Information Quality Assessment and Effects on Inventory Decision-Making

Mouzhi Ge

A dissertation submitted in fulfilment of the requirements for the award of Doctor of Philosophy (Ph.D.)

Supervisor: Dr. Markus Helfert

School of Computing
Dublin City University

September 2009

Declaration

I hereby certify that this material, which I now submit for assessment on the programme of study leading to the award of Doctor of Philosophy is entirely my own work, that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge breach any law of copyright, and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of my work.

Signed: _________________ (Candidate) ID No.: ___________ Date: ______________

Acknowledgements

First of all, I would like to sincerely thank my supervisor, Dr. Markus Helfert, for his guidance, support and encouragement. Without his supervision this thesis would have never existed. As an old Chinese proverb is saying "once a teacher for me, always a teacher for me", I am so gratifying and enjoying to have ironed his "always a teacher" in my entire life.

I really want to thank my family without whom I would never have gone this far. In the rest of my life, I can never finish paying back the love and care I owed to my parents. Besides, my sincerest thanks go to my beloved sister, Dr. Ge Wei, who has all the time supported me in different walks of my life. I am so lucky to have a sister like her.

Many thanks also go to all my dear friends, who are Xu Qing, Thoa Pham, Howard Duncan, Mary Levis, Owen Foley, Ge Jun, Shi Jie, Jiao Decai, Ma Yanjun, Guo Yuqing, Wang Shuo, and Luan Haiying without whom life is never so easy and colourful in Ireland, in particular to sweetie Xu Qing who is always there for me.

Last but not least, I would like to thank the School of Computing at the Dublin City University, who funded this research and provided me a great working atmosphere. It provided me a precious opportunity to learn not only the knowledge but also the western culture.

Table of Content

Abstract

Information quality has become a critical concern to the success of organisations. Numerous business initiatives have been delayed or even cancelled, citing poor-quality information as the main reason. Previous research indicated that understanding the effects of information quality is critical to the success of organisations. However, little research has been done to analyse the effects of information quality in an organisational context.

In order to address this drawback, the objective of this thesis is to systematically analyse the effects of information quality on decision-making. In order to achieve this objective, we propose a practical assessment framework that allows us to measure information quality dimensions and categories. This framework was validated using a real-world database. With the help of this assessment framework we gained in-depth insights into the effects of information quality on decision quality. Our results showed that the categories of intrinsic and contextual information quality are positively related to decision quality. However decision quality is not significantly affected by representational information quality. It is also found that in contrast to consistency, increasing information accuracy and completeness can significantly improve decision quality.

From our results we concluded that not all the aspects of information quality are equally effective for the improvement of decision quality. Decision-makers could decide to pay little or no attention to the improvement of representational information quality and information consistency. This finding will directly reduce the cost of information quality improvement. For practical implementations, our results concluded a validated framework that allows software engineers to implement assessment. Comparing our framework with other common frameworks that require software engineers to understand information quality theory, our framework helps software engineers to follow a step-by-step procedure to build an application of information quality assessment. It will directly increase software engineers' work efficiency.

List of Figures

List of Tables

Chapter 1: Introduction

In an age characterised by information, information quality is a critical concern for a wide range of organisations. More than ever before, many organisations focus their main business on the provision of valuable information. In order to assure that such information is of a high standard, organisations face the considerable challenge of controlling information quality. Therefore information quality management has become an essential component of organisational management.

1.1 Importance of Information Quality

Case studies concerning information quality problems are frequently documented. Such information quality issues may not only cause inconvenience in everyday life but also potentially generate harmful disasters. For example, on the 28th January 1986, NASA launched the space shuttle *Challenger*. Seconds after lift-off, the shuttle exploded and killed the seven astronauts on board. The Presidential Commission investigated the *Challenger* accident and found that NASA's decision-making process was based on incomplete and misleading information. Just two years later in July 1988, U.S. Navy Cruiser USS *Vincennes* shot down an Iranian commercial aircraft and killed its 290 passengers. Officials who investigated the *Vincennes* accident admitted that poor quality information was a major factor in the flawed *Vincennes* decision-making process. Fisher and Kingma (2001) carried out an in-depth analysis of the *Challenger* accident and the *Vincennes* accident and concluded that the explosion of space shuttle Challenger and the shooting down of an Iranian airbus by the USS Vincennes were the result of information quality problems and information quality management errors. Yet not only in the space and military industries but also in our daily life, information quality problems can be found to be severe. For instance, Pirani (2004) reported that one piece of wrong biopsy information caused a patient's death in an Australian hospital. The above case studies demonstrate that information quality is a vital issue in both industry and everyday life.

Information quality problems have been investigated in a substantial body of literature. The results of a number of investigations are summarised in the following: More than 60% of 500 medium-size firms were found to suffer from information quality problems (Wand and Wang 1996); it is estimated that 1% to 10% of data in organisational databases is inaccurate (Klein et al. 1997); an industrial information error rate up to 30% is considered typical and it is often reported that the error rate rises to 75% (Redman 1996); 70% of manufacturing orders are assessed as of poor information quality (Wang et al. 2001); 40% of the credit-risk management database in a New York bank was found to be incomplete (Wang and Strong 1996); and between 50% and 80% of computerised U.S. criminal records are estimated to be inaccurate, incomplete and ambiguous (Strong et al. 1997). From the figures above, we are able to observe that information quality problems are pervasive in many organisations.

In addition, these information quality problems are often associated with high costs. For example, poor information quality management costs more than $1.4 billion annually in 599 surveyed companies (Olson 2003): a telecommunications company lost $3 million because of the poor information quality in customer bills (Huang et al. 1999); it is estimated that poor information quality results in 8% to 12% loss of revenue in a typical enterprise, and informally estimated to be responsible for 40% to 60% of expense in service organisations (Redman 1998). These figures demonstrate that managing information quality really can be beneficial to organisations.

Observing the case studies and figures above, we are able to conclude that information quality problems are pervasive, costly and even disastrous. High quality information is not only important to organisations but to our everyday existence.

1.2 Fundamental Concepts of Information Quality

Since more people have become aware of the importance of information quality, research in the field has developed extensively over the last two decades. Based on the foremost works, in this section we introduce the fundamental concepts used in this thesis. As information quality research is a branch of information system research, first it is necessary to illuminate the notion of information systems. In some literature, the information system is often referred to as a computerised database system. In this thesis, the term ‘information system’ covers collecting, processing, distributing and using data by organisational processes or people (Strong et al. 1997). It is also referred to as “larger information system’s context” by Huang et al. (1999) and Strong et al. (1997).

Within the context of information systems above, a novel approach in information quality research is to observe information manufacturing as product manufacturing (Ballou et al. 1998). Product manufacturing is a processing system that transforms raw materials into physical products. Analogously, information manufacturing can be considered as a process transforming raw data into information products. Therefore, information products can be defined as the results of manufacturing raw data into valuable information (Pierce 2004). In this thesis, we define information manufacturing system as a system that operates on raw data to create information products. The analogy between product manufacturing and information manufacturing is described in Figure 1.

	Product Manufacturing	Information Manufacturing
Input	Raw Materials	Raw Data
Process	Assembly Line	Information System
Output	Physical Products	Information Products

Figure 1: Analogy between product manufacturing and information manufacturing (Huang et al. 1999)

Using the information manufacturing concept, treating information as a product can be a key strategy for organisations attempting to obtain a competitive advantage. However, some organisations still consider information as a mere by-product and are not aware of the importance of treating information as a product. Consequently this may lead to a variety of organisational losses such as making incorrect decisions and losing business opportunities (Wang et al. 1998). For example, an investment company sent a large amount of direct mail to the wrong target customers. This company lost its market share and incurred the need for significant reworking in their customer service. Once this company treated the information required for direct marketing as an information product, they were able to detect the root cause of the mail list problem and subsequently were able to improve their customer service (Huang et al. 1999). Therefore when companies cultivate the concept of information product into their organisational cultures, it can generate business growth and competitive advantage in the marketplace. In order to manage information as a product, it is proposed one follows the four principles outlined by Wang et al. (1998).

1. Understand consumer’s information needs.
2. Manage information as the product of a well defined production process.
3. Manage information as a product with a lifecycle.
4. Appoint an information product manager to manage the information processes and the resulting product.

To manage the quality of information products, Wang (1998) proposed the total data quality management (TDQM) model to deliver high quality information products. This model is adapted from Deming's *plan*, *do*, *check* and *act* cycle (Deming 1982) and consists of four phases: define, measure, analyse and improve. The model is shown in Figure 2.

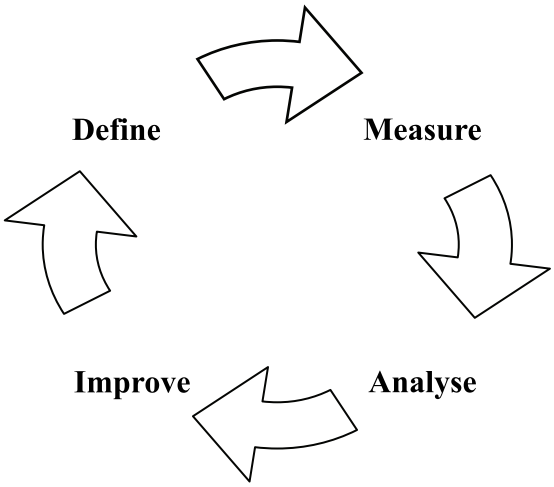


Figure 2: Total data quality management cycle (Wang 1998)

The definition phase is for determining the characteristics of information products, information quality requirements, and how information products are produced in the information manufacturing system. In this phase, we need to identify who assesses the quality of information products and which information quality dimensions are used in such assessment. The measurement phase is for assigning numerical or categorical values to information quality dimensions in a given setting (Ge and Helfert 2007). This phase consists of different measuring methods that can be used to assess information quality. According to the assessment result, the analysis phase is for discovering the root cause of information quality problems and strategising an effective scheme for information quality improvement. Once the analysis phase is finished, considering budgetary constraints and resource allocation, the improvement phase is concerned with improving the quality of information products for intended use. As shown in Figure 2, the four phases constitute a continuous information quality management cycle, indicating that organisations need to continuously implement the TDQM cycle and cultivate information quality concepts into their organisational culture.

In the TDQM cycle, the measurement phase is critical, because one cannot manage information quality without having measured it effectively and meaningfully (Stvilia et al. 2007). In order to measure information quality, information quality dimensions must be determined. To this end, Wang and Strong (1996) used an exploratory factor analysis to derive 15 information quality dimensions, which are widely accepted in the field of information quality research. In this thesis, we define information quality dimensions as a set of attributes that represent different constructs of information quality.

In order to organise these information quality dimensions, Wang and Strong (1996) classify information quality into 4 categories: intrinsic, contextual, representational and accessibility. Intrinsic information quality consists of context-independent dimensions. In contrast to intrinsic information quality, contextual information quality highlights the dimensions, which need to be considered in an application context. Representational information quality concerns whether the information is presented in an easily interpretable, understandable, concise and consistent way. Accessibility information quality emphasises that the data needs to be accessible yet still secure. Each information quality category contains 2 to 5 dimensions which form a hierarchical framework of information quality, as described in Figure 3.

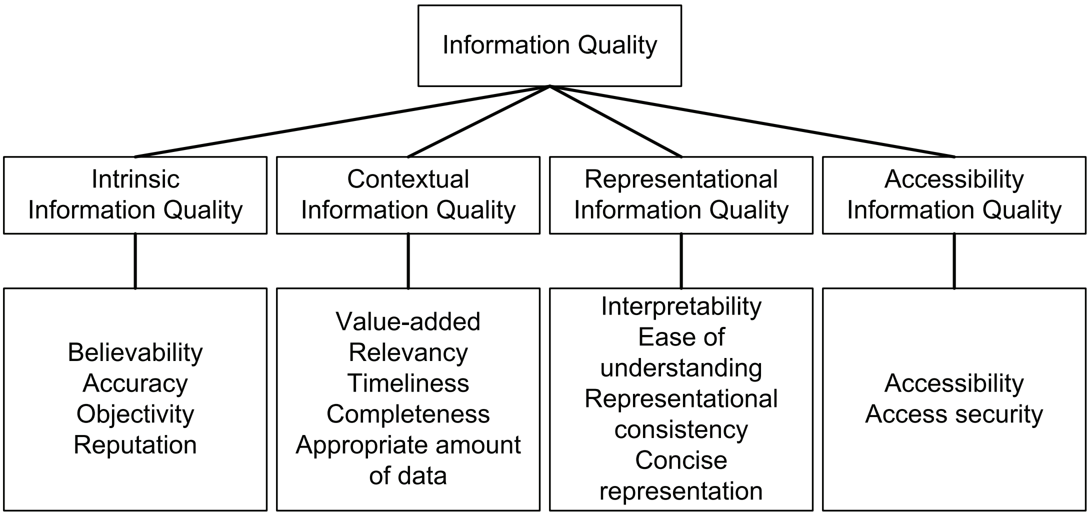


Figure 3: Information quality framework (Wang and Strong 1996)

Intrinsic information quality comprises 4 dimensions: believability, accuracy, objectivity and reputation. Contextual information quality contains 5 dimensions: value-added, relevancy, timeliness, completeness and appropriate amount of data. Representational information quality includes 4 dimensions: interpretability, ease of understanding, representational consistency and concise representation. Finally, Accessibility information quality consists of 2 information quality dimensions: accessibility and access security. In this thesis, we define information quality category as a group of dimensions that capture the similar essence of information quality.

In summary, 4 fundamental concepts are introduced in this section: information manufacturing, information products, the TDQM cycle and a hierarchical framework of information quality.

Compared with physical product manufacturing, information manufacturing specifies the context for the construction of information products. This concept is used to introduce the idea of information products and distinguish them from raw data. In Chapter 5, we consider information products and raw data as two separate measurable objects and propose individual assessment approaches. To manage information products, the TDQM cycle is introduced. This is used to provide an overview of information quality management and address the position of assessment in information quality management in Chapters 2 and 5. Finally, a hierarchical framework of information quality is introduced to specify information quality categories and dimensions. This framework allows us to refine the effects of information quality to the effects of information quality categories and dimensions in Section 1.3.

1.3 Research Question and Objective

After introducing the fundamental concepts for this research, we detail the research questions. Based on our literature review in Chapter 2, we found that more research needs to be done to investigate the relationship between information quality and decision-making.

Recent information system literature indicates an increasing trend in studying the relationship between information quality and decision-making. In these studies, information quality is found to be a crucial determinant in the making of high-quality decisions. In other words, low-quality information may result in incorrect decisions and in turn may cause the failure of business initiatives. Although making correct decisions is clearly dependent upon information quality (Belardo and Pazer 1995, Ballou and Pazer 1995, Raghunathan 1999), how exactly information quality affects decision-making is not completely known (Chengalur-Smith et al. 1998, Fisher et al. 2003). *Therefore it would be valuable to further investigate the relationship between information quality and decision-making.*

Based on the discussion above, we derived our primary research question:

- **How does information quality affect decision-making?**

Since information quality can be divided into different categories and each category contains several dimensions, this major research question can be refined to:

- How do information quality categories affect decision-making?
- How do information quality dimensions affect decision-making?

In order to investigate the effects of information quality on decision-making, we first need to measure and control levels of information quality. However, in current information quality research, a variety of results has arisen when deriving information quality dimensions. For example, Wang and Strong (1996) used an exploratory factor analysis to tailor 118 information metrics to 15 information quality dimensions. The loading of each dimension was 0.5 or greater (with a sample size of 355). This result indicated the independence of information quality dimensions and supported both convergent validity and discriminant validity. Based on Wang and Strong (1996)'s work, Lee et al. (2002) carried out a correlation analysis of the 15 information quality dimensions with 261 subjects. Their result found high correlations among the dimensions, indicating that information quality dimensions are not inherently independent. With this result, they concluded that information quality is a multi-dimensional concept but a single phenomenon. *Due to these mixed results of dependency among information quality dimensions, a confirmatory factor analysis is needed to confirm the correlations between information quality dimensions.*

Once the correlations between information quality dimensions have been confirmed, a practical assessment methodology needs to be developed. However, most information quality assessments fall into either objective or subjective methodology (Pipino et al. 2002). Objective assessment methodologies frequently employ software to automatically measure information quality using a set of quality rules. On the other hand, subjective assessment methodologies typically use surveys or interviews to measure information quality by information consumers. The advantage of objective assessment is to allow the automatic processing of large datasets and the obtaining of a single result, whereas subjective assessment typically focuses on a data sample and generates different assessment results from different users. The advantage of subjective assessment is that it allows the measurement of information quality by a comprehensive set of dimensions. For example, certain dimensions such as believability and reputation are not suitable for objective assessments. *Therefore, considering the* *advantages of both objective and subjective assessments, a comprehensive framework is needed for information quality assessment.*

Based on the above discussion, we need to investigate the following two research questions to control information quality.

- What are validated information quality dimensions?
- How can one effectively assess information quality?

In light of the research questions above, the main objective of this thesis is to clarify the relationship between information quality and decision-making. To achieve this principal objective, we need to provide a valid and reliable set of information quality dimensions, and to propose a practical framework for information quality assessment.

1.4 Thesis Organisation

The remainder of this thesis is organised as follows: Chapter 2 presents an extensive review addressing the limitations of current information quality research and highlighting the significance of our proposed research questions. To investigate our research questions, Chapter 3 presents the research methodologies employed in this thesis. Based on the review and research methodologies, Chapter 4 proposes research hypotheses and the experimental design. In the experimental design, the crucial component is information quality assessment. Thus Chapter 5 proposes and validates an information quality assessment framework. Following data collection in the experiment, Chapter 6 provides data analysis for the effects of information quality on decision-making. Finally, Chapter 7 concludes the thesis by summarising the research findings and outlining possibilities for future research. The specific organisation of each chapter is presented as follows.

Chapter 2: In Chapter 2 we present an extensive review in the field of information quality research. In order to provide an overview, we first provide an analysis of information quality research within information systems and decision-making. Subsequently, we analyse the definitions of information quality. In order to provide a detailed review, we identify three periods of information quality development. We focus on the third period and give an overview of information quality research within this period. Being a key element in information quality management, information quality assessment is reviewed by analysing information quality problems, dimensions and assessment methodologies. Based on our review, we highlight the research questions of this research and propose a research agenda for future investigations.

Chapter 3: In order to investigate the research questions, Chapter 3 discusses the research methodology and design used in this study. In presenting the nature of research, we first introduce the concept of research, research classifications and research processes. Following on from that, we evaluate the research methodologies in information system research and address the methodologies used in this thesis. To further describe our research design, we then discuss survey design, experimental design, and data analysis methods. As crucial factors in research design, validity and reliability are discussed in both measurement development and experimental design. Finally, this chapter concludes by outlining our solutions to typical methodological issues.

Chapter 4: Following the research process outlined in Chapter 3, this chapter proposes our research hypotheses. The hypotheses address the effects of information quality categories and dimensions on decision-making. In order to test the hypotheses, two experiments are proposed. One is to investigate the relationship between information quality categories and decision-making, whilst the second is to investigate the relationship between information quality dimensions and decision-making.

Chapter 5: In our experimental design, one central objective is to measure and control levels of information quality. Therefore, Chapter 5 proposes a framework that consists of two major components: validated dimensions and assessment methodology. In order to validate information quality dimensions, we adopt 17 dimensions and perform a survey with these dimensions. The survey result is analysed by a confirmatory factor analysis and a Cronbach calculation. The analysis derives a set of valid and reliable dimensions. As information quality assessment may be carried out in a single database or multiple databases, we propose assessment methodologies that can measure information quality for either. In order to validate our assessment approaches, we apply the methodology to assessing information quality in an example dataset and also a multi-database simulation

Chapter 6: Following the data collection in the experiments, Chapter 6 analyses the effects of intrinsic, contextual and representational information quality on decision-making and similarly the effects of accuracy, completeness and consistency. In order to confirm the experimental results, we organise a workshop to demonstrate the experiment and confirm our analysis result through interviews.

Chapter 7: In the final chapter, we summarise the results of our work and provide guidance for academic research and practical application. We present in detail our conclusions relating to each aspect of our work. The critical review and limitations of the work are also discussed. Finally, we outline the possible extensions for future research.

Chapter 2: Literature Review

In this chapter, we review information quality research over recent decades. Since information quality research is a branch of information system research, we first analyse the relationship between information quality research and information system research. Next we summarise a variety of contexts in which information quality research is applied. Following our research objective, we focus on literature related to information quality research in decision-making. After that, we discuss the definitions of information quality and the development of information quality research. Our literature review shows that information quality research has experienced three major phases. We focus on reviewing related works in the third phase. In this phase, we provide an overview of information quality management. Following this overview, we review information quality assessment as an important component in information quality management. Finally, we propose a research agenda to outline the result of this review and discuss the significance of our research.

2.1 Information Quality and Information Systems

Most influential information quality research originated from information system research. Information system researchers initially identify and employ a set of dimensions to address the information quality problems within information systems. As information quality awareness and requirements have increased, researchers have begun to focus on information quality frameworks (Ballou and Pazer 1985; Wang et al. 1995), information quality dimensions (Wang and Strong 1996; Pipino 2002), information quality assessment (Wand and Wang 1996; English 1999) and information quality management (Wang 1998; Huang et al. 1999).

In this section, we select 12 of the most cited articles for our review. The 12 articles are organised along a timeline in Figure 4. The works on information systems and information quality are distinguished by the research objective and discussion focus of the article. As illustrated in Figure 4, information quality research is developed from information system research. After information quality research matured into its own research field, information quality theory was applied back to the context of information systems.

