts by noting that "poor data quality hurts employee morale, breeds organizational mistrust, and makes it more difficult to align the enterprise" (p. 80).

Campbell et al. (2004) discussed data strategy from the practitioner's point of view, but grounded their work in scholarly literature and provided a direct linkage between strategy and organizational outcomes. Their paper focused on a concept called a data quality scorecard (DQS), proposed as part of a broader data quality strategy.

Conceptually, the DQS focal point is a data factory, which has identifiable sources of input, repositories, processes, and products. The scorecard is basically a table depicting various information products evaluated objectively and ranked according to a specific set of quality dimensions. The products evaluated with the DQS can be either sources into or outputs of the data factory.

Campbell et al. (2004) discussed the DQS in the context of a case study involving a company that sells data products commercially. The scorecard was used to evaluate sources of data purchased on an ongoing basis for inclusion in the company's products. The evaluation, conducted by a third party to minimize bias, considered the dimensions of accuracy, completeness, accessibility, and consistency. The results indicated that the source under consideration (referred to as Source X) did not contribute positively to the company's product, hence its use could be discontinued. Because this case involved decisions affecting both the costs and revenue of real products, the researchers were able to provide a direct linkage to organizational outcome by calculating the return on investment.

Interestingly, the company "soon discovered that cost was not the compelling factor for change" (Campbell et al., 2004, p. 159). Instead, they noted that the company's own product was "'middle of the road' in the data quality arena, with or without Source X" (p. 159). In addition, they were able to determine that the use of Source X provided no competitive advantage. Moreover, "the cost savings realized from the cancellation of Source X was so significant, it could be used to replace the external data source as well as improve efficiencies in data quality internally" (p. 161). By the conclusion of their research, the organization had shifted its perspective from one of merely trying to determine the level of its data quality to asking "how does the quality of our data providers impact our product?" (p. 162).

Kerr and Norris (2004) discussed information quality strategy within the context of a case study at the New Zealand Ministry of Health. The authors also identified a number of factors with which they related information quality to organizational outcomes, paying particular attention to those related to organizational performance and to those upon which continued success depended. With respect to organizational outcomes, they provided quantitative measurement by assessing the costs and benefits of the initiative, and by calculating an estimated return on investment. Qualitatively, they provided the ministry with an analysis of "risks of omission and commission" (p. 226), developed a governance model, and incorporated the information quality strategy into the organization's information systems strategic plan. Continued success of the information quality improvement activities was deemed to depend on implementing consistent processes, fostering a culture of continual assessment, maintaining a current set of priorities, and educating stakeholders.

The connection between the information quality strategies discussed above and their impact on organizational outcomes is scattered at best. Only two studies discussed a direct, quantifiable connection with organizational outcomes. Redman (1995) made note of cost reductions attained through improved quality, and Kerr and Norris (2004) discussed cost-benefit analysis. Both discussed applying cost aspects to the calculation of return on investment. Qualitative discussions of the impact on organizational outcome are more broadly discussed in these studies; however, with the singular exception of improved customer/supplier relationships (Campbell et al., 2004; Pierce, 2004; Redman), the connections drawn vary widely from study to study. Kerr and Norris mentioned performing an organizational risk assessment in their study, and discussed how their strategy was incorporated into the organization's information system strategic plan. Redman suggested that improved information quality affects the ability to implement business strategies. Campbell et al. had perhaps the strongest linkage, mentioning the ability to improve product quality, improve operational efficiency, and gain competitive advantage.

Summary of Information Quality Strategy Research Literature

Taken together, the strategies discussed above paint a rather broad picture, if a bit disjointed. The following paragraphs describe the topics covered, including the aspects of information quality considered, the approaches used in addressing the issue, and the connections drawn between information quality management and organizational outcomes.

Two broad aspects of information quality are considered in the context of these studies: information quality dimensions and stakeholder perspectives. Although all dimensions are covered, coverage is highly inconsistent among the studies. One focused exclusively on accuracy (Redman, 1995), one selected four specific dimensions (Campbell et al., 2004), and the other two left the identification of relevant dimensions up to the customers (Kerr & Norris, 2004; Pierce, 2004). Similarly, each stakeholder perspective is considered, but their consideration is highly inconsistent among the different studies. The most broadly covered perspective is that of the customer, receiving attention from each of the studies (Campbell et al.; Kerr & Norris; Pierce; Redman). The provider perspective was addressed in two studies (Kerr & Norris; Redman). Only one considered the perspective of the organization's management (Campbell et al.).

In terms of the approach to developing an information quality strategy, two angles are considered: the perspective from which the authors developed their approach, and the techniques they recommended. The perspective most commonly mentioned was quality management theory (Kerr & Norris, 2004; Pierce, 2004; Redman, 1995). Information quality literature, somewhat surprisingly, was mentioned in only two of the studies (Campbell et al., 2004; Pierce), as was practical experience (Kerr & Norris; Redman). Marketing theory played a significant role in Pierce's study.

A Research Framework for Information Quality Strategy

This final section of the chapter presents a framework for information quality strategy. The section begins by presenting an organizational context within which to frame information quality strategy research. The section then presents the conceptual framework used for this research, centered on the concept of a strategic relationship between information quality and organizational outcomes. Finally, the section presents the research model used for this research.

Organizational Context

Returning to the three lenses proposed by Melville et al. (2004), coupled with Barney's (1991) finding that linked sustained advantage and management skills, the relationship between information quality and organizational outcomes can now be seen as a set of management choices leading to competitive advantage. The question thus becomes one of framing those choices in an organizational context. Toward this end, Chung, Fisher, and Wang (2005) considered information quality research within the broad context of general systems theory, basing their work on that of Boulding, who defined nine levels of organization, ranging from static structures to transcendental systems, creating what he called a "hierarchy of complexity" (Boulding, 1956, p. 202).

Chung et al. (2005) simplified Boulding's hierarchy into three broad levels: mechanical systems, open systems, and human systems. They then related various aspects of information quality work, from practice and research, to these three levels. The concept of the information product, for example, fits at the mechanical level, as does work to improve the accuracy of that product. Work related to adaptability, such as understanding customer needs, is seen as fitting within the open systems level, and work related to interpretation of information (Lee, 2003-2004; Lee & Strong, 2003-2004) fits at the human level. This study provides valuable structure for organizing information quality literature and practices, and appears to fit the original intent of Boulding (1956), most notably addressing his observation that "it is all too easy for the interdisciplinary to degenerate into the undisciplined" (p. 200).

When combined with the lenses proposed by Melville et al. (2004), a two-dimensional matrix emerges in the form of a three-by-three grid, as shown in Figure 10. The vertical axis is taken directly from Chung et al.'s (2005) simplification of Boulding's (1956) hierarchy, with mechanical systems on the bottom row, open systems in the middle, and human systems along the top. The horizontal axis reflects the three lenses of Melville et al.'s (2004) integrative framework, taking into account the general theories of business strategy (Barney, 1991, 2001; Porter, 1991, 1996; Wade & Hulland, 2004). From left to right, the columns represent the firm, the competitive environment, and the macro environment. Each of the nine cells thus provides a context within which to analyze strategic relationships between particular aspects of information quality and organizational outcomes. The aspects can include dimensions, improvement techniques, or other considerations such as critical success factors (Xu & Al-Hakim, 2005). Each aspect impacts one or more organizational outcome, described in terms of specific strategic benefit, such as cost reduction, product differentiation, or the ability for others to imitate.

Conceptual Framework

The concepts discussed above form the basis of a conceptual framework. This section presents such a framework, which was used in this study for evaluating the relationship between information quality and organizational outcomes in support of an information quality strategy.

The central element of this framework is the strategic relationship as shown generically in Figure 11. An example of such a strategic relationship is shown in Figure 12, illustrating how improvements to the accuracy dimension can be applied in support of an information quality strategy. Such a relationship would logically fit within the lower left cell of the matrix in Figure 10, representing the mechanical level and the focal firm's resources. As determined by Redman (1995, 1998), efforts to improve accuracy can reduce cost, hence improving the organization's competitive position.

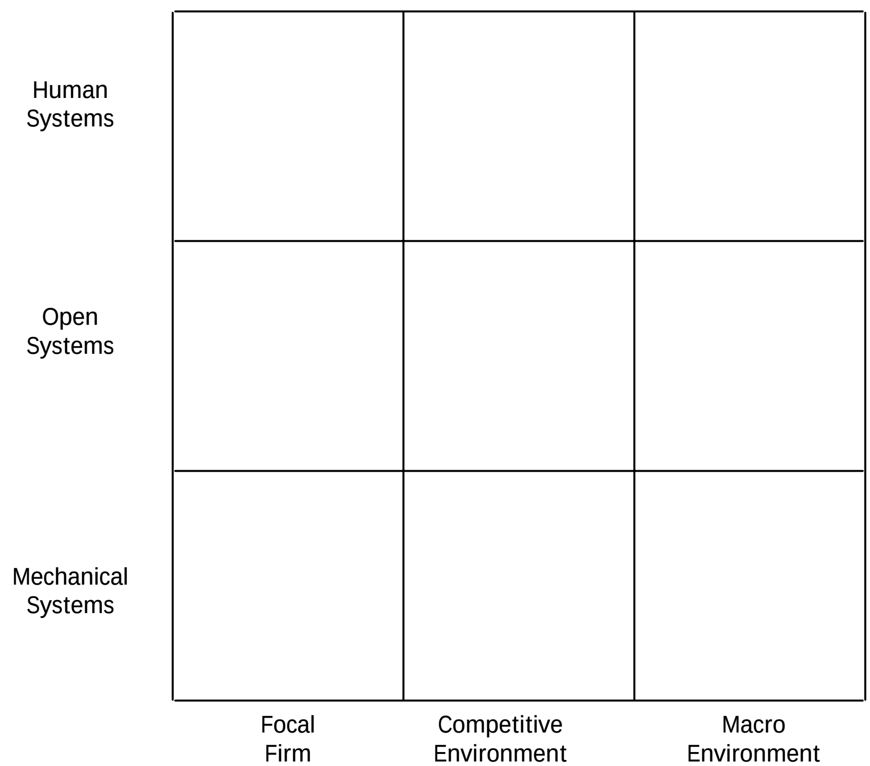

Figure 10. A contextual framework for information quality strategy research

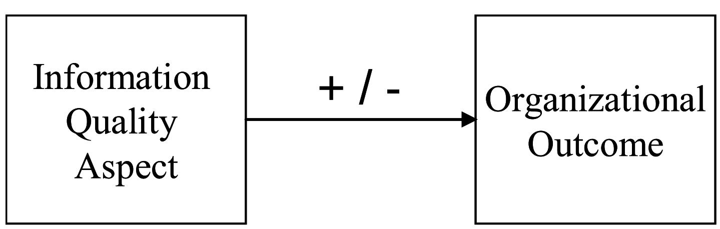

Figure 11. Strategic relationship between information quality aspect and organizational outcome

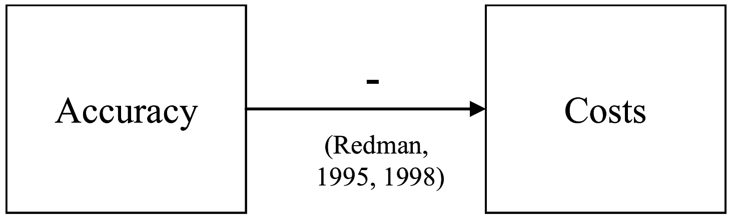

Figure 12. Information accuracy and cost reduction

Similarly, Campbell et al. (2004) found that accuracy can positively affect the firm's ability to differentiate its product. In this case, the relationship logically fits within the lower middle cell of the matrix, representing mechanical systems in the competitive environment. This strategic relationship is shown in Figure 13.

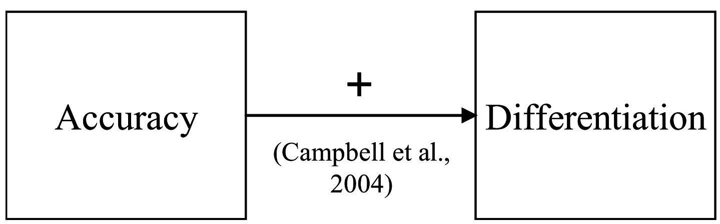

Figure 13. Information accuracy and product differentiation

Operationalizing the Variables

To operationalize the variables for this research, it was necessary to precisely define and measure the information quality aspects and organizational outcomes, to define the operating assumptions, and to frame the research within a broader context. The following paragraphs describe how this was done.

Operationalizing Information Quality Aspects

As discussed earlier in the chapter, and as summarized in Table 1, researchers have established that information quality can be measured in a variety of ways, including both subjectively and objectively. Of the measurement techniques available, the most comprehensive measurement is the subjective measurement provided by the Information Quality Assessment (IQA) instrument, which uses 69 survey items to measure 16 dimensions. The PSP/IQ Model then balances this comprehensiveness with parsimony, by reducing the 16 dimensions to four quadrants (Kahn et al., 2002; Lee et al., 2002). Operationalizing the information quality aspects, therefore, was accomplished through a straightforward adaptation of the IQA instrument and the PSP/IQ Model. Each quadrant was thus used to represent an information quality aspect in the strategic relationship shown in Figure 11.

Operationalizing Organizational Outcomes

The instrument developed by Mirani and Lederer (1998) measures a set of organizational outcomes in each of the categories illustrated in Figure 9 using two to four survey items per category. A closer look at the items that measure informational benefits reveals that each items maps to one or more of the dimensions identified by Wang and Strong's (1996) factor analysis. Table 3 summarizes the results of this examination: the column on the left shows the informational benefit categories and related survey items from Mirani and Lederer, the middle column shows the Wang and Strong dimension and corresponding information quality category, and the column on the right shows the quadrant names from Lee et al. (2002). Most of the matches are self-evident, whereas others were matched to the data quality attributes that loaded on a particular factor in Wang and Strong's study. It should be noted that the final row identifies a dimension that was not categorized (i.e., ease of operation). This dimension was identified in Wang and Strong's phase 1 factor analysis, but was dropped in their phase 2 sorting process due to inconsistent results from the participants. However, this dimension is included in the PSP/IQ model's usability quadrant (Lee et al.). Moreover, it should be noted that each of the items maps to a quadrant, and each of the four quadrants of the PSP/IQ is represented in this list.

Table 3.
Comparison of Mirani and Lederer's (1998) Informational Benefits, Wang and Strong's (1996) Dimensions, and Lee et al's (2002) PSP/IQ quadrants

Mirani and Lederer (1998)	Wang and Strong (1996)	Lee et al. (2002)
Info. Access: Faster retrieval/delivery	Timeliness (Contextual)	Dependability
Info. Access: Easier access	Accessibility (Accessibility)	Usability
Info. Quality: Improve management information for strategic planning	Value-added (Contextual)	Usability
Info. Quality: Improve accuracy/reliability	Accuracy (Intrinsic)	Soundness
Info. Quality: Improve information for operational control	Relevancy (Contextual)	Usefulness
Info. Flexibility: More concise/better format	Concise representation (Representational)	Soundness
Info. Flexibility: Flexibility of requests	Ease of operation (Not categorized)	Usability

Given that each of Mirani and Lederer's (1998) items for informational benefits maps to a dimension associated with a PSP/IQ quadrant, it follows that each one is measured by the IQA. For this reason, those particular items were able to be dropped from the Mirani and Lederer instrument for the purposes of this research without any loss of information. The remaining items, addressing the other two organizational benefit categories, are all relevant, and were thus retained. Each of the two organizational benefit categories, strategic benefits and transactional benefits, was used to represent an organizational outcome as shown in Figure 11.

Operational Assumptions and Framing the Research

The key operating assumption for this research model was that affirmative steps have been taken to positively affect the information quality aspect in question. The choice of steps is outside the scope of this research. As such, they may have been based on one or more of the information quality management approaches discussed earlier in the chapter (Ballou et al., 1998; Davidson et al., 2004; Shankaranarayanan et al., 2000; Wang, 1998; Wang et al., 1998), or they may have been based on some other means, such as the deployment of a new or updated information system (DeLone & McLean, 1992, 2003; Elmorshidy, 2005).

With respect to the contextual framework presented in Figure 10, this research was positioned within the first two columns. The organizational outcomes considered address both the focal firm (i.e., transactional benefits) and the competitive environment (i.e., competitive benefits). The information quality dimensions addressed by this research make sense within each of the three rows of the matrix, with soundness being primarily focused at the mechanical level, and the remaining aspects being distributed across the open systems and human levels.

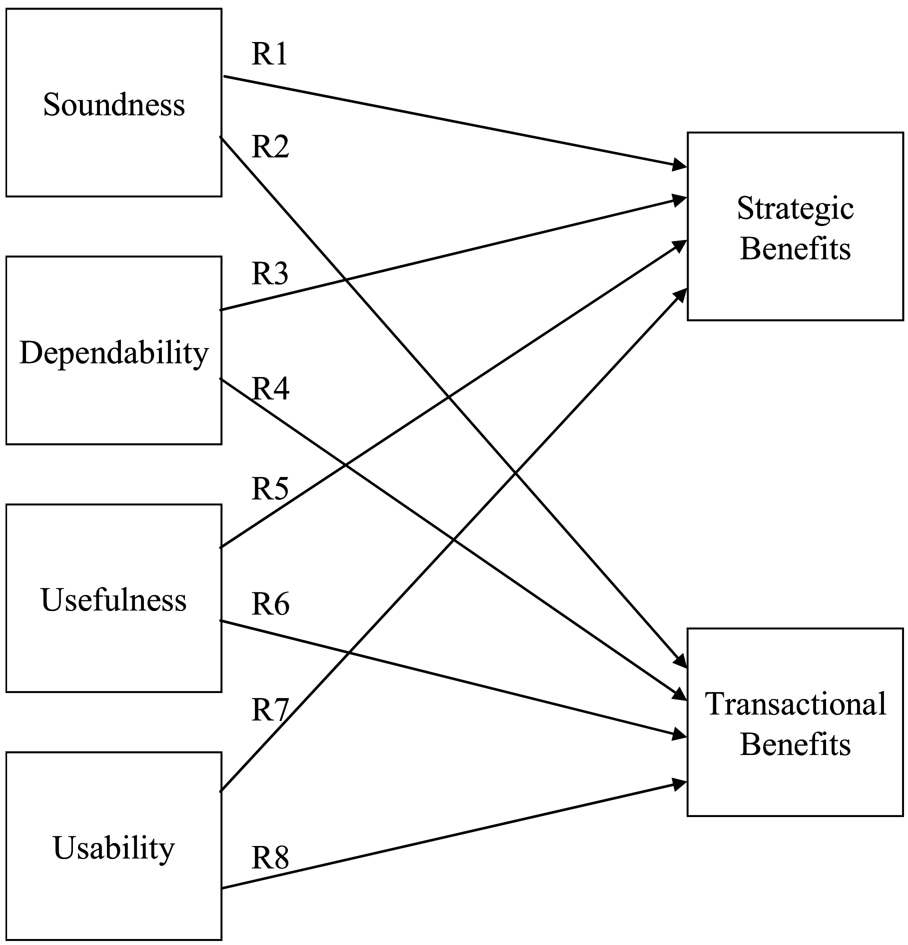

Figure 14. The research model

The Research Model

The conceptual framework and variables defined above were thus combined to form the research model as shown in Figure 14. The four information quality quadrants from the PSP/IQ model (Lee et al., 2002) are shown on the left, and the two categories of organizational benefits defined by Weill (1992) and Mirani and Lederer (1998) are shown on the right. Taken together, eight strategic relationships (R1 through R8) result, and were the focus on this research.

The relationships above describe the main effects. In addition, the moderating effect of information intensity was considered. Porter and Millar (1985) suggested that the strategic benefit organizations could derive from information systems was related to information content of the organization's products and processes and to the information intensity of the value chain. Other researchers have subsequently expanded the concept to include the information content of services (Teo & King, 1997) and have found empirical evidence confirming the role of information intensity in creating competitive advantage (Kearns & Lederer, 2003) and as a moderator of the relationship between information systems infrastructure and organizational performance (Dejnaronk, 2000). This relationship is depicted in Figure 15.

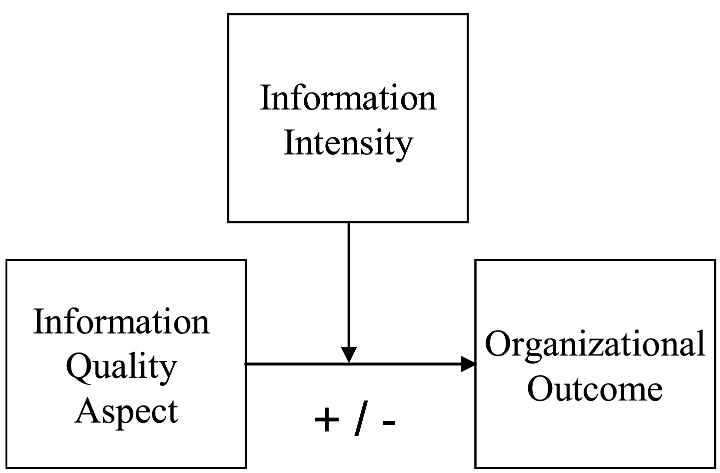

Figure 15. Information intensity as a moderator.

Conclusions

This chapter has presented research literature relevant to this research from both an information quality perspective and a strategy perspective. Each of these perspectives was examined in terms of its theoretical grounding and current lines of research. A gap in the literature is evident in that the linkage between these two perspectives has thus far only minimally been examined, with relatively little theoretical grounding.

To address this gap, a contextual framework for information quality strategy research was set forth, a conceptual model was developed for researching the relationship between information quality aspects and organizational outcomes, and a research model was defined with operationalized variables. Eight relationships were identified within this model as the focus of this research.

CHAPTER 3. METHODOLOGY

Introduction

This chapter presents the methodology used to conduct the research for this study. The first section presents the research philosophy underlying the study and places that philosophy within the broader context of research in general. The next section describes the theoretical framework within which the research was conducted and presents the hypotheses that were tested. The next two sections describe the research design and sampling design, followed by detailed discussions of the measures used in the study, the data collection procedures, the pilot study, and the data analysis procedures. The chapter concludes by describing the limitations of the study.

Research Philosophy

Arbnor and Bjerke (1997) described a number of closely related factors that shape the methods used by researchers to create knowledge. Among these factors are the researcher's background assumptions, beliefs, and paradigms. Making reference to the works of Kuhn and Törnebahm, they described a paradigm as including such notions as one's conceptions of reality, science, and scientific ideals, as well as one's sense of ethics and aesthetics.

Kuhn (1996) and Arbnor and Bjerke (1997) have provided two significantly different views of paradigm formation in the social sciences. Kuhn first discussed how various natural science paradigms emerged, beginning with mathematics and astronomy, continuing through more recently developed scientific fields such as motion, heat, historical geology, and biology, and then stated, “it remains an open question what parts of social science have yet acquired such paradigms at all” (p. 15). Arbnor and Bjerke, on the other hand, stated, “we (and many others) have found Kuhn’s type of analysis rewarding at the same time that we note an important difference between the natural and social sciences. In the natural sciences, old paradigms are replaced by new ones; in the social sciences, old paradigms usually survive alongside new ones” (p. 13).

The information quality literature exemplifies the notion that multiple paradigms can and do survive alongside one another. Research on the topic spans multiple paradigms and includes the use of a wide assortment of methodological approaches, depending on the research question under consideration. As such, within the information quality field, the ability for an individual researcher to flexibly select from among multiple paradigms and multiple methodologies is seen as advantageous. This ability fits well within what Greene, Kreider, and Mayer (2005) referred to as the “pragmatic stance” (p. 275), which they described as “an inclusive philosophical framework within which multiple assumptions and diverse methods can comfortably reside” (p. 275).

With the pragmatic stance as a backdrop, this research was conducted from the perspective of the post-positivist paradigm. This paradigm is firmly rooted in the positivist paradigm, which employed empirical means and deductive logic in the quest for an objectively knowable truth. Positivism was rooted in Descartes’ view that it was possible to realize a “dualism between mind and matter” (Crook & Garratt, 2005, p. 208), and that researchers should observe facts from a perspective free of theory and free of values, such that the knowledge attained would be “immunized and protected from the unwarranted intrusion of subjective ideas” (p. 208). While still relying on empirical means and deductive logic to develop and test hypotheses, post-positivism does so from a more enlightened and nuanced stance in which the attainment of an objectively knowable truth is no longer seen as a guiding principle. Instead, post-positivism seeks to find logically defensible affirmations in support of hypotheses that can be generalized (Crook & Garratt). This research, therefore, was undertaken with the objective of finding logically defensible affirmations in support of the set of hypotheses defined later in this chapter, with the full recognition that such findings represent only partial knowledge of the subject matter as viewed from a single, limited perspective.

Theoretical Framework

Evidence in the literature establishing the relationship between the management of information quality and organizational outcomes has to this point been limited and sparse, with much of that evidence being anecdotal. A research model was proposed for investigating this relationship. Hypotheses based on this model are discussed in the following sections.

Strategic Benefits of Information Quality

Strategic benefits include competitive advantage, alignment between the business and information systems, and customer relations improvement (Mirani & Lederer, 1998). It was hypothesized that improvement in various aspects of information quality would positively affect these strategic outcomes. Accurate, relevant, and timely information can help an organization respond to changes in its competitive environment. Information that is relevant, timely, and accessible across organizational units can assist in aligning the organization's information systems with its business objectives. Customer data that is free of errors can help an organization improve its customer service. Therefore, the five hypotheses stated below address the relationships between information quality and strategic benefits. The first four hypotheses address the relationships with individual quadrants in the PSP/IQ model, and the fifth addresses information quality as a whole.

H1: Improvements in the soundness of information will be associated with increased strategic benefits.

H2: Improvements in the dependability of information will be associated with increased strategic benefits.

H3: Improvements in the usefulness of information will be associated with increased strategic benefits.

H4: Improvements in the usability of information will be associated with increased strategic benefits.

H5: Improvements in information quality will be associated with increased strategic benefits.

Transactional Benefits

Transactional benefits include communications efficiency, systems development efficiency, and business efficiency (Mirani & Lederer, 1998). It was hypothesized that various aspects of information quality improvement would positively affect these transactional outcomes. Accurate, timely, and believable information can improve an organization's communications efficiency. Information that is consistently and concisely represented can improve the efficiency with which the organization develops and deploys new systems. Information that is presented in the right amount and in a fashion that is easy to use and understand can improve the business efficiency of the organization. Therefore, the five hypotheses stated below address the relationships between information quality and transactional benefits. The first four hypotheses address the relationships with individual quadrants in the PSP/IQ model, and the fifth addresses information quality as a whole.