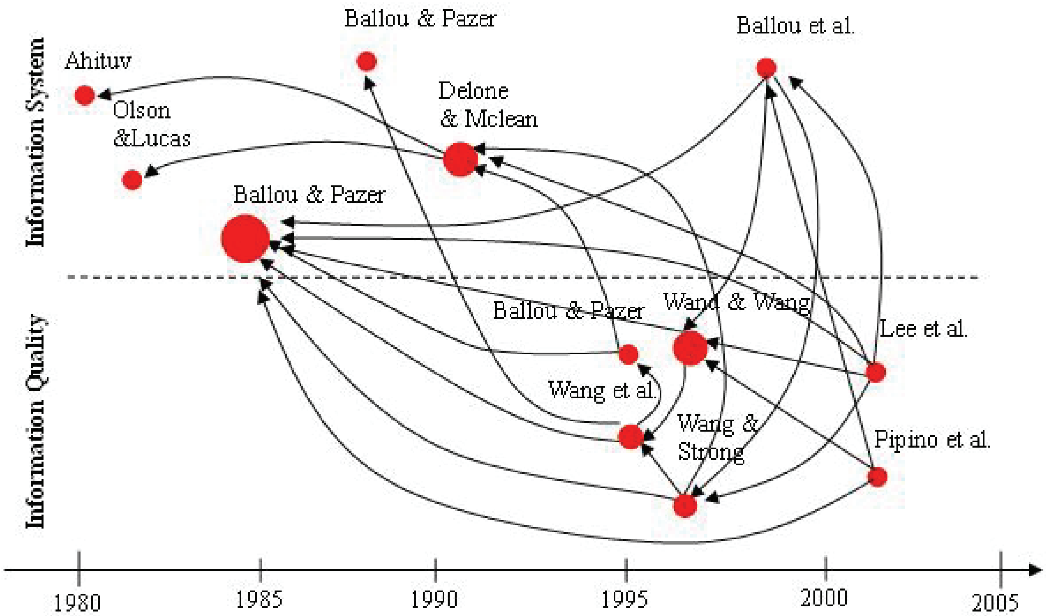


Figure 4: Citation relationship between information quality research and information system research

The number of citations is represented graphically in Figure 4. The size of the node represents the number of citations for this particular article. We can observe that the work of Ballou and Pazer (1985) is cited by all the selected information quality articles. Ballou and Pazer contributed a pioneering work for information quality research and they are the authors with the greatest impact on the study of information quality in information system research. The work of Delone and Mclean (1992) reviewed information system literature and is cited by subsequent information quality research. Hence Delone and Mclean have done a significant work that further derives information quality from information systems. Figure 4 also has shown that extensive information quality research was produced around 1995.

2.2 Contextual Information Quality

With the development of information quality research, researchers have applied information quality to various organisational contexts. In Table 1, we summarise representative publications for each application context within the period 1996-2006.

Application Context	Publications
Database	Redman (1996)
Information Manufacture Systems	Ballou et al. (1998)
Accounting	Kaplan et al. (1998)
Marketing	Teflian (1999)
Data Warehouse	English (1999)
Decision-Making	Chengalur-Smith et al. (1999)
Healthcare	Berndt et al. (2001)
Enterprise Resource Planning	Xu et al. (2002)
Customer Relationship Management	Helfert and Heinrich (2003)
Finance	Amicis and Batini (2004)
E-business	Xu and Koronios (2004)
World Wide Web	Knight and Burn (2005)
Supply Chain Management	Li and Lin (2006)

Table 1: Information quality application contexts (1996-2006)

From a variety of application contexts, we decided to focus on one particular context related to our research: decision-making and supply chain management. Previous research has reported that high-quality decisions are dependent on the quality of information contributing to those decisions. However, aside from information quality, other factors may also influence decision-making. In order to provide further insight into factors influencing decision making, in the following paragraph we analyse influencing factors affecting the relationship between information quality and decision-making.

Whilst considering other influencing factors on decision-making, many researchers have demonstrated the influence of information quality on decision-making. For example, regarding the factor of information overload, Keller and Staelin (1987) investigate how decision effectiveness is affected by both information quality and information quantity. By employing social interaction and decision aids, Sage (1991) imply that the major purpose of social interaction and decision aids is to enhance information quality, and through this, the quality of decision-making. Based on crisis decision environments and decision aids, Belardo and Pazer (1995) propose a model to present the relationship between information quality and decision quality. Considering decision strategy and decision costs, Ballou and Pazer (1995) analyse the trade-off between two information quality dimensions (accuracy and timeliness) in decision-making. Taking task complexity and decision strategy into account, Chengalur-Smith et al (1999) show that including information regarding information quality can impact upon the decision-making process. Considering the impact of time, expertise, task complexity and decision strategy, Fisher et al (2003) develop an experiment to address the utility of information quality information in decision-making. In a dynamic decision environment, Shankaranarayan (2003) proposes a virtual business environment to address the role of information quality management in dynamic decision environments. From the perspective of task complexity and information quality categorisation, Jung and Olfman (2005) find that contextual information quality significantly contributes to decision performance. From the above literature, we can observe that various factors are considered when researchers study the effects of information quality on decision-making. This means that when we investigate the effects of information quality on decision-making, various decision scenarios can be developed through the control of various influencing factors. By illustrating the literature and the influencing factors, we have structured the above discussion into Table 2.

Author \ Factor	IQ	Information overload	Decision aids	Decision strategy	Task complexity	Expertise	Time	Environment
Keller and Staelin (1987)	×	×						
Sage (1991)	×		×					
Ballou and Pazer (1995)	×	×						×
Belardo and Pazer (1995)	×	×	×					×
Ahituv et al. (1998)	×					×	×	
Chengalur-Smith et al. (1999)	×			×	×			
Shankaranarayan et al. (2003)	×							×
Fisher et al. (2003)	×			×	×	×	×	
Jung and Olfman (2005)	×				×			

Table 2: Factors influencing decision-making in information quality research

Table 2 provides an overview of extraneous variables that need to be considered in the experiments of information quality research. Methods for controlling these variables are outlined in Chapter 3.

2.3 Definition of Information Quality

To further review the topic of information quality, we need firstly to define the concept of information quality. Before defining information quality, the notions of *data*, *information* and *quality* must initially be determined. As there has been a long discussion of defining data, information and quality, we mainly focus on the literature that is used in information quality research.

Different definitions of data can be found in literature. Redman (2001) proposes that data consists of two components: data model and data value. *Data model* represents entities in the real world, such as STUDENT, COURSE and TEACHER. These entities could contain different attributes and relationships with other entities. For example, STUDENT may contain NAME, GENDER and CLASS as its attributes, and STUDENT may choose some COURSE as its relationship. *Data value* is the representation of facts and can be assigned to the attributes of entity (English 1999). For instance, JAMES can be the data value of STUDENT NAME. Therefore for a single data, data can be represented by <entity, attribute, value>. Since the value of the data reflects facts in real world, Tsichritzis and Lochovsky (1982) propose that data is based on a modelling of the real world. In this thesis, we define data as model of the real-world facts and are typically represented by <entity, attribute, value> in the database.

In information quality research, data is often considered to be the raw material for information manufacturing. This indicates that information is the finished product resulting from data processing. Data processing is effectively to present the data with meanings in contexts (English 1999, Fisher et al. 2006). These meanings can be established or intensified by relationships between different data. For example, one data is “100 students chose this course” and the other data is “10 students chose this course”. When the two data are combined to “100 students chose this course last semester and only 10 students chose this course this semester”, more meanings can be ascertained. In this thesis, we define information as a presentation of the processed data in context. According to our definitions, the relationship between the real world, data and information is shown by Figure 5.

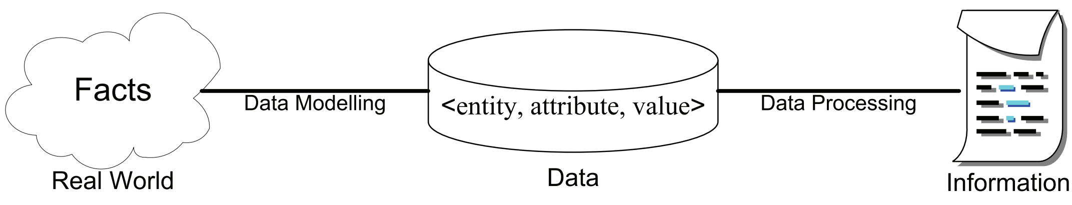


Figure 5: Relationship between real world, data and information

From the information manufacturing point of view, data and information are not absolute concepts since one piece of information can be simultaneously a finished product and also raw data material. Therefore we decided to use the context as the key determinant in distinguishing data and information. Typically, data can exist without any context but information needs to exist in a concrete context.

The concept of quality has been discussed extensively over recent decades. Various definitions are to be found in a wide range of literature. Gilmore (1974) defines quality as *conformance to specifications*. This definition is relatively straightforward and frequently used in manufacturing industries. It facilitates measurement and increases measuring efficiency. Organisations can determine the quality of products by measuring how well the product conforms to an established specification. Also, the measuring procedure can be automatically implemented. However it fails to capture the customer's view on product performance. To compensate for the disadvantage of this definition, Gronroos (1983) defines quality as *meeting and/or exceeding customer's expectations*. This definition is especially prevalent in marketing research and the service industries. Following this definition, researchers posit that it is the customer who is the ultimate judge of the quality of a product/service. Thus organisations can make a quick response to market changes. However, it is difficult to measure the extent to which a product/service meets and/or exceeds the customer's expectation. Since different customers may assign different values to product/service attributes, coordinating and unifying the various quality results are the principal difficulties facing this definition. To provide a general definition, Juran (1974) introduces the definition of quality as *fitness for use*, which is used to measure the extent to which a product successfully serves its intended use. As this is the definition widely used in information quality research, we adopt *fitness for use* as the definition of quality in our research.

Based on the discussion of data, information and quality, information quality can be generally defined from two perspectives: information and user. From the information perspective, information quality is defined as information that meets specifications or requirements (Kahn and Strong, 1998). Some researchers argue that ultimately it is the information consumer who will judge whether or not an information product is fit for use. Thus from the user's perspective, information quality is defined as information that is fit for use by information consumers (Wang and Strong, 1996). In this thesis, we include both perspectives and define information quality as the information that meets the specifications and is also fit for use by information consumers.

Reviewing literature on information quality, we have found the following definitions:

- "Information quality is defined as information that is fit for use by data consumers" (Wang and Strong, 1996)
- "Information quality is defined as the information that meets specifications or requirements" (Kahn and Strong, 1998)
- "Information has quality if it satisfies the requirements of its intended use " (Olson, 2003)
- "Information quality can be thought of as information's inherent usefulness to customers in assessing the utility" (Keller and Staelin, 1987)
- "Information is of high quality if it is fit for its intended uses in operations, decision-making, and planning. Information is fit for use if it is free of defects and possesses desired features. (Redman, 2001)
- "Information quality is defined as information which consistently meets knowledge worker's and end-customer's expectations" (English, 1999)
- "Information quality is defined as the degree to which information has content, form and time characteristics which give it value to specific end users" (Brien, 1991)
- "Information quality is the characteristic of information to meet the functional, technical, cognitive and aesthetic requirements of information producers, administrators, consumers and experts" (Eppler, 2006)

- "Information quality is defined as information which satisfies criteria of appreciation specified by the user, together with a certain standard of requirements" (Salaün and Flores, 2001)

To organise these definitions, we classify them according to the information and user perspective in Table 3.

	information perspective	user perspective
Wang and Strong (1996)		x
Kahn and Strong (1998)	x	
Olson (2003)		x
Keller and Staelin (1987)		x
Redman (2001)		x
English (1999)		x
Brien (1991)		x
Eppler (2006)		x
Salaün and Flores (2001)	x	

Table 3: Analysis of information quality definitions

From Table 3, we could observe that most selected definitions are derived from the user perspective. In these definitions, we found that Wang and Strong (1996)'s definition is widely accepted by subsequent research. Their definition stems from user perspective and can directly capture the user's opinion on the information. However, definitions from the information perspective are practical for objective assessment. As we differentiate information products from raw data, we consider the quality of information products from the user perspective and the quality of raw data from the information perspective.

2.4 Information Quality Development

In order to thoroughly understand the concepts of information quality, we decide to review the development of information quality research. According to our review, we found three major phases in information quality research.

The first phase is prior to 1980. Most information quality research in this phase remains isolated. In the late 1960s, statisticians were the first to study information quality problems (Batini and Scannapieco 2006, Lee et al. 2006). They investigated data duplication issues in statistical datasets using mathematical theories. The major researchers were accounting professionals. For example, Yu and Neter (1973) developed a quantitative model to measure data errors in financial information systems. Cushing (1974) proposed a mathematical approach to detect and prevent data errors in accounting internal control systems. Hilton (1979) provided an illustrative analysis to show the effect of information accuracy and timeliness on accounting. In the mean time, psychology and management researchers also began to consider the concept of information quality. Streufert (1973) used an experimental methodology to investigate the effects of information relevance on decision-making. Zmud (1978) carried out an empirical study to derive certain information dimensions - relevant, accurate, factual, quantity, reliable, timely, arrangement, readable and reasonable. Most of these dimensions are confirmed as information quality dimensions by subsequent research (e.g. Wang and Strong 1996, Lee et al. 2002). Reviewing the information quality research in this phase, these pioneering works have either stated or implied the concept of information quality and investigated certain information quality dimensions.

The second phase approximately lasts from the 1980s to the early 1990s. In this phase, information quality research is widespread but not yet systematic. Researchers begin to focus on exploring information quality dimensions, assessment methodologies and improvement strategies.

In this phase, different sets of information quality dimensions were explored. For example, Brodie (1980) proposed that information quality contained three distinct components: information reliability, logical integrity and physical integrity. Olson and Lucas (1982) used appearance and accuracy to measure data quality in office automation information systems. Morey (1982) considered information quality to be information accuracy and proposed three information accuracy measures in the context of information systems. O' Reilly (1982) investigated the effects of information quality on the use of information sources. In his study, information quality is measured by accessibility, accuracy, specificity, timeliness, relevance and amount of information. Ballou and Pazer (1985) considered accuracy, completeness, timeliness and consistency in the measurement of information quality in multi-input and multi-output information systems. Laudon (1986) identified completeness, accuracy and ambiguity as information quality dimensions for criminal-record systems. Observing the works above, it is found that different information quality dimensions can be derived from different contexts.

Additionally, in this phase different assessment methodologies were proposed. For example, Paradice and Fuerst (1991) developed a quantitative measure to formulate the error rate of stored records in information systems. O' Reilly (1982) proposed 18 questions with which users could determine information quality. One prominent work in this period is by Agmon and Ahituv (1987). They proposed an approach to assess information quality consisting of three elements: (1) internal assessment, to determine the intrinsic characteristics of data using widely accepted criteria, (2) assessment from the users' perspective, which focuses on the expectations and requirements of the user, and (3) assessment by comparison between database and reality. This element is designed to locate data deficiencies by comparing database systems and real world systems. This approach provides an important indicator for our research in information quality assessment, which is discussed in Chapter 5.

In addition to assessing information quality, a number of researchers also investigated the methods for information quality improvement. Whilst some researchers attempted to completely eradicate data errors from databases (Morey 1982, Janson 1988), others suggest allocating limited resources in an optimal way to enhance information quality (Ballou and Tayi 1989, Paradice and Fuerst 1991). All these methods have provided directions for the subsequent research.

As evidenced by the discussion above, in the second phase researchers concluded that information quality is a key determinant for information system success (DeLone and McLean 1992). Since different sets of dimensions are developed according to different contexts, such as reporting system (Ahituv 1980) and office automation information system (Olson and Lucas 1982), there is no comprehensive set of information quality dimensions in this phase. Furthermore, there appears to be no single accepted definition of information quality. Some researchers (e.g. Morey 1982) consider information quality to be information accuracy, while other researchers (e.g. Keller and Staelin 1987) define information quality in terms of usefulness to consumers. Finally, most of the works in this phase are not validated by practical application.

The third phase can be identified from the middle 1990s to the present day. In this phase, information quality research becomes intensive, systematic and empirical. Therefore, the amount of information quality papers significantly increases, across a wide range of journals and conferences. From 1995 to 2008, more than 15 information quality books were published. These books have addressed different aspects of information quality research. Three information quality journals have been launched so far: the Data Quality Journal in 1995, the International Journal of Information Quality in 2007 and the ACM Journal of Data and Information Quality in 2008. In addition, many leading database and information system conferences such as Special Interest Group on Management of Data (SIGMOD), Very Large Data Bases (VLDB) and Conference on Advanced Information Systems (CAiSE) have included information quality as one of their themes. Furthermore, since 1996, the International Conference on Information Quality (ICIQ) is held annually to provide a forum for researchers and practitioners to present research findings and exchange knowledge in the field of information quality. Beyond research developments in academia, industry and government have also begun to pay attention to information quality issues. For example, Navesink Consulting Group was formed in 1996 and provided information quality solutions and services to other organisations. In 2001, the US president signed information quality legislation into law (Batini and Scannapieco 2006). These newly founded companies and government operations clearly indicate the empirical application of information quality research. Due to the significant research development in this phase, the reminder of this chapter focuses on analysing literature of this phase.

2.5 Overview of Information Quality Management

As most research in the third phase focuses on contributing knowledge to information quality management, this section provides an overview of information quality management. From our review on information quality literature, we found that information quality management has merged three realms of management: quality management, information management and knowledge management. We describe the merging trend of the three research domains by Figure 6.

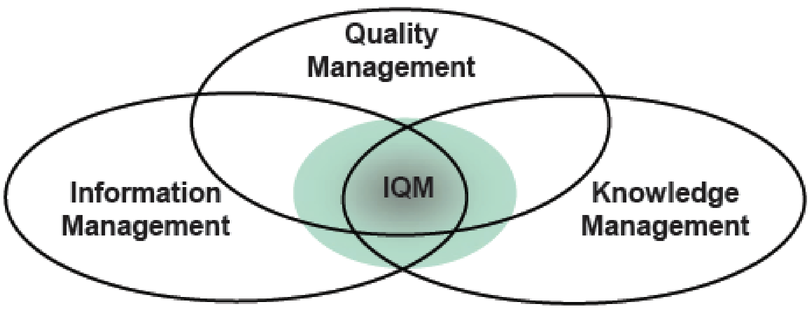


Figure 6: Information quality management from quality, information and knowledge management perspective

In order to analyse the current research on information quality management, we have selected three frequently cited literatures. These literatures introduce quality management, information management and knowledge management into information quality management.

- Quality Perspective: Using the principle "manage your information as a product", Wang (1998) proposed a total data quality management framework. It consists of four stages: define, measure, analyse and improve. The objective of TDQM is to deliver high-quality information products to information consumers.

- Information Perspective: With the principle "integration, validation, contextualisation, activation", Eppler (2006) proposed a framework comprising four steps: identification, evaluation, allocation and application. The objective of this framework is to structure information quality handling and value adding activities.

- Knowledge Perspective: Following the principle "know-what, know-how, know-why", Huang et al. (1999) proposed a framework, which comprises three processes: improve quality of information, make tacit knowledge explicit, and create organisational knowledge. The objective of this framework is to transform high-quality information into organisational knowledge.

Based on the relationship between information quality and other research fields, the above three literatures investigate information quality management from differing perspectives. Wang (1998) connected information quality management and quality management by considering information as product. Eppler (2006) connected information quality management and information management by employing the information usage cycle. Huang et al. (1999) connected information quality management and knowledge management through the relationship between information and knowledge. Each of these different perspectives has its advantage. In this thesis, we follow Wang's management framework because this framework is validated by practical application and widely accepted in information quality research.

2.6 Review of Information Quality Assessment

As a key determinant in information quality management, evidently information quality assessment is crucial, since one cannot manage information quality without meaningful measurement (Stvilia et al. 2007). Adapting a general definition for assessment (Gertz et al. 2004), information quality assessment can be defined as the process of assigning numerical or categorical values to information quality dimensions in a given setting. Based on the literature relating to information quality assessment, we organise information quality assessment into three layers: the metric layer, dimension layer, and methodology layer, as described in Figure 7.

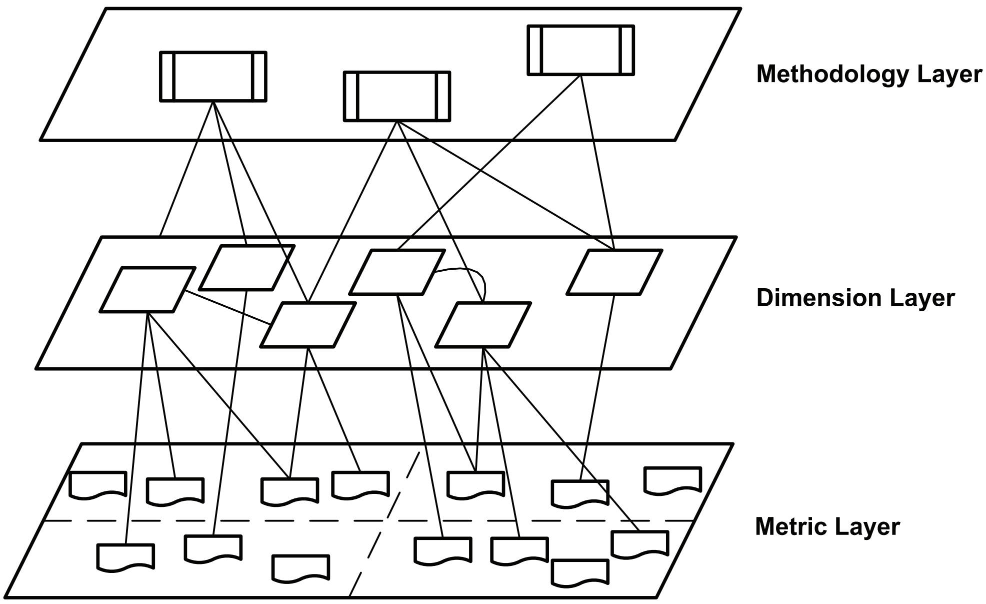


Figure 7: Framework for information quality assessment review

The metric layer includes information quality metrics representing different information quality problems. These information quality problems are classified by a "2 contextual views × 2 assessment views". This classification will be further discussed in Section 2.6.1.

The dimension layer comprises information quality dimensions, which are characteristics of the information. These information quality dimensions are connected to corresponding information quality metrics. One dimension can be linked to multiple metrics and conversely one metric can be linked to multiple dimensions. For example, whilst accuracy (information quality dimension) is linked to incorrect data (information quality metric) and out-of-date data (information quality metric), out-of-date data (information quality metric) can be linked to accuracy (information quality dimension) and timeliness (information quality dimension). Once an information quality metric is linked to multiple information quality dimensions, it will generate dependencies amongst these information quality dimensions. We will in detail discuss this layer in Section 2.6.2.

The methodology layer contains information quality assessment models, frameworks and methodologies. Components in this layer use a set of information quality dimensions to measure information quality. Overall, information quality assessment methodology employs a set of information quality dimensions which are linked to different information quality metrics. We will review different assessment methodologies in Section 2.6.3.

2.6.1 Information Quality Metric

Many researchers have contributed work on the identification of information quality problems. For example, Garvin (1988) pointed out three types of information quality problems: biased information, outdated information, and massaged information. Biased information means the content of the information is inaccurate or distorted in the transformation process. Outdated information is information that is not sufficiently up to date for the task. Massaged information refers to different representations of the same information. Lesca and Lesca (1995) classified information quality problems into the product and process views. The product view focuses on the deficiencies of the information itself, such as incompleteness and inconsistency, whilst the process view concentrates on the problems that are caused in the information production and distribution process.

Based on the literature relating to information quality problems we classify information quality problems by a two-by-two conceptual model. The columns capture information quality problems from both a information perspective and a user perspective, and the rows capture information quality problems as context-independent and context-dependent. Using this model, we classify typical information quality problems in Table 4. This table lists relevant research contributions that provide a comprehensive discussion of information quality problems (Garvin 1988[(1)], Lesca and Lesca 1995[(2)], Huang et al. 1999[(3)], Pipino et al. 2002[(4)], Oliveira et al. 2005[(5)], and Eppler 2006[(6)]).

	Information Perspective	User Perspective
Context-independent	Spelling error [(5), (6)] Missing data [(5), (6)] Duplicate data [(5), (6)] Incorrect value [(5), (1), (6)] Inconsistent data format [(5), (2), (6)] Outdated data [(5), (1), (6)] Incomplete data format [(5), (2)] Syntax violation [(5)] Unique value violation [(5)] Violation of integrity constraints [(5)] Text formatting [(5),(1)]	The information is inaccessible [(3)] The information is insecure [(3)] The information is hardly retrievable [(3)] The information is difficult to aggregate [(3)] Errors in the information transformation [(3)]
Context-dependent	Violation of domain constraint [(5), (2)] Violation of organisation's business rules [(5), (2)] Violation of company and government regulations [(5), (2)] Violation of constraints provided by the database administrator [(5)]	The information is not based on fact [(3), (1)] The information is of doubtful credibility [(3)] The information is irrelevant to the work [(3)] The information consists of inconsistent meanings [(3), (1), (2)] The information is hard to manipulate [(3), (6)] The information is hard to understand [(3)]

Table 4: Classification of information quality problems

Four quadrants in the table above are described as follows:

- The Information Perspective/Context-independent quadrant indicates information quality problems in the database. These information quality problems can be applied to any dataset.
- The Information Perspective/Context-dependent quadrant indicates information quality problems that violate the business specifications. These information quality problems can be detected by contextual rules.
- The Information Perspective/Context-independent quadrant indicates information quality problems that may occur when processing information.
- The Information Perspective/Context-dependent quadrant indicates information quality problems that render information not fit for its intended use by information consumers.

With regards to information quality problems, various methods have been applied to resolve these issues. From the information perspective, information quality problems can be resolved through data cleansing algorithms (Olson 2003), data mining rules (Savchenko 2003), statistical process control (Redman 1996) or dictionary matching routines (Strong et al.1997). From the user perspective, information quality problems cannot often be resolved by automated processes (Eppler 2006). These problems require the optimisation of resource allocation (Ballou and Tayi 1989), analysis of business issues (Wang 1998), re-engineering processes (Redman 1996), or aligning information flow with the corresponding information manufacturing system (Wang et al. 1998).

2.6.2 Information Quality Dimension

Many studies have confirmed that information quality is a multi-dimensional concept (Ballou and Pazer 1985; Redman 1996; Wand and Wang 1996; Wang and Strong 1996; Huang et al. 1999). Over the last two decades, different sets of information quality dimensions have been identified from both database and management perspectives. We review information quality dimensions from four angles: identification, definition, classification, and dependency.

2.6.2.1 Identification of Information Quality Dimensions

Wang and Strong (1996) propose three approaches to study information quality: the intuitive, theoretical and empirical approaches. We adopt these approaches in our analysis of the identification of information quality dimensions. The intuitive approach derives information quality dimensions from the researchers' experience or from the requirements of particular cases. In this approach, information quality dimensions are identified according to specific application contexts. For example, O'Reilly (1982) used accessibility, accuracy, specificity, timeliness, relevance, and the amount of information to assess information quality in the context of decision-making. Ballou and Pazer (1985) employed accuracy, timeliness, completeness and consistency to model information quality deficiencies in multi-input, multi-output information systems. The theoretical approach generates information quality dimensions on the basis of data deficiencies in the data manufacturing process. For example, Wand and Wang (1996) used an ontological approach to derive information quality dimensions by observing inconsistencies between the real-world system and the information system. The empirical approach provides information quality dimensions by focusing on whether the information is fit for use by information consumers. For example, Wang and Strong (1996) captured 15 information quality dimensions of importance to information consumers. Kahn et al. (2002) selected 16 information quality dimensions for delivering high quality information to information consumers. From the discussion above, we found that varying sets of information quality dimensions can be identified using different approaches.

2.6.2.2 Definition of Information Quality Dimensions

The three approaches above can also be considered as three perspectives for defining information quality dimensions. The intuitive approach defines information quality dimensions from the data perspective. For example, Ballou and Pazer (1985) defined completeness as a situation in which all values for a certain variable are recorded. The theoretical approach defines information quality dimensions from the real-world perspective. For example, Wand and Wang (1996) defined completeness as the ability of an information system to represent every meaningful state of the represented real world system. The empirical approach defines information quality dimensions from the user's perspective. For example, Wang and Strong (1996) defined completeness as the extent to which data are of sufficient breath, depth, and scope for the task at hand. We describe the approaches and perspectives for defining information quality dimensions in Figure 8.

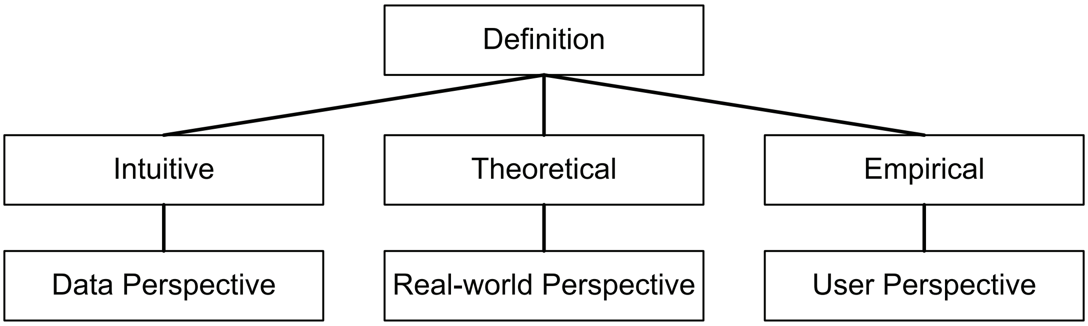


Figure 8: Approaches and perspectives for defining information quality dimensions

The advantage of using a data perspective is that information quality can be automatically assessed and objectively controlled. The advantage of employing a real-world perspective is that the referencing quality specification is theoretically perfect. However, both perspectives fail to capture the expectations of data consumers. From a user perspective, a comprehensive set of dimensions can be used to measure information quality. Also, underlying this perspective is the idea that information quality can be improved according to the intended use. Yet this perspective fails to measure information quality automatically, making it difficult to negotiate a large amount of data. Considering the advantage of each perspective, we use data perspective to define the dimensions and assess the quality of raw data. When assessing the quality of information products, we define the dimensions from user perspective.

2.6.2.3 Classification of Information Quality Dimensions

Based on the identification and the definition of information quality dimensions, different kinds of approaches have been proposed to classify information quality dimensions. Wang and Strong (1996) proposed a hierarchical framework that consists of four information quality categories: intrinsic information quality, contextual, representational and accessibility. Wand and Wang (1996) used an ontological approach to derive information quality dimensions and categorised them by internal view and external view. Internal view is use-independent and contains a set of information quality dimensions that are comparable across applications. External view is concerned with the use and effect of information systems, which represent the real-world system. Naumann and Rolker (2000) organised information quality dimensions with three main factors that influence information quality: the perception of the user, the information itself, and the process of accessing information. These three factors can be considered as subject, object and process. Helfert (2001) classified information quality dimensions by employing semiotics and two elements of quality, which are quality of design and quality of conformance. Semiotics comprises three levels: syntactic, semantic and pragmatic. The syntactic level considers the basic representation of information. The semantic level focuses on information related to real world objects. Finally the pragmatic level deals with information processes and information users. Kahn et al. (2002) developed a two-by-two conceptual model to describe information quality dimensions. Whilst the two rows are product quality and service quality, the two columns are conformance to specifications and meeting and exceeding consumer expectations. Therefore information quality dimensions are located in four quadrants: sound, dependable, useful, and usable. Bovee et al. (2003) presented a categorisation of information quality dimensions with the sequence of using information. The sequence includes the following four steps: obtaining the information (accessibility), understanding the information (interpretability), connecting the information with the given context (relevance), and assuring the information is free from error (integrity). We summarise the above discussion in Table 5.

Literature	Taxonomy	Category			
Wang and Strong (1996)	Hierarchical	Intrinsic	Contextual	Representational	Accessibility
Wand and Wang (1996)	Ontological	External	Internal	System-related	Data-related
Naumann and Rolker (2000)	Source for information quality metadata	Subject		Object	Process
Helfert (2001)	Semiotics	Syntactic		Semantic	Pragmatic
Kahn et al. (2002)	Product and service	Sound	Dependable	Useful	Usable
Bovee et al. (2003)	Sequence of using data	Accessibility	Interpretability	Relevance	Integrity

Table 5: Classifications of information quality dimensions

Using the above classifications and widely accepted information quality dimensions, we review the characteristics of each classification in Table 6. This classification can draw a description of the characteristics of each dimension. It provides a facility for understanding and selecting information quality dimensions.

Author / Dimension	Wang and Strong (1996)	Wand and Wang (1996)	Naumann and Rolker (2000)	Helfert (2001)	Kahn et al. (2002)	Bovee et al. (2003)	Eppler (2006)
Accuracy	Intrinsic	Internal view + Data related	Process	Semantic + Quality of Conformance	Sound	Integrity	Com-munica-tion level
Completeness	Contextual	Internal view + Data related	Objective	Semantic/ Pragmatic + Quality of Conformance	Sound	Integrity	Com-munica-tion level
Consistency	Represen-tational	Internal view + Data related	Process	Syntac-tic/Semantic + Quality of Design/ Con-formance	Sound	Integrity	Product level
Timeliness	Contextual	Internal view + Data re-lated, External view + Data/System related	Objective	Pragmatic + Quality of Conformance	Depend-able	Rele-vancy	Process level
Interpretabil-ity	Represen-tational	External view + Data related	Subjective	Semantic + Quality of	Useful	Inter-pretabil-	

				Conformance		ity	
Relevancy	Contextual	External view + Data related		Pragmatic + Quality of Design	Useful	Rele-vancy	Com-munica-tion level
Reliability		Internal view + Data/System related	Objective	Semantic + Quality of Conformance			
Accessibility	Accessibil-ity		Process	Syntactic + Quality of Conformance	Useable	Accessi-bility	Infra-struc-ture level
Security	Accessibil-ity		Objective	Syntactic + Quality of Conformance	Depend-able		Infra-struc-ture level
Ease of Un-derstanding	Represen-tational	External view + Data related	Subjective	Semantic + Quality of Design	Useful	Rele-vancy	
Reputation	Intrinsic		Subjective		Useable		
Believability	Intrinsic	Internal view + Data/System related	Subjective	Semantic + Quality of Conformance	Useable		
Value-Added	Contextual		Subjective		Useable		
Ease of Ma-nipulation		External view + Data related			Useable		Infra-struc-ture level
Objectivity	Intrinsic	External view + Data related	Objective	Semantic + Quality of Design	Useful		
Amount of Information	Contextual	External view + Data related	Process		Useful		
Conciseness	Represen-tational	Inter-nal/External view + Data related	Subjective	Semantic + Quality of Design	Sound	Rele-vancy	Product level

Table 6: Dimension-oriented classification review

2.6.2.4 Dependency of Information Quality Dimensions

Besides classifying the dimensions, a number of literatures have analysed the dependencies between information quality dimensions. For example, Ballou and Pazer (1995) proposed a framework to investigate the trade-offs between accuracy and timeliness in the context of decision-making. Redman (1996) pointed out that timeliness has an impact on accuracy. Ballou and Pazer (2003) modelled the utility and trade-offs between completeness and consistency. Olson (2003) gestures to the relationship between accuracy and completeness and states that consistency is a component of accuracy. Cappiello et al. (2004) analysed time-related accuracy and time-related completeness in multi-channel information systems. Amicis et al. (2006) proposed a data-driven approach to analyse the dependency between syntactic accuracy and timeliness as well as the dependency between completeness and timeliness. Figure 9 presents the important dependencies between information quality dimensions and structures the literatures discussed above.

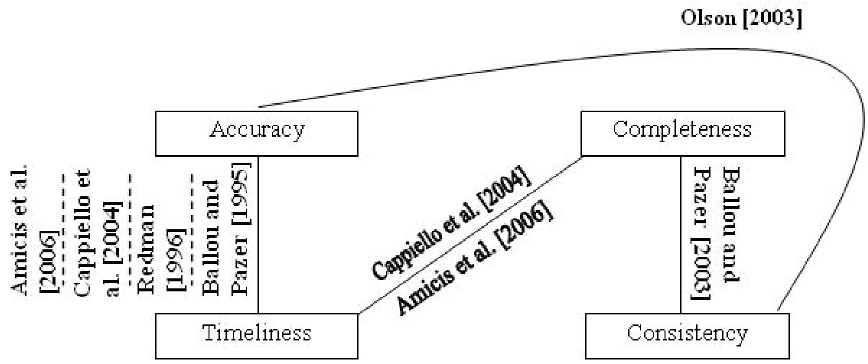
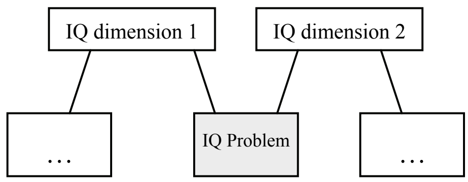


Figure 9: Dependency of information quality dimensions

From the literature review we found that the dependency of information quality dimensions can be divided into two categories: negative correlation and positive correlation. Negative correlation refers to how the improvement of one information quality dimension may lead to a decreasing goal value in another dimension (often also referred as information quality dimension trade-offs). For example, when introducing new information to improve completeness, the new introduced information may be inconsistent with the existing information. In this manner, completeness and consistence are negatively correlated. For the negative correlation, we identify two types of trade-offs between information quality dimensions: the faster the information is delivered, the less time is available to check other information quality dimensions, and when new information is introduced to improve certain information quality dimensions, such new information may lead to a decreasing goal value in other dimensions.

Positive correlation means two information quality dimensions are mutually responsible for and share a set of information quality problems. For example, when timeliness and accuracy are sharing outdated data as a mutual information quality problem, the improvement of timeliness may lead to an increasing value in accuracy. In this way, timeliness and accuracy are positively correlated. According to the discussion above, we summarise correlations of information quality dimensions in Table 7.

<table>
<tr><th>Negative Correlation</th><th>Positive Correlation</th></tr>
<tr><td>The faster the information is delivered, the less time is available to check other information quality dimensions: trade-offs between (1) Timeliness and other information quality dimensions, (2) Currency and other information quality dimensions. When the new information is introduced to improve certain information quality dimension, the new information may lead to a decreasing goal value in other dimensions: trade-offs between completeness and other dimensions, accessibility and other dimensions, security and other dimensions, relevancy and other dimensions.</td><td>

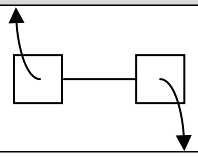
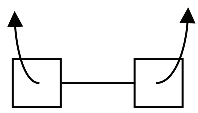
When we improve information quality dimension 1, information quality dimension 2 may be improved or stay at the same quality value. It depends on whether we fix the mutual information quality problem.</td></tr>
<tr><th>Negative Correlation Model</th><th>Positive Correlation Model</th></tr>
<tr><td></td><td></td></tr>
</table>

Table 7: Negative and positive correlation of information quality dimensions

2.6.3 Information Quality Assessment Methodology

As described in Figure 7, information quality assessment methodologies use a set of dimensions to measure information quality. Over the last decade, a variety of assessment methodologies have been proposed. We select five typical methodologies (Redman 1996; Huang et al. 1999, Lee et al. 2002, Pipino et al. 2002, and Stvilia et al. 2006) and evaluate them against the following criteria: definition of information quality dimensions, classification of information quality dimensions, model, tool, and case study. The definition of information quality dimensions is for identifying which information quality dimensions are defined from which perspective. Classification of information quality dimensions is used to compare the classification of dimensions in each methodology. Model is to demonstrate the theoretical basis of the methodology. Tool is used to validate the implementation of the methodologies. Case study concentrates on empirical feasibility of these methodologies. Using the criteria above, we can obtain the characteristics of each methodology. The criteria also provide a guidance to develop our methodology in Chapter 5.

Information quality research can be broadly classified into two research communities: database and management. The database community follows a technical and data schema oriented approach. Most research arising from this community defines information quality on the basis of data values or instances of data models that are consistent to specifications in data schema (e.g. Naumann and Rolker 2000, Oliveira et al. 2005 follow this definition). Research related to the management community is business and management focused. Approaches proposed from this community follow the concepts and principles of total quality management in the way that researchers regard information quality in the light of information that is fit for use by information consumers (e.g. Wang and Strong 1996, Bovee et al. 2003 follow this definition).

If the methodology is only applied to one information quality community, it is considered as specific methodology. If the methodology can be applied to both information quality communities, it is a generic methodology. If a case study or an application is provided in the literature, we regard it as a practical study, otherwise it is theoretical. We summarise the five methodologies and their characteristics in Table 8.

	Redman (1996)	Huang et al. (1999)	Lee et al. (2002)	Pipino et al. (2002)	Stvilia et al. (2006)
Definition	12 information quality dimensions are defined from the database community	16 information quality dimensions are defined from management community	15 information quality dimensions are defined from both communities	16 information quality dimensions are defined from both communities	22 information quality dimensions are defined from both communities
Classification	Conceptual view, data value and representation	Classification of Wang and Strong (1996)	Classification of Kahn et al. (2002)	Without classifications	Classification of Wang and Strong (1996)
Model	A step by step procedure adapted from statistical process control	Adopt Deficiency model of Wand and Wang (1996)	Adopt PSP/information quality model of Kahn et al. (2002)	The model combines subjective and objective assessment	The model consists of activity types, information quality Problems, and information quality taxonomy
Tool	DCI system	information quality assessment survey	information quality assessment survey	information quality assessment software	information quality assessment survey
Case Study	Telstra Co. Ltd.	Appliance Company		1, Global Consumer Goods, Inc., 2, Data Product Manufacturing, Inc.	1, Simple Dublin Core 2, English Wikipedia
Conclusion	Specific, practical	Specific, practical	Generic, Theoretical	Generic, practical	Generic, practical

Table 8: Comparison of information quality assessment methodologies

Due to the two research communities, information quality assessments are differentiated into objective and subjective assessment (Pipino et al. 2002). Objective information quality assessments reveal information quality problems in the database whilst subjective information quality assessments reflect the needs and experiences of data consumers. We follow this classification to discuss information quality assessment from both objective and subjective perspectives.

Objective information quality assessment is used to measure the extent to which information conforms to quality specifications and references. We sub-divide objective information quality assessment into two categories: intrinsic and real-world information quality assessment. Intrinsic information quality assessment accords with the data perspective and focuses on the quality of the data in the database. For example, Savchenko (2003) developed item frequency rules and regular expression patterns to facilitate automated intrinsic information quality assessment. Real-world assessment follows an ontological perspective and focuses on information quality deficiencies that can take place during the system design and data production. For example, Wand and Wang (1996) identified data mapping deficiencies between the real world state and information system representation. Overall, objective information quality assessment can be considered as the procedure of comparing current data value with an ideal data value.

Subjective information quality assessment is used to measure the extent to which information is fit for use by information consumers. Information consumers assess information quality according to their demands and expectations. Subjective information quality assessment follows the user perspective and focuses on any discrepancy between the current quality of information and the user's expectation. In order to indicate the differences between objective and subjective information quality assessments, we provide a comparison in Table 9.

Method / Feature	Objective Assessment	Subjective Assessment
Tool	Software	Survey
Measuring Object	Data	Information Products
Criteria	Rules, Patterns	Fitness for use
Process	Automated	User Involved
Assessing Result	Single	Multiple
Data storage	Databases	Business Contexts

Table 9: Comparison of objective and subjective information quality assessment

Objective information quality assessment uses software to automatically measure the data in a database by a set of quality rules, whereas subjective information quality assessment employs surveys or an interview approach to measure contextual information by data consumers. A single assessment result can be obtained from objective assessment. However, it is more than probable that different information consumers will generate different assessment results.

With the development of both objective and subjective information quality assessment, researchers suggest the combination of these two assessment methodologies. For example, Pipino et al. (2002) provided a framework to combine objective and subjective information quality assessments. Kahn et al (2002) proposed the PSP/IQ model, in which two views of quality are assigned: conforming to specifications (objective) and meeting or exceeding consumer expectations (subjective). Our review shows that it is beneficial to combine objective and subjective information quality assessment.

2.7 Research Agenda

From the review above, we revealed the potential research questions concerning contextual information quality, information quality management and information quality assessment. In this section, we briefly highlight the research themes to bridge gaps in information quality research, and outline a future research agenda.

Information quality research has been applied to various application contexts. Following an overview of these application contexts, we reviewed the two most cited contexts to analyse contextual information quality: information systems and decision-making. Because information quality research mostly stems from information system research, understanding the relationship between information quality and information systems is valuable for conducting further research. Our analysis has shown the citation relationship between information quality research and information system research. In the study of information quality in a decision-making context, we have shown that researchers consider different influencing factors when investigating the relationship between information quality and decision-making. However, the challenge we are still facing is how to control these influencing factors in information quality experiments. This problem indicates that controlling extraneous variables is crucial in the design of information quality experiments. In this thesis, we discuss a method for controlling extraneous variables in Chapter 3. Considering these extraneous variables, we further investigate the effects of information quality on decision-making in Chapter 4. The research challenges and questions surrounding contextual information quality are summarised in Table 10.