H6: Improvements in the soundness of information will be associated with increased transactional benefits.

H7: Improvements in the dependability of information will be associated with increased transactional benefits.

H8: Improvements in the usefulness of information will be associated with increased transactional benefits.

H9: Improvements in the usability of information will be associated with increased transactional benefits.

H10: Improvements in information quality will be associated with increased transactional benefits.

Information Intensity as a Moderator

The hypotheses above describe the main effect relationships between information quality and organizational outcomes. It was also hypothesized that information intensity would have a moderating effect on the relationship between information quality and organizational outcomes. Specifically, it was hypothesized that higher degrees of information intensity would increase the strength of that relationship. Information intensity is a measure of the information content of an organization's products and services and of the degree of dependence upon information in the organization's value chain. As such, an organization with a high degree of information intensity is more likely to experience the benefits of improved information quality than an organization with a low degree of information intensity. Therefore, the following ten hypotheses address various aspects of this moderating effect.

H11. The effect of improvements in the soundness of information on strategic benefits will be stronger in organizations that have a higher degree of information intensity than in organizations that have a lower degree of information intensity.

H12. The effect of improvements in the dependability of information on strategic benefits will be stronger in organizations that have a higher degree of information intensity than in organizations that have a lower degree of information intensity.

H13. The effect of improvements in the usefulness of information on strategic benefits will be stronger in organizations that have a higher degree of information intensity than in organizations that have a lower degree of information intensity.

H14. The effect of improvements in the usability of information on strategic benefits will be stronger in organizations that have a higher degree of information intensity than in organizations that have a lower degree of information intensity.

H15. The effect of improvements in information quality on strategic benefits will be stronger in organizations that have a higher degree of information intensity than in organizations that have a lower degree of information intensity.

H16. The effect of improvements in the soundness of information on transactional benefits will be stronger in organizations that have a higher degree of information intensity than in organizations that have a lower degree of information intensity.

H17. The effect of improvements in the dependability of information on transactional benefits will be stronger in organizations that have a higher degree of information intensity than in organizations that have a lower degree of information intensity.

H18. The effect of improvements in the usefulness of information on transactional benefits will be stronger in organizations that have a higher degree of information intensity than in organizations that have a lower degree of information intensity.

H19. The effect of improvements in the usability of information on transactional benefits will be stronger in organizations that have a higher degree of information intensity than in organizations that have a lower degree of information intensity.

H20. The effect of improvements in information quality on transactional benefits will be stronger in organizations that have a higher degree of information intensity than in organizations that have a lower degree of information intensity.

Research Design

This study used an electronically administered survey to obtain data measuring individual stakeholder perceptions of information quality, organizational outcomes, and information intensity of their organizations. Statistical analysis was conducted on the data to test the null alternatives to the hypotheses presented above.

The survey items were based on existing items from validated instruments found in the research literature. Many of the survey items had been widely validated in a variety of populations and organizational settings, while others had been validated in more limited contexts. In addition, these survey items had not previously been used together within a single instrument. For these reasons, statistical tests were conducted to validate the instrument in the context of this study's population and to test the reliability and validity of the combined instrument. The survey was administered to a probabilistic sample selected from a population representing information systems stakeholders in multiple organizations. Survey notifications and follow-up reminders were sent by electronic mail and responses were collected via a Web server. Response/non-response statistics were collected to determine the response rate and to test for evidence of a response/non-response bias.

Data collected from the survey were examined, described, and cleansed as discussed below, and were analyzed using SPSS Graduate Pack 13.0 for Windows. A series of multiple regression analyses were conducted to test the main effect hypotheses, each with multiple independent variables and a single dependent variable. A series of moderated regression analyses and subgroup analyses were conducted to test the moderating effect hypotheses, each with a single independent variable, a single dependent variable, and a single moderator variable.

Sampling Design

The target population for this study was individuals who work in for-profit, non-profit, governmental, or academic organizations and who regularly use at least one computer-based information system in the normal course of their work. Such use was intended to cover the three primary roles of information provider/collector, information systems professional, and information consumer. Participants meeting these criteria were eligible for inclusion; all others were excluded.

Given the size of the population, selection of a representative sample is the preferred approach for efficiently gathering data about the population (Cooper & Schindler, 2003; Lewin, 2005). This study was designed to use a systematic sampling strategy. A vendor- and technology-neutral industry consortium agreed to allow its database of contacts to be used as the sampling frame for this study. Participants were recruited from within this contacts database in two different ways. First, a set of potential participants were selected to receive an electronic mail invitation. Second, attendees at the consortium's summer 2006 conference were selected to receive an invitation at the conference.

For the electronic mail set, the plan originally called for the starting position in this database to be selected at random and then every *n*th record to be selected until the desired number of records had been selected. However, after screening the database based on an unexpected field that indicated whether the contact information had been verified (i.e., the field was able to be used to separate the records known to be reliable from those with an unknown reliability), it was determined that he total number of known reliable records was only fractionally larger than the set intended for selection. Given this new development, the original plan would only be acceptable for $n < 2$ or for n to be non-integer. Further, given that values of n greater than or equal to 2 would have resulted in an unacceptably small selection, and given the added complexity of using a non-integer value to select the records, it was determined that the best path forward was to select all of the records that passed the reliability screen, in effect following the original plan with $n = 1$. The total records selected in this fashion was 3,210.

In addition to the electronic mail invitations discussed above, each of the 228 attendees at the consortium's summer 2006 quarterly conference and member meeting was invited through a combination of verbal announcements and a flyer inserted in their registration packets. This brought the total number of invitees to 3,438.

The size of the sample used for data analysis is critical to the success of any research study. The minimum sample size rule of thumb recommended by Mertler and Vannatta (2005) is fifteen per independent variable in a multiple regression equation. The highest number of independent variables in any one equation for this research was five, therefore, a minimum sample size of 75 was required to adequately perform the planned analysis.

Robson (2002) recommends that up to three follow-up reminders are useful at increasing the response rate, hence the plan called for reminders to be sent if needed after two weeks and after three weeks. One reminder was sent after two weeks as planned, but the second reminder was not needed to reach the desired number of responses and was not sent.

Measures

This section identifies the different variables measured in this study and describes how those variables were measured. The section is divided into three major sub-sections: operationalizing the variables, design of the data collection instrument, and validation of the data collection instrument.

Operationalizing the Variables

Three types of variables were operationalized for this study: independent variables measuring various aspects of information quality, dependent variables measuring organizational outcomes, and moderating variables measuring information intensity. A set of demographic variables was also collected to facilitate grouping of responses.

Independent Variables

The independent variables for this study were those used to measure information quality. These variables were operationalized at two levels: the dimension level and the PSP/IQ quadrant level. The dimension level was measured directly by using the 69 survey items from the Information Quality Assessment (IQA) instrument (Lee et al., 2002; Najjar, 2002) as listed in Tables 4a through 4d. This instrument utilizes a scale from 0 to 10, where 0 represents *not at all* and 10 represents *completely*, and the midpoint is identified with the label *average*. One independent variable per information quality dimension was calculated as the mean value of the response items measuring that particular dimension. The PSP/IQ quadrant level variables were each calculated as the

Table 4a.
Information Quality Measurement Items

Dimension	Item – items labeled with "(R)" are reverse coded
Accessibility	This information is easily retrievable.
	This information is easily accessible.
	This information is easily obtainable.
	This information is quickly accessible when needed.
Appropriate Amount	This information is of sufficient volume for our needs.
	The amount of information does not match our needs. (R)
	The amount of information is not sufficient for our needs. (R)
	The amount of information is neither too much nor too little.
Believability	This information is believable.
	This information is of doubtful credibility. (R)
	This information is trustworthy.
	This information is credible.
Completeness	This information includes all necessary values.
	This information is incomplete. (R)
	This information is complete.
	This information is sufficiently complete for our needs.
	This information covers the needs of our tasks.
	This information has sufficient breadth and depth for our task.

Table 4b.
Information Quality Measurement Items

Concise representation	This information is formatted compactly.
	This information is presented concisely.
	This information is presented in a compact form.
	The representation of this information is compact and concise.
Consistent representation	This information is consistently presented in the same format.
	This information is not presented consistently. (R)
	This information is presented consistently.
	This information is represented in a consistent format.
Ease of Operation	This information is easy to manipulate to meet our needs.
	This information is easy to aggregate.
	This information is difficult to manipulate to meet our needs. (R)
	This information is difficult to aggregate. (R)
	This information is easy to combine with other information.
Free of Error	This information is correct.
	This information is incorrect. (R)
	This information is accurate.
	This information is reliable.

Table 4c.
Information Quality Measurement Items

Interpretability	It is easy to interpret what this information means.
	This information is difficult to interpret. (R)
	It is difficult to interpret the coded information. (R)
	This information is easily interpretable.
	The measurement units for this information are clear.
Objectivity	This information was objectively collected.
	This information is based on facts.
	This information is objective.
	This information presents an impartial view.
Relevancy	This information is useful to our work.
	This information is relevant to our work.
	This information is appropriate for our work.
	This information is applicable to our work.
Reputation	This information has a poor reputation for quality. (R)
	This information has a good representation.
	This information has a reputation for quality.
	This information comes from good sources.
Security	This information is protected against unauthorized access.
	This information is not protected with adequate security. (R)
	Access to this information is sufficiently restricted.
	This information can only be accessed by people who should see it.

Table 4d.
Information Quality Measurement Items

Timeliness	This information is sufficiently current for our work.
	This information is not sufficiently current for our work. (R)
	This information is sufficiently timely.
	This information is not sufficiently timely. (R)
	This information is sufficiently up-to-date for our work.
Understandability	This information is easy to understand.
	The meaning of this information is difficult to understand. (R)
	This information is easy to comprehend.
	The meaning of this information is easy to understand.
Value-added	This information provides a major benefit to our work.
	This information does not add value to our work. (R)
	Using this information increases the value of our work.
	This information adds value to our tasks.

mean value of the dimension values corresponding to that particular quadrant (Kahn et al., 2002; Lee et al.).

Dependent Variables

The dependent variables for this study were those used to measure organizational outcomes. These variables were operationalized at two levels: the dimension level and the category level. The dimension level was measured directly by using the 18 relevant survey items from the Organizational Benefits of IS Projects instrument developed by

Mirani and Lederer (1998) as listed in Tables 5a and 5b; seven items from that instrument were discarded because they are used to measure the dimensions associated with a category not included in this study. This instrument utilizes a scale from 1 to 7, where 1 represents *not a benefit* and 7 represents *very important.* The midpoint is not identified for this scale. One dependent variable per organizational outcome dimension was calculated as the mean value of the response items measuring that particular dimension. The category variables were then calculated as the mean value of the dimension values corresponding to that particular category.

Intervening Variables

Information intensity was hypothesized as a moderator variable for this study. This variable was operationalized at two levels: the aspect level and the aggregate level. The aspect level was measured directly by using survey items from two separate studies. The items from the two studies were segregated such that each set measured a separate aspect of information intensity, resulting in the item list shown in Table 6. With one exception, all the items used by Dejnaronk (2000) to measure information intensity of the product or service were used to measure the same aspect in this study. All the items used by Kearns and Lederer (2003) to measure information intensity of the value chain were used to measure that aspect in this study. One item was dropped from Dejnaronk's list due to nearly identical wording with an included item from Kearns and Lederer's; based on the wording of that item, it was deemed more closely aligned with the latter and was thus retained in that set. This instrument utilizes a scale from 1 to 7, where 1 represents

Table 5a.
Organizational Benefits Measurement Items

Category/Dimension	Item – none are reverse coded
	"Use of this information will . . ."
Strategic Benefits/ Competitive Advantage	Enhance competitiveness or create strategic advantage.
	Enable the organization to catch up with competitors.
Strategic Benefits/ Alignment	Align well with stated organizational goals.
	Help establish useful linkages with other organizations.
	Enable the organization to respond more quickly to change.
Strategic Benefits/ Customer Relations	Improve customer relations.
	Provide new products or services to customers.
	Provide better products or services to customers.
Transactional Benefits / Communications Efficiency	Save money by reducing travel costs.
	Save money by reducing communication costs.
Transactional Benefits / Systems Development Efficiency	Save money by reducing system modification or enhancement costs.
	Allow other applications to be developed faster.
	Allow previously infeasible applications to be developed faster.
	Provide the ability to perform maintenance faster.

Table 5b.
Organizational Benefits Measurement Items

Transactional Benefits / Business Efficiency	Save money by avoiding the need to increase the work force.
	Speed up transactions or shorten product cycles.
	Increase return on financial assets.
	Enhance employee productivity or business efficiency.

strongly disagree and 7 represents *strongly agree.* The midpoint is not identified for this scale.

A variable per information intensity aspect was calculated as the mean value of the response items measuring that particular aspect. A single aggregate variable, used as the moderator variable, was then calculated as the mean value of the two aspect variables. Although operationalized as a single moderating variable, this two-stage approach was used to give equal weight to each aspect of information intensity, given the asymmetry in the number of survey items and the absence of a theoretical basis for assigning more weight to one or the other.

Instrument Design

It is recommended that the answers to four types of questions be collected when utilizing surveys to conduct research: administrative questions, filtering questions, target questions, and classification questions (Cooper & Schindler, 2003). The following sections describe the use of these types of question in this study.

Table 6.
Information Intensity Measurement Items

Aspect	Item – none are reverse coded
Product/service	We have many product/service varieties within a line of product/service.
	Our product/service is complex (i.e., containing many parts that must work together).
	Cycle time from the initial order to the delivery of our product/service is long.
	Our product/service mainly provides information.
	Our product/service operation involves substantial information processing.
	Our product/service requires extensive user training.
	Customers need a lot of information related to our product/service before purchasing the product/service.
Value chain	Information is used to a great extent in our production or service operations.
	Information used in our production or service operations is frequently updated.
	Information used in our production or service operations is usually accurate.
	Many steps in our production or service operations require the frequent use of information.

Administrative Questions

Administrative questions, which provide basic identifying information regarding the participants, "are rarely asked of the participant but are necessary to study patterns within the data and identify possible error sources" (Cooper & Schindler, 2003, p. 361). For this study, the value of assuring confidentiality of respondents outweighed the value of tracing responses to a specific respondent, thus administrative questions were not asked of the participants. However, a minimal set of variables related to either the list source or method of invitation was collected to permit evaluation of any response/non-response bias and to minimize the chance of duplicate responses from a single participant.

Filtering Questions

Filtering questions may be used to screen respondents with respect to their qualifications for participating in a study (Cooper & Schindler, 2003). At a minimum, participants in this study were required to work for an organization and to regularly interact with an information product or service (e.g., application, database, or report) in the course of their role with the organization. Such interaction could be as a provider/collector, operator/custodian, or consumer of the information product or service.

To the extent practical, potential participants were screened prior to being solicited for participation. To ensure the qualification of those who were solicited, the survey included questions to assess such qualification, terminating participation for those clearly unqualified.

Target Questions

Target questions are those which "address the investigative questions of a specific study" (Cooper & Schindler, 2003, p. 362). The specific questions identified for this study are those included above in Table 4 through Table 6. Cooper and Schindler suggest that target questions be arranged logically with more general questions asked early in the survey and specific ones asked later, and that they be grouped logically with clear transitions between groups. Finally, it is suggested that the sequence of questions be carefully considered, taking into account such factors as motivating continued participation and minimizing participant bias (Robson, 2002).

The first group of target questions were those that measure information intensity. This is the most general of the three sets, and it was used to motivate the participants to think about how information is used in their organizations, their organizations' products and services, and throughout the value chains associated with those products and services. Placing this set of questions first allowed information intensity to be measured without first causing the participants to reflect upon the quality of that information. Items from this set were arranged randomly to reduce the apparent relationships between the items.

The second set of target questions were those that measure organizational benefits. Participants were instructed to consider an information system that they regularly use at their organization. The notion of *information system* was clarified by giving a list of examples of information products and services. The notion of *use* was defined as involving any of the three major stakeholder roles: information provider/collector, information consumer, or one who develops, deploys, operates, or maintains the information system. These notions were clarified by asking respondents to select from a list of statements best describing the nature of the information system and nature of their role in the context of that system.

Having introduced the set of questions in this fashion, participants were asked to reflect upon the benefits that use of the selected system provides to their organization. Placing the organizational benefits questions second provided a narrowing of perspective following the initial set of questions, while again avoiding the potential of unduly influencing responses with notions of information quality. Items from the set were modified to reflect a present continuing tense, rather than future tense as stated in the original, and were randomly arranged to reduce the apparent relationships between the items.

The third and final set of target questions continued the focus on a specific information system selected by the participant. For these questions, the participants were asked to reflect upon various aspects of that system's information quality. The questions were presented in the same order as suggested by Najjar (2002).

Classification Questions

Classification questions are those which allow responses to be grouped for analysis according to demographic criteria or other categories (Cooper & Schindler, 2003). Classification questions were included in this survey to collect the participant's job title/function, plus information about the participant's organization, including industry categorization and size. The list of these items and their possible answers is shown in Table 7.

Validating the Instrument

Any instrument used to measure a phenomenon must be assessed with respect to its content (or face) validity and its construct validity. Doing so provides assurance that

Table 7.
Demographic and Categorical Information

Item	Possible answers
Type of organization	for-profit corporation, non-profit corporation, government agency, other
Industry identification	Manufacturing, engineering, transportation, health care, financial services, other
Primary business activity at participant's location	Banking, insurance, research and development, manufacturing, transportation, hospitality, retail, other
Number of employees at participant's location	under 100, 101 to 1000, 1001 to 10000, 10,000+
Number of employees in the organization (corporation or governmental agency)	under 100, 101 to 1000, 1001 to 10000, 10,000+
Annual sales for corporation or annual budget for governmental agency (in millions)	< 1, >= 1 and < 10, >= 10 and < 100, >= 100 and < 1000, >= 1000
Participant's years with this organization	under 1, 1 to 5, 6 to 10, 10 to 20, over 20
Participant's years in this industry	under 1, 1 to 5, 6 to 10, 10 to 20, over 20
Participant's job title or function	Executive, Management, Consultant, Engineer, Researcher, IT Professional, Administration, Other
Participant's highest level of education	High school or equivalent, Technical school certification, Associate's degree, Bachelor's degree, Master's or Specialist's degree, Doctoral degree or beyond
Participant's gender	Male or female
(Conference invitees only) Part of the world in which the participant resides	Africa, Asia (other than Middle East), Australia/Pacific, Europe, Middle East, North America, South America

the instrument actually measures what it is intended to measure and that it does not inadvertently measure anything else (Churchill, 1979). Content validity, which describes the extent to which the instrument covers the appropriate concepts, is measured subjectively (Saraph, Benson, & Schroeder, 1989), whereas construct validity, which describes both the convergent and discriminant aspects of sets of items included in the instrument, is measured using statistical means (Churchill; Fiske & Campbell, 1992; Moore & Benbaset, 1991; Saraph et al.).

Convergent validity of an instrument is typically assessed by calculating the Cronbach alpha value of a set of items (Cooper & Schindler, 2003; Moore & Benbaset, 1991; Nunnally, 1978; Saraph et al., 1989). Alpha values range from zero to one, with higher numbers representing greater degrees of convergence among the items. Acceptable alpha values vary according to the purpose of the study. In the early stages of research, values as low as .50 or .60 are acceptable (Moore & Benbaset), although a minimum of .70 is more commonly considered the acceptable threshold (Dejnaronk, 2000; Moore & Benbaset; Nunnally; Saraph et al.). Citing Nunnally, Moore and Benbaset suggested that alpha values above .80 are “often wasteful” (p. 205).

Assessing discriminant validity is less straightforward than assessing convergent validity, and there are differences of opinion with respect to what constitutes an appropriate method (Fiske & Campbell, 1992; Shemwell & Yavas, 1999). The debate centers around the issue of whether multidimensional constructs are appropriate for measuring phenomena. Shemwell and Yavas argue in favor of including such constructs, noting that they are a reality in many domains. They argue that such a construct is useful and meaningful when the constructs are distinct at one level, yet share common variance at another level. They describe such a construct as having a "weak form of discriminant validity" (p. 68). Regardless whether multidimensional constructs are deemed permissible, factor analysis, either exploratory or confirmatory, is a common method for assessing discriminant validity. An alternative approach is to examine a correlation matrix of all the major variables. When using this approach, "discriminant validity is present when items that belong to a construct display lower correlation with other constructs" (Dejnaronk, 2000, p. 104).

The instrument used for this study was drawn directly from prior studies for which the validity had already been determined in varying degrees. To assess the validity of this instrument, it was deemed appropriate to consider the steps that had been completed with respect to the separate portions of the instrument, and to determine what additional steps were needed to assure the validity of the instrument as a whole (Robson, 2002).

When designing new instruments, Churchill (1979) recommended a seven-step development approach. The first step is to specify the domain of the construct based on a literature search. Second, the researcher should generate a sample of items, drawing on knowledgeable individuals' opinions and experiences. Third, the researcher should collect a set of data using those items, and fourth, should use the data collected to purify the measure using an iterative process of conducting factor analysis to group items, calculating the coefficient alpha, and removing items that contribute relatively little to the alpha value. The fifth step is to collect additional data using the modified measurement, followed by an assessment of the reliability and then an assessment of the validity.

As discussed above, the three main portions of this instrument assessed information intensity, organizational benefits, and information quality. Steps 1 and 2 had been conducted previously on all three portions (Dejnaronk, 2000; Kearns & Lederer, 2003; Lee et al., 2002; Mirani & Lederer, 1998). Given that the information intensity portion was drawn from two separate studies, steps 3 and 4 had not been conducted on the combination of items. Steps 3 through 7 had been conducted previously on the organizational benefits portion (Mirani & Lederer) and the information quality portion (Lee et al.), both in the original studies and in studies other than the ones in which the instruments were developed (Bendoly & Kaefer, 2004; Kaefer & Bendoly, 2004; Kahn et al., 2002; Najjar, 2002; Pipino et al., 2002; Pipino et al., 2005).

Based on the above, it was deemed necessary to determine both convergent and discriminant validity of the information intensity portion of the instrument. After screening the data for outliers and missing or invalid values, principal components analysis, without rotation and with VARIMAX rotation, was conducted using SPSS to assess the dimensionality of the construct. Two factors were expected to result, representing information intensity of the organization's products/services and its value chain. The Cronbach alpha of the items loading on each factor were assessed. For any factors with an alpha less than .70, the item loading coefficients were to be considered. Low-loading items were to be examined for their contributions by dropping them one at a time, beginning with the lowest loading value, followed by calculation of a new

Cronbach alpha. This process was to be repeated until an alpha value of .70 was attained or until only two items remained for that factor. If no combination of items could be found to result in an alpha of .70, the data were to be reexamined from the beginning using a threshold of .60. Factors for which no combination resulted in an alpha of at least .60 were to be dropped from further consideration. If no combination could be found resulting in an alpha of .60 for any of the factors, then the hypotheses examining the moderator effect (H11 through H20) would be considered unsupportable by the data and would not be tested further. The detailed results of this analysis are provided in chapter 4.

The other two portions of the instrument had been widely validated, thus it was not deemed necessary to revalidate either portion for the purpose of this study. Nonetheless, Cronbach alpha values were calculated for each of the constructs as a way of identifying unexpected patterns in the data. Any alpha values below .70 was investigated for the purpose of understanding the cause and determining whether any modifications to the study were warranted. The detailed results of this analysis are also provided in chapter 4.

Prior studies have indicated fairly strong correlation among information quality dimensions (Lee et al., 2002). For this reason, it was hypothesized that the instrument would exhibit the weak form of discriminant analysis, thus a second order analysis, modeled after the analyses conducted by Shemwell and Yavas (1999) and Dejnaronk (2000), was deemed appropriate. This analysis was to be used to test for both convergent and discriminant validity at a level of abstraction higher than that discussed above, that is, at the level of the PSP/IQ quadrants, organizational benefit dimensions, and information intensity aspects.

Data Collection Procedures

This section presents the procedures used for collecting the data, for securing and storing the data once collected, and for protecting the human participants providing the data.

Procedures for Data Collection

Data for this research were collected by means of a Web-based survey. The survey was hosted on a server operated by a commercial service provider using an account subscribed to by the researcher. Persons identified in the sample were invited to participate via a combination of electronic mail, flyers, and conference announcements. Those who chose to participate were provided with the Web address of the starting point for the survey. After accessing that Web address, participants indicated their responses by making selections on a series of screens. Upon completion of the survey, participants submitted their responses to the server, where they were collected and stored until retrieved by the researcher.

Data Security and Storage

Data were initially collected on the Web server hosting the survey. The server was protected using industry standard security practices, including but not limited to firewalls, password-protected accounts, and access controls. Access to the data collected on the server was available only to the researcher upon presentation of appropriate login credentials. Upon completion of the survey, the data were retrieved from the server in the form of an Excel spreadsheet, which was downloaded to the researcher's personal computer (*QuestionPro policies and procedures*, 2006). The researcher's personal computer was protected from unauthorized access and other exploits through the use of multiple layers of security, including hardware and software firewalls, an encrypted local network, and anti-virus software regularly and frequently updated through automated processes. Data were deleted from the server following completion of the survey at the time the account subscribed to by the researcher was closed. The data have backed up from the personal computer onto compact disk, and a copy has been stored in a locked facility at a separate location. The data will be retained for a minimum of seven years following publication of these research results.

Protection of Human Participants

This survey was conducted in accordance with procedures specified in the approval granted by Capella University's Institutional Review Board. Participants were recruited using non-coercive means involving an initial invitation and follow-up reminders as recommended by Robson (2002). An informed consent notification was provided at the beginning of the survey. In this notification, participants were given essential information about the research, were advised of the expected time commitment, were advised of any expected risks of participation, were advised that their participation was voluntary and that they could discontinue participation at any time. No personally identifiable information was collected as part of the survey. Participants were not paid for their participation.

Pilot Testing

Pilot tests are recommended for all fixed design research studies as a way to gauge the appropriateness of the overall study design and instrument design (Cooper & Schindler, 2003; Robson, 2002). In particular, Moore and Benbaset (1991) recommend that participants in a pilot study be asked to comment on the length, wording, and instructions for using the instrument.