Research Issues Concerning Contextual Information Quality
Research Question 1: What is the relationship between information quality and application contexts?
Research Question 1a: How does information quality impact application contexts?
Research Question 1b: What is the relationship between information quality research and information system research?
Research Question 1c: How does one control extraneous variables in an information quality experiment?

Table 10: Research issues concerning contextual information quality

From our overview of information quality management, we found that information quality management has merged three domains of management: quality management, information management and knowledge management. One drawback in this broad view of information quality management is the difficulty of deploying information quality management in organisations. Thus in empirical applications of information quality management, organisations are still facing issues on how to manage information quality effectively. Concerning this problem, organisations need to understand the costs and benefits of information quality management, how to deploy it and how to build information quality culture in organisations. Hence a comprehensive framework is required for information quality management in future research. In this thesis, we contribute the practical assessment methodology in Chapter 5 and the analysis of the effects of information quality on decision-making in Chapter 6 to information quality management. The research challenges and questions regarding information quality management are listed in Table 11.

Research Issues Concerning Information Quality Management
Research Question 2: How to manage information quality in organisations?
Research Question 2a: How to analyse cost and benefit of information quality management?
Research Question 2b: How to evaluate the maturity of information quality management?
Research Question 2c: How to deploy information quality management in organisations?
Research Question 2d: How to build information quality awareness and information quality culture in organisations?

Table 11: Research issues concerning information quality management

Following our review of information quality management, we focus on information quality assessment. This is an important component in information quality management. The assessment section analyses information quality problems, information quality dimensions and assessment methodologies. Our result has shown that information that is of high quality in one context may be considered to be of low quality in another, information that is considered to be of high quality by one person may be considered to be of low quality by another, and information that is of high quality according to the conformity of one specification may be of low quality according to the conformity to another. Therefore most researchers and organisations are facing the following issue: how to assess and evaluate information quality effectively. In the attempt to resolve this dilemma, subsequent issues have been generated, such as how to identify potential information quality problems, how to define and select information quality dimensions and how to connect information quality dimensions to information quality problems. Addressing these challenges, we propose a set of validated dimensions and practical assessment methodologies in Chapter 5. The research questions concerning information quality assessment are summarised in Table 12.

Research Issues Concerning Information Quality Assessment
Research Question 3: How to assess information quality effectively?
Research Question 3a: How to identify potential information quality problems?
Research Question 3b: How to define and select information quality dimensions?
Research Question 3c: What is the relationship between information quality problems and information quality dimensions?
Research Question 3d: How are information quality dimensions dependent upon each other? And how does one deal with the dependencies of information quality dimensions?
Research Question 3e: How to choose the most suitable information quality assessment methodology for organisations?

Table 12: Research issues concerning information quality assessment

2.8 Summary

Based on our review of information quality research, several conclusions can be drawn:

- A solution for assessing information quality effectively is needed in organisations. Based on the review and discussion of information quality assessment, three components can be found to be significant for information quality assessment: information quality problems, information quality dimensions and information quality assessment methodology. Using our review, in Chapter 5 we propose assessment models and algorithms to measure information quality.
- Relationships between information quality and organisational contexts need to be investigated. An overview of information quality research in context is

provided in our work. Specifically we focus on reviewing the relationship between information quality and decision-making. Based on our review, we systematically investigate the effects of information quality on decision-making in Chapter 4 and 6.

- A comprehensive framework for information quality management is needed in organisations. Three perspectives of information quality management can be found within our review. Further information quality management needs to recognise quality management, information management and knowledge management. To extend information quality management, information quality assessment and an understanding of the effects of information quality are indispensable. Therefore Chapter 4, 5 and 6 attempt to contribute fresh knowledge to information quality management.

This chapter provides an overview of the roots and foremost research in this domain as well as indicating challenges for future research. Although over the last decade researchers have contributed a large amount on models, frameworks, assessment and management approaches, as our review shows, there remains many open research questions. In this thesis, we focus mainly on issues concerning information quality assessment and contextual information quality. Relating the issues to our research questions discussed in Chapter 1, we can describe the roadmap of this research in Figure 10.

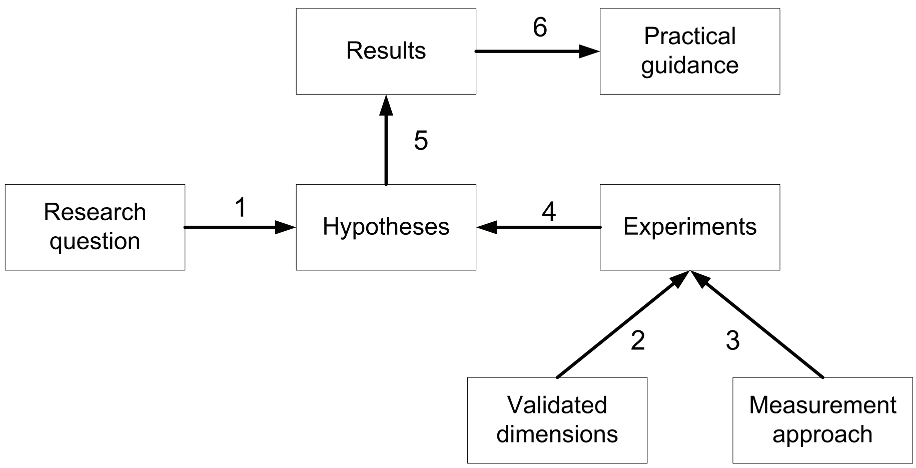


Figure 10: Roadmap of this research

Firstly, we will derive a set of hypotheses from our research question and implement experiments to test the hypotheses. As the experimental design is facing the problem that how to measure information quality, we will confirm the dimensions used to measure information quality and then propose a practical measurement approach. The dimensions will be validated by surveys and interviews, and the measurement approach will be validated by simulations and real-world applications. Based on the research on information quality measurement, the experiment will be designed in the context of supply chain management. From the experiment, we can collect and analyse the experimental data to examine our hypotheses. Using the analysis results, we will provide practical guidance and indications for information quality management.

Chapter 3: Research Methodology

In this chapter, we will analyse the methodologies used to conduct our research. The first section introduces the concept of research, research classifications and research processes. The second section presents research methodologies in information system research and the methods we used in this thesis. The third section proposes our research design, including detailed discussions of survey development, experimental design, and data analysis method. The next section discusses the validity and reliability of measurement development and experimental design. Finally, this chapter concludes by offering solutions to methodological issues in this research.

3.1 Introduction

Research involves using valid and reliable procedures, methods and techniques to obtain objective answers to questions. Grinnell (1993) defines research as the "careful, systematic, patient study and investigation in some field of knowledge taken to establish facts or principles". Generally, research can be considered as a way of thinking and creating new knowledge (Kumar 1996, Oates 2006).

Kumar (1996) describes 6 inherent characteristics in the research process. (1) Research must be controlled. For example, when we examine the casual relationship between two variables, exogenous factors affecting this relationship must be controlled and minimised. In Section 3.3.3, we review the exogenous factors that may affect the relationship between information quality and decision-making. According to the review, we have controlled each factor in our experiments. (2) Research must be rigorous. That means the research procedure must be accurate, relevant and appropriate to the research question. Based on our review, we found 5 research methodologies that are used to examine our research questions. The methodologies are listed in Table 14. (3) Research must be systematic, indicating that the research procedure must follow a logical sequence. Following a structure of experimental research, we carried out a systematic investigation to our research questions. (4) Research must be empirical. It means that the data collected in the research must be based on real-world evidence. In our study, the assessment framework is validated by a practical application and data are collected from the real-world users. We also organize a workshop used to confirm our results. (5) Research results must be valid and verifiable, meaning that conclusions must be drawn from reliable research findings and verified by different steps in the research process. Following the validation procedure described in Figure 12, we validate our results by a number of statistical methods. (6) Research must be critical. This requires that the research can withstand critical examination and review. Besides a critical review in Chapter 7, our results are published in a variety of conferences and journals.

In information system research, research methodologies can be classified into quantitative and qualitative research. Quantitative research is used to classify features, construct statistical models and objectively analyse observation by scientific methods. It involves the analysis of numerical data. For example, doctors measure patients' body temperature rather than ask them. On the other hand, qualitative research is used to explore people's opinions, attitudes and experiences. This type of research aims to describe a situation or a phenomenon through subjective interpretations. It generates descriptive conversations instead of numerical data. For example, doctors ask patients to describe their pain rather than use a measure of pain.

As discussed above, both quantitative and qualitative methods have their advantages. We found that a mixed research method is most suitable for addressing our research question. Thus in Chapter 4, we use laboratory experiments to analyse relationships. In Chapters 5, we use survey and simulation to validate our framework. These methodologies involve scientific data analysis and therefore are classified as quantitative research. In Chapters 6, we use interviews to confirm the results of our data analysis. The interview is a typical qualitative research methodology. As this research follows the falsification principle (Popper 1980) to examine the hypotheses, and we rely on a host of scientific methods to produce and analyse numerical data, this research follows the positivist research philosophy to interpret our results.

Jenkins (1985) proposes the following process for information system research: initially, a general idea needs to be generated. Based on the general idea, a related literature review is carried out to refine the idea and propose a clear statement of the research objective. In our research, this was discussed in Chapter 2. To achieve the research objective, appropriate research methodologies need to be selected. Using the selected research methodology, detailed scenarios and procedures are designed for the research study as a whole. According to this design, researchers implement the design and collect data. Following data collection, data analysis is used to report and interpret the research findings. Finally, researchers can summarise and publish the research results. From the above steps, we can identify 4 crucial components in the research process - research objective, research methodology, research design and data analysis. In line with the research objectives of this thesis, we outline the 4 components in Table 13.

Research objective	Clarify the relationship between information quality and decision quality
Research methodology	Laboratory experiment, interview
Research design	Experimental design, software design, sampling design
Data analysis	Analysis of variance (ANOVA), descriptive analysis

Research objective	Assess information quality
Research methodology	Survey, interview, simulation, math modeling
Research design	Survey design, simulation design, sampling design
Data analysis	Factor analysis, descriptive analysis

Table 13: Crucial components in the research process

3.2 Research Methodology

A number of research methodologies have been identified in information system research. Reviewing the previous literature on research methodology (Jenkins 1985, Galliers 1991, Mingers 2001, Oates 2006), we list 14 prevalent methodologies in information system research and mark the methodologies we used in this thesis (Table 14). In these research methodologies, no single research methodology is inherently superior to all others (Benbasat 1987). All the research methodologies have strengths and weaknesses but are found to be valuable if used appropriately (Galliers 1991). Thus Kaplan and Duchon (1988) proposed to combine different research methodologies to improve the quality of the research.

Laboratory experiment	√	Math modelling	√
Field experiment		Simulation	√
Survey	√	Action research	
Interview	√	Forecasting	
Case study		Subjective/Argumentative	
Role/Game Playing		Grounded theory	
Content analysis		Ethnography	

Table 14: Research methodologies in information system research

In order to strengthen our research, we use 5 research methodologies in this thesis: survey, math modelling, simulation, interview and laboratory experiment. To validate our information quality assessment framework, we use a survey to derive a set of valid and reliable dimensions. Then we use math modelling to investigate the algorithms of the dimensions and these algorithms are examined in a simulation. Based on our assessment framework, we further use laboratory experiments to investigate the effects of information quality on decision-making. After analysing the experimental results, we use interviews to confirm the results of our data analysis. Next, we shall discuss each research methodology in detail.

Survey is used to obtain practical data from a large group of people. After data collection, quantitative analysis can be used to identify patterns from the collated data. The use of surveys enables researchers to study a greater number of variables than in the case of experimental approaches. In our study, the survey is an effective method to validate information quality dimensions by factor analysis.

An interview is a particular conversation between the researcher and the interviewee. This conversation is based on a strong research purpose. The use of interviews permits researchers to obtain detailed information and more explanation regarding an issue. Interviews can be divided into three types: structured, semi-structured and unstructured interviews. Structured interviews employ pre-determined questions for every interviewee. In the interviewing procedure, both researcher and interviewee only focus on the standardised questions. A semi-structured interview is based on a list of themes. Compared to structured interviews, more questions and explanations can be raised if needed. Unstructured interview is the freest way to carry out the interview. Based on a proposed topic, the procedure of an unstructured interview is a discussion between researchers and interviewee. In our study, as we have obtained a clear goal of the interview, structured interviews are used to test face and content validity and confirm the data analysis results in Section 6.3.

Math modelling is used to model real world situations through mathematical equations. This methodology aims to objectively determine the relationship between independent and dependent variables. The results of this methodology are mathematical formulations, which can be used as grounded indications for software development. In our study, math modelling is used to measure information quality dimensions.

Simulation involves mirroring a segment of the real world or copying the behaviour of a system. This methodology is used in a situation that is difficult or impossible to analyse. Based on the formulations of information accuracy and completeness, we simulate a system to determine information accuracy and completeness in the information chain.

Laboratory experiments involve directly manipulating a small number of variables and identifying the relationship between these variables. Using quantitative analysis, the experiment is designed to prove or disprove the hypotheses. An ideal experiment is designed to control all other possible factors affecting the experimental outcome and show how independent variables affect dependent variables (Kerlinger and Lee 2000). It has been found that laboratory experiments are an effective methodology in addressing the cause and effect relationship (Campbell and Stanley 1963, Jarvenpaa et al. 1985, Field and Hole 2003), especially in investigating the cause and effect relationship between attributes of the decision environment, characteristics of information system and decision performance (Dickson et al. 1977). In our study, experimental research methodology is used to investigate the effects of information quality on decision-making.

3.3 Research Design

Thyer (1993) defines research design as "a blueprint or detailed plan for how a research study is to be completed – operationalising variables so that they can be measured, selecting a sample of interest to study, collecting data to be used as a basis for testing hypotheses, and analysing the results". Our research design is divided into two major sections: survey design and experimental design. The survey design is to develop an environment for validating information quality dimensions. Experimental design concentrates on the procedure of the experimental task, the consideration of exogenous factors and methodological issues arising in experimental research.

3.3.1 Survey Design

When developing measurements, Churchill (1979) recommends a seven-step procedure. The first step is to specify the domain of construct from the literature review. In the specified domain, the second step is to generate a sample of measuring items based on the researchers' experience. Based on the measuring items, the third step involves collecting a set of data for the items. Fourth, the collected data are used for conducting factor analysis to group items. The fifth to seventh steps are collecting new data for the purified measurement and assessing its reliability and validity. The final step is to develop norms for the measurement.

In this thesis, we used a survey to confirm the validation of information quality dimensions. In accordance with Churchill's (1979) measurement development, we first adopt 15 constructs of information quality based on our literature review. For each construct, we propose 2 to 5 measuring items. Thus a total of 50 measuring items are proposed for the constructs. Using the measuring items, we invite subjects to evaluate the significance of the measuring items. An 11-point Likert type scale is used to measure this significance. Following the collection of data, we implement a confirmatory factor analysis and a Cronbach's α calculation to validate the constructs. Based on the validation results, we classify the items into valid and reliable constructs. Finally, we develop the norm for each information quality construct. The detail of our survey design is discussed in Chapter 5.

3.3.2 Experimental Design

Experimental research methodology is used to investigate the relationship between information quality and decision-making. This methodology consists mainly of three components: model construction, experimental design, and data analysis. Model construction provides the basis for the experimental design. Data analysis methods offer indications for data collection in the experiment. Thus experimental design connects model construction and data analysis. With the data analysis results, we could examine the hypotheses and validate the theoretical model. The relationships between the above three components are described in Figure 11.

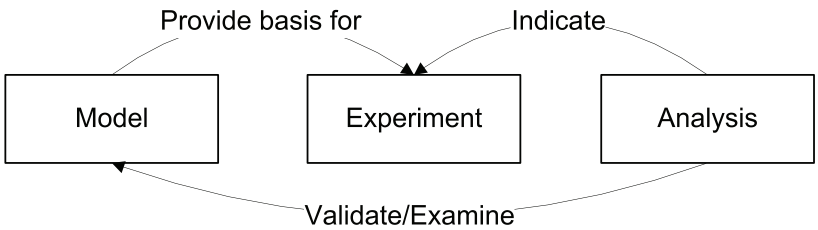


Figure 11: Relationship between model construction, experimental design and data analysis

In constructing the model, we firstly need to determine the independent and dependent variables. As discussed in Chapter 1, two relationships are investigated in this study: (1) information quality categories and decision quality, (2) information quality dimensions and decision quality. In the first relationship, we focus on three widely accepted information quality categories as our independent variables: intrinsic, contextual and representational information quality. Decision quality is considered to be a dependent variable. In the second relationship, another three widely used dimensions are considered as independent variables: accuracy, completeness and consistency. Again, decision quality is used as the dependent variable.

Before designing the experiment, we need to determine the methods used for data analysis. Statistical methods for data analysis can be generally divided into descriptive statistics and inferential statistics. Descriptive statistics deal with the collection, tabulation and summarisation of the data in order to provide meaningful information. Inferential statistics analyse and interpret data to develop meaningful inferences. In order to test the hypotheses, inferential statistics are frequently employed in data analysis. There are two types of inferential statistic: parametric and non-parametric statistics. The selection of inferential statistics depends on the scales of the data. There are four basic data scales: nominal, ordinal, interval and ratio data. While parametric statistics are used to analyse interval and ratio data, non-parametric statistics are used to analyse nominal and ordinal data. We summarise the data scales and corresponding statistics in Table 15.

Scale	Statistics	Descriptions
Nominal Data	Non-parametric	This scale is used to compare things. Numbers are used only as names for categories.
Ordinal Data		This scale is used to rank things in terms of some property such as size, length etc.
Interval Data	Parametric	Data are spaced at equal intervals but there is no true zero point.
Ratio Data		The same as interval data, except that there is a true zero point on the scale.

Table 15: Scales of data in data analysis

As indicated in Table 16, possible methods can be used for data analysis. However, considering the type of data we collected in our experiment as well as our factorial design, we found the analysis of variance (ANOVA) to be the most suited method. In our experiment, since two proposed relationships respectively contain three independent variables and participants independently contribute interval or ratio data, a three-way independent ANOVA can be employed to analyse the experimental results. We outline the procedure of selecting methods for data analysis in Table 16.

<table>
<tr><td>Descriptive Statistics</td><td colspan="5">Inferential Statistics</td></tr>
<tr><td></td><td>Non-parametric</td><td colspan="4">Parametric</td></tr>
<tr><td></td><td></td><td>One independent variable</td><td colspan="3">More than one independent variable</td></tr>
<tr><td></td><td></td><td></td><td colspan="3">three-way ANOVA</td></tr>
<tr><td></td><td></td><td></td><td>Repeated</td><td>Mixed</td><td>Independent</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>×</td></tr>
</table>

Table 16: Procedure of selecting methods for data analysis

In order to use the three-way independent ANOVA, four assumptions should be considered in sampling:

A1, Normally distributed data: it is assumed that data are from normally distributed populations. This assumption can be tested by histogram or the Komogorov-Smirnov and Shapiro-Wilk test.

A2, Homogeneity of variance: this assumption indicates that variance throughout the data should be roughly consistent. This assumption can be tested by Levene's test.

A3, Interval/Ratio data: in the experiment, each participant should contribute one piece of interval or ratio score data.

A4, Independence: this assumption means that the behaviour of one participant will not influence the behaviours of others.

In the above four assumptions, assumptions 1 and 2 are not utterly inflexible. That means ANOVA may be still robust even if the two assumptions are found to be violated. In the case of the violation of assumption 1, Gravetter and Wallnau (2000) stated that this violation would not cause any major problems given sufficiently large sample sizes (i.e. 30 or more). When assumption 2 is violated, Stevens (1996) points out that ANOVA remains fairly robust when sample sizes are equal. Assumptions 3 and 4 can be controlled by appropriate task design and interaction constraint.

When designing the experiment, we need to control exogenous factors that influence the relationship between information quality and decision quality. Besides information quality, other factors may also influence decision quality, such as information overload (Eppler and Mengis 2003), decision aids (Neumann and Hadass, 1980), decision models (Janis and Mann 1977), decision strategy (Payne et al. 1988), task complexity (Campbell 1988), expertise (Fisher et al 2003), decision time (Svenson et al. 1990), decision environment (Shankaranarayan 2003), interaction (Burleson et al. 1984), and information presentation (Remus 1984). In our research, we review the methods of controlling these influencing factors. Based on our review, we considered each factor in the experiments. These exogenous factors are summarised in Figure 12.

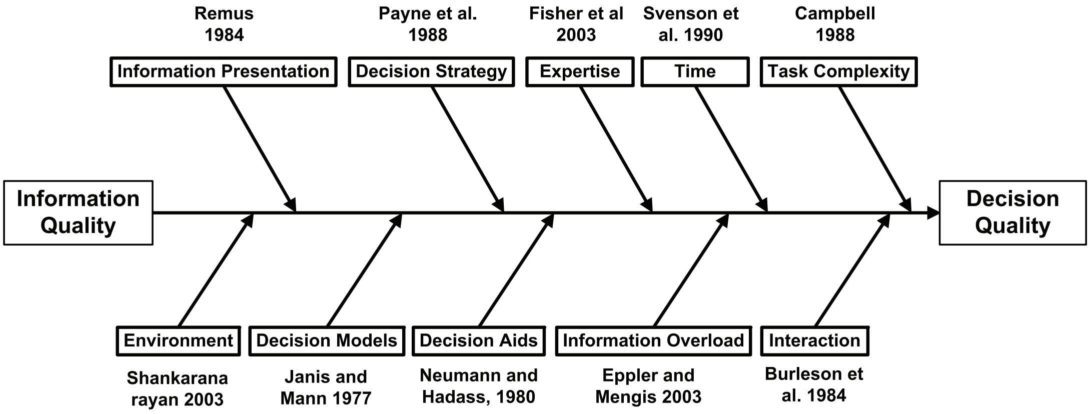


Figure 12: Exogenous factors influencing the relationship between information quality and decision quality

3.3.3 Consideration of Exogenous Factors

We can deal with these influencing factors using two major approaches: considering the influencing factors as independent variables, or fixing the influencing factors to one status. In this research, we attempt to control the influencing factors and thus effectively investigate the relationship between information quality and decision quality. In this section, we will explain how to measure and control the influencing factors.

Task Complexity

Since different task complexities can vary the decision results, we need to find an effective method to set an appropriate task complexity for our experiment. Different kinds of tasks could achieve complexity through differing paths. For example, in a decision-making task, complexity is a function of the number of alternatives and the number of relevant dimensions (Campbell 1988). In problem tasks, complexity is a function of the number of potential paths to the desired outcome (Taylor 1981). Therefore complexity should be defined according to concrete tasks. For a decision-making task, task complexity could be measured from two perspectives. First, from the subjective perspective, task complexity could be examined by participant's experience, familiarity and interest (Shaw 1976). Secondly, from the objective perspective, task complexity could be examined by task alternatives, task constraints and information interrelationships (Schroder et al. 1967). In order to measure the task complexity from an objective view, Simon (1964) identified two main contributors to task complexity. One is the number of elements in a system; the other is the degree and nature of the interactions among elements. By adopting Simon's (1964) definition, Jarvenpaa (1985) utilised a number of information distractors to control task complexity. A distractor in the experiment is irrelevant information which distracts the decision-maker. Payne (1976) defined task complexity as a function of both the number of alternatives facing the decision-maker and the number of attributes on which each alternative is compared. From this definition, we could observe that task complexity can be controlled by two elements: (1) the number of decision alternatives and (2) the number of criteria used to evaluate alternatives. By adopting Payne (1976)'s definition, two decision experiments were proposed. One experiment (Chengalur-Smith et al. 1999) uses 4 alternative × 5 criteria presenting a simple task and 6 alternative × 7 criteria presenting a complex task. The other experiment (Jung 2005) uses 3 alternative × 3 criteria for the simple task and 5 alternative × 5 criteria for the complex task.

According to the above review, we can use three elements to control task complexity in the decision experimental design: alternatives, criteria and distractors. Decision-makers will process information, with distractors, to make a decision from a set of alternatives. For each alternative, a certain number of criteria will be used for the alternative evaluation. Therefore, based on these three elements, we are able to develop tasks into varyingly complex levels.

Decision Time

When we design our decision-related experiment, the time used for making a decision is considered as a crucial factor (Wright 1974, Bronner 1982, Svenson 1989). Svenson et al. (1990) pointed out that decision quality depends heavily on time, and that variable time should be controlled in any decision-making experiment. Fisher et al. (2003) categorised decision time into time constraint and time pressure. Time constraint is an allocated period of time for decision-making. It could be controlled by allocating a certain amount of time for the making of a decision. Time pressure is a subjective reaction to the allocated time and it occurs when available time is less than required time (Svenson and Edland 1987). In the experimental design, time pressure can be controlled by setting time constraint well below the standard time. Regarding time pressure, two decision experiments are proposed. One experiment (Ordonez and Benson, 1997) set time constraint below the mean time of completing tasks to increase time pressure. The other experiment (Bruggen et al. 1998) set time constraint to 75% of the median average time to activate time pressure. In summary, time constraint and time pressure are the two key elements in controlling the decision time.

Expertise

From the previous study (Fisher et al. 2003), we found that user's expertise is an important variable in the decision-making experiments. It can reduce the task complexity if the user's expertise is relevant to the task (Buckland and Florian 1991). In decision-making studies, researchers have suggested both positive and negative effects of expertise (Yates at al 1990, Sanbonmatsu et al 1992, Fisher et al 2003): Experts may be better at using relevant information (Sanbonmatsu et al, 1992) whereas novices may be more sensitive to new information (Yates et al, 1991). Regarding the effects of expertise on decision-making, two decision experiments were proposed. One experiment (Mackay and Elam, 1992) uses 2 different domains and 2 level domain-specific experiences to control the expertise levels. The other experiment (Fisher et al, 2003) uses 2 level working years, 2 level domain-specific experience experts and 2 level managerial experience experts to control the expertise levels.

According to the above literature and experiment history, expertise can be controlled by (1) working experience and (2) domain-specific knowledge. In order to simplify expertise control, two levels of expertise are often used in the experiment: expert and novice.

Decision Strategy

In the experiment, decision-makers may use a variety of strategies to evaluate decision problems (Abelson and Levi 1985). Decision strategy can be observed as a sequence of events (Payne et al. 1988), such as evaluating alternatives, comparing alternatives and so forth. For example, one person may determine the most important attribute and select the alternative which qualifies most highly with regards to that attribute. Another person may examine all the alternatives and assign attribute values for each alternative. Therefore, decision strategies are various methods which use available information to make a decision. Regarding decision strategy, two decision experiments (Payne et al. 1988; Jarvenpaa 1989) demonstrate that decision quality is affected by the decision strategy selection. In Table 17, 9 typical decision strategies are listed based on Payne et al. (1988)'s work.

Strategy	Abbreviation
Weighted Additive	WADD
Equal Weight Heuristic	EQW
Satisfying Heuristic	SAT
Lexicographic Heuristic	LEX
Elimination by Aspects	EBA
Majority of Confirming Dimensions	MCD
Elimination by Aspects plus Weighted Additive	EBA + WADD
Elimination by Aspects plus Majority of Confirming Dimensions	EBA + MCD

Table 17: Decision strategies adopted from Payne et al. (1988)

Interaction

When we design our experiment, we can create an individual or a group decision environment. To consider this option, we review the interaction in the decision-making. The individual decision-maker is undoubtedly important in effective decision-making, however many decision problems are solved by a set of decision-makers working together as a group. Groups possess substantially more effective advantages for improving decision quality. For instance, groups are more successful than individuals in understanding problems, detecting alternative flaws and collecting broad relevant knowledge (Almeida and Marreiros, 2005). However, there are also some disadvantages in group decision-making. For example, group members consume more time and cost than individuals do (Wetmore and Summers, 2003). Burleson et al. (1984) conducted an experiment to compare the decision quality achieved by interacting groups and non-interacting groups. Their results have shown that interacting groups produce better quality decisions than non-interacting groups but ineffective interaction may also result in relatively poor decisions. To facilitate group decision-making, a group decision support system is designed to provide an approach to deal with increasingly complex decision environments. It also enables group members to structure problems, share information and integrate knowledge. Overall, group decision-making represents an important role in organisations (Almeida and Marreiros, 2005). In an experiment, we can control the interaction variable by constraining or fostering interaction amongst decision-makers.

Information Overload

Controlling the load of the information is crucial in designing a decision-making experiment. If the information is overloaded to the user, it can directly influence the decision results (Chewning and Harrell 1990). For example, Keller and Staelin (1987) pointed out that under a fixed information quality environment decision effectiveness first increases and then deceases as the quantity of the given information rises. From Keller and Staelin (1987)'s work, we can observe that the amount of the given information positively affects decision-making up to a certain point. Once the quantity of the given information is beyond this turning point, decision effectiveness declines. To investigate information overload, Galbraith (1974) defined information overload with the following relationship: "information processing requirements > information processing capacity". The term requirement represents the amount of information that has to be processed under a time constraint. The term capacity represents the ability of a human to process a certain amount of information within a time constraint. Information overload occurs when the information supply exceeds the information capacity (Eppler and Mengis 2003). Therefore information overload can be controlled by the two elements: information load and processing time.

Decision Model

Apart from using decision strategy, decision-makers also can employ models to make a decision. Many researchers have proposed different decision-making models to solve decision-making issues. Janis and Mann (1977) proposed the conflict model to consider decisions. In this model they consider decision-making as a competition process between decisions. Schoemaker (1982) reviewed the expected utility model (EU model) and related empirical studies. This model is based on the utility theory in order to solve decision problems. For the experiment, decision models can be used to facilitate the decision-making process. The selection of the decision model is according to different decision scenarios.

Decision Aids

When we design the environment of our decision, we also considered the issue of decision aids. Over the last decade, decision aids have been developed to assist in the solving of decision problems. These aids are collectively known as decision support systems (DSS). Though various definitions of DSS are proposed, DSS can be considered as an interactive and adaptable computer-based information system made to facilitate the solution of a non-structured problem in decision-making (Neumann and Hadass, 1980; Sprague, 1980; Turban, 1995). Several DSS experiments (Beauclair, 1987; Zigurs et al. 1987; George et al. 1988) have implied that with the use of decision aids, decision-makers could achieve higher decision quality.

Information Presentation

One of the most important factors affecting decision-making is the way in which the information is presented (Remus, 1984). Generally, information presentation can be divided into graphical and tabular displays. There have been inconsistent results with regards to the effects on decision-making. For example, whilst Lusk and Kersnick (1979), Lucas (1981) and Ghani (1981) found that graphical displays are more difficult to use than tubular displays, Carter (1947), Feliciano et al. (1963) and Ives

(1982) reported positive effects of graphics on decision-making. Hence based on different situations, information presentations may impact decision-making differently.

Environment

We investigate eight environmental parameters to further control the decision environment.

- Dynamics: Shankaranarayan (2003) pointed out that the construction of a dynamic decision environment is characterised by levels of data granularity, variety of decision tasks and multiple stakeholders such as data providers, decision-makers, and data custodians.
- Risk: Kahneman and Tversky (1979) pointed out that decision-making under risk can be viewed as a choice between prospects or gambles.
- Crisis: Belardo and Pazer (1995) pointed out that crisis decision-making is characterised by three distinguishing features: frequency of occurrence, scope of impact, and time pressure.
- Demographics: Demographics are population characteristics, which include age, gender, education level, etc. Buckland and Florian (1991) and Fisher et al (2003) stated that the demographical variations could generate different decision outcomes.
- Cheating: Callahan (2004) points out that cheating is characterised by creating an unfair advantage in one's own interest. During the experiment, participants may employ unfair approaches to complete the task. Thus experimenters should guide and supervise the participants to effectively accomplish the experiment.
- Human Learning: Ballinger et al. (2003) pointed out that decision skill can accumulate over sequences of people. Thus, social learning of decision skills could take place when a series of similar decision scenarios happen.
- Incentives/Reward Effects: This is in order to guarantee that all the experiment participants will seriously consider the relevant information. Incentives

such as cash or other rewards can be used to motivate participants to achieve high-quality experimental results.

- Cost: Pingle (1992) presented two decision cost factors: the cost of making errors and the cost of the decision-making process. Ballou and Pazer (1987, 1995) modelled the cost of data quality in decision-making and the cost of decision strategy. Cost is an important factor that should be considered in the decision environment.

Based on the review of influencing factors, we summarise the influencing factors and the parameters of each factor in Table 18.

<table>
<tr><th>Influencing factors</th><th colspan="12">Parameters of each influencing factor</th></tr>
<tr><td>Task Complexity</td><td colspan="4">Simple</td><td colspan="8">Complex</td></tr>
<tr><td>Decision Time</td><td colspan="4">Time Constraint</td><td colspan="8">Time Pressure</td></tr>
<tr><td>Expertise</td><td colspan="4">Expert</td><td colspan="8">Novice</td></tr>
<tr><td>Decision Strategy</td><td>WADD</td><td>EQW</td><td>SAT</td><td>LEX</td><td colspan="2">EBA</td><td colspan="2">MCD</td><td colspan="2">EBA + WADD</td><td colspan="2">EBA + MCD</td></tr>
<tr><td>Interaction</td><td colspan="4">Group Decision-Making</td><td colspan="8">Individual Decision-Making</td></tr>
<tr><td>Information Overload</td><td colspan="4">Overload</td><td colspan="8">Non-Overload</td></tr>
<tr><td>Information Presentation</td><td colspan="4">Graphical Display</td><td colspan="8">Tabular Display</td></tr>
<tr><td>Decision Aids</td><td colspan="4">Decision Support System</td><td colspan="8">Manual</td></tr>
<tr><td>Decision Models</td><td colspan="3">Conflict</td><td colspan="4">EU</td><td colspan="5">Accountability</td></tr>
<tr><td>Environment</td><td>Dynamics</td><td>Risk</td><td>Crisis</td><td colspan="2">Demographics</td><td colspan="2">Cheating</td><td colspan="2">Human Learning</td><td colspan="2">Incentives</td><td>Cost</td></tr>
</table>

Table 18: Influencing factors and the status of each factor

Although we might not employ all the factors in the experiment, we are able to address these factors and keep these potential influences to a minimum. Among these factors, some may interact with each other, for example, time pressure may generate information overload while information overload may make subjects feel time pressure. Therefore the dependency between the influencing factors should be considered during development of the decision scenario.

Based on the above influencing factors, we are able to develop a decision scenario through setting the status of each factor. Table 19 provides an example of designing the scenario. In Table 19, "√" indicates we set the influencing factor into this status. "No" means we will not employ this factor or status. Using this approach, we investigate selected scenarios in a specific domain in Chapter 4. This provides a thorough understanding in the relationship between information quality and decision-making.

Task Complexity		**Time**		**Expertise**		**Interaction**		**Information Presentation**	
Simple	Complex	Time Constraint	Time Pressure	Expert	Novice	Group	Individual	Tabular	Graphical
√	No	No	No	No	√	No	√	√	√

Information Overload	**Decision Model**	**Decision Aids**
No	No	No

Decision Strategy							
WADD	EQW	SAT	LEX	EBA	MCD	EBA + WADD	EBA + MCD
No	No	No	No	No	√	No	No

Environment							
Dynamics	Risk	Crisis	Demographics	Cheating	Human Learning	Incentives	Cost
No	No	No	√	√	√	√	No

Table 19: One example of decision scenario

In the above scenario, each subject will complete a simple decision task individually without time constraint. The subjects have no or only little experience with this decision issue. We provide both tabular and graphical information to the subjects and the information is not overloaded. No decision models and aids are used to facilitate decision-making. In the decision-making process, we select MCD decision strategy, which is for processing pairs of alternatives, to solve a two-option decision task. In the decision environment, a no-risk, no-crisis and static task is employed. Cheating and the learning behaviours of the decision-maker are considered. Decision-makers' characteristics such as age, gender, education level will be recorded in the experiment. Monetary incentives will be awarded to the subjects after the experiment is completed. Based on the consideration of exogenous factors, we can then detail the concrete decision task, specify the experimental procedure and implement the experiment.

3.4 Reliability and Validity

Two types of reliability are considered in this research: experiment reliability and measurement reliability. Experiment reliability is the ability to produce the same experimental results under the same conditions (Field and Hole 2003). One crucial factor for achieving experiment reliability is to measure the dependent variable as precisely as possible, meaning a clear, precise and objective definition is needed for the dependent variable. Measurement reliability deals with the consistency and reproducibility of the measurement (Litwin 1995). This reliability can be tested by a variety of methods, such as test-retest, alternative form, internal consistency and inter-observer. Since internal consistency is commonly used in quantitative information system research, this research employs internal consistency as the method for testing measurement reliability. The internal consistency test aims to measure the different aspects of a single issue and its result indicates how well different items measure that same issue.

In addition to reliability, two types of validity are considered in this research: experiment validity and measurement validity.

Experiment validity can be divided into internal validity and external validity. The lack of internal validity means the experimental result is affected by uncontrolled factors. To improve internal validity, Field and Hole (2003) proposed eight factors potentially threatening internal validity: group threats, regression to mean, time thread, history, maturation, instrument change, different mortality, reactivity, and experimenter effects. The above threats can be resolved or minimised by experimental controls, such as providing monetary incentive to subjects and selecting appropriate subjects at random. External validity tests how well the research findings generalise to other populations and circumstances. Two threats are associated with external validity: over-use of the special participants and restricted numbers of participants (Field and Hole 2003). Considering the two threats, external validity can be increased by carrying out empirical tests across different participants and situations.

Measurement validity indicates how well a measuring instrument measures what it is intended to measure (Litwin 1995, Hogan 2003). Four types of validity are typically measured when assessing measurement validity: face, content, criterion and construct. Face validity refers to whether the construct looks like measuring the target. The test of face validity is based on a review of the constructs by untrained judges. It empirically helps to test whether this instrument appears to be a valid measurement. Content validity deals with the extent to which a measurement represents a well-defined domain or concept. The assessment of content validity involves a review of the measurement content by a set of reviewers who possess specific knowledge of this domain. It fundamentally strengthens the rigorousness of the measurement validity. Criterion validity is the measure of how well a proposed instrument is correlated with a classic instrument or predictor. Since there is as yet no classic measurement instrument available in information quality research, this type of validity is not used in this study. Construct validity aims to assess how meaningful a measurement is for practical use. This validity mainly consists of convergent and discriminant validity. Convergent validity implies different measuring items of a given construct produce similar results. The assessment of convergent validity is analogous to testing measurement reliability. Discriminant validity, alternatively termed as divergent validity, evaluates the distinctions of the different constructs. Strong construct validity is exhibited when both convergent and discriminant validity are simultaneously supported. Various analyses can be used to support construct validity. Commonly used methods are the multitrait-multimethod matrix (MTMM), factor analysis and coefficient correlation. In this research, factor analysis is used to assess construct validity.

Reliability is a necessary but not a sufficient condition for validity (Churchill, Jr 1979). Hence if a measure is valid, it is reliable. However, no measuring instrument is perfect (Litwin 1995) and it must be acknowledged that developing a reliable and valid measure is a highly difficult process (Jarvenpaa et al. 1985). Based on the discussion concerning validity, we summarise classifications, testing methods and possible solutions of validity in Figure 13.

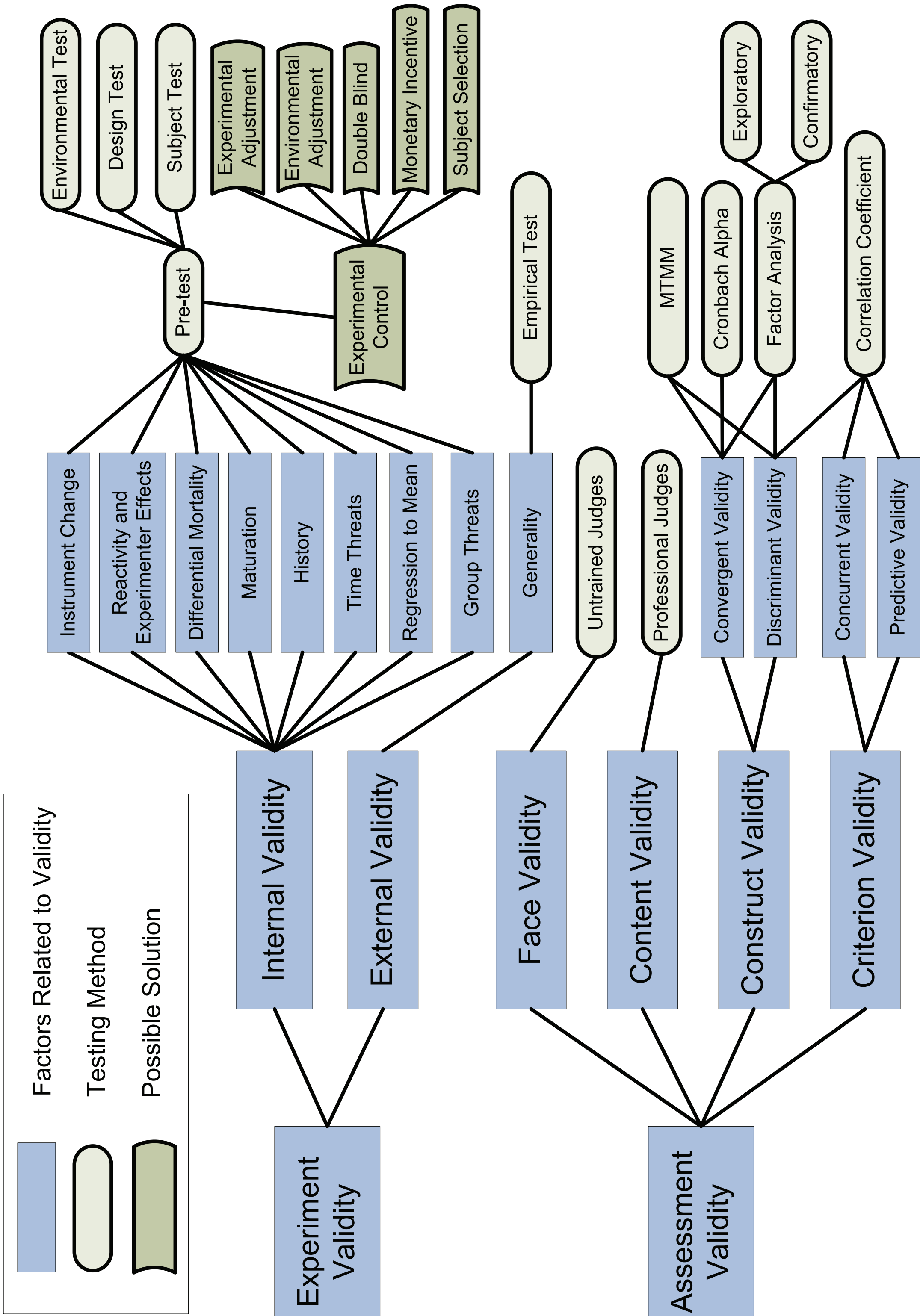


Figure 13: Experiment and measurement validity

3.5 Methodological Issues in Experimental Information System Research

Jarvenpaa et al. (1985) proposed four open methodological issues in experimental information system research: research strategy, measuring instruments, research design, and experimental task. Research strategy emphasises that the research program should be performed under a theory, a model or a framework. Two issues are related to the research strategy: a lack of theories for guiding the research (Taylor and Benbasat 1980), and studies which fail to build upon the work of others (Jarvenpaa et al. 1985). Measuring instruments focus on the reliability and validity of the measurements. Research designs concentrate on two issues: the importance of the research and the absence of experimental control (Jarvenpaa et al. 1985). Experimental task refers to a work that is taken by subjects in the experiment. The task is considered inappropriate when it is ambiguous or excessively complex. An ambiguous task might consist of inconsistent, incomplete and incorrect problems. An overly complex task may foster in the subjective influences such as preference, experience and even gambling.