For this study, a pilot study was conducted by administering the instrument to a small subset of the study's population and then by asking those participants to comment on those aspects of the instrument suggested by Moore and Benbaset (1991). The pilot study sampling frame was non-probabilistic, in that it included participants personally known by the researcher as well as individuals known to reflect an assortment of organization types and organizational roles, and to reflect varying degrees of knowledge about information quality. The quantitative data collected in the pilot were analyzed in accordance with the procedures specified for the full study. Qualitative data collected in response to the request for feedback were assessed to determine whether adjustments to the instrument design or to its administration are warranted prior to proceeding with the full study. The only adjustments made at this time were to the lists of choices answers for demographic questions related to industry, primary business, and job title.

Data Analysis Procedures

The following sections provide detail regarding the data handling and analysis. The first section describes the exploratory data analysis process used, as well as how the data were screened and cleansed with respect to missing data and extreme values. The second section describes how the data were analyzed for the main effect hypotheses. The third section describes the processes used for analyzing the moderator effect. The detailed results of these procedures are provided in chapter 4.

Exploratory Data Analysis

Prior to testing any of the hypotheses, it is essential that data be examined, screened, and cleansed if necessary to meet the assumptions associated with the statistical techniques employed. This section describes the procedures used for such exploratory data analysis.

The data were first screened for missing data. Where data were found to be missing, the data set was examined to determine the best approach for handling the missing data. There are two fundamental approaches to handling missing data; remove the cases or variables, or substitute values for the missing data. Mertler and Vannatta (2005) recommend a series of considerations to assist the researcher in determining which of these two approaches and their many variations is most appropriate to the situation at hand. If the number of cases with missing data is small, then deleting those cases is generally appropriate. If the number missing is not small, then substitution should be considered.

After the missing data were handled, the data used in evaluating each hypothesis were screened for extreme values. Multivariate outliers are those cases which represent unusual or extreme combinations of values. They can be identified through the use of the Mahalanobis distance procedure, which "is evaluated as a chi-square (χ^2) statistic with degrees of freedom equal to the number of variables in the analysis" (Mertler & Vannatta, 2005, p. 29). According to Mertler and Vannatta, outlier cases for which the Mahalanobis distance is significant at $p < .001$ should be investigated. If it appears that the case represents an error, it should be dropped. If it appears legitimate, the researcher should consider whether to analyze the results with and without the case in question and should assess options such as transforming the data as a way of reducing its impact.

In addition to missing data and outliers, the use of multiple regression is based on three basic assumptions regarding the data: normality, linearity, and homoscedasticity. The tests for these assumptions include both graphical and statistical examinations. For each hypothesis, a scatterplot matrix of the dependent variable and each independent variable were generated as a first indication. The ideal shape of each plot is an ellipse. Where the plot was not elliptical, each variable was assessed individually for normality using the Kolmorgov-Smirnov test, as well as determining the skewness and kurtosis of each variable. To the extent these tests revealed problems, transformations such as square roots, logarithms, reflections, and inverses were considered as appropriate for the particular normality problem detected. Linearity and homoscedasticity were examined by plotting the standardized predicted values and standardized residuals against each other. If the assumptions are met, the plot should fit a roughly rectangular pattern for linearity, and to indicate homoscedasticity, the values should be distributed fairly evenly above and below the plotted reference line. As with linearity, problems revealed through these plots were examined and, to the extent necessary, were addressed through transformations. Finally, it should be noted that while conformance to these assumptions is the ideal, some departure from the ideal was expected. Moreover, slight to moderate violations of the assumptions "merely weaken the regression analysis, but do not invalidate it" (Mertler & Vannatta, 2005, p. 174).

Main Effect Hypothesis Testing

Each of the main effect hypotheses was analyzed using stepwise multiple regression with stepwise selection. Stepwise multiple regression is considered appropriate for exploratory studies. Stepwise selection adds variables in the order of their contributions, yet tests the significance of each variable already added, removing them if it is determined that they no longer provide a significant contribution, resulting in the potential for a more parsimonious regression model. The end result of each regression is an equation of the form:

$$Y_j = \beta_0 + \beta_1 + \ldots + \beta_i + \varepsilon \qquad \text{(Equation 1)}$$

where β_i = a particular independent variable, and Y_j = an instance of a dependent variable.

Each beta coefficient (β) represents the standardized weighted contribution of a particular independent variable in predicting the value of a dependent variable (Mertler & Vannatta, 2005).

Before interpreting a multiple regression equation, it is important to consider the tolerance value, which is a measure of multicollinearity ranging from 0 to 1. Values of less than 0.1 are indicative of a multicollinearity problem. An alternative test for multicollinearity is the variance inflation factor, for which values greater than 10 are cause for concern. Two acceptable approaches for dealing with multicollinearity problems are to remove one of the problem variables or to combine two problem variables into one. The latter approach is recommended when the variables have an intercorrelation of .80 or higher (Mertler & Vannatta, 2005).

The output of the regression analysis consisted of three parts: the model summary, an ANOVA table, and a set of coefficients. In the model summary, the values for multiple correlation (R), the squared multiple correlation (R^2), and the adjusted squared multiple correlation (R^2_{adj}) were reviewed to assess how well the model predicted the dependent variable. In particular, R^2 and its adjusted variant (R^2_{adj}) were used to assess the total contribution of the independent variables. Both R and R^2 tend to overestimate the contribution, especially with small sample sizes, in which cases R^2_{adj} is considered to be more representative of the true contribution. In addition, since this analysis used a stepwise method, the change in the value of R^2 (ΔR^2) was reported for each step generated (Mertler & Vannatta, 2005).

The ANOVA table presented the F-test and level of significance for each step generated, reporting the degree to which the relationship was linear. A significant F-test is indicative of a linear relationship, hence a significant prediction. Finally, the set of coefficients was examined to consider the unstandardized coefficients (B), the standardized coefficients (β), the t values, significance values, and a set of correlation indices (Mertler & Vannatta, 2005).

Moderator Effect Hypothesis Testing

A moderator variable “systematically modifies either the form and/or strength of the relationship between a predictor and a criterion variable” (Sharma, Durand, & Gur-

Arie, 1981, p. 291). Sharma et al. refer to the former as a “pure moderator” (p. 293) if it is unrelated to either the predictor or criterion, or as a “quasi moderator” (p. 293) otherwise. They refer to variables that modify the strength of the relationship as “homologizer variables” (p. 292). Carte and Russell (2003) refer to the effects of pure and homologizer variables as “differential prediction” (p. 482) and “differential validity” (p. 482), respectively, while not mentioning quasi moderators. The terminology of Sharma et al. is used for the remainder of this section.

There are two basic approaches to testing for moderator variables: subgroup analysis and moderated regression analysis. In subgroup analysis, the data set is first divided into homogenous subgroups based on the value of the hypothesized variable. Regression analysis is then run on the groups separately, and the difference in R^2 values is determined. If the variable is a moderator, then R^2 will differ markedly (Sharma et al., 1981).

Moderated regression analysis involves the examination of the coefficients from the following three regression equations:

$$y = a + b_1x \quad \text{(Equation 2)}$$

$$y = a + b_1x + b_2z \quad \text{(Equation 3)}$$

$$y = a + b_1x + b_2z + b_3xz \quad \text{(Equation 4)}$$

According to Sharma et al. (1981), if b_3 is zero, but b_2 is not, then equations 3 and 4 do not differ, hence z is not a moderator variable. If b_2 is zero, but b_3 is not, then z is a pure moderator. If z is a quasi moderator, then b_2 must differ from b_3 and both coefficients must be nonzero.

Combining these two approaches, Sharma et al. (1981) recommend a step-by-step procedure for testing moderator variables. First, use moderated regression analysis to determine whether there is a significant interaction between the hypothesized moderator and the predictor variable. If so, then determine whether the hypothesized moderator is related to the criterion variable. A relationship in this case indicates a quasi moderator, and a lack of relationship indicates a pure moderator. If there was not a significant interaction in the first step, then determine whether the hypothesized moderator is related to either the criterion or predictor variables. If so, then it is not a moderator; else, conduct a subgroup analysis and test for significance in the differences in predictive validity. A significant difference in this test is indicative of a homologizer variable. Lack of a significant difference indicates that the variable is not a moderator.

This study applied the approach described above. Where subgroup analysis was called for by this procedure, two subgroups were used in the subgroup analysis based on a median split of the hypothesized moderator variable.

Limitations of Methodology

This section discusses limitations identified for the methodology used in this study. Four broad categories of such limitations have been identified: limitations of survey research, limitations of Web-based surveys, limitations of the statistical analysis techniques used in this study, and limitations associated generalizing the results to the population.

Limitations of Survey Research

Survey research is limited by the extent to which the responses accurately reflect the perspectives of the participants, and the extent to which those perspectives reflect the real-world situation under investigation. These limitations can be mitigated through rigorous attention to the design of the survey instrument and the extent of the limitation can be assessed by analyzing the construct validity of the instrument (Cooper & Schindler, 2003; Robson, 2002). The instrument used for this study was developed using accepted practices and the majority of the items used in the instrument had been validated previously. Further tests were conducted to assess the validity of the remaining items, as described earlier in this chapter.

Limitations of Web-Based Surveys

Because of the prevalence of unsolicited electronic mail, also known as “spam”, Web-based surveys tend to have very low response rates (Bullen, 2005). This creates the risk of obtaining small sample sizes, thus reducing the statistical power of the data and increasing the likelihood of both Type I and Type II errors (i.e., detecting relationships where none exist failing to detect relationships that do exist, respectively) (Mertler & Vannatta, 2005). A combination of approaches were taken to minimize the effect of the low response rates. First, electronic mail invitations to participate were sent from an individual familiar to many of those being invited. The second approach to minimizing the effect was to increase the number of invitations to participate. A low-end estimate of the response rate was used to select the desired size of this group. In addition, when the data set was found to include an indicator that certain records had been verified, the selection algorithm was adjusted, increasing the size of the selected set further.

Limitations of the Statistical Analysis Techniques in this Study

Multiple regression analysis was the primary technique used in this study. As discussed earlier in this chapter, this technique is based on a number of assumptions regarding the data. Each of these assumptions was tested for as described earlier, and to the extent feasible, data transformations were employed to meet the assumptions. In those cases where the assumptions could not be met through such transformations, the statistical power of the analysis was reduced, and any interpretations were limited accordingly.

This study also made extensive use of tests for moderator variables. A limitation associated with the analysis of moderator variables that is particularly relevant to this study is the problem of granularity of scale. In particular, the use of moderated regression analysis involves multiplying two variables together to create a third. When the first variable's scale has m values and the second variable's scale has n values, the new variable thus has $m \cdot n$ values. This limitation creates a situation in which an effect is measured with a scale considerably courser than the effect itself. This creates the "risk of severely attenuating observed ΔR^2" (Carte & Russell, 2003, p. 490), which increases the likelihood of Type II errors.

The instruments from which this survey instrument's items were taken measured information intensity and organizational outcomes using seven values each, and measured information quality using eleven values. Although more than five to seven values on a Likert scale are not deemed to significantly increase measurement capability (Carte & Russell, 2003), the choice of eleven values for the information quality instrument was a deliberate choice on the part of its designers, based on their experience with early versions of the instrument (Lee et al., 2002). To alleviate this problem, Carte and Russell recommend that the scale of the dependent variable be adjusted to a number of values equal the product of the values used to measure the other factors. In the case of this study, that would have required changing the scale of the organizational benefits to 77 values. This number was deemed unreasonably high for a Likert-type instrument. Even if the other variables were reduced to scales ranging from 1 to 5, the dependent variable would still need 25 values to meet this guideline. Rather than change the scale to such a high number of values, this research retained the scales used in the original instruments, and the increased risk of Type II errors was accepted and explicitly acknowledged.

Population-Based Limitations

The population for this study was defined rather broadly; however, it was still limited to persons working in an organization of a given size and who use information regularly. The ability to generalize the results is limited to that population, and is further limited by the characteristics of those who actually participated. The use of electronic mail and the Web to administer the survey is also likely to have the effect of disproportionately representing certain types of workers as compared to others. For example, knowledge workers who regularly use electronic mail and the Web were much more likely to be targeted than other information collectors and creators, such as those who use point of sale terminals or those who provide customer support services. This limitation was accepted as such, hence limiting the interpretation of the study's findings.

CHAPTER 4. DATA ANALYSIS

Introduction

This chapter presents an analysis of the data collected in support of this research. A total of 3,438 individuals were invited to participate in a Web-based survey, and 110 responses were received. Data were then prepared, examined, and screened for outliers and missing values. The hypotheses were then tested using a combination of multiple regression analysis, moderated regression analysis, and subgroup analysis. Support was found for all the main-effect hypotheses, as well as for many of the sub-hypotheses that were developed to address systematic differences uncovered during the data examination. Support was not found for the moderator-effect hypotheses.

Survey Administration

The variables identified in the research model were operationalized through a self-administered Web-based survey. The sample frame for the survey was the contacts database of an industry consortium that focuses on the business and technical aspects of the development and promotion of information systems interoperability. Participants were recruited in two ways. First, invitations to participate in the survey were sent to 3,210 individuals in the form of an electronic mail message sent by the Chief Executive Officer of the consortium. Second, each of the 228 attendees at the consortium's summer 2006 quarterly conference and member meeting was invited through a combination of verbal announcements and a flyer inserted in their registration packets. In each case, the invitation identified the purpose of the survey, indicated that the survey was consistent with the purposes of the consortium, encouraged participation, and assured participants of the confidentiality of responses. A URL was provided, directing participants to the first page of the survey. Separate instances of the survey, each with its own URL, were maintained to keep the two sets of responses separate, thus permitting analysis of any potential differences between the response sets

The total number of responses received between July 10, 2006, and August 7, 2006 (a four-week period) in response to the e-mail invitation was 86, representing a response rate of 2.7%. Tables 8a and 8b provide a summary of the responses received each day during that period.

. The total number of responses received between July 16, 2006, and August 7, 2006 (a three-week period) in response to the meeting announcement and flyer was 24, representing a response rate of 10.5%. Tables 9a and 9b provide a summary of the responses received each day during that period.

Taking these two sets together, a total of 246 surveys were started. Of those, 110 were completed, representing a combined completion rate of 44.7% and a completed response rate of 3.2%.

Table 8a.
Day-by-day Responses to E-mail Invitation

Event	Date	Surveys Started	Surveys Completed
Invitation sent	July 10	42	18
	July 11	17	7
	July 12	4	1
	July 13	1	1
	July 14	2	2
	July 15	1	1
	July 16	0	0
	July 17	2	0
	July 18	1	0
	July 19	1	1
	July 20	1	0
	July 21	0	0
	July 22	0	0
	July 23	0	0
	July 24	1	1
Reminder sent	July 25	56	26
	July 26	35	13
	July 27	23	9
	July 28	6	3

Table 8b.
Day-by-day Responses to E-mail Invitation

	July 29	1	1
	July 30	3	1
	July 31	4	0
	August 1	4	0
	August 2	0	0
	August 3	3	0
	August 4	3	1
	August 5	0	0
	August 6	0	0
	August 7	0	0
Total	All dates	211	86

Data Coding

Responses were collected on a Web server for each instance of the survey, and were subsequently downloaded as an Excel spreadsheet. Parts I and II of the survey used 7-point scales with values from 1 through 7 and were coded automatically using this scale. Part III of the survey used an 11-point scale with values from 0 through 10 and included several reverse-coded items. Data for this part of the survey were collected on a scale of 1 to 11, hence required adjustment for proper coding prior to importing into SPSS. Part IV of the survey collected categorical data for analysis purposes. Each of these items was assigned an integer code in Excel, and that integer was associated with an

Table 9a.
Day-by-day Responses to Invitation at Conference

Event	Date	Surveys Started	Surveys Completed
Registration opened	July 16	0	0
Opening day of conference	July 17	4	2
	July 18	1	1
Announcements made in member break-out meetings	July 19	12	9
	July 20	11	5
Last day of meetings	July 21	4	4
	July 22	0	0
	July 23	0	0
	July 24	0	0
	July 25	1	1
	July 26	0	0
	July 27	0	0
	July 28	0	0
	July 29	0	0
	July 30	0	0
	July 31	1	1
	August 1	0	0
	August 2	1	1
	August 3	0	0

Table 9b.
Day-by-day Responses to Invitation at Conference

	August 4	0	0
	August 5	0	0
	August 6	0	0
	August 7	0	0
Total	All dates	35	24

appropriate text label in SPSS. Some of these items also included an "other" response field in which participants could enter free-form text. These items were examined and mapped to either one of the existing codes or to a new code as deemed appropriate by the researcher. In addition to the survey items, certain administrative data items were collected automatically by the Web server, and were included as string variables in the Excel spreadsheet. Of these, only the country code was subsequently used.

Each of the variables used in hypothesis testing was associated with a set of survey items. Following instrument validation, the values for these variables were calculated as the statistical mean of the retained items associated with each variable.

Response Analysis

This section presents a response analysis, reviewing general characteristics of the data set and the respondents. Following the screening of data for coding errors and unusual patters, the data were examined to assess the general characteristics of the respondents and to assess the extent to which there were systematic differences based on respondent characteristics. Case numbers included in this discussion represent the sequential position within the complete set of started responses, regardless of completion status.

Data Screening

Univariate analysis was conducted on all the variables to ensure proper coding and proper recording of all values and to examine the data for any unusual patterns that could be problematic to the analysis. Some minor errors were noted and corrected as a result of this analysis.

The maximum number of target item values possible from the 110 completed responses was 10,780. An examination of the data revealed that only 72 of the 110 responses had values for all 98 target items, leaving 38 cases with at least one missing value. A frequency analysis indicated a total of 172 missing target item values, representing 1.6% of those possible. Closer examination of the data set revealed that three cases (#55, #143, and #183) together contributed more than half the total missing values, thus those cases were excluded and the data were reexamined. After excluding those three cases, the missing values were reduced to 83, which represented 0.79% of the 10,486 total values possible. The remaining missing values appeared to be randomly distributed across the items and cases, and no single item was missing more than 5% of its possible values. Based on this analysis, it was determined that the 107 remaining cases would be useful for subsequent analysis and that the missing data among those cases would not pose a systematic problem.

Each target item was screened for outliers. According to Mertler and Vannatta (2005), for sample sizes greater than 100, the likelihood of finding a few cases with values more than three standard deviations from the mean is very high. As such, they suggest that four standard deviations is a better rule of thumb for this size sample. Toward that end, standardized scores were calculated for each variable, and any value 4 or greater was treated as an outlier. Only two cases (#72 and #125) met this criterion, both with respect to a single item in the information intensity portion of the survey. To address this issue, rather than drop the cases outright, a new variable was created, recoding those two as system-missing, thus permitting the cases to be used in calculations except for those in which the affected data item was involved. After recoding, it was determined that no values even exceeded three standard deviations from the mean.

To test for multivariate outliers, the Mahalanobis distance was calculated for each case, taking into account all 98 target items, and those distance values were compared against the chi square critical values for 98 degrees of freedom at $p = .001$. The upper and lower bounds were determined to be 141.01 and 60.36, respectively. No cases were found to exceed the upper bound; however, one case (#200) had a value below the lower bound. An examination of this case indicated nothing unusual except for a relatively narrow range of response selections. As such, it was decided that this represented a legitimate case and that it should be retained.

Participant Characteristics

The consortium whose contacts database was used as the sampling frame has approximately 250 member organizations located throughout the world. The historical focus of the consortium as an organization is the intersection between business and information technology requirements. As such, a diverse set of participants was expected.

Geographically, a majority of participants who completed the survey (56.07%) were from North America. The next largest group was from Europe (30.85%). The remaining participants were from Asia (5.61%), Australia (3.74%), Africa (2.80%), and South America (0.93%).

Participants were overwhelmingly male, with males submitting 86.92% of the completed responses.

Participants were generally highly educated. Those with master's or specialist's degrees constituted 48.60% of the sample. The next largest group (31.78%) held bachelor's degrees, and 15.88% held doctorates. Only 3.74% of the participants indicated that their highest level of education was a high school diploma or equivalent.

Participants also have considerable experience in their respective jobs and industries. A majority (53.27%) have been in their industries for 20 years or more. Other responses are shown in Table 10. Participants were fairly evenly distributed in terms of the number of years with their current organizations. Roughly one quarter each reported 5 to 10 years, 10 to 20 years, and more than 20 years, as shown in Table 11.

In terms of job title or function, the largest group of participants was made up of IT professionals, making up 43.14% of the sample. Consultants were next at 19.61%. The remaining participants were fairly evenly distributed across several different job functions, as shown in Table 12.

Table 10.
Years Experience in Industry

	Frequency	Percent	Cumulative Percent
Less than 1 year	1	.93	.93
1 to 5 years	9	8.41	9.34
5+ to 10 years	10	9.34	18.69
10+ to 20 years	30	28.04	46.73
Greater than 20 years	57	53.27	100.00
Total	107	100.00	

Table 11.
Years With Current Organization

	Frequency	Percent	Cumulative Percent
Missing	1	.93	.93
Less than 1 year	7	6.54	7.48
1 to 5 years	20	18.69	26.17
5+ to 10 years	26	24.30	50.47
10+ to 20 years	27	25.23	75.70
Greater than 20 years	26	24.30	100.00
Total	107	100.00	

In terms of the organizations represented by the participants, roughly three-quarters (75.47%) were for-profit organizations. Non-profit, governmental, and academic organizations made up the balance with 13.21%, 8.49%, and 2.83%, respectively. More than half the organizations (52.34%) are in the information technology industry. Another 17.76% are in the aerospace and defense industry. The remaining organizations are distributed fairly evenly across a wide assortment of industries as shown in Table 13. The primary business activity at participant sites is distributed similarly, but not identically; details are provided in Table 14. With respect to organization size, a majority (59.81%) of participants were from very large organizations, with 10,000 or more employees, although other organization sizes are also represented. Small organizations of fewer than 100 employees make up 16.82% of the sample, medium organizations (101 to 1000) are the smallest group at 7.48%, and large organizations of 1,001 to 10,000 employees make up 15.89%.

Table 12.
Job Title or Function

	Frequency	Percent	Cumulative Percent
Consultant	20	19.61	19.61
Education	2	1.96	21.57
Engineer	7	6.86	28.43
Executive	7	6.86	35.29
IT Professional	44	43.14	78.43
Management	10	9.80	88.23
Professional (other than IT)	6	5.89	94.12
Researcher	4	3.92	98.04
Sales	1	.98	99.02
Missing	1	.98	100.0
Total	102	100.00	

Finally, in terms of stakeholder roles with respect to information, all three of the targeted groups are represented. Information consumers make up nearly half the sample with 48.60%. Information providers/collectors constitute 29.91% of the sample, and information custodians make up the remaining 21.49%.

Response Bias Analysis

The sample for this study was selected in a manner consistent with the privacy requirements of the organization providing the sampling frame. As such, the researcher was not permitted to examine the list of invited participants to assess the characteristics of the sample for the purpose of evaluating any response/non-response bias. Nonetheless, it is possible to examine the responses received to determine whether any systematic bias was indicated.

Independent *t* tests were conducted on all the target items in the survey to determine whether there were significant differences in the responses for those who completed the survey than for those who started the survey but dropped out before completing it. In those cases where the participant started but did not complete the survey, there was a noticeable drop in responses after the first set of target questions measuring information intensity; many items after that point did not have any data at all. However, for those items for which data were available in the partially completed set, none had significant differences at $p = .01$. As such, survey attrition was not considered to pose a systematic difference in response.

Independent *t* tests were also conducted on all the target items in the survey to determine whether there were significant differences in the responses for those who

Table 13.
Industries Represented in the Sample

	Frequency	Percent	Cumulative Percent
Aerospace/defense	19	17.76	17.76
Banking/financial	2	1.87	19.63
Customer service	1	.93	20.56
Education	2	1.87	22.43
Energy	4	3.74	26.17
Government contracting	2	1.87	28.04
Health care	1	.93	28.97
Information technology	56	52.34	81.31
Insurance	1	.93	82.24
Manufacturing	1	.93	83.18
Research and development	5	4.67	87.85
Telecommunications	1	.93	88.78
Transportation	4	3.74	92.52
Public Relations	2	1.87	94.39
Consulting	5	4.67	99.07
Non-profit	1	.93	100.00
Total	107	100.00	

Table 14.
Primary Business at Participant Site

	Frequency	Percent	Cumulative Percent
Aerospace/defense	13	12.15	12.15
Banking/financial	2	1.87	14.02
Customer service	2	1.87	15.89
Education	3	2.80	18.69
Energy production/distribution	4	3.74	22.43
Engineering	5	4.67	27.10
Health care	1	.93	28.03
Information systems/technology	57	53.27	81.31
Insurance	1	.93	82.24
Research and development	10	9.35	91.59
Retail	1	.93	92.52
Transportation	1	.93	93.46
Public relations	2	1.87	95.33
Non-profit	1	.93	96.26
Administration and management	1	.93	97.20
Professional Services	1	.93	98.13
Missing	2	1.87	100.00
Total	107	100.00	

responded to the e-mail invitation than for those who responded to the invitation received in person at the conference. Of the 98 target items, one (1.02% of all items) was found to have significant differences at $p = .01$. Given this small percentage, the manner of invitation was not considered to pose a systematic difference in response.

Independent *t* tests were conducted on all the target items in the survey to determine whether there were significant differences in the responses between the sexes. None of the target items were found to have significant differences at $p = .01$. As such, gender was not considered to pose a systematic difference in response.

To determine whether there were significant differences in responses from different parts of the world, it was first necessary to adjust the way world regions were grouped. Because most of the regions represented had very small samples, a new variable was created to group all regions other than Europe and North America into a single "Rest of the World" group. A one-way ANOVA was then conducted using this variable as the grouping variable. No items had a significant difference at $p = .01$. As such, world region was not considered to pose a systematic difference in response.

To determine whether there were significant differences in responses from different industries, it was first necessary to adjust the way industries were grouped. Because most of the industries represented had very small samples, a new variable was created to group all industries other than the two with reasonably large samples (Information Technology and Aerospace/Defense) into a single "Other" group. A one-way ANOVA was then conducted using this variable as the grouping variable. No items had a significant difference at p = .01. As such, industry was not considered to pose a systematic difference in response.

A one-way ANOVA was conducted on all the target items in the survey to determine whether there were significant differences in the responses from different types of organization. None of the target items were found to have significant differences at $p = .01$. As such, organization type was not considered to pose a systematic difference in response.

A one-way ANOVA was conducted on all the target items in the survey to determine whether there were significant differences in the responses from different job titles or functions. Two (2.04%) of the target items were found to have significant differences at $p = .01$. Given the low percentage, job title or function was not considered to pose a systematic difference in response.

A one-way ANOVA was conducted on all the target items in the survey to determine whether there were significant differences in the responses from people in different stakeholder roles with respect to the information system considered for the context of the survey. Six (6.1%) of the target items were found to have significant differences at $p = .01$. Closer inspection revealed a systematic pattern in which information custodians rated the quality of information in their systems higher than either information providers or information consumers. Because of this pattern, it was determined that separate analysis would be required to assess the implications of these differences.

Construct Analysis

This section details the steps that were followed to conduct the construct analysis of the survey instrument. Each of the three parts of the survey containing target questions was evaluated separately. Upon completion of the construct validity analysis, variables to be used in hypothesis testing were constructed from the survey item responses and were then screened.

Part I – Information Intensity

As stated in chapter 3, it was necessary to determine both convergent and discriminant validity of Part I (the information intensity portion) of the survey instrument. To screen for multivariate outliers in this portion, the Mahalanobis distance was assessed using only the 11 items from this part of the survey. The chi-square critical values at $p = .001$ for 11 degrees of freedom are 1.83 for the lower bound and 31.26 for the upper bound. No cases had Mahalanobis distances below the lower bound, and three cases (#4, #10, and #161) exceeded the upper bound. These cases were investigated and found to be made up largely of response values at each extreme. As such, these cases were dropped from further consideration.

Discriminant analysis was performed on these items using principle components analysis. A commonly used approach to identifying the number of principle components or factors to retain is to rely upon Kaiser's rule, in which those with Eigenvalues of 1 or greater are accepted, while all others are rejected. However, Mertler and Vanatta (2005) suggest that this approach loses reliability as sample sizes decrease below 300, and characterize the reliability for sample sizes of approximately 100 cases as "poor"

(p. 260). To compensate for this weakness when evaluating smaller sample sizes, they suggest that several criteria should be considered concurrently to determine the appropriate number of factors. In particular, they suggest consideration of communalities (seeking to find all items above 0.7), percent of total variance explained (seeking 70% or greater), and residuals (seeking to find few residuals above 0.05). This analysis is to be conducted iteratively, overriding Kaiser's rule and increasing the number of factors until the criteria are satisfied. They also recommend considering Bartlett's test of sphericity to assess the adequacy of the sample size, treating significance as indicative of an adequate sample. For small samples, Garson (2006a) also recommends assessing the Kaiser-Meyer-Olkin (KMO) statistic, accepting only samples that produce values of 0.6 or greater.

Principle component analysis for Part I of the survey was conducted using this process. Both Bartlett's test of sphericity (p = .000) and the KMO statistic (.74) indicated that the sample size was sufficient. Kaiser's rule was used for the first iteration, yielding four factors. However, none of the criteria specified by Merlter and Vanatta (2005) was satisfied, with four of the communalities being less than .7, a total explained variance of only 69.84%, and with more than half (29 of 55) the residuals exceeding .05. By increasing the number of factors iteratively, a solution with seven factors was deemed to meet the criteria sufficiently. At seven factors, there were no communalities below .7, 88.04% of the variance was explained, and the number of residuals above .05 had been reduced to 19. Although seven factors were identified, only the first three had more than one item loading at .4 or greater. Given that the concept of convergent validity of a single item is not a meaningful concept, it was decided to drop all items other than the ones that loaded on the first three factors, accounting for 49.20% of the variance explained. The resulting factors and their loadings are presented in Table 15.

Table 15.
Information Intensity Factors and Their Loadings.

	Component						
	1	2	3	4	5	6	7
IIVC	.88						
IIPS5	.78						
IIVC4	.76						
IIPS6		.93					
IIPS2		.84					
IIPS7			.79				
IIPS3			.72				
IIPS4				.94			
IIVC2					.93		
IIVC3						.96	
IIPS1							.96

To determine the convergent validity of these factors, Cronbach alpha was calculated for each set of items. The first factor, with three items, has an alpha value of .75, which is above the target threshold of .7. The second, with two items, has an alpha of .88, placing it well above the target. The third, however, only has an alpha of .57, suggesting that those two items should be dropped. To confirm this suggestion, the survey wording for each of these seven items was reviewed. The items associated with the first factor clearly have to do with the complexity of product or service operations. The items associated with the second factor clearly have to do with the complexity of the organization's product or service. However, the items associated with the third factor are not so readily categorized.

Based on this analysis, it was decided that the five items loading on these two factors would be retained for subsequent analysis. Although these two factors only explain 36.67% of the total variance, this solution is intuitively meaningful and has acceptably high degrees of both discriminant and convergent validity.

Part II – Organizational Benefits

As stated in chapter 3, discriminant and convergent validity tests have been conducted in prior research studies on the items in Part II (the organizational benefits portion) of the survey instrument. Nonetheless, it was decided that convergent validity would be reassessed to screen for unusual data patterns.

First, however, to screen for multivariate outliers in this portion, the Mahalanobis distance was assessed using only the 18 items from this part of the survey. The chi-square critical values at p = .001 for 18 degrees of freedom are 4.91 for the lower bound and 42.31 for the upper bound. Three cases (#144, #186, and #223) had Mahalanobis distances below the lower bound, and three cases (#31, 32, and 62) exceeded the upper bound. These cases were investigated to assess whether deletion was appropriate. The case with the lowest value was found to have all 6's and was dropped. The other two low-distance cases were found to have responses in a narrow range, but were not otherwise unusual. Similarly, the high-distance cases were found to have responses over a wide range, but were not otherwise unusual. Based on this assessment, all of these cases, with the exception of #144, were retained.

Table 16.
Organizational Benefits Item Convergence.

Category	Dimension	Number of items	α
Strategic Benefits	Alignment	3	.68
Strategic Benefits	Competitive Advantage	2	.82
Strategic Benefits	Customer Relations	3	.73
Transactional Benefits	Business Efficiency	4	.80
Transactional Benefits	Communications Efficiency	2	.52
Transactional Benefits	Systems Development Efficiency	4	.81

Cronbach alpha values were calculated for each set of items in Part II of the study. These values are listed in Table 16. Examination of those dimensions with alphas below .7 indicated that no adjustments could be made to improve the alpha. With respect to Alignment, each of the three items had item-to-total correlations of approximately .5, and removing any of them would have lowered the alpha rather than raise it. With respect to Communications Efficiency, there were only two items, thus removal would result in the inability to calculate a new alpha. For these reasons, it was decided that these two dimensions would be removed from further consideration, leaving each category with two dimensions.

Table 17.
Information Quality Item Convergence

PSP/IQ Quadrant	Dimension	Number of items	α
Soundness	Completeness	6	.91
Soundness	Concise Representation	4	.86
Soundness	Consistent Representation	4	.87
Soundness	Free of error	4	.92
Dependability	Security	4	.86
Dependability	Timeliness	5	.89
Usefulness	Appropriate Amount	4	.73
Usefulness	Interpretability	5	.76
Usefulness	Objectivity	4	.83
Usefulness	Relevancy	4	.88
Usefulness	Understandability	4	.88
Usability	Accessibility	4	.89
Usability	Believability	4	.88
Usability	Ease of operation	5	.84
Usability	Reputation	4	.86
Usability	Value-added	4	.92

Part III – Information Quality

As stated in chapter 3, discriminant and convergent validity tests have been conducted in prior research studies on the items in Part III (the information quality portion) of the survey instrument. Nonetheless, it was decided that convergent validity would be reassessed to screen for unusual data patterns.

First, however, to screen for multivariate outliers in this portion, the Mahalanobis distance was assessed using only the 69 items from this part of the survey. The chi-square critical values at p = .001 for 69 degrees of freedom are 38.30 for the lower bound and 111.06 for the upper bound. All cases were found to have distances within this range.

Cronbach alpha values were calculated for each set of items in Part III of the study. These values are listed in Table 17. Given that each of these alpha values was well above the target of .7, no further analysis was necessary. All 69 items making up the 16 dimensions were retained.

Variables Construction and Screening

Using the results of the analysis described above, new variables were constructed at two levels. At the lower level, survey items were used to construct a set of dimension-level variables. At the upper level, dimension-level variables were used to construct a set of category or quadrant-level variables. In each case, the statistical mean of the variables at one level was used to construct a single variable at the next level.

For information intensity, two variables were constructed in this fashion to represent the factors identified above, and then a single variable was constructed to represent information intensity. For organizational benefits, four variables were constructed to represent the remaining organizational benefits dimensions, and then two were constructed to represent strategic benefits and transactional benefits. For information quality, sixteen variables were constructed to represent the dimensions, and then four were constructed to represent the quadrants in the PSP/IQ model.

These variables were then screened for outliers and normality. Two cases (#72 and #175) had values more than four standard deviations away from the mean in the information intensity operations complexity variable. These were addressed by creating a new variable in which these two cases were coded as missing. No other outliers were identified.

Having addressed the outliers, the systematic difference by participants in different stakeholder roles was readdressed by conducting a one-way ANOVA on the constructed variables. Differences significant at $p = .05$ were found for the following variables: operational complexity (information intensity), customer relations (strategic benefit), concise representation (soundness), consistent representation (soundness), and value-added (usability). At the higher level, strategic benefits and soundness each had significant differences at $p = .05$. Because of these differences, it was determined that analyses involving these variables would be considered both separately and in aggregate.

Initial screening for normality was conducted by calculating the Kolmogorov-Smirnov statistic, looking specifically at the Lilliefors significance correlation. For the information intensity variables, this test indicated that none of the variables were normal. For the organizational benefits variables, the test indicated that business efficiency and competitive advantage were non-normal. For the information quality variables, the test indicated that believability, concise representation, ease of operation, security, and value-added were non-normal. Normality was also checked separately for each stakeholder role for the variables identified above as having significant role-based differences. In this analysis, operational complexity was non-normal within each role, concise representation was non-normal for information providers, and value-added was non-normal for information consumers. All other role-specific distributions were normal. The conclusions of this initial screening were confirmed by visually examining the histograms and Normal Q-Q plots for each variable.

The normality problem for each of the information intensity variables was characterized by a strong negative skew. Several transformations were examined, but none resolved or substantially improved the normality problem.

Two adjustments were made to address the business efficiency normality problem. First, case #24 was excluded, as this case appeared to be unduly influencing the distribution. Second, although the Kolmogov-Smirnov statistic was still significant, the use of a square root transformation reduced the significance level and improved the appearance of the histogram and Normal Q-Q plot, thus the square root transformation was accepted.

Believability was transformed by reflecting the variable (by subtracting the value from the maximum plus one) and taking the square root. This transformed variable resulted in a non-significant Kolmogorov-Smirnov statistic.

Table 18a.
Summary of Normality Resolution

Variable	Description	Level	Normality	Transformations	Special Considerations
II	Information intensity	Upper	No	None	
IIOC	Operational Complexity	Lower	No	None	
IIPC	Product/Service Complexity	Lower	No	None	
OBSB	Strategic Benefits	Upper	Yes		Roles significant
OBSBCA	Competitive Advantage	Lower	No	None	
OBSBCR	Customer Relations	Lower	Yes		Role significant
OBTB	Transactional Benefit	Upper	Yes		
OBTBBE	Business Efficiency	Lower	No	Square root	
OBTBSDE	Systems Development Efficiency	Lower	Yes		
IQSD	Soundness	Upper	Yes		Role significant
IQSDCC	Concise Representation	Lower	Conditional	None	Role significant. Not normal for information providers; normal for other roles

Table 18b.
Summary of Normality Resolution

IQSDCP	Completeness	Lower	Yes		
IQSDCR	Consistent Representation	Lower	Yes		Role significant
IQSDFE	Free of error	Lower	Yes		
IQDP	Dependability	Upper	Yes		
IQDPS	Security	Lower	No	None	
IQSDT	Timeliness	Lower	Yes		
IQUF	Usefulness	Upper	Yes		
IQUFAA	Appropriate Amount	Lower	Yes		
IQUFI	Interpretability	Lower	Yes		
IQUFO	Objectivity	Lower	Yes		
IQUFRL	Relevance	Lower	Yes		
IQUFU	Understandability	Lower	Yes		
IQUB	Usability	Upper	Yes		
IQUBAC	Accessibility	Lower	Yes		
IQUBB	Believability	Lower	Transformed only	Reflect and square root	
IQUBEO	Ease of operation	Lower	No	None	
IQUBRP	Reputation	Lower	Yes		
IQUBVA	Value-added	Lower	Yes		Role significant

The value-added normality problem for the information consumer role was determined to be due to the influence of two extreme values; these were recoded into a new variable as missing, resulting in a sufficiently normal distribution.

Transformations were also considered for the competitive advantage, ease of operations, and security variables, but none of the transformations improved the normality. As such, no changes were made to these variables.

Tables 18a and 18b provide a variable-by-variable summary of these normality resolutions.After screening for multivariate outliers, a dimension-level factor analysis was conducted, using the procedure described above. One case (#208) had a Mahalanobis distance value substantially above the maximum and was excluded from the analysis. The resulting model had eight factors that explained 87.40% of the variance. In this model, the information intensity variables each loaded on a factor by themselves, all the organizational benefits variables loaded on a single factor on which no other variables loaded, and the information quality variables loaded on the remaining factors. Based on this analysis, the discriminant validity of the overall model was confirmed.

Hypothesis Testing – Main Effect

Multiple regression analysis was conducted to test the main-effect hypotheses, H1 through H10. In each case, stepwise multiple regression was conducted to determine which of the independent variables associated with information quality were predictors of the organizational benefits dependent variable. Residuals analysis was conducted in each case to determine whether there were systematic violations of the assumptions of multivariate linearity, normality, and homoscedasticity.

Hypothesis 1

H1: Improvements in the soundness of information will be associated with increased strategic benefits.

The independent variables associated with this hypothesis include completeness, concise representation, consistent representation, and freedom from error. The dependent variable, strategic benefits, represents the statistical mean of the variables for competitive advantage and customer relations. Of these variables, three have significant differences between stakeholder roles. As such, the following three sub-hypotheses were also evaluated:

H1a: Improvements in the soundness of information will be associated with increased strategic benefits as measured by information producers/information collectors.

H1b: Improvements in the soundness of information will be associated with increased strategic benefits as measured by information custodians.

H1c: Improvements in the soundness of information will be associated with increased strategic benefits as measured by information consumers.

To evaluate H1, stepwise multiple regression analysis was conducted to determine which of the independent variables (completeness, concise representation, consistent representation, and freedom from error) were predictors of strategic benefits. The descriptive statistics for these variables are shown in Table 19. Regression results indicate two predictive models.

Model 1, which has a tolerance of 1.00, indicates Completeness as a significant predictor of Strategic Benefits, $R^2 = .12$, $R^2_{adj} = .11$, $F(1,99) = 12.94$, $p = .001$. This model accounted for 11% of the variance in Strategic Benefits. A summary of the regression model is presented in Table 20. The bivariate and partial correlation coefficients between the predictor and the dependent variable are presented in Table 21.

Model 2, which has a tolerance of .40, indicates Completeness and Consistent Representation as predictors of Strategic Benefits, $R^2 = .17$, $R^2_{adj} = .15$, $F(1,98) = 10.05$, $p < .001$. This model accounted for 15% of the variance in Strategic Benefits. A summary of the regression model is presented in Table 22. The bivariate and partial correlation coefficients between the predictor and the dependent variable are presented in Table 23.

Table 19.
Descriptive Statistics for H1 Variables.

Variable	*M*	*SD*	*N*
Strategic Benefits	4.78	1.26	101
Completeness	6.17	1.91	101
Concise Representation	5.80	1.93	101
Consistent Representation	6.56	2.08	101
Free of Error	6.99	1.86	101

Analysis of the residuals revealed no evidence of violations of the assumptions of linearity, normality, or homoscedasticity, hence the results of the multiple regression analysis are accepted as tenable and the null hypothesis $H1_{null}$ is rejected.

Table 20.
Model Summary for H1 Model 1.

Predictor	R	R^2	R^2_{adj}	ΔR^2	F_{chg}	p	df_1	df_2
Completeness	.34	.12	.11	.12	12.94	.001	1	99

Table 21.
Coefficients for H1 Model 1.

Predictor	B	β	t	Bivariate r	Partial r
Completeness	.22	.34	3.60	.34	.34

Table 22.
Model Summary for H1 Model 2.

Predictors	R	R^2	R^2_{adj}	ΔR^2	F_{chg}	p	df_1	df_2
Completeness and Consistent Representation	.41	.17	.15	.06	6.45	.001	1	98

Table 23.
Coefficients for H1 Model 2.

Predictor	B	β	t	Bivariate r	Partial r
Completeness	.41	.62	4.31	.34	.40
Consistent Representation	-.22	-.37	-2.54	.11	-.25

Sub-hypothesis H1a

To evaluate H1a, stepwise multiple regression analysis was conducted to determine which of the independent variables (completeness, concise representation, consistent representation, and freedom from error) were predictors of strategic benefits as measured by information producers/information collectors. The descriptive statistics for these variables are shown in Table 24. Regression results indicate an overall model with one predictor (Completeness) that significantly predicts Strategic Benefits, $R^2 = .17$, $R^2_{adj} = .14$, $F(1,27) = 5.47$, $p = .027$. This model, which has a tolerance of 1.00, accounted for 14% of the variance in Strategic Benefits. A summary of the regression model is presented in Table 25. The bivariate and partial correlation coefficients between the predictor and the dependent variable are presented in Table 26.

Table 24.
Descriptive Statistics for H1a Variables.

Variable	*M*	*SD*	*N*
Strategic Benefits	4.56	1.33	29
Completeness	6.18	2.07	29
Concise Representation	5.94	1.95	29
Consistent Representation	6.93	1.80	29
Free of Error	7.00	1.83	29

Analysis of the residuals revealed no evidence of violations of the assumptions of linearity, normality, or homoscedasticity, hence the results of the multiple regression analysis are accepted as tenable and the null hypothesis $H1a_{null}$ is rejected.

Table 25.
Model Summary for H1a.

Predictor	R	R^2	R^2_{adj}	ΔR^2	F_{chg}	p	df_1	df_2
Completeness	.41	.17	.14	.17	5.47	.027	1	27

Table 26.
Coefficients for H1a.

Predictor	B	β	t	Bivariate r	Partial r
Completeness	.26	.41	2.34	.41	.41

Sub-hypothesis H1b

To evaluate H1b, stepwise multiple regression analysis was conducted to determine which of the independent variables (completeness, concise representation, consistent representation, and freedom from error) were predictors of strategic benefits as measured by information custodians. The descriptive statistics for these variables are shown in Table 27. Regression results indicate an overall model with one predictor (Consistent Representation) that significantly predicts Strategic Benefits, $R^2 = .31$, $R^2_{adj} = .27$, $F(1,20) = 8.91$, $p = .007$. This model, which has a tolerance of 1.00, accounted for 27% of the variance in Strategic Benefits. A summary of the regression model is presented in Table 28. The bivariate and partial correlation coefficients between the predictor and the dependent variable are presented in Table 29.

Table 27.
Descriptive Statistics for H1b Variables.

Variable	*M*	*SD*	*N*
Strategic Benefits	5.33	1.26	22
Completeness	6.90	2.04	22
Concise Representation	6.59	2.14	22
Consistent Representation	7.39	2.16	22
Free of Error	7.64	1.97	22

Analysis of the residuals revealed no evidence of violations of the assumptions of linearity or normality, but there is slight evidence that heteroscedasticity may be a problem. However, Garson (2006b) notes that moderation violations of the assumption of homoscedasticity have only a minor impact on the regression estimates. As such, the results of this multiple regression analysis are accepted as tenable and the null hypothesis $H1b_{null}$ is rejected.

Table 28.
Model Summary for H1b.

Predictor	R	R^2	R^2_{adj}	ΔR^2	F_{chg}	p	df_1	df_2
Consistent Representation	.56	.31	.27	.31	8.91	.007	1	20

Table 29.
Coefficients for H1b.

Predictor	B	β	*t*	Bivariate *r*	Partial *r*
Consistent Representation	.32	.56	2.99	.56	.56

Sub-hypothesis H1c

To evaluate H1c, stepwise multiple regression analysis was conducted to determine which of the independent variables (completeness, concise representation, consistent representation, and freedom from error) were predictors of strategic benefits as measured by information consumers. The descriptive statistics for these variables are shown in Table 30. Regression results indicate no significant predictors of Strategic Benefits. As such, the null hypothesis $H1c_{null}$ is not rejected.

Table 30.
Descriptive Statistics for H1c Variables.

Variable	*M*	*SD*	*N*
Strategic Benefits	4.67	1.18	50
Completeness	5.84	1.70	50
Concise Representation	5.38	1.71	50
Consistent Representation	5.98	2.06	50
Free of Error	6.70	1.80	50

Hypothesis 2

H2: Improvements in the dependability of information will be associated with increased strategic benefits.

The independent variables associated with this hypothesis include security and timeliness. The dependent variable, strategic benefits, represents the statistical mean of the variables for competitive advantage and customer relations. Of these variables, only one has significant differences between stakeholder roles. To address these differences, the following three sub-hypotheses were also evaluated:

H2a: Improvements in the dependability of information will be associated with increased strategic benefits as measured by information producers/information collectors.

H2b: Improvements in the dependability of information will be associated with increased strategic benefits as measured by information custodians.

H2c: Improvements in the dependability of information will be associated with increased strategic benefits as measured by information consumers.

To evaluate H2, stepwise multiple regression analysis was conducted to determine which of the independent variables (security and timeliness) were predictors of strategic benefits. The descriptive statistics for these variables are shown in Table 31. Regression results indicate an overall model with one predictor (Timeliness) that significantly predicts Strategic Benefits, $R^2 = .10$, $R^2_{adj} = .09$, $F(1,99) = 10.55$, $p = .002$. This model, which has a tolerance of 1.00, accounted for 9% of the variance in Strategic Benefits. A summary of the regression model is presented in Table 32. The bivariate and partial

correlation coefficients between the predictor and the dependent variable are presented in Table 33.

Table 31.
Descriptive Statistics for H2 Variables.

Variable	*M*	*SD*	*N*
Strategic Benefits	4.78	1.26	101
Timeliness	6.90	1.75	101
Security	7.08	2.15	101

Table 32.
Model Summary for H2.

Predictor	R	R^2	R^2_{adj}	ΔR^2	F_{chg}	p	df_1	df_2
Timeliness	.31	.10	.09	.10	10.55	.002	1	99

Analysis of the residuals revealed no evidence of violations of the assumptions of linearity, normality, or homoscedasticity, hence the results of the multiple regression analysis are accepted as tenable and the null hypothesis $H2_{null}$ is rejected.

Table 33.
Coefficients for H2.

Predictor	B	β	*t*	Bivariate *r*	Partial *r*
Timeliness	.22	.31	3.25	.31	.31

Sub-hypothesis H2a

To evaluate H2a, stepwise multiple regression analysis was conducted to determine which of the independent variables (timeliness and security) were predictors of strategic benefits as measured by information producers/information collectors. The descriptive statistics for these variables are shown in Table 34. Regression results indicate an overall model with one predictor (Timeliness) that significantly predicts Strategic Benefits, $R^2 = .14$, $R^2_{adj} = .11$, $F(1,27) = 4.33$, $p = .047$. This model, which has a tolerance of 1.00, accounted for 11% of the variance in Strategic Benefits. A summary of the regression model is presented in Table 35. The bivariate and partial correlation coefficients between the predictor and the dependent variable are presented in Table 36.

Table 34.
Descriptive Statistics for H2a Variables.

Variable	*M*	*SD*	*N*
Strategic Benefits	4.56	1.33	29
Timeliness	6.85	1.90	29
Security	7.12	1.88	29

Analysis of the residuals revealed no evidence of violations of the assumptions of linearity, normality, or homoscedasticity, hence the results of the multiple regression analysis are accepted as tenable and the null hypothesis $H2a_{null}$ is rejected.

Table 35.
Model Summary for H2a.

Predictor	R	R^2	R^2_{adj}	ΔR^2	F_{chg}	p	df_1	df_2
Timeliness	.37	.14	.11	.14	4.33	.047	1	27

Table 36.
Coefficients for H2.

Predictor	B	β	t	Bivariate r	Partial r
Timeliness	.26	.37	2.08	.37	.37

Sub-hypothesis H2b

To evaluate H2b, stepwise multiple regression analysis was conducted to determine which of the independent variables (timeliness and security) were predictors of strategic benefits as measured by information custodians. The descriptive statistics for these variables are shown in Table 37. Regression results indicate no significant predictors of Strategic Benefits. As such, the null hypothesis $H2b_{null}$ is not rejected.

Table 37.
Descriptive Statistics for H2b Variables.

Variable	M	SD	N
Strategic Benefits	5.33	1.25	22
Timeliness	7.39	2.07	22
Security	7.28	2.17	22

Sub-hypothesis H2c

To evaluate H2c, stepwise multiple regression analysis was conducted to determine which of the independent variables (timeliness and security) were predictors of strategic benefits as measured by information consumers. The descriptive statistics for these variables are shown in Table 38. Regression results indicate no significant predictors of Strategic Benefits. As such, the null hypothesis $H2c_{null}$ is not rejected.

Table 38.
Descriptive Statistics for H2c Variables.

Variable	*M*	*SD*	*N*
Strategic Benefits	4.67	1.18	50
Timeliness	6.70	1.48	50
Security	6.96	2.31	50

Hypothesis 3

H3: Improvements in the usefulness of information will be associated with increased strategic benefits.

The independent variables associated with this hypothesis include appropriate amount, interpretability, objectivity, relevance, and understandability. The dependent variable, strategic benefits, represents the statistical mean of the variables for competitive advantage and customer relations. Of these variables, only one has significant differences between stakeholder roles. To address these differences, the following three sub-hypotheses were also evaluated:

H3a: Improvements in the usefulness of information will be associated with increased strategic benefits as measured by information producers/information collectors.

H3b: Improvements in the usefulness of information will be associated with increased strategic benefits as measured by information custodians.

H3c: Improvements in the usefulness of information will be associated with increased strategic benefits as measured by information consumers.

Table 39.
Descriptive Statistics for H3 Variables.