Recognising the above four issues, we provide the solutions on the basis of discussions in this chapter, shown in Table 20:

Issues	Solutions
Research strategy • Lack of theories for guiding the research • Studies without building upon the work of others	The research model is based on a solid literature foundation, from both industry and academia. We have provided an extensive literature review on the relationship between information quality and decision-making in Chapter 2.
Measuring instrument • Reliability • Validity	We adopt the validated information quality dimensions from Wang and Strong (1996). The reliability and validity of these dimensions are tested by Cronbach's alpha and factor analysis in Chapter 5.
Research design • The importance of the research • The lacking of the experimental control	Previous literatures state that it is valuable to address information quality effects on decision-making. 10 influencing variables are considered in Chapter 3: task complexity, time, interaction, expertise, information presentation, information overload, decision model, strategy, aids and environment.
Experimental task • Ambiguous • Overly complex	This issue has been examined by the pilot study in Chapter 4

Table 20: Solutions to methodological issues

3.6 Summary

This chapter first provides an overview of the concept of research, including definition, characteristics, classification and process. Then 5 research methodologies employed in this thesis are discussed: survey, simulation, interview, math modelling, and laboratory experimentation. Based on the methodologies, a detailed research design is presented, which primarily consists of survey design and experimental design. As crucial issues in this research, reliability and validity are assured. Also, their solutions and testing methods are presented. Using the above discussion, this chapter addresses the solutions to four important methodological issues in our experiments and presents the detailed research methodology used in this work.

Chapter 4: Research Model and Hypothesis

In Chapter 2 and 3, we discussed the literature review and the research methodologies we used in this research. This provides us with a conceptual foundation for detailing the research model and the hypothesis. In this chapter, we propose the research model and hypotheses to investigate the effects of information quality on decision-making in more detail. To validate the model and examine the hypotheses, two experiments are designed. One is in order to investigate the effects of information quality categories on decision-making and the other is to investigate the effects of information quality dimensions on decision-making. In the experimental design, a main challenge is the information assessment, which will be discussed in Chapter 5.

4.1 Introduction

One important factor concerning information quality in today's organisations is that it directly influences decision-making. Owing to this, recent information quality research shows an increasing tendency to study the relationship between information quality and decision-making. Although their research findings confirmed that making correct decisions is dependent upon high quality information, exactly how information quality affects decision-making is still not entirely understood (Fisher et al 2003). To systematically examine the effects of information quality on decision-making, this chapter details the effects of information quality to the effects of its categories and dimensions. In our study, intrinsic, contextual and representational are used as categories of information quality. Accuracy, completeness and consistency are used as dimensions. Additionally, experimental methodologies will be used to study the relationship between information quality and decision-making. We describe the general research model in Figure 14.

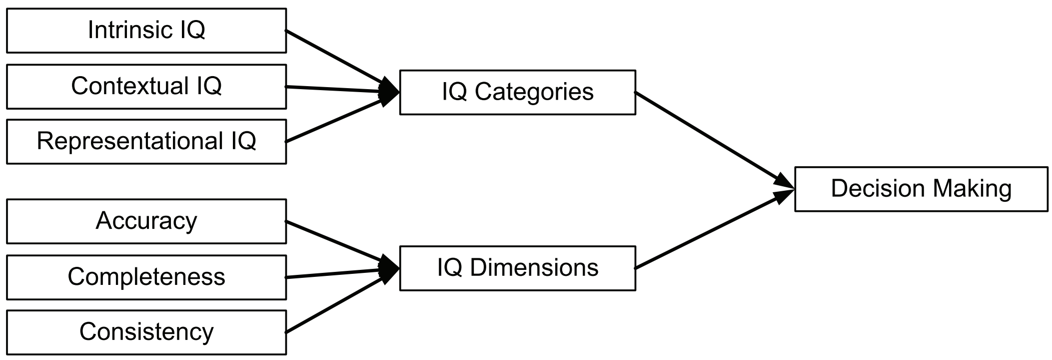


Figure 14: General research model for investigating the effects of information quality on decision-making

In previous research, many studies have confirmed the causal relationship between information quality and decision-making. For example, Keller and Staelin (1987) concluded that decision effectiveness can be improved by increasing total information quality. Belardo and Pazer (1995) discussed two causal curves between information quality and decision quality. Jung and Olfman (2005) found that decision-makers can expect improvements in decision performance through enhancing contextual information quality. Hence, generally we can observe the effects of information quality on decision-making. In terms of a research model, information quality can be identified as an independent variable, and decision-making depends on information quality.

As introduced in Chapter 1, Wang and Strong (1996) proposed a hierarchical framework to organise information quality dimensions. This framework consists of three layers, which are information quality, information quality categories and information quality dimensions. Structuring our research along these 3 layers, the elements in each layer can be considered as independent variables in the relationship between information quality and decision-making.

In the information quality layer, total information quality can be used as one of the independent variables. For example, Raghunathan (1999) investigated the impact of total information quality and the decision-maker's quality on decision quality. Keller and Staelin (1987) examined the effects of total information quality and information quantity on decision effectiveness.

In the information quality category layer, the four categories can be considered as independent variables. For example, Jung and Olfman (2005) conducted an experiment to study the effects of contextual information quality and task complexity on decision performance.

In the information quality dimension layer, different dimensions such as completeness and relevance can be included as independent variables. For example, Ahituv et al. (1998) investigated the effects of completeness of information and time pressure on decision performance. Letzring et al. (2006) studied how relevance of information and information quantity affect decision accuracy.

As discussed in Chapter 2, information quality dimensions can be expressed through concrete information quality problems. Therefore information quality problems such as duplicate data and missing data are often regarded as independent variables. For example, Ballou and Pazer (1990) proposed a model to analyse the impact of information quality problems and decision strategy on decision accuracy.

In addition to the information quality hierarchy, a number of researchers employ *data quality information* as the independent variable. For example, considering decision strategy, task complexity, expertise and decision time, Chengalur-Smith et al. (1999) and Fisher et al. (2003) explored the effects of *data quality information* on decision outcomes. We summarise the above discussion in Figure 15.

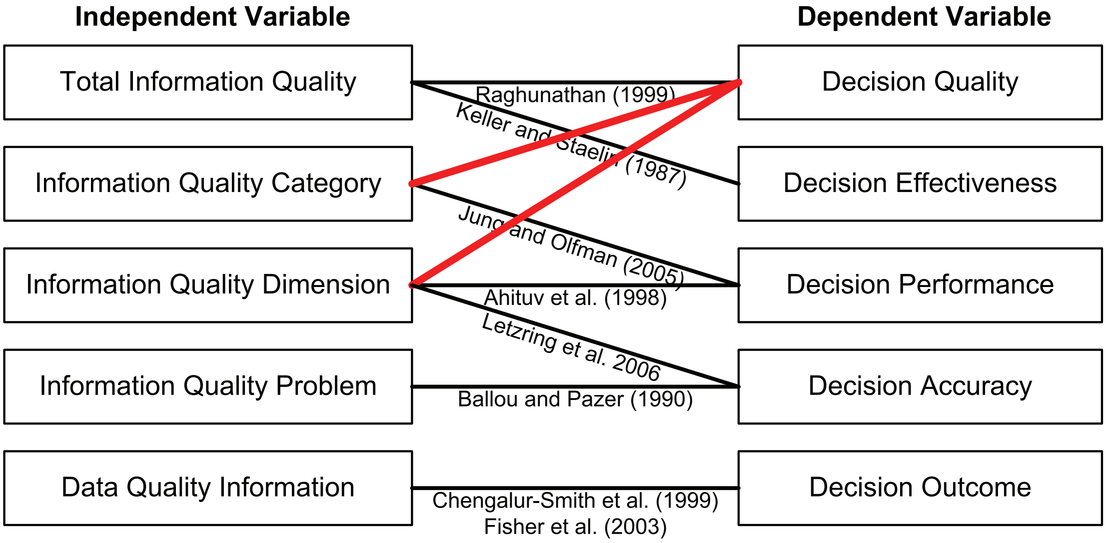


Figure 15: Review of information quality effects on decision-making

As illustrated in Figure 15, we show that different dependent variables have been used to study the relationship between information quality and decision-making, for example, decision quality (Raghunathan 1999), decision effectiveness (Keller and Staelin 1987) and decision performance (Jung and Olfman 2005). Although these terms are literally different, on closer examination the primary measurement used in the research is usually the correctness of the decision. In this research, we select the term decision quality, which can be found in a wide range of literature. Therefore the research objectives in Chapter 1 are detailed as investigating the effects of information quality categories and dimensions on decision quality.

4.2 Model and Hypothesis

From the previous research findings, we are able to conclude that decision-making clearly depends on information quality. However, besides information quality there are also other factors affecting decision quality, such as time pressure and task complexity (refer to Chapter 3). Therefore, we can specify our general research model as relationship between information quality, decision quality and exogenous factors. In principle, decision quality can be expressed as a function of information quality and a function of other influencing factors.

$$\text{Decision Quality} = \text{f (IQ)} + \sum_{n=1}^{m} g_n\ (F_n) \qquad \textbf{(4-1)}$$

Where f (.) is the effect function of information quality and $g_n(.)$ is the effect function of other influencing factors F_n. Note that $g_n(.)$, and F_n are domain specific. They have to be derived in a concrete domain.

Now we can adopt this formula to our research and express the effects of intrinsic, contextual and representational information quality on decision quality. Equation 4-1 can be re-formed to:

$$\text{Decision Quality} = f_i\,(\text{IIQ}) + f_c\,(\text{CIQ}) + f_r\,(\text{RIQ}) + \sum_{n=1}^{m} g_n\ (F_n) \qquad \textbf{(4-2)}$$

In equation 4-2, IIQ, CIQ and RIQ respectively represent intrinsic information quality, contextual information quality and representational information quality. f_i (.), f_c (.) and f_r (.) are the effect functions for the information quality category. To further investigate equation 4-2, we will elaborate this equation to a graphical model and hypotheses.

When investigating the effects of information quality categories on decision quality, we select supply chain management as our domain since inventory control is a typical business operation and it is also well developed in previous experimental research. In the inventory management, we focus on two typical inventory control policies, which are Just In Time (JIT) and Economical Order Quatity (EOQ). As the two policies can be used to establish the best decision, and thus can be used to measure decision quality. The dependent variables are detailed as JIT decision quality and EOQ decision quality. Accordingly two scenarios are respectively designed. This graphical model can be described in Figure 16.

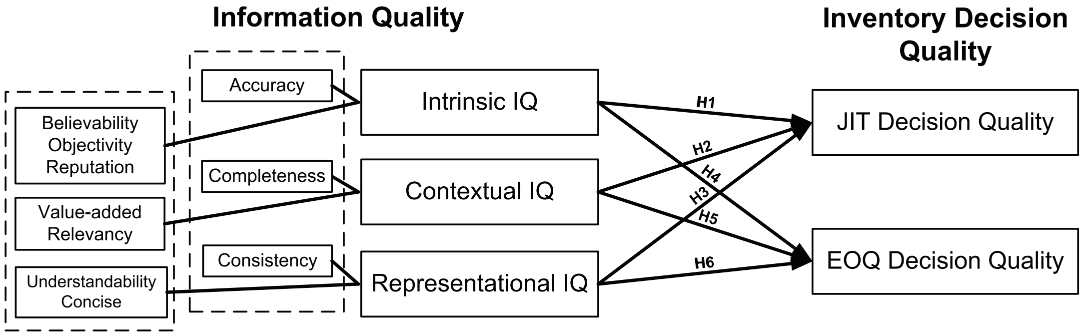


Figure 16: Model of the relationship between information quality categories and inventory decision quality

In a first step, we distinguish between decision quality for JIT and for EOQ. The two scenarios differ in the following. In inventory control, the aim of JIT is to eliminate the need for holding inventory items. It is designed to reduce the inventory to the ideal status of zero inventories (Hall 1983). The usage of JIT policy can eliminate holding cost, material waste and production bottlenecks (Rao and Scheraga 1988). In contrast to JIT, EOQ is a popular traditional inventory control technique and aims to determine an optimal order quantity by balancing holding cost and ordering cost. EOQ centres on minimising total costs instead of minimising inventories. The benefit of EOQ is exhibited when inventory purchasing is at high levels of annual demand (Fazel 1997).

As discussed in Chapter 1, we can classify information quality into categories and each category contains several dimensions. However, some of the dimensions are excluded within our study: firstly, since no real time information is used in our experiment, timeliness is dropped. Secondly, the dimension regarding appropriate amount of data may confuse the correlation between information quality and decision-making with information quality and information overload. Consequently, appropriate amount of data is dropped as a dimension in order to clarify the concept of information quality and information overload. Thirdly, within our experiment we do not address interpretability of symbols and characters, and consequently interpretability is not considered.

Considering these revisions (Figure 16), we include the following information quality dimensions in our research model: intrinsic information quality consists of 4 information quality dimensions: accuracy, believability, objectivity and reputation. Contextual information quality contains 3 information quality dimensions: completeness, value-added and relevancy. Representational information quality includes 3 information quality dimensions: consistency, understandability and concise representation. Concerning measurement, accuracy, completeness, and consistency can be measured by both objective and subjective approaches. Other dimensions are only measured subjectively.

From information quality category perspective, we refine our research question into hypotheses. As our review shows a positive correlation between information quality and decision quality, we propose the following hypotheses:

> **H1**, Intrinsic information quality is positively associated with JIT decision quality.
>
> **H2**, Contextual information quality is positively associated with JIT decision quality.
>
> **H3**, Representational information quality is positively associated with JIT decision quality.
>
> **H4**, Intrinsic information quality is positively associated with EOQ decision quality.
>
> **H5**, Contextual information quality is positively associated with EOQ decision quality.
>
> **H6**, Representational information quality is positively associated with EOQ decision quality.

Since little research and practice assesses information quality in the category level, no sub-hypotheses are proposed for each main hypothesis. In order to examine the hypotheses 1-6, we design experiment 1 to examine the relationship between information quality categories and decision quality. This experiment includes two scenarios. One scenario is for JIT decision quality and the other scenario is for EOQ decision quality.

After proposing the model and hypotheses for testing the effects of information quality categories on decision quality, we then investigate the hypotheses regarding the effects of information quality dimensions on decision quality. Using equation 4-1, the effects of information quality dimensions on decision quality can be expressed as follows.

$$\text{Decision Quality} = \sum_{k=1}^{i} \mathrm{f_k}\,(\mathrm{D}_k) + \sum_{n=1}^{m} \mathrm{g}_n\,(\mathrm{F}_n) \tag{4-3}$$

Where D_k represents certain information quality dimension and $f_k()$ represents its effect function. Since we considered accuracy, completeness and consistency as the selected dimensions, equation 4-3 can be re-formed to.

$$\text{Decision Quality} = \mathrm{f}_1\,(\text{Accuracy}) + \mathrm{f}_2\,(\text{Completeness}) + \mathrm{f}_3\,(\text{Consistency}) + \sum_{n=1}^{m} \mathrm{g}_n\,(\mathrm{F}_n) \tag{4-4}$$

Accuracy of information can be defined as "the extent to which data are correct, reliable, and certified free of error" (Wang and Strong 1996). The intrinsic characteristic of accuracy is the correctness of data values stored for an object (Olson 2003). To further explain the meaning of correctness, Ballou and Pazer (1985) defined accuracy as a recorded value that conforms to the actual value. This conformity consists of two aspects: form and content (Olson 2003). Therefore high-accuracy information will both contain the correct values and be represented in a tangible form. O'Reilly (1982) and Ballou and Pazer (1995) found that decision-makers can benefit from high-accuracy information because it improves information processing and retrieval activities. Thus we can propose the following hypotheses.

H7: Accuracy of information is positively related to decision quality.

H7a: When information is complete and consistent, accuracy of information is positively related to decision quality.

H7b: When information is complete and inconsistent, accuracy of information is positively related to decision quality.

H7c: When information is incomplete and consistent, accuracy of information is positively related to decision quality.

H7d: When information is incomplete and inconsistent, accuracy of information is positively related to decision quality.

Completeness of the information can be defined as "the extent to which data are of sufficient breadth, depth, and scope for the task at hand" (Wang and Strong 1996). This definition is task-centred and derived from the intended use of the information consumer. According to this objective and data-centred view, completeness is defined as all values for a certain variable are recorded (Ballou and Pazer 1985). From these two major definitions, we can observe two components that are vital to completeness: content and structure (Ballou and Pazer 2003). Therefore high-completeness information is achieved when information content and structure are both at a high-quality level. That means high-completeness information must contain no NULL values and carry the full meanings for its intended task. Ahituv et al. (1998) found that decision performance can be improved by increasing completeness of information. Hence it is possible to infer the following hypotheses.

H8: Completeness of information is positively related to decision quality.

H8a: When information is accurate and consistent, completeness of information is positively related to decision quality.

H8b: When information is inaccurate and consistent, completeness of information is positively related to decision quality.

H8c: When information is accurate and inconsistent, completeness of information is positively related to decision quality.

H8d: When information is inaccurate and consistent, completeness of information is positively related to decision quality.

Consistency of the information can be defined as the extent to which multiple data are conflicting with each other in the areas of view, value and format (Ge and Helfert 2008). Redman (1996) refined consistency into view, value and representation consistency. View consistency refers to the consistency of different components of the view and the consistency of attributes amongst entity types. This level of consistency is often required in the data schema design. Value consistency examines inconformity between data values. It concerns inconsistent contents of overlapping entities. Representation consistency is defined as when the formats of the data value are the same in all cases (Ballou and Pazer 1985). This level of consistency is frequently discussed in information quality research and termed as representational consistency by Wang and Strong (1996). Since view consistency is not practically used with all types of data, this study only focuses on value and representational consistency. Therefore, inconsistency in this study refers to the value and format differences of the actual data. Based on the implications of previous research (Redman 1996, Ballou and Pazer 2003), we are able to propose the following hypotheses.

H9: Consistency of information is positively related to decision quality.

H9a: When information is complete and accurate, consistency of information is positively related to decision quality.

H9b: When information is complete and inaccurate, consistency of information is positively related to decision quality.

H9c: When information is incomplete and accurate, consistency of information is positively related to decision quality.

H9d: When information is incomplete and inaccurate, consistency of information is positively related to decision quality.

To summarise, a total of 21 hypotheses are proposed, consisting of 9 major hypotheses and 12 sub-hypotheses. Hypotheses 1-6 are proposed to study the effects of intrinsic, contextual and representational information quality on decision quality. Hypotheses 7-9 are proposed to investigate the effects of accuracy, completeness and consistency on decision quality.

4.3 Experimental Design

In order to examine the hypotheses, we designed two laboratory experiments. With experiment 1, we aim to investigate effects of information quality categories on decision quality. This experiment is used to examine hypotheses 1-6. With experiment 2, we aim to further understand the relationship between information quality dimensions and decision quality. This experiment is used to test hypotheses 7-9.

4.3.1 Experiment 1

Experiment 1 is based on the popular and well known management game, the Beer Game. This game is a role-playing simulation, which involves managing supply and demand in a beer supply chain. The concept for this game was first developed at the Massachusetts Institute of Technology in the 1960s. Since then, several extensions and modifications have been suggested. Kaminsky and Simchi-Levi (1998) identified several weaknesses in this traditional game and consequently developed the computerised Beer Game. Based on the computerised Beer Game, we provide various quality levels of marketing and selling information to subjects. Using the given information, subjects are asked to make inventory control decisions.

Experiment 1 is implemented in 2 scenarios. Although both scenario 1 and 2 use the concept of the computerised Beer Game (Kaminsky and Simchi-Levi 1998), the objective and configuration of each scenario are different. One difference is the order complexity: in scenario 1, subjects order one identical brand of beer in 10 weeks and make one decision in each week. In scenario 2, subjects order 10 different brands of beer and make one decision within each brand. The second difference is the objective of the experiment: scenario 1 uses JIT inventory control policy. Its goal is to minimise inventory. Therefore the optimal decision in scenario 1 results in zero inventory. Scenario 2 uses EOQ inventory control. The goal of EOQ policy is to minimise the total inventory costs. Thus the optimal decision in scenario 2 is the decision which results in the lowest total cost. The differences between scenario 1 and scenario 2 are summarised in Table 21.

	Order Complexity	Objective	Optimal Decision
Scenario 1	• One identical brand of beer over 10 *weeks* • One decision in each *week*.	Minimise inventory	Zero inventory
Scenario 2	• Order 10 different brands of beer • Make one decision for each brand.	Minimise total costs	Minimal total costs

Table 21: Differences between scenario 1 and scenario 2

We use the following design to control the variables in both scenarios. In each scenario we manipulated three independent variables. Thus this study implemented two $2 \times 2 \times 2$ factorial designs which controlled two levels for the independent variable: low or high. The measurement of decision quality was identical for both scenarios but establishment of the best decision was according to the goal of the scenario. Both scenarios can mutually confirm experimental results and test external validity of the experimental results.

Measurement

One fundamental problem in designing such experiments lies in determining or measuring the quality level of information. To simplify the measurement, we only considered two levels for each information quality category, high and low respectively. To divide the quality levels, both objective and subjective measures are used.

High information quality for one category is achieved when all the dimensions in this category are at the level of high quality. For the dimensions only assessed by a subjective approach, information quality professionals specifically design these information quality dimensions to be of high quality. For the dimensions assessed by objective and subjective approach, we first objectively set information quality value as high quality. Then information quality professionals are asked to confirm the objective information quality level. If the information quality professionals disagree with the objective information quality level, we adjust the objective information quality value until the information quality professionals agree. For example, accuracy, believability, objectivity and reputation are the 4 information quality dimensions in intrinsic information quality. Believability, objectivity and reputation are the dimensions only assessed by subjective approach, and accuracy can be assessed by objective and subjective approach.

When we design the high intrinsic information quality, we for example temporarily set accuracy to 80% as a high-quality accuracy. If the information quality professionals do not consider this value as high-quality accuracy, we adjust the value until the information quality professionals consider the objective value to be high quality. Likewise, low information quality of one category is achieved when all the dimensions in this category are at the level of low quality.

One problem in designing objective information quality levels is that 0% and 100% are not necessarily best to represent low information quality and high information quality. The ideal division of information quality level is that: (1) both high-quality information and low-quality information can be used for decision-making. (2) The largest difference between low information quality and high information quality is achieved. (3) Subjects are not aware of the information quality changes between low information quality and high information quality.

Since the measurement of information quality is an important aspect in this study, we will further discuss the methodology of considering objective and subjective measurement in Chapter 5.

As discussed in Chapter 3, decision quality can be calculated by the ratio between the number of correct decisions to the number of total decisions. This study presents two experiments. In each experiment every subject is asked to complete 10 inventory control decisions. Decision quality is then computed by the number of correct decisions out of 10. Correct decisions are the ones which conform to the optimum decision.