Variable	*M*	*SD*	*N*
Strategic Benefits	4.78	1.26	101
Appropriate Amount	6.24	1.67	101
Interpretability	6.25	1.77	101
Objectivity	6.74	1.78	101
Relevance	7.63	1.59	101
Understandability	6.45	1.94	101

To evaluate H3, stepwise multiple regression analysis was conducted to determine which of the independent variables (appropriate amount, interpretability, objectivity, relevance, and understandability) were predictors of strategic benefits. The descriptive statistics for these variables are shown in Table 39. Regression results indicate an overall model with one predictor (Appropriate Amount) that significantly predicts Strategic Benefits, $R^2 = .11$, $R^2_{adj} = .10$, $F(1,99) = 12.61$, $p = .001$. This model, which has a tolerance of 1.00, accounted for 10% of the variance in Strategic Benefits. A summary of the regression model is presented in Table 40. The bivariate and partial correlation coefficients between the predictor and the dependent variable are presented in Table 41.

Analysis of the residuals revealed no evidence of violations of the assumptions of linearity, normality, or homoscedasticity, hence the results of the multiple regression analysis are accepted as tenable and the null hypothesis $H3_{null}$ is rejected.

Table 40.
Model Summary for H3.

Predictor	R	R^2	R^2_{adj}	ΔR^2	F_{chg}	p	df_1	df_2
Appropriate Amount	.34	.11	.10	.11	12.61	.001	1	99

Table 41.
Coefficients for H3.

Predictor	B	β	t	Bivariate r	Partial r
Appropriate Amount	.25	.34	3.55	.34	.34

Sub-hypothesis H3a

To evaluate H3a, stepwise multiple regression analysis was conducted to determine which of the independent variables (appropriate amount, interpretability, objectivity, relevance, and understandability) were predictors of strategic benefits as measured by information producers/information collectors. The descriptive statistics for these variables are shown in Table 42. Regression results indicate an overall model with one predictor (Appropriate Amount) that significantly predicts Strategic Benefits,

$R^2 = .14$, $R^2_{adj} = .10$, $F(1,27) = 4.24$, $p = .049$. This model, which has a tolerance of 1.00, accounted for 10% of the variance in Strategic Benefits. A summary of the regression model is presented in Table 43. The bivariate and partial correlation coefficients between the predictor and the dependent variable are presented in Table 44.

Table 42.
Descriptive Statistics for H3a Variables.

Variable	*M*	*SD*	*N*
Strategic Benefits	4.56	1.33	29
Appropriate Amount	6.37	1.74	29
Interpretability	6.57	1.69	29
Objectivity	6.50	1.59	29
Relevance	7.34	1.69	29
Understandability	6.71	2.13	29

Analysis of the residuals revealed no evidence of violations of the assumptions of linearity, normality, or homoscedasticity, hence the results of the multiple regression analysis are accepted as tenable and the null hypothesis $H3a_{null}$ is rejected.

Table 43.
Model Summary for H3a.

Predictor	R	R^2	R^2_{adj}	ΔR^2	F_{chg}	p	df_1	df_2
Appropriate Amount	.37	.14	.10	.14	4.24	.049	1	27

Table 44.
Coefficients for H3a.

Predictor	B	β	*t*	Bivariate *r*	Partial *r*
Appropriate Amount	.28	.37	2.06	.37	.37

Table 45.
Descriptive Statistics for H3b Variables.

Variable	*M*	*SD*	*N*
Strategic Benefits	5.33	1.25	22
Appropriate Amount	6.88	1.73	22
Interpretability	6.68	2.22	22
Objectivity	7.24	1.98	22
Relevance	8.14	1.57	22
Understandability	6.93	2.18	22

Sub-hypothesis H3b

To evaluate H3b, stepwise multiple regression analysis was conducted to determine which of the independent variables (appropriate amount, interpretability, objectivity, relevance, and understandability) were predictors of strategic benefits as measured by information custodians. The descriptive statistics for these variables are shown in Table 45. Regression results indicate an overall model with one predictor (Appropriate Amount) that significantly predicts Strategic Benefits, $R^2 = .22$, $R^2_{adj} = .18$, $F(1,20) = 5.49$, $p = .030$. This model, which has a tolerance of 1.00, accounted for 18% of the variance in Strategic Benefits. A summary of the regression model is presented in

Table 46. The bivariate and partial correlation coefficients between the predictor and the dependent variable are presented in Table 47.

Analysis of the residuals revealed no evidence of violations of the assumptions of linearity, normality, or homoscedasticity, hence the results of the multiple regression analysis are accepted as tenable and the null hypothesis $H3b_{null}$ is rejected.

Table 46.
Model Summary for H3b.

Predictor	R	R^2	R^2_{adj}	ΔR^2	F_{chg}	p	df_1	df_2
Appropriate Amount	.46	.22	.18	.22	5.49	.030	1	20

Table 47.
Coefficients for H3b.

Predictor	B	β	t	Bivariate r	Partial r
Appropriate Amount	.34	.47	2.34	.47	.47

Sub-hypothesis H3c

To evaluate H3c, stepwise multiple regression analysis was conducted to determine which of the independent variables (appropriate amount, interpretability, objectivity, relevance, and understandability) were predictors of strategic benefits as measured by information consumers. The descriptive statistics for these variables are shown in Table 48. Regression results indicate no significant predictors of Strategic Benefits. As such, the null hypothesis $H3c_{null}$ is not rejected.

Table 48.
Descriptive Statistics for H3c Variables.

Variable	*M*	*SD*	*N*
Strategic Benefits	4.67	1.18	50
Timeliness	6.70	1.48	50
Security	6.96	2.31	50

Hypothesis 4

H4: Improvements in the usability of information will be associated with increased strategic benefits.

The independent variables associated with this hypothesis include accessibility, believability, ease of operation, reputation, and value-added. The dependent variable, strategic benefits, represents the statistical mean of the variables for competitive advantage and customer relations. Of these variables, two have significant differences between stakeholder roles. To address these differences, the following three sub-hypotheses were also evaluated:

H4a: Improvements in the usability of information will be associated with increased strategic benefits as measured by information producers/information collectors.

H4b: Improvements in the usability of information will be associated with increased strategic benefits as measured by information custodians.

H4c: Improvements in the usability of information will be associated with increased strategic benefits as measured by information consumers.

To evaluate H4, stepwise multiple regression analysis was conducted to determine which of the independent variables (accessibility, believability, ease of operation, reputation, and value-added) were predictors of strategic benefits. The descriptive statistics for these variables are shown in Table 49. Regression results indicate an overall model with one predictor (Value-Added) that significantly predicts Strategic Benefits, $R^2 = .23$, $R^2_{adj} = .22$, $F(1,99) = 28.93$, $p < .001$. This model, which has a tolerance of 1.00, accounted for 22% of the variance in Strategic Benefits. A summary of the regression model is presented in Table 50. The bivariate and partial correlation coefficients between the predictor and the dependent variable are presented in Table 51.

Analysis of the residuals revealed no evidence of violations of the assumptions of linearity, normality, or homoscedasticity, hence the results of the multiple regression analysis are accepted as tenable and the null hypothesis $H4_{null}$ is rejected.

Table 49.
Descriptive Statistics for H4 Variables.

Variable	*M*	*SD*	*N*
Strategic Benefits	4.78	1.26	101
Accessibility	6.65	1.99	101
Believability*	1.90	.46	101
Ease of Operation	5.79	1.90	101
Reputation	6.80	1.85	101
Value-Added	7.46	1.71	101

* Believability was transformed by reflecting and calculating the square root

Table 50.
Model Summary for H4.

Predictor	R	R^2	R^2_{adj}	ΔR^2	F_{chg}	p	df_1	df_2
Value-Added	.48	.23	.22	.23	28.93	.000	1	99

Table 51.
Coefficients for H4.

Predictor	B	β	t	Bivariate r	Partial r
Value-Added	.35	.48	5.38	.48	.48

Sub-hypothesis H4a

To evaluate H4a, stepwise multiple regression analysis was conducted to determine which of the independent variables (accessibility, believability, ease of operation, reputation, and value-added) were predictors of strategic benefits as measured by information producers/information collectors. The descriptive statistics for these variables are shown in Table 52. Regression results indicate an overall model with one predictor (Value-Added) that significantly predicts Strategic Benefits, $R^2 = .27$, $R^2_{adj} = .24$, $F(1,27) = 9.89$, $p = .004$. This model, which has a tolerance of 1.00, accounted for 24% of the variance in Strategic Benefits. A summary of the regression model is presented in Table 53. The bivariate and partial correlation coefficients between the predictor and the dependent variable are presented in Table 54.

Analysis of the residuals revealed no evidence of violations of the assumptions of linearity, normality, or homoscedasticity, hence the results of the multiple regression analysis are accepted as tenable and the null hypothesis $H4a_{null}$ is rejected.

Table 52.
Descriptive Statistics for H4a Variables.

Variable	*M*	*SD*	*N*
Strategic Benefits	4.56	1.33	29
Accessibility	6.68	2.00	29
Believability*	1.92	.42	29
Ease of Operation	5.97	1.96	29
Reputation	7.03	1.55	29
Value-Added	7.12	1.76	29

* Believability was transformed by reflecting and calculating the square root

Table 53.
Model Summary for H4a.

Predictor	R	R^2	R^2_{adj}	ΔR^2	F_{chg}	p	df_1	df_2
Value-Added	.52	.27	.24	.27	9.89	.004	1	27

Table 54.
Coefficients for H4a.

Predictor	B	β	t	Bivariate r	Partial r
Value-Added	.39	.52	3.14	.52	.52

Sub-hypothesis H4b

To evaluate H4b, stepwise multiple regression analysis was conducted to determine which of the independent variables (accessibility, believability, ease of operation, reputation, and value-added) were predictors of strategic benefits as measured by information custodians. The descriptive statistics for these variables are shown in Table 55. Regression results indicate an overall model with one predictor (Value-Added) that significantly predicts Strategic Benefits, $R^2 = .29$, $R^2_{adj} = .26$, $F(1,20) = 8.29$, $p = .009$. This model, which has a tolerance of 1.00, accounted for 26% of the variance in Strategic Benefits. A summary of the regression model is presented in Table 56. The bivariate and partial correlation coefficients between the predictor and the dependent variable are presented in Table 57.

Table 55.
Descriptive Statistics for H4b Variables.

Variable	*M*	*SD*	*N*
Strategic Benefits	5.33	1.25	22
Accessibility	7.34	2.11	22
Believability*	1.73	.52	22
Ease of Operation	6.20	1.96	22
Reputation	7.32	2.27	22
Value-Added	8.14	1.51	22

* Believability was transformed by reflecting and calculating the square root

Analysis of the residuals revealed no evidence of violations of the assumptions of linearity, normality, or homoscedasticity, hence the results of the multiple regression analysis are accepted as tenable and the null hypothesis $H4b_{null}$ is rejected.

Table 56.
Model Summary for H4b.

Predictor	*R*	R^2	R^2_{adj}	ΔR^2	F_{chg}	*p*	df_1	df_2
Value-Added	.54	.29	.26	.29	8.29	.009	1	20

Table 57.
Coefficients for H4b.

Predictor	B	β	*t*	Bivariate *r*	Partial *r*
Value-Added	.45	.54	2.88	.54	.54

Sub-hypothesis H4c

To evaluate H4c, stepwise multiple regression analysis was conducted to determine which of the independent variables (accessibility, believability, ease of operation, reputation, and value-added) were predictors of strategic benefits as measured by information consumers. The descriptive statistics for these variables are shown in Table 58. Regression results indicate an overall model with one predictor (Value-Added) that significantly predicts Strategic Benefits, $R^2 = .13$, $R^2_{adj} = .12$, $F(1,20) = 7.44$, $p = .009$. This model, which has a tolerance of 1.00, accounted for 12% of the variance in Strategic Benefits. A summary of the regression model is presented in Table 59. The bivariate and partial correlation coefficients between the predictor and the dependent variable are presented in Table 60.

Table 58.
Descriptive Statistics for H4c Variables.

Variable	*M*	*SD*	*N*
Strategic Benefits	4.67	1.18	50
Accessibility	6.34	1.89	50
Believability*	1.97	.45	50
Ease of Operation	5.50	1.83	50
Reputation	6.44	1.78	50
Value-Added	7.36	1.71	50

* Believability was transformed by reflecting and calculating the square root

Table 59.
Model Summary for H4c.

Predictor	R	R^2	R^2_{adj}	ΔR^2	F_{chg}	p	df_1	df_2
Value-Added	.37	.13	.12	.13	7.44	.009	1	20

Analysis of the residuals revealed no evidence of violations of the assumptions of linearity, normality, or homoscedasticity, hence the results of the multiple regression analysis are accepted as tenable and the null hypothesis $H4c_{null}$ is rejected.

Table 60.
Coefficients for H4c.

Predictor	B	β	*t*	Bivariate *r*	Partial *r*
Value-Added	.25	.37	2.72	.37	.37

Hypothesis 5

H5: Improvements in information quality will be associated with increased strategic benefits.

The independent variables associated with this hypothesis include soundness, dependability, usefulness, and usability. The dependent variable, strategic benefits, represents the statistical mean of the variables for competitive advantage and customer relations. Of these variables, two have significant differences between stakeholder roles. As such, the following three sub-hypotheses were also evaluated:

H5a: Improvements in information quality will be associated with increased strategic benefits as measured by information producers/information collectors.

H5b: Improvements in information quality will be associated with increased strategic benefits as measured by information custodians.

H5c: Improvements in information quality will be associated with increased strategic benefits as measured by information consumers.

To evaluate H5, stepwise multiple regression analysis was conducted to determine which of the independent variables (soundness, dependability, usefulness, and usability) were predictors of strategic benefits. The descriptive statistics for these variables are shown in Table 61. Regression results indicate an overall model with one predictor (Usability) that significantly predicts Strategic Benefits, $R^2 = .16$, $R^2_{adj} = .15$, $F(1,99) = 18.36$, $p < .001$. This model, which has a tolerance of 1.00, accounted for 15% of the variance in Strategic Benefits. A summary of the regression model is presented in Table

62. The bivariate and partial correlation coefficients between the predictor and the dependent variable are presented in Table 63.

Table 61.
Descriptive Statistics for H5 Variables.

Variable	*M*	*SD*	*N*
Strategic Benefits	4.78	1.26	101
Soundness	6.38	1.75	101
Dependability	6.99	1.67	101
Usefulness	6.66	1.45	101
Usability	6.77	1.54	101

Table 62.
Model Summary for H5.

Predictor	R	R^2	R^2_{adj}	ΔR^2	F_{chg}	p	df_1	df_2
Usability	.40	.16	.15	.16	18.36	.000	1	99

Analysis of the residuals revealed no evidence of violations of the assumptions of linearity, normality, or homoscedasticity, hence the results of the multiple regression analysis are accepted as tenable and the null hypothesis $H5_{null}$ is rejected.

Table 63.
Coefficients for H5.

Predictor	B	β	*t*	Bivariate *r*	Partial *r*
Usability	.32	.40	4.29	.40	.40

Sub-hypothesis H5a

To evaluate H5a, stepwise multiple regression analysis was conducted to determine which of the independent variables (soundness, dependability, usefulness, and usability) were predictors of strategic benefits as measured by information producers/information collectors. The descriptive statistics for these variables are shown in Table 64. Regression results indicate an overall model with one predictor (Dependability) that significantly predicts Strategic Benefits, $R^2 = .16$, $R^2_{adj} = .13$, $F(1,27) = 5.28$, $p = .030$. This model, which has a tolerance of 1.00, accounted for 13% of the variance in Strategic Benefits. A summary of the regression model is presented in Table 65. The bivariate and partial correlation coefficients between the predictor and the dependent variable are presented in Table 66.

Table 64.
Descriptive Statistics for H5a Variables.

Variable	*M*	*SD*	*N*
Strategic Benefits	4.56	1.33	29
Soundness	6.51	1.69	29
Dependability	6.98	1.73	29
Usefulness	6.70	1.53	29
Usability	6.79	1.54	29

Analysis of the residuals revealed no evidence of violations of the assumptions of linearity or normality, but there is slight evidence that heteroscedasticity may be a problem. However, Garson (2006b) notes that moderation violations of the assumption of homoscedasticity have only a minor impact on the regression estimates. As such, the results of this multiple regression analysis are accepted as tenable and the null hypothesis $H5a_{null}$ is rejected.

Table 65.
Model Summary for H5a.

Predictor	R	R^2	R^2_{adj}	ΔR^2	F_{chg}	p	df_1	df_2
Dependability	.40	.16	.13	.16	5.28	.030	1	27

Table 66.
Coefficients for H5a.

Predictor	B	β	t	Bivariate r	Partial r
Dependability	.31	.40	2.30	.40	.40

Sub-hypothesis H5b

To evaluate H5b, stepwise multiple regression analysis was conducted to determine which of the independent variables (soundness, dependability, usefulness, and usability) were predictors of strategic benefits as measured by information custodians. The descriptive statistics for these variables are shown in Table 67. Regression results indicate an overall model with one predictor (Soundness) that significantly predicts Strategic Benefits, R^2 = .27, R^2_{adj} = .23, F(1,20) = 7.24, p = .014. This model, which has a tolerance of 1.00, accounted for 23% of the variance in Strategic Benefits. A summary of the regression model is presented in Table 68. The bivariate and partial correlation coefficients between the predictor and the dependent variable are presented in Table 69.

Table 67.
Descriptive Statistics for H5b Variables.

Variable	*M*	*SD*	*N*
Strategic Benefits	5.33	1.25	22
Soundness	7.13	1.88	22
Dependability	7.34	1.92	22
Usefulness	7.17	1.66	22
Usability	7.35	1.63	22

Analysis of the residuals revealed no evidence of violations of the assumptions of linearity or normality, but there is slight evidence that homoscedasticity may be a problem. However, Garson (2006b) notes that moderation violations of the assumption of homoscedasticity have only a minor impact on the regression estimates. As such, the results of this multiple regression analysis are accepted as tenable and the null hypothesis $H5b_{null}$ is rejected.

Table 68.
Model Summary for H5b.

Predictor	R	R^2	R^2_{adj}	ΔR^2	F_{chg}	p	df_1	df_2
Soundness	.52	.27	.23	.27	7.24	.014	1	20

Table 69.
Coefficients for H5b.

Predictor	B	β	*t*	Bivariate *r*	Partial *r*
Soundness	.34	.52	2.69	.52	.52

Sub-hypothesis H5c

To evaluate H5c, stepwise multiple regression analysis was conducted to determine which of the independent variables (soundness, dependability, usefulness, and usability) were predictors of strategic benefits as measured by information consumers. The descriptive statistics for these variables are shown in Table 70. Regression results indicate two predictive models.

Model 1, which has a tolerance of 1.00, indicates Usability as a significant predictor of Strategic Benefits, $R^2 = .09$, $R^2_{adj} = .07$, $F(1,48) = 6.98$, $p = .002$. This model accounted for 7% of the variance in Strategic Benefits. A summary of the regression model is presented in Table 71. The bivariate and partial correlation coefficients between the predictor and the dependent variable are presented in Table 72.

Model 2, which has a tolerance of .26, indicates Usability and Soundness as predictors of Strategic Benefits, $R^2 = .23$, $R^2_{adj} = .20$, $F(1,47) = 8.77$, $p = .005$. This model accounted for 20% of the variance in Strategic Benefits. A summary of the regression model is presented in Table 73. The bivariate and partial correlation coefficients between the predictor and the dependent variable are presented in Table 74.

Table 70.
Descriptive Statistics for H5c Variables.

Variable	*M*	*SD*	*N*
Strategic Benefits	4.67	1.18	50
Soundness	5.97	1.63	50
Dependability	6.83	1.53	50
Usefulness	6.41	1.27	50
Usability	6.52	1.47	50

Table 71.
Model Summary for H5c Model 1.

Predictor	R	R^2	R^2_{adj}	ΔR^2	F_{chg}	p	df_1	df_2
Usability	.29	.09	.07	.09	4.47	.040	1	48

Table 72.
Coefficients for H5c Model 1.

Predictor	B	β	*t*	Bivariate *r*	Partial *r*
Usability	.23	.29	2.11	.29	.29

Table 73.
Model Summary for H5c Model 2.

Predictors	R	R^2	R^2_{adj}	ΔR^2	F_{chg}	p	df_1	df_2
Usability and Soundness	.48	.23	.20	.14	8.77	.005	1	47

Table 74.
Coefficients for H5c Model 2.

Predictor	B	β	*t*	Bivariate *r*	Partial *r*
Usability	.71	.89	3.72	.29	.48
Soundness	-.51	-.71	-2.96	.04	-.40

Analysis of the residuals revealed no evidence of violations of the assumptions of linearity or normality, but there is slight evidence that homoscedasticity may be a problem. However, Garson (2006b) notes that moderation violations of the assumption of homoscedasticity have only a minor impact on the regression estimates. As such, the results of this multiple regression analysis are accepted as tenable and the null hypothesis $H5c_{null}$ is rejected.

Hypothesis 6

H6: Improvements in the soundness of information will be associated with increased transactional benefits.

The independent variables associated with this hypothesis include completeness, concise representation, consistent representation, and freedom from error. The dependent variable, transactional benefits, represents the statistical mean of the variables for business efficiency and systems development efficiency. Of these variables, two have significant differences between stakeholder roles. As such, the following three sub-hypotheses were also evaluated:

H6a: Improvements in the soundness of information will be associated with increased transactional benefits as measured by information producers/information collectors.

H6b: Improvements in the soundness of information will be associated with increased transactional benefits as measured by information custodians.

H6c: Improvements in the soundness of information will be associated with increased transactional benefits as measured by information consumers.

To evaluate H6, stepwise multiple regression analysis was conducted to determine which of the independent variables (completeness, concise representation, consistent representation, and freedom from error) were predictors of strategic benefits. The descriptive statistics for these variables are shown in Table 75. Regression results indicate two predictive models.

Model 1, which has a tolerance of 1.00, indicates Completeness as a significant predictor of Transactional Benefits, $R^2 = .14$, $R^2_{adj} = .13$, $F(1,99) = 15.40$, $p < .001$. This model accounted for 13% of the variance in Transactional Benefits. A summary of the regression model is presented in Table 76. The bivariate and partial correlation coefficients between the predictor and the dependent variable are presented in Table 77.

Model 2, which has a tolerance of .40, indicates Completeness and Consistent Representation as predictors of Transactional Benefits, $R^2 = .23$, $R^2_{adj} = .22$, $F(1,98) = 15.00$, $p < .001$. This model accounted for 22% of the variance in Transactional Benefits. A summary of the regression model is presented in Table 78. The bivariate and partial

correlation coefficients between the predictor and the dependent variable are presented in Table 79.

Table 75.
Descriptive Statistics for H6 Variables.

Variable	*M*	*SD*	*N*
Transactional Benefits	4.41	1.19	101
Completeness	6.17	1.91	101
Concise Representation	5.80	1.93	101
Consistent Representation	6.56	2.08	101
Free of Error	6.99	1.86	101

Table 76.
Model Summary for H6 Model 1.

Predictor	R	R^2	R^2_{adj}	ΔR^2	F_{chg}	p	df_1	df_2
Completeness	.37	.14	.13	.14	15.40	.000	1	99

Table 77.
Coefficients for H6 Model 1.

Predictor	B	β	t	Bivariate r	Partial r
Completeness	.23	.37	3.92	.37	.37

Analysis of the residuals revealed no evidence of violations of the assumptions of linearity, normality, or homoscedasticity, hence the results of the multiple regression analysis are accepted as tenable and the null hypothesis $H6_{null}$ is rejected.

Table 78.
Model Summary for H6 Model 2.

Predictors	R	R^2	R^2_{adj}	ΔR^2	F_{chg}	p	df_1	df_2
Completeness and Consistent Representation	.48	.23	.22	.10	12.77	.001	1	98

Table 79.
Coefficients for H6 Model 2.

Predictor	B	β	*t*	Bivariate *r*	Partial *r*
Completeness	.47	.75	5.40	.37	.48
Consistent Representation	-.29	-.50	-3.57	.08	-.34

Sub-hypothesis H6a

To evaluate H6a, stepwise multiple regression analysis was conducted to determine which of the independent variables (completeness, concise representation, consistent representation, and freedom from error) were predictors of transactional benefits as measured by information producers/information collectors. The descriptive statistics for these variables are shown in Table 80. Regression results indicate an overall model with one predictor (Free of Error) that significantly predicts Transactional Benefits, $R^2 = .19$, $R^2_{adj} = .16$, $F(1,27) = 6.46$, $p = .017$. This model, which has a tolerance of 1.00, accounted for 16% of the variance in Transactional Benefits. A summary of the regression model is presented in Table 81. The bivariate and partial

correlation coefficients between the predictor and the dependent variable are presented in Table 82.

Table 80.
Descriptive Statistics for H6a Variables.

Variable	*M*	*SD*	*N*
Transactional Benefits	4.34	1.35	29
Completeness	6.18	2.07	29
Concise Representation	5.94	1.95	29
Consistent Representation	6.93	1.80	29
Free of Error	7.00	1.83	29

Table 81.
Model Summary for H6a.

Predictor	R	R^2	R^2_{adj}	ΔR^2	F_{chg}	p	df_1	df_2
Free of Error	.44	.19	.16	.19	6.46	.017	1	27

Table 82.
Coefficients for H6a.

Predictor	B	β	*t*	Bivariate *r*	Partial *r*
Free of Error	.32	.44	2.54	.44	.44

Analysis of the residuals revealed no evidence of violations of the assumptions of linearity, normality, or homoscedasticity, hence the results of the multiple regression analysis are accepted as tenable and the null hypothesis $H6a_{null}$ is rejected.

Sub-hypothesis H6b

To evaluate H6b, stepwise multiple regression analysis was conducted to determine which of the independent variables (completeness, concise representation, consistent representation, and freedom from error) were predictors of transactional benefits as measured by information custodians. The descriptive statistics for these variables are shown in Table 83. Regression results indicate two predictive models.

Table 83.
Descriptive Statistics for H6b Variables.

Variable	*M*	*SD*	*N*
Transactional Benefits	4.73	1.16	22
Completeness	6.90	2.04	22
Concise Representation	6.59	2.14	22
Consistent Representation	7.39	2.16	22
Free of Error	7.64	1.97	22

Table 84.
Model Summary for H6b Model 1.

Predictor	R	R^2	R^2_{adj}	ΔR^2	F_{chg}	p	df_1	df_2
Completeness	.42	.18	.14	.18	4.40	.049	1	20

Table 85.
Coefficients for H6b Model 1.