Exogenous Factors

As discussed in Chapter 3, in addition to information quality other factors may also influence decision quality. These exogenous factors are difficult to control in experiments, and thus we need to keep the influence of exogenous factors to a minimum. In order to control the exogenous factors we configure experiment 1 as follows: All the subjects have no or very little experience with inventory decision issues. Each subject completes two decision tasks individually without time constraint. Both tabular and graphical information is provided to subjects. The information provided would not typically be regarded as information overload. No decision models, calculators or any other aids are used to facilitate decision-making. The decision task is non-risk, non-crisis and non-dynamic. Cheating and interacting behaviours of decision-makers are forbidden and ensured against by the presence of supervisors. The demography of the decision-makers, such as age, gender, education level, is recorded as part of the experiment. Monetary incentives are awarded to subjects after completion of the experiments.

Scenario 1

For this scenario we use a four-component beer supply chain: manufacturer, distributor, retailer and customer. One round of the experiment includes 10 weeks. In each week, the order of events is as follows: (1) Manufacturer fills the distributor's demands of last week. (2) Distributor fills the retailer's demands of last week and places an order with the manufacturer for next week. (3) Retailer fills the customer's demands for this week and places an order with the distributor for next week. If these demands are not catered for, unsatisfied demands are recorded as back orders. The manufacturer is guaranteed to provide enough products for the distributor. Therefore there is no back order with the manufacturer. In the beginning of the game, there are no back orders in any component and the demands of the last week are perfectly satisfied.

A software-based system is developed to deliver the experimental scenario (Figure 17). Subjects play the role of distributors who place orders to manufacturers and meet the demands of retailers. The other three roles are taken by the computer. To simplify the design of the JIT inventory control, no lead time is set between distributors and manufacturers. This is to encourage subjects not to stock any product, in other words to achieve zero inventory. In each week, we provide the marketing information and selling history to subjects. According to the given information, subjects are able to make more reliable and rational inventory decisions. In one round, subjects are asked to place 10 orders to the manufacturers. Orders which conform to the optimum decision are recorded as correct inventory decisions. Since the goal of this experimental scenario is to minimise the inventory down to zero, the best decision is determined by the order which equals the retailer's need plus meets existing back orders.

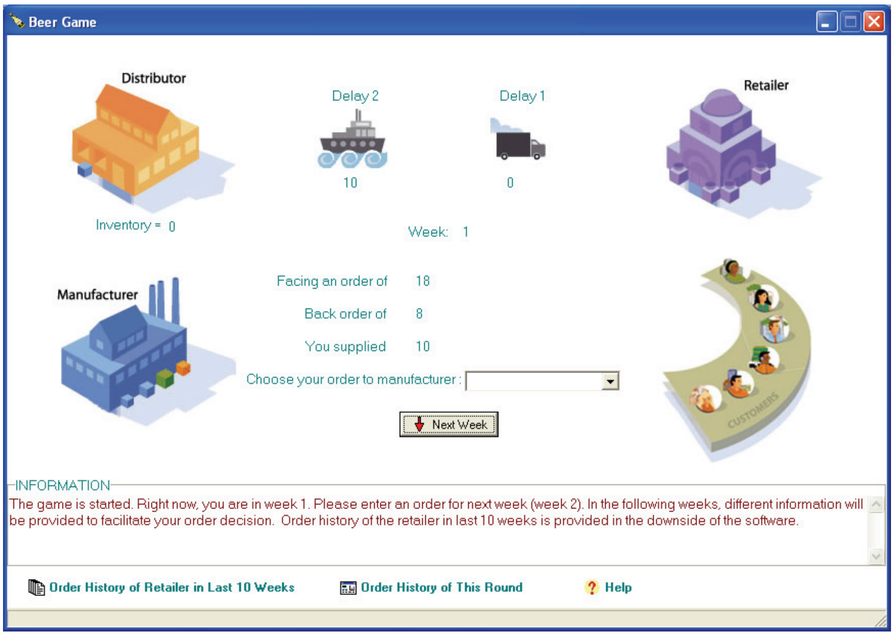


Figure 17: Experiment of JIT inventory control

Scenario 2

This scenario is also based on the Beer Game. However, all the ordering decisions are made in the first week. The decisions are used as an ordering scheme for the following weeks. One decision includes two primary elements: how many products should be ordered and how often orders should be placed. One round of the experiment is counted by 10 ordering decisions. Two costs are associated with this scenario: the cost of placing an order and the holding cost per week. The goal of this experiment is to identify the minimum cost to meet the demands. According to the EOQ model, the best decision can be calculated by $\sqrt{\dfrac{2 \times \text{weekly demand} \times \text{ordering cost}}{\text{weekly holding cost}}}$

To meet the assumptions of the EOQ model, we specify the scenario as follows:

- Weekly demand for each brand of beer is known, deterministic and constant.
- Lead time is not taken into account. This means replenishment can be effected instantaneously.
- Replenishment cycle is repetitive. This means ordering conditions remain the same over using EOQ.
- No stock outs and quantity discounts are considered.

Subjects are also asked to play the role of distributor. In contrast to scenario 1, subjects obtain information about the selling history all at one time. Accordingly they make ordering decisions for the remaining weeks. This scenario is implemented in the following software-based system (Figure 18).

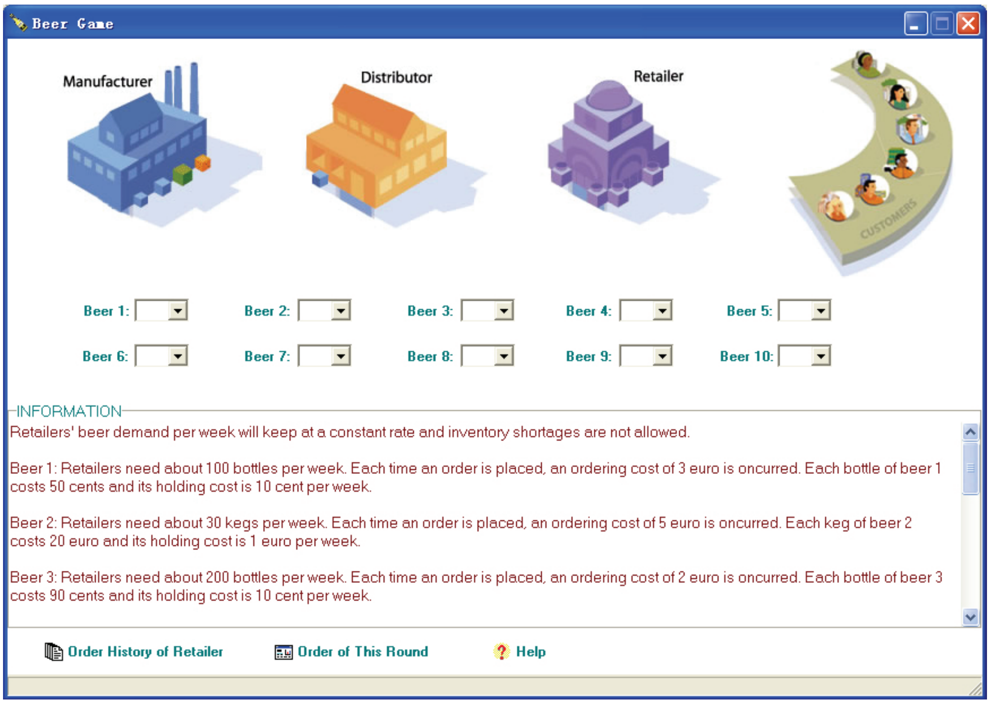


Figure 18: Experiment of EOQ inventory control

Pilot Study

To test the experimental procedure and to show the feasibility of manipulating factors, a pilot study was conducted. The pilot study involved 16 subjects. 8 experimental treatments were implemented: high/low intrinsic information quality × high/low contextual information quality × high/low representational information quality. Subjects were divided into eight groups. Each group was assigned to one of experimental treatments. After the experiment, we collected the results and feedback from the subjects.

The pilot study led to three revisions of scenario 1. First, we shortened and clarified the instruction. A diagram of the order of events and an illustrative example were provided to help the subjects to understand the experimental scenario. Secondly we adjusted the schema of decision cost. The cost of holding each unit in the inventory was adjusted to equal the cost of each back order. This adjustment was to prevent the preference of subjects for holding extra units in inventory. Third we reduced the degree of task complexity. Instead of entering an order number by subjects, five options were provided to facilitate the order decision.

For scenario 2, two revisions were incorporated. Firstly we highlighted the differences in instructions between scenario 1 and scenario 2, because each subject primarily needed to finish scenario 1 and then complete scenario 2. This revision made it easier for subjects to comprehend the aim of scenario 2. Secondly, we declared that calculation was forbidden in the experiment. This revision was to prevent the subjects' from learning in the experiment.

4.3.2 Experiment 2

While experiment 1 aims to investigate the effect of information quality categories on decision quality, we designed a second experiment to detail the effect of selected dimensions on decision quality. This experiment is also adopted from the computerised Beer Game (Kaminsky and Simchi-Levi 1998). In this experiment, the beer supply chain once again consists of four components: manufacturer, distributor, retailer and customer. Participants are asked to play the role of distributor. The other three roles are taken by the computer.

One round of the game is 10 weeks. In each week, the distributor fills retailer's demand as much as possible and places an order to the manufacturer. If the retailer's demand is not catered for, the unsatisfied demands are recorded as back orders. The cost of holding each item is €0.5 per week and the cost of each back order is €1 per week. The objective of the game is to minimise total cost. Therefore, this forms a trade-off between minimum inventory and minimum back orders. Game participants need to decide the appropriate ordering number to maintain an optimal inventory.

Additionally, there is no initial cost applied at the beginning of the game.

Between each two components in the supply chain, there are two delays that represent the shipment between the components. Suppose the distributor places an order to the manufacturer in week w. This order arrives with the manufacturer in week w+1 and then the manufacturer attempts to fill the order using the available inventory. These filled orders are shipped in delay 1. Next, in week w+2, the filled orders are shipped in delay 2. Finally the distributor could obtain the supplied items at the start of week w+3. Therefore the leading time in this game is two weeks.

To clarify the supply and order procedure, we provide the order of events occurring in each week. A total of 5 events are specified as follows:

1) The items of delay 2 are moved to the inventory of the downstream component.
2) The items of delay 1 are moved to delay 2.
3) Orders from the downstream component are filled to the maximum extent possible. The supplied items are moved to delay 1. For example, when the manufacturer receives the orders from the distributor, using available items, the manufacturer fills the orders as much as possible. Then these supplied items are shipped to delay 1, which is between the manufacturer and distributor. The unfilled orders are accumulated to the manufacturer's back order. The same procedure also occurs between the retailer and the distributor.
4) The cost of inventory and back order is calculated and added to the total cost. Note that the cost of shipping items in the delay is considered as holding cost of the upstream component. For example, distributor's holding item is the items in distributor's inventory plus the items in delay 1 and delay 2 which are between distributor and retailer.
5) Orders are placed to upstream component. This final step requires game participants to place an order to the manufacturer.

Considering the subject's role is that of the distributor, we describe the 5 steps in Figure 19.

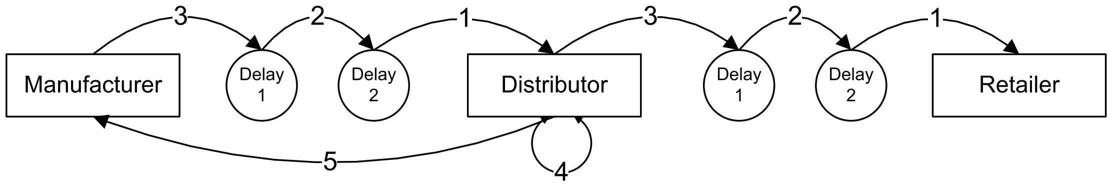


Figure 19: Order of events in each week

The game is started at the end of week 0 and ended at the beginning of week 10. That means in the beginning of the game, Participants are in the event 5 of week 0 and are asked to place an order for the following weeks. At the end of the game, the final event is the event 4 of week 10. The other 9 weeks, week 1 to week 9, consist of all the 5 events. We describe all the events occurred in the 10 weeks by Figure 20.

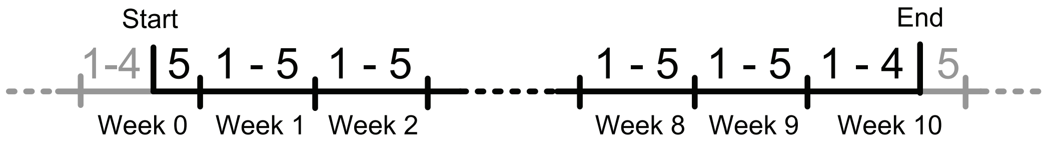


Figure 20: Total events in 10 weeks

In each week, the order from the retailer and the finished products from the manufacturer are generated at random. The participants, who are playing distributor, need to manage the supply to the retailer and the order to the manufacturer. Considering the cost, they also need to coordinate the inventory and the back orders. This game is based on a software system, as shown in Figure 21.

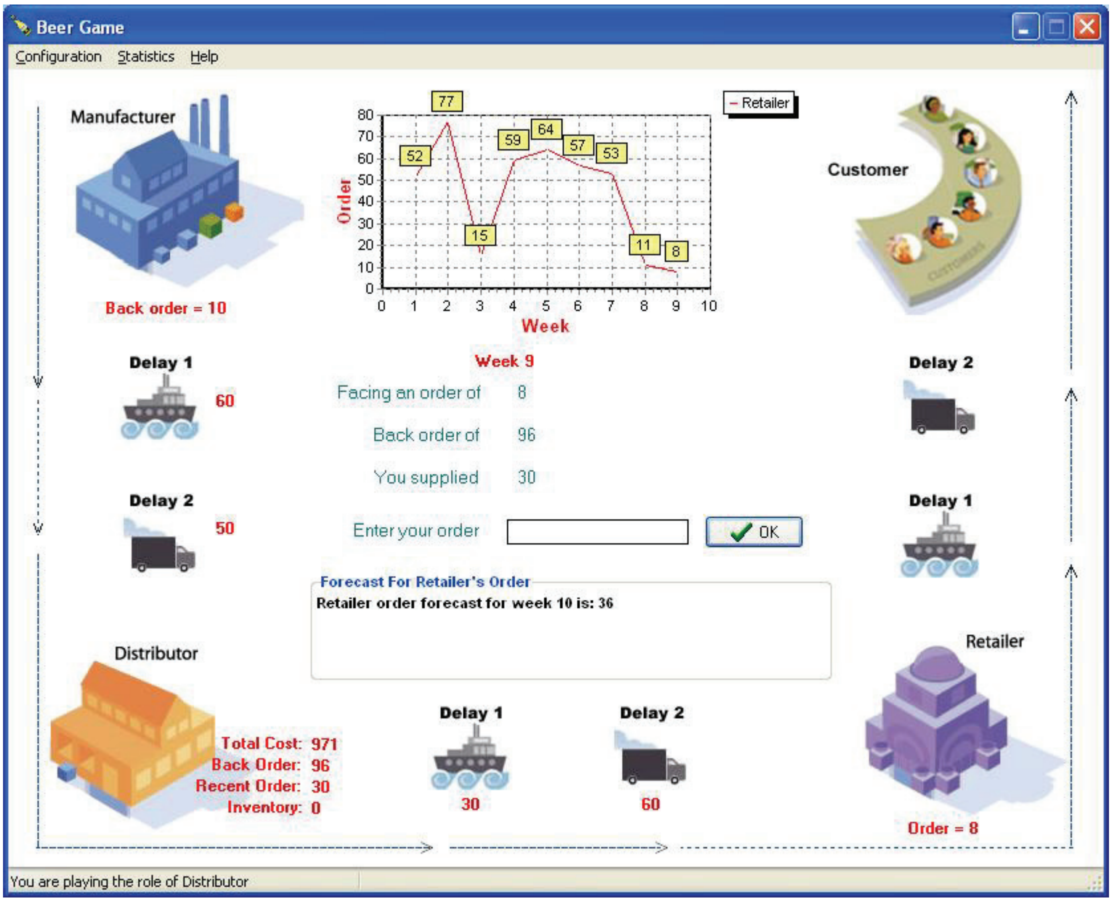


Figure 21: Screenshot of the software in experiment 2

To facilitate the participant's ordering decisions, we provide the forecast information about the retailer's future order. This information is varied at different quality levels to different participants. Since three information quality dimensions, accuracy, completeness and consistency, are used as independent variables, a 2 × 2 × 2 factorial design is implemented in the experiment. Thus 2 quality levels are applied to each information quality dimension and 8 treatments are used to test the hypotheses.

To further develop the experiment, we need to consider three crucial issues: (1) How to measure independent and dependent variables? This is a core issue in the experimental research. (2) How to keep the independent variables not correlated? As higher correlation between independent variables could increase the sampling error of the partials (Blalock 1963), we need to specify which information quality problems are linked to which information quality dimensions. (3) How to control the exogenous factors that may affect experimental results? Kerlinger and Lee (2000) state that an ideal experiment controls all other possible factors which may affect the experimental results and show how independent variables affect dependent variables. We consider the solutions of the three issues as follows.

Measurement

To keep low correlations between the three information quality dimensions, completeness needs to be firstly considered - in that if information is missing, other dimensions are rendered obsolete. Completeness can be measured by equation 4-5:

$$\text{Completeness} = \frac{\sum_{k=1}^{N} \text{Completeness}(D_k)}{N} \qquad \textbf{(4-5)}$$

Where N is the total number of the data item D_k. If D_k is not null, completeness (D_k) is equal to 1. Otherwise completeness (D_k) is equal to 0. The procedure of determining completeness (D_k) is discussed in Section 5.4.1. In the experiment, 10 pieces of information are used to forecast the retailer's order. If completeness is designed to 60%, that means 4 pieces of forecast information are missing.

Based on the complete information, accuracy can be measured by equation 4-6. In this equation, N represents the total number of complete data item D_k. for each data item, accuracy (D_k) measures the distance between the current value and the correct value (CV).

$$\text{Accuracy} = \frac{\sum_{k=1}^{N} \text{Accuracy}(D_k)}{N} \qquad \textbf{(4-6)}$$

To calculate accuracy (D_k), a bottom value (BV), which means the most distant value from the correct value, needs to be predetermined. Once the correct value and bottom value are determined, we could compute accuracy (D_k) by equation 4-7 and 4-8. In the two equations, x represents the current value of D_k.

$$\text{Accuracy}(D_k) = \frac{x}{CV - BV} \, (BV \le x \le CV) \qquad \textbf{(4-7)}$$

$$\text{Accuracy }(D_k) = 2 - \frac{x}{CV - BV} \; (CV < x \leq BV+2|CV-BV|) \qquad \textbf{(4-8)}$$

For example, based on 6 pieces of complete information, we set the bottom value to 0. That means in the data items, the values 0 and 2CV represent the lowest level of accuracy. If accuracy is designed to 80%, one possible solution is to set the value of every data item to 80% of correct value.

Based on the accuracy level of complete information, we could define the value of consistency by equation 4-9. In this equation, N is the total number of the data item D_k. If D_k contains inconsistent information, consistency (D_k) is equal to 0. Otherwise consistency (D_k) is equal to 1. The procedure of determining consistency (D_k) is discussed in Section 5.4.1.

$$\text{Consistency} = \frac{\sum_{k=1}^{N} \text{Consistency}(D_k)}{N} \qquad \textbf{(4-9)}$$

Two types of consistency are used in this experiment: value and format consistency. Value inconsistency means D_k consists of two or more conflicting values for the same data entity. For example, when the accuracy of the information is 80%, we could provide two different values for this information. One is achieved by setting the value to 80% of the correct value and the other by setting the value to 120% of the correct value. The two values are both on 80% accuracy, but conflicting. Format inconsistency indicates that one data entity is represented by two or more formats. For instance, one figure can be represented by the form of Arabic numerals or plain English. When both forms are used, it generates format inconsistency. In this experiment, we combine value and format inconsistency to organise the consistency. That means if D_k is inconsistent, its value and format are simultaneously inconsistent. To further describe the design of measurements, we show the three dimensions in Figure 22.

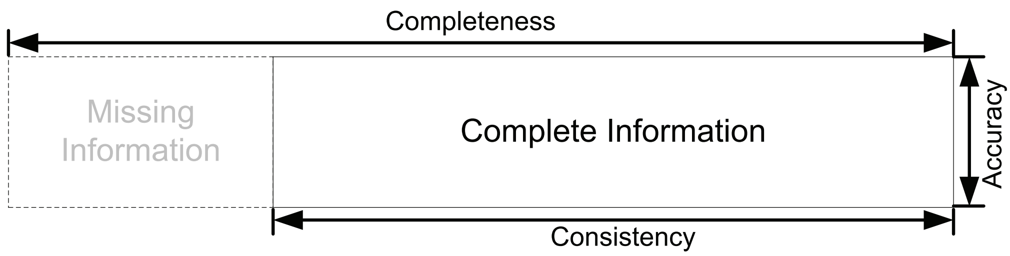


Figure 22: Design of measurements in experiment 2

In this experiment, we require two quality levels for each dimension: high information quality and low information quality. To designate the quality level, two ideal objectives are defined: both high and low information quality can be used for decision-making, and the largest difference between low information quality and high information quality is achieved. Based on the measurements of information quality dimensions, the quality level division is determined by a discussion of information quality professionals.

As the measurement of information quality dimensions is critical in this study, we will systematically discuss how to measure and control information quality dimensions in Chapter 5.

As a dependent variable, decision quality is defined as the relationship between the alternative selected by the subject and the correct or best alternative (Jarvenpaa 1989, Kaiser et al. 1992, Gonzalez and Kasper 1997). Consequently, decision quality can be measured by the proximity of the user's decision to the optimum decision. However, it proves difficult to find an objective best decision in certain decision scenarios, such as unstructured decision-making. Alternatively, Citera (1998) proposes that decision quality can be calculated by the difference between a subject's decision and the expert's decision. Nevertheless, experts may evince bias on their evaluations (Jacoby 1977). Considering the above limitations, Pingle (1992) and Kocher and Sutter (2006) propose to use decision cost as a surrogate for the measurement of decision quality. Since decision cost is practical and fit for our experiment, we adopt this viewpoint and measure decision quality by the cost of the decision's effects. That means higher decision quality can be indicated by lower decision cost.