Predictor	B	β	*t*	Bivariate *r*	Partial *r*
Completeness	.24	.42	2.10	.42	.42

Model 1, which has a tolerance of 1.00, indicates Completeness as a significant predictor of Transactional Benefits, $R^2 = .18$, $R^2_{adj} = .14$, $F(1,99) = 4.40$, $p = .049$. This model accounted for 14% of the variance in Transactional Benefits. A summary of the regression model is presented in Table 84. The bivariate and partial correlation coefficients between the predictor and the dependent variable are presented in Table 85.

Model 2, which has a tolerance of .25, indicates Completeness and Consistent Representation as predictors of Transactional Benefits, $R^2 = .35$, $R^2_{adj} = .28$, $F(1,98) = 5.10$, $p < .017$. This model accounted for 28% of the variance in Transactional Benefits. A summary of the regression model is presented in Table 86. The bivariate and partial correlation coefficients between the predictor and the dependent variable are presented in Table 87.

Table 86.
Model Summary for H6b Model 2.

Predictors	R	R^2	R^2_{adj}	ΔR^2	F_{chg}	p	df_1	df_2
Completeness and Consistent Representation	.59	.35	.28	.17	4.93	.039	1	98

Table 87.
Coefficients for H6b Model 2.

Predictor	B	β	t	Bivariate r	Partial r
Completeness	.64	1.13	3.07	.42	.58
Consistent Representation	-.44	-.82	-2.22	.16	-.45

Analysis of the residuals revealed no evidence of violations of the assumptions of linearity, normality, or homoscedasticity, hence the results of this multiple regression analysis are accepted as tenable and the null hypothesis $H6b_{null}$ is rejected.

Sub-hypothesis H6c

To evaluate H6c, stepwise multiple regression analysis was conducted to determine which of the independent variables (completeness, concise representation, consistent representation, and freedom from error) were predictors of transactional benefits as measured by information consumers. The descriptive statistics for these variables are shown in Table 88. Regression results indicate no significant predictors of Strategic Benefits. As such, the null hypothesis $H6c_{null}$ is not rejected.

Table 88.
Descriptive Statistics for H6c Variables.

Variable	*M*	*SD*	*N*
Transactional Benefits	4.30	1.11	50
Completeness	5.84	1.70	50
Concise Representation	5.38	1.71	50
Consistent Representation	5.98	2.06	50
Free of Error	6.70	1.80	50

Hypothesis 7

H7: Improvements in the dependability of information will be associated with increased transactional benefits.

The independent variables associated with this hypothesis include timeliness and security. The dependent variable, transactional benefits, represents the statistical mean of the variables for business efficiency and systems development efficiency. Of these variables, none have significant differences between stakeholder roles, hence no sub-hypotheses were evaluated.

To evaluate H7, stepwise multiple regression analysis was conducted to determine which of the independent variables (timeliness and security) were predictors of transactional benefits. The descriptive statistics for these variables are shown in Table 89. Regression results indicate an overall model with one predictor (Timeliness) that significantly predicts Transactional Benefits, $R^2 = .11$, $R^2_{adj} = .10$, $F(1,99) = 11.59$, $p = .001$. This model, which has a tolerance of 1.00, accounted for 10% of the variance in Transactional Benefits. A summary of the regression model is presented in Table 90. The bivariate and partial correlation coefficients between the predictor and the dependent variable are presented in Table 91.

Analysis of the residuals revealed no evidence of violations of the assumptions of linearity or normality, but there is slight evidence that heteroscedasticity may be a problem. However, Garson (2006b) notes that moderation violations of the assumption of homoscedasticity have only a minor impact on the regression estimates. As such, the results of this multiple regression analysis are accepted as tenable and the null hypothesis $H7_{null}$ is rejected.

Table 89.
Descriptive Statistics for H7 Variables.

Variable	*M*	*SD*	*N*
Transactional Benefits	4.41	1.19	101
Timeliness	6.90	1.75	101
Security	7.08	2.15	101

Table 90.
Model Summary for H7.

Predictor	R	R^2	R^2_{adj}	ΔR^2	F_{chg}	p	df_1	df_2
Timeliness	.32	.11	.10	.11	11.59	.001	1	99

Table 91.
Coefficients for H7.

Predictor	B	β	*t*	Bivariate *r*	Partial *r*
Timeliness	.22	.32	3.40	.32	.32

Hypothesis 8

H8: Improvements in the usefulness of information will be associated with increased transactional benefits.

The independent variables associated with this hypothesis include appropriate amount, interpretability, objectivity, relevance, and understandability. The dependent variable, transactional benefits, represents the statistical mean of the variables for business efficiency and systems development efficiency. Of these variables, none have significant differences between stakeholder roles, hence no sub-hypotheses were evaluated.

To evaluate H8, stepwise multiple regression analysis was conducted to determine which of the independent variables (appropriate amount, interpretability, objectivity, relevance, and understandability) were predictors of transactional benefits. The descriptive statistics for these variables are shown in Table 92. Regression results indicate an overall model with one predictor (Appropriate Amount) that significantly predicts Transactional Benefits, $R^2 = .09$, $R^2_{adj} = .08$, $F(1,99) = 9.64$, $p = .002$. This model, which has a tolerance of 1.00, accounted for 8% of the variance in Transactional Benefits. A summary of the regression model is presented in Table 93. The bivariate and partial correlation coefficients between the predictor and the dependent variable are presented in Table 94.

Table 92.
Descriptive Statistics for H8 Variables.

Variable	*M*	*SD*	*N*
Transactional Benefits	4.41	1.19	101
Appropriate Amount	6.24	1.67	101
Interpretability	6.25	1.77	101
Objectivity	6.74	1.76	101
Relevance	7.63	1.59	101
Understandability	6.45	1.94	101

Analysis of the residuals revealed no evidence of violations of the assumptions of linearity, normality, or homoscedasticity, hence the results of this multiple regression analysis are accepted as tenable and the null hypothesis $H8_{null}$ is rejected.

Table 93.
Model Summary for H8.

Predictor	R	R^2	R^2_{adj}	ΔR^2	F_{chg}	p	df_1	df_2
Appropriate Amount	.30	.09	.08	.09	9.64	.002	1	99

Table 94.
Coefficients for H8.

Predictor	B	β	t	Bivariate r	Partial r
Appropriate Amount	.21	.30	3.11	.30	.30

Hypothesis 9

H9: Improvements in the usability of information will be associated with increased transactional benefits.

The independent variables associated with this hypothesis include accessibility, believability, ease of operation, reputation, and value-added. The dependent variable, transactional benefits, represents the statistical mean of the variables for business efficiency and systems development efficiency. Of these variables, one has significant differences between stakeholder roles. To address these differences, the following three sub-hypotheses were also evaluated:

H9a: Improvements in the usability of information will be associated with increased transactional benefits as measured by information producers/information collectors.

H9b: Improvements in the usability of information will be associated with increased transactional benefits as measured by information custodians.

H9c: Improvements in the usability of information will be associated with increased transactional benefits as measured by information consumers.

To evaluate H9, stepwise multiple regression analysis was conducted to determine which of the independent variables (accessibility, believability, ease of operation, reputation, and value-added) were predictors of transactional benefits. The descriptive statistics for these variables are shown in Table 95. Regression results indicate two predictive models.

Model 1, which has a tolerance of 1.00, indicates Value-Added as a significant predictor of Transactional Benefits, $R^2 = .15$, $R^2_{adj} = .14$, $F(1,99) = 17.56$, $p < .001$. This model accounted for 14% of the variance in Transactional Benefits. A summary of the regression model is presented in Table 96. The bivariate and partial correlation coefficients between the predictor and the dependent variable are presented in Table 97.

Model 2, which has a tolerance of .75, indicates Value-Added and Ease of Operation as predictors of Transactional Benefits, $R^2 = .19$, $R^2_{adj} = .17$, $F(1,98) = 11.22$, $p < .001$. This model accounted for 17% of the variance in Transactional Benefits. A summary of the regression model is presented in Table 98. The bivariate and partial

correlation coefficients between the predictor and the dependent variable are presented in Table 99.

Analysis of the residuals revealed no evidence of violations of the assumptions of linearity, normality, or homoscedasticity, hence the results of this multiple regression analysis are accepted as tenable and the null hypothesis $H9_{null}$ is rejected.

Table 95.
Descriptive Statistics for H9 Variables.

Variable	*M*	*SD*	*N*
Transactional Benefits	4.41	1.19	101
Accessibility	6.65	1.99	101
Believability*	1.90	.46	101
Ease of Operation	5.79	1.90	101
Reputation	6.80	1.85	101
Value-Added	7.46	1.71	101

* Believability was transformed by reflecting and calculating the square root

Table 96.
Model Summary for H9 Model 1.

Predictor	R	R^2	R^2_{adj}	ΔR^2	F_{chg}	p	df_1	df_2
Value-added	.39	.15	.14	.15	17.56	.000	1	99

Table 97.
Coefficients for H9 Model 1.

Predictor	B	β	*t*	Bivariate *r*	Partial *r*
Value-added	.27	.39	4.19	.39	.39

Table 98.
Model Summary for H9 Model 2.

Predictors	R	R^2	R^2_{adj}	ΔR^2	F_{chg}	p	df_1	df_2
Value-added and Ease of Operation	.43	.19	.17	.04	4.29	.041	1	98

Table 99.
Coefficients for H9 Model 2.

Predictor	B	β	*t*	Bivariate *r*	Partial *r*
Value-Added	.20	.28	2.65	.39	.26
Ease of Operation	.14	.22	2.07	.36	.21

Sub-hypothesis H9a

To evaluate H9a, stepwise multiple regression analysis was conducted to determine which of the independent variables (accessibility, believability, ease of operation, reputation, and value-added) were predictors of transactional benefits as measured by information producers/information collectors. The descriptive statistics for these variables are shown in Table 100. Regression results indicate an overall model with one predictor (Value-Added) that significantly predicts Transactional Benefits, $R^2 = .19$,

R^2_{adj} = .16, $F(1,27) = 6.21$, p = .019. This model, which has a tolerance of 1.00, accounted for 16% of the variance in Transactional Benefits. A summary of the regression model is presented in Table 101. The bivariate and partial correlation coefficients between the predictor and the dependent variable are presented in Table 102.

Analysis of the residuals revealed no evidence of violations of the assumptions of linearity, normality, or homoscedasticity, hence the results of the multiple regression analysis are accepted as tenable and the null hypothesis $H9a_{null}$ is rejected.

Table 100.
Descriptive Statistics for H9a Variables.

Variable	*M*	*SD*	*N*
Transactional Benefits	4.34	1.35	29
Accessibility	6.68	2.00	29
Believability*	1.92	.42	29
Ease of Operation	5.97	1.96	29
Reputation	7.03	1.55	29
Value-Added	7.12	1.76	29

* Believability was transformed by reflecting and calculating the square root

Table 101.
Model Summary for H9a.

Predictor	R	R^2	R^2_{adj}	ΔR^2	F_{chg}	p	df_1	df_2
Value-Added	.43	.19	.16	.19	6.21	.019	1	27

Table 102.
Coefficients for H9a.

Predictor	B	β	*t*	Bivariate *r*	Partial *r*
Value-Added	.33	.43	2.49	.43	.43

Sub-hypothesis H9b

To evaluate H9b, stepwise multiple regression analysis was conducted to determine which of the independent variables (accessibility, believability, ease of operation, reputation, and value-added) were predictors of transacational benefits as measured by information custodians. The descriptive statistics for these variables are shown in Table 103. Regression results indicate no significant predictors of Strategic Benefits. As such, the null hypothesis $H9b_{null}$ is not rejected.

Table 103.
Descriptive statistics for H9b variables.

Variable	*M*	*SD*	*N*
Transactional Benefits	4.73	1.16	22
Accessibility	7.34	2.11	22
Believability*	1.73	.52	22
Ease of Operation	6.20	1.96	22
Reputation	7.32	2.27	22
Value-Added	8.14	1.51	22

* Believability was transformed by reflecting and calculating the square root

Sub-hypothesis H9c

To evaluate H9c, stepwise multiple regression analysis was conducted to determine which of the independent variables (accessibility, believability, ease of operation, reputation, and value-added) were predictors of transactional benefits as measured by information consumers. The descriptive statistics for these variables are shown in Table 104. Regression results indicate an overall model with one predictor (Value-Added) that significantly predicts Transactional Benefits, $R^2 = .14$, $R^2_{adj} = .12$, $F(1,20) = 7.54$, $p = .008$. This model, which has a tolerance of 1.00, accounted for 12% of the variance in Transacational Benefits. A summary of the regression model is presented in Table 105. The bivariate and partial correlation coefficients between the predictor and the dependent variable are presented in Table 106.

Table 104.
Descriptive Statistics for H9c Variables.

Variable	*M*	*SD*	*N*
Transactional Benefits	4.30	1.11	50
Accessibility	6.34	1.89	50
Believability*	1.97	.45	50
Ease of Operation	5.50	1.83	50
Reputation	6.44	1.78	50
Value-Added	7.36	1.71	50

* Believability was transformed by reflecting and calculating the square root

Table 105.
Model Summary for H9c.

Predictor	*R*	R^2	R^2_{adj}	ΔR^2	F_{chg}	*p*	df_1	df_2
Value-Added	.37	.14	.12	.14	7.54	.008	1	20

Table 106.
Coefficients for H9c.

Predictor	B	β	*t*	Bivariate *r*	Partial *r*
Value-Added	.24	.37	2.75	.37	.37

Analysis of the residuals revealed no evidence of violations of the assumptions of linearity, normality, or homoscedasticity, hence the results of the multiple regression analysis are accepted as tenable and the null hypothesis $H9c_{null}$ is rejected.

Hypothesis 10

H10: Improvements in information quality will be associated with increased transactional benefits.

The independent variables associated with this hypothesis include soundness, dependability, usefulness, and usability. The dependent variable, transactional benefits, represents the statistical mean of the variables for business efficiency and systems development efficiency. Of these variables, one has significant differences between stakeholder roles. As such, the following three sub-hypotheses were also evaluated:

H10a: Improvements in information quality will be associated with increased transactional benefits as measured by information producers/information collectors.

H10b: Improvements in information quality will be associated with increased transactional benefits as measured by information custodians.

H10c: Improvements in information quality will be associated with increased transactional benefits as measured by information consumers.

To evaluate H10, stepwise multiple regression analysis was conducted to determine which of the independent variables (soundness, dependability, usefulness, and usability) were predictors of transactional benefits. The descriptive statistics for these variables are shown in Table 107. Regression results indicate two predictive models.

Model 1, which has a tolerance of 1.00, indicates Usability as a significant predictor of Transactional Benefits, $R^2 = .14$, $R^2_{adj} = .13$, $F(1,99) = 15.97$, $p < .001$. This model accounted for 13% of the variance in Transactional Benefits. A summary of the regression model is presented in Table 108. The bivariate and partial correlation coefficients between the predictor and the dependent variable are presented in Table 109.

Model 2, which has a tolerance of .19, indicates Usability and Usefulness as predictors of Transactional Benefits, $R^2 = .19$, $R^2_{adj} = .18$, $F(1,98) = 11.81$, $p < .001$. This model accounted for 18% of the variance in Transactional Benefits. A summary of the regression model is presented in Table 110. The bivariate and partial correlation coefficients between the predictor and the dependent variable are presented in Table 111.

Analysis of the residuals revealed no evidence of violations of the assumptions of linearity, normality, or homoscedasticity, hence the results of this multiple regression analysis are accepted as tenable and the null hypothesis $H10_{null}$ is rejected.

Table 107.
Descriptive Statistics for H10 Variables.

Variable	*M*	*SD*	*N*
Transactional Benefits	4.41	1.19	101
Soundness	6.38	1.75	101
Dependability	6.99	1.67	101
Usefulness	6.66	1.45	101
Usability	6.77	1.54	101

Table 108.
Model Summary for H10 Model 1.

Predictor	R	R^2	R^2_{adj}	ΔR^2	F_{chg}	p	df_1	df_2
Usability	.37	.14	.13	.14	19.78	.000	1	99

Table 109.
Coefficients for H10 Model 1.

Predictor	B	β	*t*	Bivariate *r*	Partial *r*
Value-added	.29	.37	4.00	.37	.37

Table 110.
Model Summary for H10 Model 2.

Predictors	R	R^2	R^2_{adj}	ΔR^2	F_{chg}	p	df_1	df_2
Usability and Usefulness	.44	.19	.18	.06	6.73	.011	1	98

Table 111.
Coefficients for H10 Model 2.

Predictor	B	β	*t*	Bivariate *r*	Partial *r*
Usability	.66	.86	4.13	.37	.39
Usefulness	-.44	-5.40	-2.60	.23	-.26

Sub-hypothesis H10a

To evaluate H10a, stepwise multiple regression analysis was conducted to determine which of the independent variables (soundness, dependability, usefulness, and usability) were predictors of transacational benefits as measured by information producers/information collectors. The descriptive statistics for these variables are shown in Table 112. Regression results indicate an overall model with one predictor (Usability) that significantly predicts Transactional Benefits, $R^2 = .15$, $R^2_{adj} = .12$, $F(1,27) = 4.76$, $p = .038$. This model, which has a tolerance of 1.00, accounted for 12% of the variance in Transactional Benefits. A summary of the regression model is presented in Table 113. The bivariate and partial correlation coefficients between the predictor and the dependent variable are presented in Table 114.

Analysis of the residuals revealed no evidence of violations of the assumptions of linearity, normality, or homoscedasticity, hence the results of the multiple regression analysis are accepted as tenable and the null hypothesis $H10a_{null}$ is rejected.

Table 112.
Descriptive Statistics for H10a Variables.

Variable	*M*	*SD*	*N*
Transactional Benefits	4.34	1.35	29
Soundness	6.51	1.69	29
Dependability	6.98	1.73	29
Usefulness	6.70	1.53	29
Usability	6.79	1.54	29

Table 113.
Model Summary for H10a.

Predictor	R	R^2	R^2_{adj}	ΔR^2	F_{chg}	p	df_1	df_2
Usability	.39	.15	.12	.15	4.76	.038	1	27

Table 114.
Coefficients for H10a.

Predictor	B	β	*t*	Bivariate *r*	Partial *r*
Usability	.34	.39	2.18	.39	.39

Sub-hypothesis H10b

To evaluate H10b, stepwise multiple regression analysis was conducted to determine which of the independent variables (soundness, dependability, usefulness, and usability) were predictors of transactional benefits as measured by information custodians. The descriptive statistics for these variables are shown in Table 115.

Regression results indicate an overall model with one predictor (Usability) that significantly predicts Transactional Benefits, $R^2 = .15$, $R^2_{adj} = .12$, $F(1,27) = 4.76$, $p = .038$. This model, which has a tolerance of 1.00, accounted for 12% of the variance in Transactional Benefits. A summary of the regression model is presented in Table 116. The bivariate and partial correlation coefficients between the predictor and the dependent variable are presented in Table 117.

Table 115.
Descriptive Statistics for H10b Variables.

Variable	*M*	*SD*	*N*
Transactional Benefits	4.73	1.16	22
Soundness	7.13	1.88	22
Dependability	7.34	1.92	22
Usefulness	7.17	1.66	22
Usability	7.35	1.63	22

Table 116.
Model Summary for H10b.

Predictor	R	R^2	R^2_{adj}	ΔR^2	F_{chg}	p	df_1	df_2
Dependability	.49	.24	.20	.24	6.27	.021	1	20

Table 117.
Coefficients for H10b.

Predictor	B	β	*t*	Bivariate *r*	Partial *r*
Dependability	.30	.49	2.51	.49	.49

Analysis of the residuals revealed no evidence of violations of the assumptions of linearity, normality, or homoscedasticity, hence the results of the multiple regression analysis are accepted as tenable and the null hypothesis $H10b_{null}$ is rejected.

Sub-hypothesis H10c

To evaluate H10c, stepwise multiple regression analysis was conducted to determine which of the independent variables (soundness, dependability, usefulness, and usability) were predictors of transactional benefits as measured by information consumers. The descriptive statistics for these variables are shown in Table 118. Regression results indicate an overall model with one predictor (Usability) that significantly predicts Transactional Benefits, $R^2 = .13$, $R^2_{adj} = .12$, $F(1,48) = 7.42$, $p = .009$. This model, which has a tolerance of 1.00, accounted for 12% of the variance in Transactional Benefits. A summary of the regression model is presented in Table 119. The bivariate and partial correlation coefficients between the predictor and the dependent variable are presented in Table 120.

Table 118.
Descriptive Statistics for H10c Variables.

Variable	*M*	*SD*	*N*
Transactional Benefits	4.30	1.11	50
Soundness	5.97	1.63	50
Dependability	6.83	1.53	50
Usefulness	6.41	1.27	50
Usability	6.52	1.47	50

Table 119.
Model Summary for H10c.

Predictor	*R*	R^2	R^2_{adj}	ΔR^2	F_{chg}	*p*	df_1	df_2
Usability	.37	.13	.12	.13	7.42	.009	1	48

Table 120.
Coefficients for H10c.

Predictor	B	β	*t*	Bivariate *r*	Partial *r*
Usability	.28	.37	2.72	.37	.37

Analysis of the residuals revealed no evidence of violations of the assumptions of linearity, normality, or homoscedasticity, hence the results of the multiple regression analysis are accepted as tenable and the null hypothesis $H10c_{null}$ is rejected.

Summary of Main Effect Hypothesis Testing

The analysis above reveals support for all ten of the originally proposed main effect hypotheses. An additional 24 sub-hypotheses were proposed to assess the differences arising from differences in responses provided by people in each of the stakeholder roles. Of those 24 sub-hypotheses, the analysis reveals support for 18. Table 121 provides a recap of the support for these hypotheses.

Including both the quadrant level and the dimension level, there were a total of 20 hypothesized predictor variables. At the quadrant level, each of the four variables is a significant predictor in at least one regression model. At the dimension level, seven of the sixteen variables are significant predictors in at least one regression model; nine are not.

Tables 122a and 122b provide a summary of these predictor variables and their significant relationships.

Table 121.
Summary of Support for Main Effect Hypotheses

Hypotheses	Overall	Information Producers	Information Custodians	Information Consumers
H1, H1a, H1b, H1c	Yes ***	Yes *	Yes **	No
H2, H2a, H2b, H2c	Yes **	Yes *	No	No
H3, H3a, H3b, H3c	Yes **	Yes *	Yes *	No
H4, H4a, H4b, H4c	Yes ***	Yes **	Yes **	Yes **
H5, H5a, H5b, H5c	Yes ***	Yes *	Yes *	Yes **
H6, H6a, H6b, H6c	Yes ***	Yes, *	Yes *	No
H7	Yes **	-	-	-
H8	Yes **	-	-	-
H9, H9a, H9b, H9c	Yes ***	Yes *	No	Yes **
H10, H10a, H10b, H10c	Yes ***	Yes *	Yes *	Yes **

* p < .05, ** p < .01, *** p < .001

Table 122a.
Summary of Predictor Variables

Predictor variable	Criterion variable	Hypotheses (*β*)
Soundness	Strategic Benefits	H5b (.52), H5c (-.71)
Concise Representation	None	
Completeness	Strategic Benefits Transactional Benefits	H1 (.34 & .62), H1a (.41) H6 (.37 & .75), H6b (.42 & 1.13)
Consistent Representation	Strategic Benefits Transactional Benefits	H1 (-.37), H1b (.56) H6 (-.50), H6b (-.82)
Predictor variable	Criterion variable	Hypotheses (*β*)
Free of error	Transactional Benefits	H6a (.44)
Dependability	Strategic Benefits Transactional Benefits	H5a (.40) H10b (.49)
Security	None	
Timeliness	Strategic Benefits Transactional Benefits	H2 (.31), H2a (.37) H7 (.32)
Usefulness	Transactional Benefits	H10 (-.54)
Appropriate Amount	Strategic Benefits Transactional Benefits	H3 (.34), H3a (.37), H3b (.46) H8 (.30)
Interpretability	None	
Objectivity	None	
Relevance	None	
Understandability	None	
Usability	Strategic Benefits Transactional Benefits	H5 (.40), H5c (.29 & .89) H10 (.37 & .86), H10a (.39), H10c (.37)

Table 122b
Summary of Predictor Variables

Accessibility	None	
Believability	None	
Ease of operation	Transactional Benefits	H9 (.22)
Reputation	None	
Value-added	Strategic Benefits	H4 (.48), H4a (.52), H4b (.54), H4c (.37)
	Transactional Benefits	H9 (.39 & .28), H9a (.43), H9c (.37)

Hypothesis Testing – Moderator Effect

Moderated regression analysis was conducted to test the moderator-effect hypotheses, H11 through H20 using the procedure recommended by Sharma et al. (1981). The first step of this analysis involves the examination of the coefficients from the following three regression equations:

$y = a + b_1x$ (Equation 5)

$y = a + b_1x + b_2z$ (Equation 6)

$y = a + b_1x + b_2z + b_3xz$ (Equation 7)

Sharma et al. (1981) prescribe a set of steps for examining the effects of these equations, and then prescribe a set of steps to be conditionally applied based on the outcome of this assessment.

Prior to conducting these analyses, each variable was centered as recommended by Cohen, Cohen, West, and Aiken (2003). To address the differences among stakeholder roles, sub-hypotheses were also evaluated wherever the variables involved produced different sets of predictors in the evaluation of the corresponding main-effect hypotheses.

In none of the analyses for hypotheses H11 through H20 was coefficient b_3 found to be statistically significant, hence no significant interaction was found for any of the moderator-effect hypotheses. From that point in each analysis, one of two distinct patterns emerged. In the first pattern, information intensity was found to be significantly correlated to the criterion variable (either strategic benefits or transactional benefits). At this point, according to Sharma et al. (1981), the analysis stops and the hypothesized variable is deemed not to be a moderator of any type.

The second pattern emerged in some of the sub-hypotheses, in that information intensity was not correlated to either the criterion or the predictor for that particular role-specific subset of cases. When this pattern emerged, the next step in the analysis was to split the cases into two sets using the median for information intensity as the cut-off. Once split in this manner, regression analysis was run again to determine whether there was any significant difference in the predictive validity of regression models produced from the two subsets. In each instance where this pattern was followed, neither subset produced a significant model, hence there was insufficient evidence to support the finding of information intensity as a homologizer variable.

In summary, the null hypothesis was not rejected for any of the ten moderator-effect hypotheses originally proposed. In addition, sub-hypotheses were analyzed for five of these (namely H11, H15, H16, H19, and H20), resulting in a total of fifteen sub-hypotheses, none of which were rejected.

Summary

Chapter 4 has presented the results of research investigating the relationship between information quality and organizational outcome, with information intensity hypothesized as a moderator. The results of a Web-based survey were analyzed in this chapter. Support was found for all the main-effect hypotheses and for most of the sub-hypotheses developed to address a systematic difference in responses. No support was found for any of the moderator-effect hypotheses.

CHAPTER 5. RESULTS, CONCLUSIONS, AND RECOMMENDATIONS

Introduction

Despite compelling evidence that lack of attention to information quality problems leads to substantial human and economic losses, the literature to this point has been devoid of conceptual models of information quality strategy or of systematic exploration of the nature of the relationship between information quality and organizational outcomes. This research was undertaken to fill each of those voids by presenting contextual and conceptual models of information quality strategy and by providing an empirical analysis of the ability to predict organizational outcomes based on information quality measurements.

Summary of the Study

This study has investigated the relationship between the management of information quality and organizational outcomes. A literature review revealed that the relationship between information and decision-making is a complex one that has been the subject of extensive research spanning several decades. Included in this history was the behavioral theory of economics, which sought to explain how people in organizations make decisions in the face of "imperfect information" (Simon, 1979, p. 503). The explanation was found in such concepts as bounded rationality and game theory (Seth & Thomas, 1994; Simon, 1979; Winter, 1971). A forerunner of the behavioral theory, known as institutional economics (Commons, 1931; Cranfill, 1940; Simon, 1979), was characterized by its focus on collective action in the presence of conflict, and assumed that the necessary information was available to decision makers (Seth & Thomas, 1994). This view, although simplistic, was more sophisticated than that of the classical economic theorists, who assumed that a single correct decision could be reached, given enough information (Commons, 1931; Cranfill, 1940; Ricardo, 1960).

These early economic theories have contemporary counterparts in two primary schools of thought in the strategy literature: the competitive environment view (Porter, 1991, 1996) and the resource-based view (Barney, 1991, 2001; Bharadwaj, 2000; Wade & Hulland, 2004; Wernerfelt, 1984). The competitive environment view draws significantly from behavioral economics, and finds its roots in institutional economics, given the focus on collective action in the presence of conflicting objectives. The resource-based view, on the other hand, focuses its attention on resources owned or controlled by the firm, and how those resources can be exploited for competitive advantage, much in the same way that classical economics focused on property and commodities and on maximizing the economic value of each.

The study of information quality evolved separately from the study of strategy and economics. Rooted in the separate disciplines of information theory (Shannon, 1948; Shannon & Weaver, 1949), semiotics (Liu, 2000; Stamper, 1996), and quality (Crosby, 1996; Deming, 1982; Juran, 1988), the study of information quality emerged as a distinct discipline in the mid-1990s (Wang et al., 1995; Wang & Strong, 1996). Since that time, research in this discipline has made significant theoretical and practical advances, including both frameworks (Ballou et al., 1998; Kahn et al., 2002; Koronios, Lin, & Gao, 2005; Wang, 1998) and management approaches (Ballou et al.; Kahn et al.; Lee, Pipino, Strong, & Wang, 2004; Lee & Strong, 2003; Lee et al., 2002). However, despite these advances, a review of the literature revealed very little evidence of understanding from either a theoretical or practical perspective of the relationship between information quality improvement activities and organizational outcomes. The few pieces of literature that addressed information quality strategy research were found to be written from a variety of perspectives with little or no commonality in approach or findings (Campbell et al., 2004; Kerr & Norris, 2004; Pierce, 2004; Redman, 1998). Consequently, this study was positioned to fill that gap by providing common contextual and conceptual frameworks for information quality strategy and by empirically investigating the relationship between information quality and organizational outcomes.

The contextual framework was developed by combining two prior frameworks that had been developed to fill similar gaps in the literature. First, Melville et al. (2004), working against the twin backdrops of strategy literature and information systems success literature, developed an integrative framework of IT business value. This framework consisted of three primary lenses: the focal firm, the competitive environment, and the macro environment. Second, Chung et al. (2005) adapted and simplified Boulding's general systems model hierarchy of complexity and developed a three-level structure for organizing various aspects of information quality, including theory, research, and practice. The three levels of this structure represent the mechanical level, the open systems level, and the human level. By combining these two frameworks, the two dimensional contextual framework for information quality strategy shown in Figure 16 emerged, providing a means of organizing information quality strategy research.

	Focal Firm	Competitive Environment	Macro Environment
Human Systems			
Open Systems			
Mechanical Systems			

Figure 16. Contextual framework for information quality strategy research

The contextual framework above provided the basis of a conceptual framework, the central element of which is the strategic relationship illustrated generically in Figure 17. Any number of such strategic relationships can thus be analyzed within the context of the contextual framework, which provides reference points for anchoring the analysis within both the information quality literature and the strategy literature.

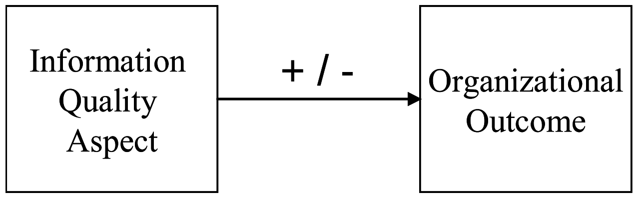

Figure 17. Strategic relationship between information quality aspect and organizational outcome

With respect to this study, ten such relationships were investigated as the set of main effect hypotheses, addressing the first two of the following three research questions:

1. What is the nature of the relationship between stakeholder perception of information quality improvement and organizational outcomes?

2. What interaction effects exist between different aspects of information quality improvement and organizational outcomes?

An additional set of ten moderated relationships, with information intensity as the moderator, were also investigated as the set of moderator effect hypotheses for the purpose of evaluating the third research question:

3. In what way does information intensity affect the relationship between information quality improvement and organizational outcomes?

Each of the twenty hypotheses was investigated using the results of a Web-based survey for which 110 responses were received. The survey items and operationalized variables were taken directly from prior literature (Dejnaronk, 2000; Kahn et al., 2002; Kearns & Lederer, 2003; Lee et al., 2002; Mirani & Lederer, 1998; Najjar, 2002) and the construct validity of the resulting instrument was confirmed through data analysis. The main effect hypotheses were investigated using stepwise multiple regression analysis, and the moderator effect hypotheses were investigated using a combination of moderated regression analysis and subgroup analysis as warranted by the data.

Discussion of the Results

The results of these analyses were mixed. Each of the main effect hypotheses was supported with statistically significant results. Moreover, in many of these analyses, a set of sub-hypotheses was evaluated to consider separately the effects as reported by people in different stakeholder roles. Some of these results proved to be especially interesting, at times resulting in substantially different regression models, depending on stakeholder role. On the other hand, none of the moderator effect hypotheses had statistically significant results. Each of these aspects is discussed in more detail in the following sections.

Discussion of Main Effect Results

As noted above, each of the ten main effect hypotheses was supported with statistically significant results. For most of these analyses, although anywhere from two to five independent variables were specified, only one variable provided a sufficient contribution to R^2 to meet the selection criteria for the stepwise analysis. Consequently, those analyses resulted in a simple regression model with a single predictor variable. The remaining analyses resulted in regression models with two predictors.

Significance of Stakeholder Roles

One of the survey items asked participants to indicate their roles as stakeholders with respect to the particular information system on which they were basing their responses. The roles considered were information collectors/providers, information custodians, and information consumers. As suggested by Lee et al. (2002), it was anticipated that stakeholder role might significantly influence the assessments of information quality. Analysis of the data collected revealed that this effect was present in the sample, in that three of the sixteen information quality dimensions had differences significant at $p = .05$. In addition to the anticipated differences, the strategic benefits variable also differed significantly across stakeholder roles. Although these four variables represent a small proportion of the total set of variables, together they were involved in eight of the ten hypotheses.

To account for these differences, each affected hypothesis was evaluated four different ways. First, the original hypothesis was evaluated using the complete data set to produce an overall predictive model. It was then evaluated three more times to test the sub-hypotheses, using only the data representing each stakeholder role separately and producing role-specific predictive models.

As noted, differences in perception of information quality were not unexpected. However, several of these analyses resulted in completely different predictive models, depending on the stakeholder role in question. In some cases, these differences included reversal of coefficient signs, suggesting a potential difference in perspective as to whether an increase in the measure of a particular quality aspect is helpful or harmful to the organizational outcome in question.

Taken together, these differences suggest that the relationship between information quality and organizational outcomes is indeed a complex one. Not only do multiple dimensions contribute to the predictive model, different segments of the population produce different predictive models.

Significance of Level of Analysis

The hypotheses for this study considered information quality at two different levels, using the PSP/IQ model (Kahn et al., 2002; Lee et al., 2002) as the basis. Eight of the hypotheses considered one PSP/IQ quadrant at a time, using the individual dimensions associated with that quadrant as independent variables. The other two hypotheses considered information quality as a whole, using the quadrants as independent variables.

An unexpected finding from this study was that these two levels of analysis produced inconsistent results that in some cases were contradictory. Considering the Soundness quadrant over the entire data set as an example, two of the dimensions were significant predictors of Strategic Benefits, yet Soundness as a whole was not a significant predictor of Strategic Benefits. As another example, one of the dimensions that contributes to the Usefulness quadrant is a significant predictor of Strategic Benefits over the entire data set and for two of the three role-specific subsets. However, the Usefulness quadrant does not significantly predict Strategic Benefits for any of the analyses. Even more interestingly, the same Usefulness dimension is a significant predictor of Transactional Benefits across the entire data set; however, Usefulness is also a predictor of Transactional Benefits, but with the coefficient sign reversed.

Taken together, these apparent discrepancies raise questions concerning the practice of aggregating measurements to produce a simpler set of information quality metrics. If the goal is merely to provide a simpler measure of overall information quality, then aggregation is a suitable mechanism. However, if the goal has to do with predicting a specific organizational outcome, then it appears that aggregation may result in a distortion of the relationship.

Predictive Models

The previous sections describe two ways in which the results of the analysis were more complex than anticipated. By contrast, the predictive models themselves were somewhat less complex than anticipated, in that most are simple linear regression models with only a single predictor variable each. The most complex models have only two predictors each. Figure 18, which is an adaptation of the PSP/IQ model, summarizes the predictive models for Strategic Benefits for the entire data set, illustrating the simplicity of these models. Dimensions that did not significantly predict the dependent variable are not included in this figure. The letters “n.s.” after a quadrant name indicate a non-significant result for the quadrant as a whole. The numbers shown after each dimension or quadrant name are the beta coefficients for the respective variables. As shown, at the dimension level, three of the four quadrants have only one dimension as a significant predictor variable, and one quadrant has two. At the quadrant level, only one aggregate quadrant measure is a predictor. Figure 19 provides a similar summary of the predictive models for Transactional Benefits using the entire data set.

	Conforms to Specifications	Meets or Exceeds Expectations
Product	Soundness (n.s.) • Completeness +.62 • Consistent representation - .37	Usefulness (n.s.) • Appropriate amount +.34
Service	Dependability (n.s.) • Timeliness +.31	Usability (+.40) • Value-added +.48

Figure 18. Summary of predictive models of strategic benefits (entire data set)

Figures 20 and 21 depict summaries of the predictive models by role for Strategic Benefits and Transactional Benefits, respectively. In these figures, the three numbers associated with each dimension or quadrant name are the beta coefficients for the roles of information collector/provider, information custodian, and information consumer, in that order. The letters "n.s." indicate a non-significant result, and the letters "N/A" indicate that no sub-hypotheses were associated with a particular quadrant.

Despite the complex nature of the overall relationship and the seemingly contradictory nature of certain predictors under certain conditions, a clearly discernable set of patterns is evident among these relationships. First, there is a small set of variables that consistently show up in the models, namely Completeness, Consistent Representation, Timeliness, Appropriate Amount, and Value-Added. Of these, the Value-Added dimension appears to be the most consistently positive predictor of organizational outcomes, suggesting that this characteristic is perhaps more important than others when focusing on organizational outcomes.

	Conforms to Specifications	Meets or Exceeds Expectations
Product	Soundness (n.s.) • Completeness +.75 • Consistent representation - .50	Usefulness (-.54) • Appropriate amount +.30
Service	Dependability (n.s.) • Timeliness +.32	Usability (+.86) • Value-added +.28 • Ease of Operation +.22

Figure 19. Summary of predictive models of transactional benefits (entire data set)

Only one additional dimension, Ease of Operation, appears in any of the models, and it appears only once, namely as a second predictor of Transactional Benefits, along with Value-Added when analyzing the entire data set. Inclusion of this dimension in this regression model has an intuitive appeal, in that when seeking to improve organizational efficiencies, such as those included in the Transactional Benefits variable, ease of operation would be of benefit, whereas difficulty of operation would be an impediment.

	Conforms to Specifications	Meets or Exceeds Expectations
Product	Soundness (n.s., +.52, -.71) • Completeness +.41, n.s., n.s. • Consistent representation n.s., +.55, .n.s.	Usefulness (n.s., n.s., n.s.) • Appropriate amount +.37, +.46, n.s.
Service	Dependability (+.40, .n.s., n.s.) • Timeliness +.37, n.s., n.s.	Usability (n.s., n.s., +.89) • Value-added +.52, +.54, +.37

Figure 20. Summary of predictive models of strategic benefits (by role)

Another pattern worth noting is that the only dimension-level variable with any negative coefficients is Consistent Representation. However, it is important to note that this dimension is not always negative. In particular, in the only model in which this dimension is the sole predictor, it has a positive coefficient, whereas it is always negative when included in a model with the Completeness dimension. In each such model, Completeness was entered first, followed by Consistent Representation, and the Completeness coefficient is always of greater magnitude than that of Consistent

Representation, resulting in a net positive relationship. This pattern suggests the presence of a discernable interaction effect among these two dimensions.

	Conforms to Specifications	Meets or Exceeds Expectations
Product	Soundness (n.s., n.s., n.s.) • Completeness n.s., +1.13, n.s. • Consistent representation n.s., - .82, .n.s.	Usefulness (n.s., n.s., n.s.) • N/A
Service	Dependability (n.s., +.49., n.s.) • N/A	Usability (+.39, n.s., +.37) • Value-added +.43, n.s., +.37

Figure 21. Summary of predictive models of transactional benefits (by role)

With respect to the dimensions that were not included in any particular model, it is important for several reasons not to over-interpret their lack of inclusion. First, the measurements of organizational outcome used in this analysis were limited to only four dimensions, namely Competitive Advantage, Customer Relations, Business Efficiency, and Systems Development Efficiency. Two dimensions were excluded as a result of the reliability analysis, and there are many other potential dimensions that were not included in the measurement instrument for reasons of parsimony. Other reasons are included in the limitations section below.

Discussion of Moderator Effect Results

As noted above, no support was found for any of the moderator effect hypotheses. As with the information quality dimensions not included in the regression models, it is important for several reasons not to over-interpret this lack of support. Most notably, the theoretical basis for considering information intensity as a moderator of the information quality-organizational outcomes relationship is sound, suggesting that the effect should be detectable if measured properly. However, there were serious normality problems with the data collected, in that the sample appeared strongly biased toward organizations with high information intensity. Although regression can be somewhat resilient to normality problems, serious normality problems with the data can lead to an inability to detect significant relationships (Mertler & Vannatta, 2005).

Additionally, scale has been demonstrated to be a problem with moderated regression analysis, in that the variable used in the analysis is produced by multiplying two other variables together, creating a situation in which an effect is measured with a scale considerably courser than the effect itself. This effect is known to increase the likelihood of Type II errors (Carte & Russell, 2003), which may be the case in this analysis. For these reasons, no implications should be drawn from the lack of support for these hypotheses.

Conclusions

Repeating the words of Porter (1991), “the reason why firms succeed or fail is perhaps the central question in strategy” (p. 95). Success or failure for an organization can take many forms and can be perceived as having any number of proximate causes. Fundamentally, however, it is the cumulative effect of many decisions made over time by management and by empowered individuals. The quality of those decisions is reflected in the organizational outcomes attained.

Decisions require information. Researchers have long recognized and studied this fundamental, but complex relationship, and have long recognized that the information available to decision-makers is often imperfect. Nonetheless, decisions have to be made, often before better information can be made available.

The research for this study has demonstrated that the relationship between the quality of information and organizational outcomes is systematically measurable and that this relationship is, for the most part, positive. Additionally, this research has set forth both contextual and conceptual models deemed useful in positioning and describing this and other research on the topic within the broader context of the body of literature.

Contribution to the Strategy Literature

Within the strategy literature, the contextual model serves as an extension of the three-lens integrative framework presented in Melville et al. (2004). The three-lens framework fits within the strategy literature in that the focal firm lens corresponds well to the resource-based view of the firm, and the competitive environment lens corresponds exactly to the competitive environment focus advocated by Porter and others. The third lens, focusing on the macro environment, also has strategy implications, most notably with respect to the study of economies, nations, or world regions. The extension provided by this research is the second axis, which expands each of those three lenses into three different levels of systems complexity, namely mechanical systems, open systems, and human systems.

The empirical results of this research contribute to the strategy literature by demonstrating that the quality of information has a quantifiable relationship to the quality of decisions as reflected in organizational outcomes. The recognition of the importance of this relationship traces all the way back to the earliest writings on institutional economics (Commons, 1931; Cranfill, 1940), yet retains its relevance in contemporary literature (Barney, 1991, 2001; DeLone & McLean, 1992, 2003; Porter, 1991; Porter & Millar, 1985).

The empirical analysis for this research was positioned across two of the three strategic lenses: the focal firm and competitive environment. In particular, the organizational outcomes measured as transactional benefits have an inward focus, thus fit within the focal firm lens, contributing to the resource based view strategy literature. Similarly, the organizational outcomes measured as strategic benefits have an outward focus, thus fit within the competitive environment lens, contributing to the competitive environment focused strategy literature.

Contribution to the Information Quality Literature

Within the information quality literature, the contextual model serves as an extension of the three-level systems complexity model set forth by Chung et al. (2005).

From this perspective, the extension is the addition of the strategy axis, which expands each of those three levels into three different strategic contexts.

The empirical results of this research also contribute to the information quality literature by demonstrating that the quality of information has a systematically quantifiable relationship to the quality of decisions as reflected in organizational outcomes. This relationship has been studied fairly extensively within the information quality literature, but primarily through the use of case studies and action research. Collectively, those studies have clearly identified that the relationship exists, but they have not undertaken a systematic examination from a quantitative perspective. As such, this research contributes to and provides validation for the existing body of case study and action research by providing a degree of triangulation.

This research also confirms and extends the findings of Lee et al. (2002) regarding the differences in perspective sometimes found among stakeholders in different roles. The extension provided by this research is the finding that different stakeholders not only view information quality differently, but in some cases also view the benefits of that information differently.

Implications for Researchers

Researchers of both strategy and information quality can benefit from this study in a number of ways. First, the contextual model presented in this study is intended to be useful to researchers interested in examining the intersection of these two disciplines. The predecessors of this model were each clearly rooted in their respective disciplines, and the model did prove useful in the context of this research. As such, researchers are encouraged to use this model and to continue testing its efficacy and explanatory capabilities.

The conceptual model presented in this study is also intended to be useful to researchers examining the intersection of the disciplines, and is intended to be particularly applicable to researchers examining various aspects of the relationship between information and decision making.

Researchers may also benefit by considering the specific empirical findings of this research in the development of research models examining this or similar phenomena. Although the interpretation of these findings should be limited as discussed below, and although the empirical results cannot be generalized beyond the population represented by this sample, this analysis has clearly demonstrated the ability to predict certain organizational outcomes based on the measurement of certain information quality characteristics. As such, these findings can provide a useful starting point for subsequent empirical examination.

Implications for Practitioners

Practitioners can also benefit from this study, although the results should be considered somewhat preliminary. In particular, this research demonstrates that attention to improving the five information quality dimensions identified as significant predictors is likely to be associated with improved organizational outcomes of the type considered in this study.

As noted, this should be considered somewhat preliminary from a practitioner's standpoint. The reason for this statement is three-fold. First, this study did not examine cause and effect, leaving open the possibility that other factors may be at play. Second, in practical terms, the measurement scale used for organizational outcomes can be considered arbitrary at best. That is, although this study predicts that one outcome may be better than another, it is not clear how this improvement translates into meaningful terms such as increased revenue or reduced cost. Finally, the fact that some information quality dimensions were not included in the list should not be interpreted as them having no meaningful, practical effect. Instead, this should be interpreted simply as a lack of evidence in this case.

Limitations of the Study

A number of limitations of this study were described in chapter 3. To the extent practical, steps were taken to minimize or mitigate the effect of these limitations. Nonetheless, some important limitations remain and are discussed in this section. Most notable among these is the fact that the main effects hypothesis testing for this research was conducted using stepwise regression. This technique is considered appropriate for exploratory research (Mertler & Vannatta, 2005), but has also been sharply criticized as being limited in power and applicability (Cohen et al., 2003). Nonetheless, it was chosen for this study due primarily to the lack of available theoretical basis for sequencing the regression analysis any differently.

The penalty for this choice is the lack of explanatory power. As Cohen et al. (2003) point out, stepwise regression is limited in its power to predictive models only, and it should not be relied upon exclusively or routinely for the development of explanatory theories. In particular, they note that stepwise regression research has been shown in some cases to omit predictors from the model that would have produced statistically significant results with other regression techniques. For this reason, the ability to draw conclusions from this research is similarly limited.

Additionally, as noted throughout the discussion above, the lack of finding for any particular analysis should not be interpreted as the lack of a corresponding effect. The number of dimensions of organizational outcomes considered in this case were limited initially for reasons of parsimony, and were reduced further for reasons of reliability. As such, it is quite likely that the predictive capabilities of the information quality dimensions would have been different had other organizational outcome dimensions been measured.

Similarly, as stated earlier, the lack of significant finding for information intensity as a moderator variable should not be interpreted as the absence of such an effect. Instead, the combined effect of the normality problem and the granularity of scale problem identified by Carte and Russell (2003) created a situation in which the risk of Type II errors was substantial. As such, it is not reasonable to interpret the results of this analysis as indicative of the lack of a moderator effect.

Finally, as noted in chapter 3, this study was conducted from the perspective of the post-positivist paradigm. As such, this study can only be used to describe the observed effects, and cannot be used to meaningfully interpret the reasons behind any of those effects.

Recommendations for Future Research

This study has revealed the presence of an empirically measurable, systematic relationship between information quality and organizational outcomes. As such, this study indicates that further research in this area is likely to yield meaningful results.

Several lines of research are recommended based on the findings of this study. First, research similar to this study, but using a different regression model or a different analytical approach, such as path analysis, is highly recommended. Such a study could build directly on the findings of this research by adding explanatory power to the analysis.

Researchers are also encouraged to conduct research similar to this study, but using different measurements of organizational outcomes. Such studies may reveal additional relationships not evident in this study.

To address the information intensity question, researchers are encouraged to replicate this study using a different sampling frame in an effort to find a more normally distributed sample for the information intensity variables. Additional work on improving the instrument used to measure information intensity is also warranted. Only upon the completion of additional studies will there be sufficient evidence to draw conclusions regarding the potential moderating effect of information intensity on the primary relationships.

In addition to the contributions and extensions identified above, this research also raises some questions. Most notably, by finding substantially different and apparently conflicting regression models at the dimension level versus the PSP/IQ quadrant level, this research draws into question the appropriateness of the pursuit of increasingly simple metrics for information quality (Lee et al., 2002; Pipino et al., 2002; Pipino et al., 2005; Wang et al., 1995). Admittedly, the evidence from this research is limited, and may be indicative of other effects not measured at an observable level within the scope of this effort. As such, further research is encouraged to better understand the effect aggregation has on the ability to predict and explain the relationship between information quality and organizational outcomes.

An additional question was raised during the pilot study phase. Participants in that phase were asked for general impressions of the survey instrument, and very consistently reported being troubled by the redundancy of the survey items. Taking these comments into consideration in the context of the very high Cronbach alpha values of the information quality constructs (greater than .9 in many cases), an examination of the instrument seems warranted, with an eye toward providing a more streamlined and parsimonious instrument without unduly diminishing its ability to measure information quality.

Finally, researchers are encouraged to seek validation of these findings through research conducted in accordance with research paradigms other than post-positivism. Such studies can provide valuable validation of these findings through triangulation, and can provide a much richer set of interpretative and explanatory capabilities.

REFERENCES

9/11 Commission. (2004). *The 9/11 Commission report: Final report of the commission on terrorist attacks upon the United States* (Publication No. 041-015-00236-8). Washington, DC: U.S. Government Printing Office.

Agmon, N., & Ahituv, N. (1987). Assessing data reliability in an information system. *Journal of Management Information Systems, 4*(2), 34-44.

Arbnor, I., & Bjerke, B. (1997). *Methodology for creating business knowledge* (2nd ed.). Thousand Oaks, CA: SAGE.

Avery, J. (2003). *Information theory and evolution*. Singapore: World Scientific Publishing Co. Pte. Ltd.

Bailey, J. E., & Pearson, S. W. (1983). Development of a tool for measuring and analyzing computer user satisfaction. *Management Science, 29*(5), 530-545.

Ballou, D. P., & Pazer, H. L. (1985). Modeling data and process quality in multi-input, multi-output information systems. *Management Science, 31*(3), 150-162.

Ballou, D. P., Wang, R. Y., Pazer, H., & Tayi, G. K. (1998). Modeling information manufacturing systems to determine information product quality. *Management Science, 44*(4), 462-484.

Barney, J. B. (1991). Firm resources and sustained competitive advantage. *Journal of Management, 17*(1), 99-120.

Barney, J. B. (2001). Is the resource-based "view" a useful perspective for strategic management research? Yes. *Academy of Management Review, 26*(1), 41-56.

Belkin, L. (2004). How can we save the next victim? In T. L. Beauchamp & N. E. Bowie (Eds.), *Ethical theory and business* (7th ed., pp. 136-146). Upper Saddle River, NJ: Pearson - Prentice Hall.

Bendoly, E., & Kaefer, F. (2004). Business technology complementarities: Impacts of the presence and strategic timing of ERP on B2B e-commerce technology efficiencies. *Omega, 32*(5), 395-405.

Bertoletti, M., Missier, P., Scannapieco, M., Aimetti, P., & Batini, C. (2005). Improving government-to-business relationships through data reconciliation and process reengineering. In R. Y. Wang, E. M. Pierce, S. E. Madnick & C. W. Fisher (Eds.), *Information quality* (pp. 151-166). New York: M. E. Sharpe.

Bharadwaj, A. S. (2000). A resource-based perspective on information technology capability and firm performance: An empirical investigation. *MIS Quarterly, 24*(1), 169-196.

Boulding, K. E. (1956). General systems theory - the skeleton of science. *Management Science, 2*(3), 197-208.

Bovee, M. W. (2004). Information quality: A conceptual framework and empirical validation. *DAI, 65* (07), 2668, (UMI 3141462)

Bullen, J. I. (2005). Information security in Brazil: Modeling and predicting outsourcing decisions. *DAI, 65* (10), 3910, (UMI 3151948)

Campbell, T., Douglass, K., & Smith-Adams, W. (2004). Using the data quality scorecard as a negotiation strategy. In S. Chengalur-Smith, J. A. Long, L. Raschid & C. E. Seko (Eds.), *Proceedings of the 2004 international conference on information quality* (pp. 154-163). Cambridge: Massachusetts Institute of Technology.

Carr, N. G. (2003). IT doesn't matter. *Harvard Business Review, 81*(5), 41-49.

Carte, T. A., & Russell, C. J. (2003). In pursuit of moderation: Nine common errors and their solutions. *MIS Quarterly, 27*(3), 479-501.

Chung, W. Y., Fisher, C. W., & Wang, R. Y. (2005). Redefining the scope and focus of information-quality work. In R. Y. Wang, E. M. Pierce, S. E. Madnick & C. W. Fisher (Eds.), *Information quality* (pp. 230-248). New York: M. E. Sharpe.

Churchill, G. A., Jr. (1979). A paradigm for developing better measures of marketing constructs. *Journal of Marketing Research, 16*(1), 64-73.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Erblaum.

Columbia Accident Investigation Board. (2003). *Columbia Accident Investigation Board report, volume 1* (033-000-01260-8). Washington, DC: U.S. Government Printing Office.

Commons, J. R. (1931). Institutional economics. *American Economic Review, 21*(4), 648-657.

Cooper, D. R., & Schindler, P. S. (2003). *Business research methods* (8th ed.). Boston: McGraw-Hill Irwin.

Cranfill, S. E. (1940). Recent contributions of John R. Commons to economic thought. *Southern Economic Journal, 7*(1), 63-79.

Crook, C., & Garratt, D. (2005). The positivist paradigm in contemporary social science research. In B. Somekh & C. Lewin (Eds.), *Research methods in the social sciences* (pp. 207-214). London: SAGE.

Crosby, P. B. (1992). *Completeness: Quality for the 21st century*. New York: Dutton.

Crosby, P. B. (1996). *Reflections on quality*. New York: McGraw-Hill.

Davidson, B. N., Lee, Y. W., & Wang, R. Y. (2004). Developing data production maps: Meeting patient discharge data submission requirements. *International Journal of Healthcare Technology and Management, 6*(2), 223-240.

Dejnaronk, A. (2000). An exploratory study of information systems infrastructure and its link to performance. *DAI, 61* (108), 3248, (UMI 9982053)

DeLone, W. H., & McLean, E. R. (1992). Information systems success: The quest for the dependent variable. *Information Systems Research, 3*(1), 60-95.

DeLone, W. H., & McLean, E. R. (2003). The DeLone and McLean model of information systems success: A ten-year update. *Journal of Management Information Systems, 19*(4), 9-30.

Deming, W. E. (1982). *Quality, productivity, and competitive position*. Cambridge, MA: MIT Center for Advanced Engineering Study.

Elmorshidy, A. (2005). Information systems (IS) success in non-organizational contexts: Examining the DeLone and McLean is success model in the context of an online stock trading environment. *DAI, 65* (07), 2410, (UMI 3139268)

Feltham, G. A. (1968). The value of information. *Accounting Review, 43*(4), 684-696.

Fisher, C. W., & Kingma, B. R. (2001). Criticality of data quality as exemplified in two disasters. *Information & Management, 39*(2001), 109-116.

Fiske, D. W., & Campbell, D. T. (1992). Citations do not solve problems. *Psychological Bulletin, 112*(3), 393-395.

Gallagher, C. A. (1974). Perceptions of the value of a management information system. *Academy of Management Journal, 17*(1), 46-55.

Garson, G. D. (2006a). *Factor analysis*. Retrieved August 9, 2006, from http://www2.chass.ncsu.edu/garson/pa765/factor.htm

Garson, G. D. (2006b). *Multiple regression*. Retrieved August 16, 2006, from http://www2.chass.ncsu.edu/garson/PA765/regress.htm

Greene, J. C., Kreider, H., & Mayer, E. (2005). Combining qualitative and quantitative methods in social inquiry. In B. Somekh & C. Lewin (Eds.), *Research methods in the social sciences* (pp. 274-281). London: SAGE.

Halloran, D., Manchester, S., Moriarty, J., Riley, R., Rohrman, J., & Skramstad, T. (1978). Systems development quality control. *MIS Quarterly, 2*(4), 1-13.

Handscombe, R. D., & Patterson, E. A. (2004). *The entropy vector: Connecting science and business*. Singapore: World Scientific.

Helfert, M., & Herrmann, C. (2005). Introducing data-quality management in data warehousing. In R. Y. Wang, E. M. Pierce, S. E. Madnick & C. W. Fisher (Eds.), *Information quality* (pp. 135-150). New York: M. E. Sharpe.

Juran, J. M. (1988). *Juran on planning for quality*. New York: The Free Press.

Kaefer, F., & Bendoly, E. (2004). Measuring the impact of organizational constraints on the success of business-to-business e-commerce efforts: A transactional focus. *Information & Management, 41*(5), 529-541.

Kahn, B. K., Pierce, E. M., & Melkas, H. (2004). IQ research directions. In S. Chengalur-Smith, J. A. Long, L. Raschid & C. E. Seko (Eds.), *Proceedings of the 2004 international conference on information quality* (pp. 326-332). Cambridge, MA: MIT.

Kahn, B. K., Strong, D. M., & Wang, R. Y. (2002). Information quality benchmarks: Product and service performance. *Communications of the ACM, 45*(4), 184-192.

Katz-Haas, R., & Lee, Y. W. (2005). Understanding interdependencies between information and organizational processes. In R. Y. Wang, E. M. Pierce, S. E. Madnick & C. W. Fisher (Eds.), *Information quality* (pp. 167-178). New York: M. E. Sharpe.

Kearns, G. S., & Lederer, A. L. (2003). A resource-based view of strategic IT alignment: How knowledge sharing creates competitive advantage. *Decision Sciences, 34*(1), 1-29.

Kerr, K., & Norris, T. (2004). The development of a healthcare data quality framework and strategy. In S. Chengalur-Smith, J. A. Long, L. Raschid & C. E. Seko (Eds.), *Proceedings of the 2004 international conference on information quality* (pp. 218-233). Cambridge: Massachusetts Institute of Technology.

Koronios, A., Lin, S., & Gao, J. (2005). A data quality model for asset management in engineering organisations. In F. Naumann, M. Gertz & S. E. Madnick (Eds.), *Proceedings of the 10th international conference on information quality* (pp. 27-51). Cambridge: Massachusetts Institute of Technology.

Kuhn, T. S. (1996). *The structure of scientific revolutions* (3rd ed.). Chicago: The University of Chicago Press.

Lattin, J., Carroll, J. D., & Green, P. E. (2003). *Analyzing multivariate data*. Pacific Grove, CA: Brooks/Cole - Thomson Learning.

Lee, Y. W. (2003-2004). Crafting rules: Context-reflective data quality problem solving. *Journal of Management Information Systems, 20*(3), 93-119.

Lee, Y. W., Pipino, L., Strong, D. M., & Wang, R. Y. (2004). Process-embedded data integrity. *Journal of Database Management, 15*(1), 87-103.Lee, Y. W., & Strong, D. M. (2003). Knowing-why about data processes and data quality. *Journal of Management Information Systems, 20*(3), 13-39.

Lee, Y. W., & Strong, D. M. (2003-2004). Knowing-why about data processes and data quality. *Journal of Management Information Systems, 20*(3), 13-39.

Lee, Y. W., Strong, D. M., Kahn, B. K., & Wang, R. Y. (2002). AIMQ: A methodology for information quality assessment. *Information and Management, 40*(2), 133-146.

Levitin, A. V., & Redman, T. C. (1998). Data as a resource: Properties, implications, and prescriptions. *Sloan Management Review, 40*(1), 89-101.

Lewin, C. (2005). Elementary quantitative methods. In B. Somekh & C. Lewin (Eds.), *Research methods in the social sciences* (215-225). London: Sage Publications.

Liu, K. (2000). *Semiotics in information systems engineering*. Cambridge, UK: Cambridge University Press.

Maffei, R. B. (1958). Simulation, sensitivity, and management decision rules. *Journal of Business, 31*(3), 177-186.

Mahoney, F. X., & Thor, C. G. (1994). The TQM trilogy: Using ISO 9000, the Deming Prize, and the Baldridge Award to establish a system for total quality management. New York: American Management Association.

McEliece, R. J. (2002). *The theory of information and coding* (2nd ed.). Cambridge, UK: Cambridge University Press.

McKinney, V., Yoon, K., & Zahedi, F. (2002). The measurement of web-customer satisfaction: An expectation and disconfirmation approach. *Information Systems Research, 13*(3), 296-315.

Melkas, H. (2004). Towards holistic management of information within service networks: Safety telephone services for ageing people. Unpublished Ph.D. Dissertation, Teknillinen Korkeakoulu (Helsinki, Finland).

Melville, N., Kraemer, K., & Gurbaxani, V. (2004). Review: Information technology and organizational performance: An integrative model of IT business value. *MIS Quarterly, 28*(2), 282-322.

Merrell, F. (1997). *Peirce, signs, and meaning*. Toronto: University of Toronto Press.

Mertler, C. A., & Vannatta, R. A. (2005). *Advanced and multivariate statistical methods* (Third ed.). Glendale, CA: Pyrczak.

Mirani, R., & Lederer, A. L. (1998). An instrument for assessing the organizational benefits of IS projects. *Decision Sciences, 29*(4), 803-838.

Moore, G. C., & Benbaset, I. (1991). Development of an instrument to measure the perceptions of adopting and information technology innovation. *Information Systems Research, 2*(3), 192-222.

Najjar, L. (2002). The impact of information quality and ergonomics on service quality in the banking industry. *DAI, 63* (09), 3258, (3064565)

National Institute of Standards and Technology. (2005). *Frequently asked questions and answers about the Malcolm Baldridge National Quality Award*. Retrieved February 8, 2005, from http://www.nist.gov/public affairs/factsheet/baldfaqs.htm

Nunnally, J. C. (1978). *Psychometric theory* (Second ed.). New York: McGraw-Hill.

Oman, R. C., & Ayers, T. B. (1988). Improving data quality. *Journal of Systems Management, 39*(5), 31-35.

Paradice, D. B., & Fuerst, W. L. (1991). An MIS data quality methodology based on optimal error detection. *Journal of Information Systems, 5*(1), 48-66.

Parssian, A. H. (2002). Assessing information quality for relational databases. *DAI, 64* (04), 1330, (UMI 3087738)

Pierce, E. M. (2004). Developing, implementing and monitoring an information product quality strategy. In S. Chengalur-Smith, J. A. Long, L. Raschid & C. E. Seko (Eds.), *Proceedings of the 2004 international conference on information quality* (pp. 13-26). Cambridge: Massachusetts Institute of Technology.

Pierce, E. M. (2005). What's in your information product inventory? In R. Y. Wang, E. M. Pierce, S. E. Madnick & C. W. Fisher (Eds.), *Information quality* (pp. 99-114). New York: M. E. Sharpe.

Pipino, L. L., Lee, Y. W., & Wang, R. Y. (2002). Data quality assessment. *Communications of the ACM, 45*(4), 211-218.

Pipino, L. L., Wang, R. Y., Kopsco, D., & Rybolt, W. (2005). Developing measurement scales for data-quality dimensions. In R. Y. Wang, E. M. Pierce, S. E. Madnick & C. W. Fisher (Eds.), *Information quality* (pp. 37-51). New York: M. E. Sharpe.

Porter, M. E. (1991). Towards a dynamic theory of strategy. *Strategic Management Journal, 12*, 95-117.

Porter, M. E. (1996). What is strategy? *Harvard Business Review, 74*(6), 61-78.

Porter, M. E. (2001). Strategy and the Internet. *Harvard Business Review, 79*(3), 62-78.

Porter, M. E., & Millar, V. E. (1985). How information gives you competitive advantage. *Harvard Business Review*(July-August), 149-160.

Priem, R. L., & Butler, J. E. (2001). Is the resource-based "view" a useful perspective for strategic management research? *Academy of Management Review, 26*(1), 22-40.

QuestionPro policies and procedures. (2006). Retrieved March 25, 2006, from http://www.questionpro.com/help.html

Redman, T. C. (1995). Improve data quality for competitive advantage. *Sloan Management Review, 36*(2), 99-107.

Redman, T. C. (1998). The impact of poor data quality on the typical enterprise. *Communications of the ACM, 41*(2), 79-82.

Redman, T. C. (2005). Measuring data accuracy: A framework and review. In R. Y. Wang, E. M. Pierce, S. E. Madnick & C. W. Fisher (Eds.), *Information quality* (pp. 21-36). New York: M. E. Sharpe.

Ricardo, D. (1960). Rent. In G. D. H. Cole (Ed.), *Classics in economics: A course of selected reading by authorities* (pp. 29-33). Port Washington, NY: Kennikat Press.

Robson, C. (2002). *Real world research* (2nd ed.). Oxford: Blackwell Publishers.

Rogers Commission. (1986). Report of the Presidential Commission on the Space Shuttle Challenger Accident. Washington, DC: U.S. Government Printing Office.

Saraph, J. V., Benson, P. G., & Schroeder, R. G. (1989). An instrument for measuring the critical factors of quality management. *Decision Sciences, 20*(4), 810-829.

Scannapieco, M., Pernici, B., & Pierce, E. M. (2005). IP-UML. In R. Y. Wang, E. M. Pierce, S. E. Madnick & C. W. Fisher (Eds.), *Information quality* (pp. 115-131). New York: M. E. Sharpe.

Seth, A., & Thomas, H. (1994). Theories of the firm: Implications for strategy research. *Journal of Management Studies, 31*(2), 165-191.

Shankaranarayanan, G. (2005). Towards implementing total data quality management in a data warehouse. *Journal of Information Technology Management, 16*(1), 21-30.

Shankaranarayanan, G., & Cai, Y. (2006). Supporting data quality management in decision-making. *Decision Support Systems, 42*(1), 302-317.

Shankaranarayanan, G., Wang, R. Y., & Ziad, M. (2000). IP-MAP: Representing the manufacture of an information product. In B. D. Klein & D. F. Rossin (Eds.), *Proceedings of the 2000 international conference on information quality* (pp. 1-16). Cambridge: Massachusetts Institute of Technology.

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal, 27*, 379-423, 623-656.

Shannon, C. E., & Weaver, W. (1949). *The mathematical theory of communication*. Urbana, IL: The University of Illinois Press.

Sharma, S., Durand, R. M., & Gur-Arie, O. (1981). Identification and analysis of moderator variables. *Journal of Market Research, XVIII*(3), 291-300.

Shemwell, D. J., & Yavas, U. (1999). Measuring service quality in hospitals: Scale development and managerial applications. *Journal of Marketing Theory and Practice, 7*(3), 65-75.

Simon, H. A. (1979). Rational decision making in business organization. *American Economic Review, 69*(4), 493-513.

Stamper, R. (1996). Signs, information, norms and systems. In B. Holmqvist, P. B. Andersen, H. Klein & R. Posner (Eds.), *Signs of work: Semiosis and information processing in organisations* (pp. 349-398). Berlin: Walter D. Gruyter.

Strong, D. M., Lee, Y. W., & Wang, R. Y. (1997). Data quality in context. *Communications of the ACM, 40*(5), 103-110.

Teo, T. S. H., & King, W. R. (1997). Integration between business planning and information systems planning: An evolutionary-contingency perspective. *Journal of Management Information Systems, 14*(1), 185-214.

Thornsbury, S., Davis, K., & Minton, T. (2003). Adding value to agricultural data: A golden opportunity. *Review of Agricultural Economics, 25*(2), 550-568.

Trueblood, R. M. (1960). Operations research - a challenge to accounting. *Journal of Accountancy, 109*(5), 47-51.

Wade, M., & Hulland, J. (2004). Review: The resource-based view and information systems research: Review, extension, and suggestions for future research. *MIS Quarterly, 28*(1), 107-142.

Wand, Y., & Wang, R. Y. (1996). Anchoring data quality dimensions in ontological foundations. *Communications of the ACM, 39*(11), 86-95.

Wang, R. Y. (1998). A product perspective on total data quality management. *Communications of the ACM, 41*(2), 58-65.

Wang, R. Y., Lee, Y. W., Pipino, L. L., & Strong, D. M. (1998). Manage your information as a product. *Sloan Management Review, 39*(4), 95-105.

Wang, R. Y., Storey, V. C., & Firth, C. P. (1995). A framework for analysis of data quality research. *IEEE Transactions on Knowledge and Data Engineering, 7*(4), 623-640.

Wang, R. Y., & Strong, D. M. (1996). Beyond accuracy: What data quality means to data consumers. *Journal of Management Information Systems, 12*(4), 5-34.

Weill, P. (1992). The relationship between investment in information technology and firm performance: A study of the valve manufacturing sector. *Information Systems Research, 3*(4), 307-333.

Wernerfelt, B. (1984). A resource-based view of the firm. *Strategic Management Journal, 5*, 171-180.

Winter, S. G. (1971). Satisficing, selection, and the innovating remnant. *Quarterly Journal of Economics, 85*, 237-261.

APPENDIX: SURVEY INSTRUMENT

Part I – General

The following items address basic information about the organization in which you work and the nature of your interaction with computer-based information systems.

1. In my work for this organization, I regularly interact with computer-based information systems in the following ways (check all that apply):
    a. Receive reports
    b. Provide information for others to use
    c. Look up information
    d. Update or modify data
    e. Perform modeling simulation or analysis
    f. Perform computer-based design or engineering
    g. Monitor status of something (e.g., shipping, manufacturing, inventory)
    h. Design or deploy information systems
    i. Manage, operate, or administer information systems

When you have made your selections, click the "continue" button.

Part II – Use of Information by your Organization

This part of the survey addresses general characteristics about your understanding of the way information is used by your organization. For each statement, select a number from 1 to 7, where 1 indicates that you *strongly disagree* with the statement, and 7 indicates that you *strongly agree* with the statement.

1. Our product or service operation involves substantial information processing.
2. We have many product or service varieties within a line of products or services.
3. Information is used to a great extent in our production or service operations.
4. Our product or service mainly provides information.
5. Many steps in our production or service operations require the frequent use of information.
6. Customers need a lot of information related to our products or services before purchasing the product or service.
7. Cycle time from the initial order to the delivery of our product or service is long.
8. Information used in our production or service operations is usually accurate.
9. Our product or service requires extensive user training.
10. Our product or service is complex (i.e., is contains many parts that must work together).
11. Information used in our production or service operations is frequently updated.

When you have made your selections, click the "continue" button.

Part III – Organizational Benefits of Information

For this section of the survey, please think of a particular information system that you currently interact with in the performance of your job. This system can be a report that you receive regularly, an interactive system that you update, a database application that you interact with, a system that you operate or are deploying, or something similar. Please select a single system and keep this system in mind as you respond to the items in this section.

1. Before proceeding, please indicate which of the following best describes the nature of your interaction with the system you have selected:
    a. I receive reports from this system
    b. I provide information to this system for others to use
    c. I use this system to look up information
    d. I update or modify the data in this system
    e. I use this system to perform modeling simulation or analysis
    f. I use this system to perform computer-based design or engineering
    g. I use this system to monitor status of something (e.g., shipping, manufacturing, inventory)
    h. I am involved in the design or deployment of this system
    i. I am responsible for managing, operating, or administering this system

Each item below addresses your understanding of the benefits your organization derives from the use of the information in this system. For each item, select a number from 1 to 7 that best completes the sentence: "Use of the information in this system . . ." The number 1 indicates that you *strongly disagree* with the statement, and 7 indicates that you *strongly agree*.

"Use of the information in this system . . ."

1. enables the organization to respond more quickly to change.
2. provides better products or services to our customers.
3. saves money by reducing system modifications or enhancement costs.
4. provides the ability to perform maintenance faster.
5. increases return on financial assets.
6. enables the organization to catch up with competitors
7. improves customer relations.
8. saves money by reducing travel costs.
9. allows other applications to be developed faster.
10. saves money by avoiding the need to increase the work force.
11. enhances employee productivity or business efficiency.
12. aligns well with stated organizational goals.
13. provides new products or services to customers.
14. saves money by reducing communication costs.
15. allows previously infeasible applications to be developed faster.
16. speeds up transactions or shortens production cycles.
17. enhances competitiveness or creates strategic advantage.

18. helps establish useful linkages with other organizations.

When you have made your selections, click the "continue" button.

Part IV – Information Quality

(Note: these items will be presented on three screens containing 25, 25, and 19 items, respectively.)

For this section of the survey (this screen and two more screens), please continue to think of the same system you considered for the previous section, but respond in terms of the information itself. For each item, select the number from 0 to 10 to indicate the extent to which you agree with the statement, where 0 indicates *not at all*, 5 indicates *average*, and 10 indicates *completely*.

1. This information is easy to manipulate to meet our needs.
2. It is easy to interpret what this information means.
3. This information is consistently presented in the same format.
4. This information includes all the necessary values.
5. This information is easily retrievable.
6. This information is formatted compactly.
7. This information is protected against unauthorized access.

8. This information is incomplete.
9. This information is not presented consistently.
10. This information has a poor reputation for quality.
11. This information is complete.
12. This information is presented concisely.
13. This information is easy to understand.
14. This information is believable.
15. This information is easy to aggregate.
16. This information is of sufficient volume for our needs.
17. This information is correct.
18. This information is useful to our work.
19. This information provides a major benefit to our work.
20. This information is easily accessible.
21. This information has a good reputation.
22. This information is sufficiently current for our work.
23. This information is difficult to interpret.
24. This information is not protected with adequate security.
25. This information is of doubtful credibility.

(Screen break here, with scale repeated on the next screen, along with the words, "Screen 2 of 3 for this section".)

26. The amount of information does not match our needs.

27. This information is difficult to manipulate to meet our needs.

28. This information is not sufficiently timely.

29. This information is difficult to aggregate.

30. The amount of information is not sufficient for our needs.

31. This information is incorrect.

32. This information does not add value to our work.

33. This information was objectively collected.

34. It is difficult to interpret the coded information.

35. The meaning of this information is difficult to understand.

36. This information is not sufficiently current for our work.

37. This information is easily interpretable.

38. The amount of information is neither too much nor too little.

39. This information is accurate.

40. Access to this information is sufficiently restricted.

41. This information is presented consistently.

42. This information has a reputation for quality.

43. This information is easy to comprehend.

44. This information is based on facts.

45. This information is sufficiently complete for our needs.

46. This information is trustworthy.

47. This information is relevant for our work.

48. Using this information increases the value of our work.

49. This information is presented in a compact form.

50. This information is appropriate for our work.

(Screen break here, with scale repeated on the next screen, along with the words, "Almost finished!".)

51. The meaning of this information is easy to understand.

52. This information is credible.

53. This information covers the needs of our tasks.

54. The representation of this information is compact and concise.

55. This information adds value to our tasks.

56. The measurement units for this information are clear.

57. This information is objective.

58. This information can only be accessed by people who should see it.

59. This information is sufficiently timely.

60. This information is easy to combine with other information.

61. This information is represented in a consistent format.

62. This information is easily obtainable.

63. This information comes from good sources.

64. This information is quickly accessible when needed.

65. This information has sufficient breadth and depth for our task.

66. This information presents an impartial view.

67. This information is applicable to our work.

68. This information is sufficiently up-to-date for our work.

69. This information is reliable.

When you have made your selections, click the "continue" button.

Part V – Classification Data

You are almost finished! Questions in the final section will be used for classification and analysis by subgroups only. Please provide the appropriate response to each item.

1. Which of the following best describes the type of organization you work for?
    a. For-profit.
    b. Non-profit.
    c. Governmental agency.
    d. Other.
2. Which of the following best describes the industry in which you work or are most closely associated?
    a. Manufacturing
    b. Engineering
    c. Transportation

d. Hospitality

e. Health care

f. Education

g. Other

3. What is the primary business activity at your location?

a. Banking

b. Insurance

c. Research and development

d. Manufacturing

e. Transportation

f. Hospitality

g. Health care

h. Retail

i. Education

j. Other

4. How many employees work at your location?

a. Under 100

b. 101 to 1,000

c. 1,001 to 10,000

d. Over 10,000

5. How many employees are there in your entire organization?

a. Under 100

b. 101 to 1,000

c. 1,001 to 10,000

d. Over 10,000

6. What are your organization's approximate annual revenues in U.S. dollars or equivalent (approximate budget if non-profit or governmental)?

   a. Under $1 million

   b. At least $1 million, less than $10 million

   c. At least $10 million, less than $100 million

   d. At least $100 million, less than $1 billion

   e. Greater than $1 billion

7. How long have you been with this organization?

   a. Less than 1 year

   b. At least 1 year, less than 5 years

   c. At least 5 years, less than 10 years

   d. At least 10 years, less than 20 years

   e. 20 years or more

8. How long have you been in this industry?

   a. Less than 1 year

   b. At least 1 year, less than 5 years

   c. At least 5 years, less than 10 years

   d. At least 10 years, less than 20 years

   e. 20 years or more

9. Which of the following best describes your job title or function?

    a. Executive
    b. Management
    c. Consultant
    d. Engineer
    e. Researcher
    f. IT Professional
    g. Professional (other than IT)
    h. Administration
    i. Other

10. Which of the following best describes your highest level of education?

    a. High school or equivalent
    b. Technical school certification
    c. Associate's degree
    d. Bachelor's degree
    e. Master's or Specialist's degree
    f. Doctoral degree or beyond

When you have made your selections, click the "submit" button.