Exogenous Factor

As stated in Chapter 3, the aim of experimental research is to show how independent variables only cause the changes of dependent variables (Oates 2005). Therefore, it is necessary to control exogenous factors that may influence the experiment. Controlling exogenous factors means eliminating such factors or holding them constant. Using the framework of controlling exogenous factors in Chapter 3, we manage the influence of exogenous factors as follows: all the subjects have no or only little experience with this decision issues. Each subject completes decision tasks individually without time constraint. Only tabular information is provided to subjects for the making of decisions. The given information is not typically regarded as information overload. No decision models, calculators or any other aids are used to facilitate decision-making. The decision task is non-risk, non-crisis and non-dynamic. Cheating and interacting behaviours of decision-makers are forbidden and ensured against by supervisors. The demography of the decision-makers, such as age, gender, education level, is recorded as part of the experiment.

As this experiment is based on experiment 1, and we collected much experience from previous experiments, pilot study of this experiment is not implemented.

4.4 Summary

In this chapter, we have proposed 9 major hypotheses. Hypothesis 1-6 are used to examine the relationship between information quality categories and decision-making. Hypothesis 7-9 are used to examine the relationship between information quality dimensions and decision-making. Along with hypothesis 7-9, 12 sub-hypotheses are proposed to investigate in detail the effects of accuracy, completeness and consistency on decision-making. In order to test these hypotheses, two experiments are proposed, both of which are based on the Beer Game. We adopt the traditional Beer Game to our experiments. In the experiment, one critical challenge is to determine the measurement of information quality. In order to address this challenge, Chapter 5 will systematically discuss the measurement of information quality.

Chapter 5: Practical-oriented Information Quality Assessment

In Chapter 4, we proposed a research model along with hypotheses and two experimental settings to examine the effect of information quality on decision making. The impact on decision making will be measured by the quality of the decision. One major challenge we identified for our experimental design is to develop a suitable measure for information quality. Therefore, this chapter presents our framework to assess information quality. This framework consists of two major components: a set of validated information quality dimensions and a methodology to assess information quality. In order to validate the framework, first, we conduct a survey to confirm a set of validated dimensions. Subsequently, based on our dimensions, we propose assessment methodologies in the scenario of a single database and multiple databases. The developed information quality framework provides a key component for conducting the experiments and analysing the data in Chapter 6.

5.1 Introduction

As outlined in Chapter 4, when we propose our experimental settings, measuring information quality is a major element. However, as we could observe from our literature review in Chapter 2, foremost research does not provide a sufficient theoretical base for conducting our research. Current frameworks lack validated dimensions and suitable assessment techniques. In order to address these limitations, we developed an information quality framework that allows us to control the measurement of information quality in our experiments.

Furthermore, the importance of information quality assessment is also demonstrated by many researchers, since one cannot manage information quality without measuring it meaningfully (Stvilia et al. 2007). However, today's organisations are still confronting various difficulties in information quality assessment. For example, what are reliable and valid dimensions? How does one distinguish different types of data sources? And how does one measure the quality of different data sources? Such difficulties in assessing information quality often create barriers for information quality assessment and even produce conflicting and erroneous assessment results.

When addressing the difficulties in information quality assessment, research has proposed several frameworks to manage these problems. However most frameworks are based on particular cases and often lack a thorough validation by empirical applications. As such, their assessment methods usually remain at a conceptual or theoretical stage. Therefore effectively assessing information quality is still a challenge for research.

5.2 Information Quality Assessment Framework

In information manufacturing systems, information is processed, transformed, produced and stored through a system of software components and human interventions (refer to Chapter 1). Datasets are considered as raw material for information manufacturing. The results, usually referred as information products, are regarded as final products of the information manufacturing system.

From this perspective, we can pinpoint at least two elements eligible for information quality assessment: raw data and information products. The raw data is stored in databases. It is usually organised as dataset comprising of a hierarchical or relational structure with attributes, tuples and relations. Raw data is often not intended for direct use and possesses a large amount of data records. However raw data is typically highly structured and well organised. The assessment of raw data is often based on quality rules. These formal quality rules (e.g. Oliveira et al. 2005) enable an automatic assessment using software tools. The second objects in information quality assessment are information products, which are manufactured and used for intended purposes. Therefore, information products should be assessed during their usage.

This involves information consumers' expectations and thus should be assessed by information consumers.

In addition to the two measuring objects (raw data or information products) and in accordance with our view of information manufacturing systems, we are able to distinguish two further components in information quality assessment: assessment operators (software or user) and measuring criteria (rules or expectations). All the three key components can be generally categorized as who measures, which dimensions and what objects. Hence following the idea of who uses which dimensions to measure what, we detail the following assessment framework outlined in Figure 23.

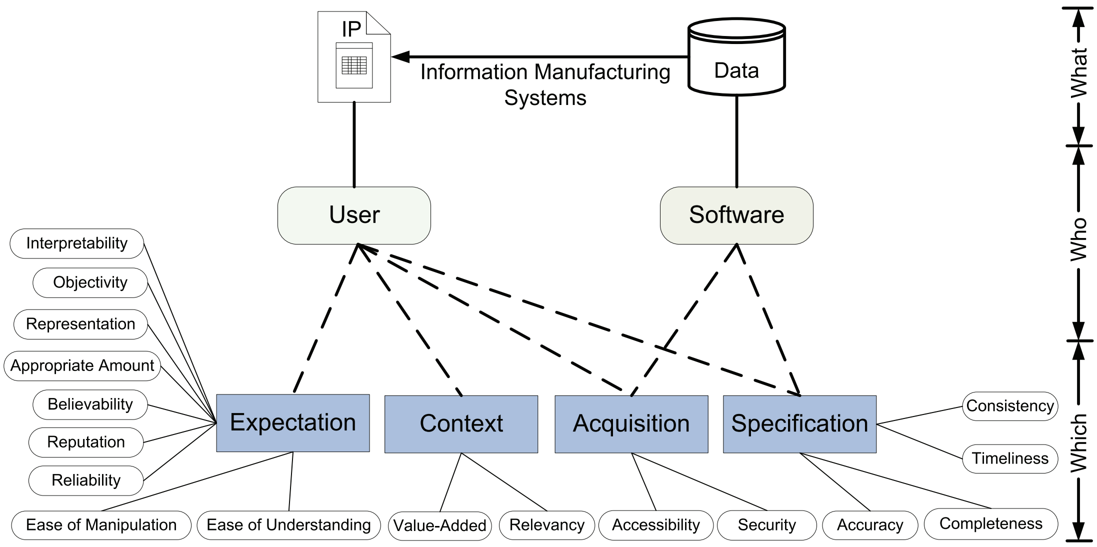


Figure 23: Information quality assessment framework

"Who" represents the evaluator of data or information that generates information quality assessment results. According to the principal categories of automatic and human-based task performer, the evaluator can be a person and/or a software program. Note that although users may design rules and operate the software program, we regard the software component as an independent assessment entity.

"What" corresponds to the objects that are assessed. We consider two measuring objects: the raw data stored in the databases and the information products manufactured by information systems.

“Which” represents the set of information quality dimensions that are used in the assessment. As discussed above, researchers such as Ballou and Pazer (1985) or Wang and Strong (1996) have derived different sets of information quality dimensions.

To organise the dimensions, most previous classifications often concentrate on the nature of the dimensions. Therefore, previous classifications are frequently unsuitable for information quality assessment, and thus we decided to present an alternative classification merely to facilitate the assessment. Our classification contains four categories: acquisition, context, specification and expectation. Based on the classification, we adopted 17 widely accepted information quality dimensions (Wang and Strong 1996, Wand and Wang 2002, Lee et al. 2002, Pipino et al. 2002, Kahn et al. 2002) and classified them in Table 22.

Acquisition	Context	Specification	Expectation
Accessibility Security	Relevancy Value-added	Accuracy Completeness Timeliness Consistency	Interpretability Objectivity Representation Believability Reliability Reputation Ease of Manipulation Ease of Understanding Appropriate Amount

Table 22: Our classification of information quality dimensions

(1) The acquisition category represents the level of information that a user can obtain. It measures the extent to which information is readily available and retrievable. Because information access is a prerequisite for any further information processing, we designate acquisition as a primary measure. *Acquisition information quality* reflects the characteristics of accessing and retrieving the information. Consequently, two

dimensions are identified in this category: accessibility and security. The acquisition dimension includes accessibility measures but also aspects of security and data protection.

(2) The context category characterises the intended use of information. It measures the extent to which information is relevant and useful. Obviously the user is the main subject involved in context-dependent information quality assessment. Based on the context-related evaluation, users evaluate the relevance of information in a particular context. We select the context dimension as second prerequisite for any further information quality evaluation because only information of high contextual value is useful for further evaluations. *Context information quality* indicates that the utilisations of dimensions are dependent on a specific context. Two dimensions are included in this category: relevancy and value-added.

(3) The specification category is constructed around dimensions which can be objectively determined. Examples of these specifications include rules and references used by the software. *Specification information quality* comprises the dimensions that can be objectively assessed by mathematical functions, four of which are found to be objectively measurable (Ballou and Pazer 1995, Lee et al. 2002). Thus, *specification information quality* contains four dimensions: accuracy, completeness, timeliness and consistency.

(4) The expectation category evaluates the user's perceived quality of information products by subjective information quality dimensions such as objectivity and believability. Obviously the expectation dimension is subjective and can only be evaluated by users. *Expectation information quality* can only be used by users in information quality assessment. Eight information quality dimensions are included in *expectation information quality*: interpretability, objectivity, representation, believability, reputation, ease of manipulation, ease of understanding and appropriate amount.

5.3 Validation of Information Quality Dimensions

In this section, we describe refining popular information quality frameworks and present at the end a list of validated information quality dimensions. From previous frameworks, we can obtain different sets of information quality dimensions. Therefore this section aims to validate a set of information quality dimensions for our research. Following Anastasi (1988) and Litwin (1995) the validation of the measurement instrument is divided into four types: criterion validity, face validity, content validity and construct validity.

- *Criterion validity* measures how well a proposed instrument correlates with a classic instrument or predictor (Anastasi 1988, Litwin 1995). Criterion validity comprises two elements of concurrent validity and predictive validity. While concurrent validity tests the correlation against a classic instrument, predictive validity tests the correlation against a predictor. As information quality research is still in its infancy, some information quality dimensions remain ambiguous for practical use. Therefore criterion validity is not fitting for this study.
- *Face validity* evaluates if the proposed items effectively measure the intended use (Anastasi 1988). Usually, this can be tested through reviewing the items with untrained judges (Litwin 1995). These untrained individuals are asked to confirm whether the measuring items are acceptable for the measurement instrument. In our study, 10 postgraduate research students reviewed the instrument for face validity. These respondents confirmed the measure. However, one respondent suggested that it was difficult for novice users to assign properties to the value-added criteria. Ultimately we decided, as only 2 out of 10 papers list this dimension in Table 23, to revise our measures and delete the dimension "value-added" from the list.
- *Content validity* is used to measure the extent to which the proposed items reflect the specific domain of content (Carmines & Zeller, 1991). Testing content validity requires structured reviews of the instrument's content by experienced professionals. These professionals should possess extensive domain knowledge in or-

der to properly evaluate the information quality dimensions. The reviewers evaluate whether the measurement instrument is complete and correct, in other words includes all necessary items and no unnecessary items. In our study, 10 information system researchers reviewed the instrument. The researchers confirmed the measurement instrument, however one respondent provided comments on the "appropriate amount" dimension. This researcher suggested that if we distinguished the concepts of information quality and information overload, the "appropriate amount" dimension could be combined with the "completeness" dimension. Discussing this point with the other researchers, we accepted their comments and consequently eliminated the dimension "appropriate amount".

- *Construct validity* consists of two elements: convergent validity and discriminant (also known as divergent) validity. To test construct validity, factor analysis can be employed to purify the measures (Litwin 1995). In the following, we mainly focus on testing the construct validity of the dimensions.

As discussed in Chapter 1, both Wang and Strong (1996) and Lee et al. (2002) have validated information quality dimensions. However their works are found to be conflicting in the aspect of discriminant validity. The different results in discriminant validity might be a result of the differing forms of the surveys. Wang and Strong (1996) assess the importance of information quality dimensions, while Lee et al. (2002) use information quality dimensions to evaluate the information. Furthermore, the characteristics of the underlying samples are different. The respondents in Lee et al. (2002) are from five organisations. The subjects used by Wang and Strong (1996) contain a more comprehensive demography.

Due to the inconsistencies between and differing interpretations of validity of selected information quality dimensions, we decide to carry out a further study and confirm valid information quality dimensions in the context of our framework. Initially we analyse the popularity and importance of the dimensions. The quantity of supporting literature is a crucial indicator for selecting dimensions. To provide the supporting literature, we selected 10 of the most influential studies in information quality research, including 8 journal papers and 2 books. For each dimension, we illustrate the supporting studies as well as the total number of supporting studies in Table 23.

	Ballou & Pazer 1985	DeLone & Mclean 1992	Goodhue 1995	Wang & Strong 1996	Wand & Wang 1996	Redman 1996	Pipino et al. 2002	Lee et al. 2002	Bovee et al. 2003	Eppler 2006	Support Literature
Accessibility		√	√	√		√	√	√	√	√	8
Security				√			√	√		√	4
Relevancy		√		√	√	√	√	√	√		7
Value-added				√			√				2
Accuracy	√	√	√	√	√	√	√	√	√	√	10
Completeness	√	√	√	√	√	√	√	√	√	√	10
Timeliness	√	√	√	√	√	√	√	√	√	√	10
Consistency	√	√		√	√	√	√	√	√	√	9
Interpretability				√	√	√	√	√	√	√	7
Objectivity		√		√	√		√	√			5
Representation		√	√	√	√	√	√	√			7
Believability				√			√	√			3
Reliability		√	√		√						3
Reputation				√			√	√			3
Ease of Manipulation			√				√	√		√	4
Ease of Understanding		√	√	√	√		√	√	√	√	8
Appropriate Amount				√		√	√	√			4

Table 23: Supporting literature of information quality dimensions

Using the frequency that is equal to or greater than half of the total selected literatures, we can obtain a set of significant information quality dimensions: accuracy, completeness, timeliness, consistency, accessibility, ease of understanding, relevancy, interpretability, representation and objectivity. This result generally conforms to the research findings of Wand and Wang (1996).

Since the dimension “value-added” and “appropriate amount” are eliminated when we test the face and content validity, we created a survey to further validate the 15 remaining dimensions. In accordance with the procedure of a critical measurement validating approach (Churchill 1979), we generated 2 to 5 measuring items for each of the 15 selected dimensions. Based on the dimensions, a total of 50 items is thus created to indicate the selected dimensions. These dimensions and the attributes of measuring items are shown in Table 24.

IQ Dimension (15)	Attributes of Items (50)
Accessibility	accessible, obtainable, retrievable, available. (4 items)
Security	secure, protected, authorized access. (3 items)
Relevancy	useful, relevant, applicable, helpful. (4 items)
Accuracy	correct, accurate, free of error, precise. (4 items)
Completeness	sufficient, complete, comprehensive, include all necessary values, detailed. (5 items)
Timeliness	current, up to date, delivered on time, timely. (4 items)
Consistency	consistent meaning, consistent structure, presented in the same format. (3 items)
Interpretability	interpretable, without inappropriate language and symbol, readable. (3 items)
Objectivity	impartial, unbiased, objective, based on facts. (4 items)
Representation	concise, compact. (2 items)
Reliability	reliable, dependable. (2 items)
Believability	believable, trustworthy, credible. (3 items)
Reputation	from good sources, of good reputation, well referenced. (3 items)
Ease of Manipulation	easy to manipulate, easy to aggregate, easy to combine. (3 items)
Ease of Understanding	easy to understand, easy to comprehend, easy to identify the key point. (3 items)

Table 24: Information quality dimensions and their measuring items

Previous Research provides us with two principal approaches to designing surveys for validating information quality dimensions: (1) assessing the importance of the dimensions (e.g. Wang and Strong 1996, McKinney et al. 2002), and (2) evaluating the given information by information quality dimensions (e.g. Lee et al. 2002, Slone

2006). Using the second approach, users must be able to evaluate the given information by all the dimensions. However some users may consider certain dimensions as not suitable for evaluating the given information. Therefore we decided that the form of assessing the importance of information quality dimensions should be utilised to design the survey. In terms of measurement, based on the observations of McKinney et al. (2002) and Lee et al. (2002), we use an 11-point Likert type scale. The number 10 is labeled as "Extremely important", while 0 is labeled "Not important at all", and 5 is labeled "Average". Most items in the survey are formulated as "the information that is <Attributes of the Item> is". For example, "the information that is accessible is". These items are listed at random in the appendix.

Data were collected by means of a web-based system and a paper-based questionnaire. The web-based system was hosted on our university server. We distributed the web address of this survey to students, researchers and organisational staff. The paper-based questionnaires were used for the postgraduate students and researchers in Dublin City University. The participants of the questionnaire were invited to complete the survey on site. The paper-based questionnaire was used to increase the response rate of the survey. The survey was carried out amongst both industry members and academia. 316 viable responses were collected from 580 participants; 52% were postgraduate students, 17% were information system researchers and a further 31% were organisational staff. The average age of the participants was 28 years of age. Based on the collected data, we carried out a confirmatory factor analysis and a Cronbach alpha calculation, both of which can indicate construct validity. The results of the factor analysis are shown in Table 25.

Constructs (number of items)	Items	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7	Factor 8	Factor 9
Accessibility (4)	Accessible		**0.726**							
	Obtainable		**0.601**							
	Retrievable		**0.851**							
	Available		**0.792**							

Security (3)	Secure							**0.818**		
	Protected							**0.787**		
	Authorized							**0.771**		
Relevancy (4)	Useful	**0.723**								
	Relevant	**0.686**								
	Applicable	**0.843**								
	Helpful	**0.848**								
Ease of Understanding Interpretability (5)	Easy to understand					**0.855**				
	Easy to comprehend					**0.758**				
	Easy to identify points					**0.805**				
	Interpretable					**0.693**				
	Readable					**0.727**				
Ease of Manipulation (3)	Easy to manipulate			**0.605**						
	Easy to aggregate			**0.842**						
	Easy to combine			**0.876**						
Objectivity (4)	Impartial								**0.782**	
	Unbiased								**0.693**	
	Objective								**0.793**	
	Based on facts								**0.868**	
Reliability Believability Reputation (6)	Reliable				**0.696**					
	Believable				**0.687**					
	Trustworthy				**0.788**					
	Credible				**0.628**					
	From good sources				**0.787**					
	Good reputation				**0.771**					
Accuracy Completeness Consistency (9)	Correct						**0.835**			
	Accurate						**0.897**			
	Free of error						**0.840**			
	Precise						**0.770**			
	Sufficient						**0.764**			
	Complete						**0.863**			
	Comprehensive						**0.622**			
	Consistent meaning						**0.707**			
	Consistent structure						**0.888**			
Timeliness (3)	Current									**0.870**
	Up to date									**0.898**
	Timely									**0.795**

Table 25: Results of factor analysis

Our confirmatory factor analysis shows that 9 information quality dimensions (loading > 0.6) have been found. The details are discussed as follows.

1. The dimension interpretability exhibited a cross loading with the ease of understanding dimension. We conflated both dimensions into the dimension understandability.
2. The dimensions reliability, believability and reputation showed a cross loading between each other. It seemed that users considered the information from credible sources or information with a good reputation as reliable and believable. In order to combine the dimensions, we decided to select one frequently used dimension representing all the three dimensions. According to Wand and Wang (1996), reliability is found to be the most frequently used in literatures. Therefore, we used reliability to represent all the above three dimensions.
3. The dimensions accuracy, completeness and consistency were grouped together. Users possibly regarded that inaccurate information consists of incomplete or inconsistent information. However, due to the common usage of these dimensions in objective information quality assessment, we decided to keep the dimension names and grouped the 3 dimensions into one category. Following Bovee et al. (2003), we titled this category “integrity”. When implementing objective information quality assessment, we still use accuracy, completeness and consistency rather than the term integrity. Integrity is only used when we carry out subjective information quality assessment.

Considering low loading factors, we dropped 9 measuring items and the dimension representation. The remaining items were showing high factor loading and no significant cross loading. The outcome of the analysis indicated and exhibited both the convergent validity and discriminant validity.

After the purification, we obtained 9 information quality dimensions with 41 measuring items. In order to test the construct reliability, Cronbach alphas were computed. The Cronbach alpha values evaluate how successfully the dimensions captured the variance of the measuring items. Table 26 lists the Cronbach alpha values for our information quality categories and constructs.

Category	Construct	Number of Items	Cronbach Alpha
Acquisition	Accessibility	4	0.91
	Security	3	0.88
Context	Relevancy	4	0.81
Specification	Integrity	10	0.75
	Timeliness	3	0.92
Expectation	Understandability	5	0.86
	Reliability	6	0.81
	Ease of Manipulation	3	0.83
	Objectivity	4	0.77

Table 26: Results of Cronbach alpha calculation

As shown in Table 26, the Cronbach alpha values of the above information quality dimensions ranged from 0.75 to 0.92. According to the acceptable rate of 0.7 (Nunnally, 1967), the analysing results also supported convergent validity. It indicated measures of each dimension with high reliability.

In this section, we have confirmed 9 valid and reliable dimensions for our assessment. Further, we structured the dimensions into the categories we proposed in Section 5.2. Acquisition category contains accessibility and security; context category contains relevancy; specification category contains integrity and timeliness; expectation category contains understandability, reliability, ease of manipulation and objectivity. This categorisation is described in Table 26.

5.4 Assessment Methodology

In the previous section, we developed a comprehensive and validated list of information quality dimensions. Using our dimensions, in this section we develop methodologies for information quality assessment. Since information manufacturing can be carried out in a single database or multiple databases, we propose different methodologies for single-database assessment and multi-database assessment.

5.4.1 Assess Information Quality in a Single Database

In the single-database assessment, our approach differs between the assessment of raw data and information products (Figure 24). Since raw data is always assessed by an automatic procedure, context and expectation information quality dimensions are not relevant for assessing the raw data. Consequently, in order to assess the quality of raw data, we only focus on the categories of acquisition and specification. The acquisition dimensions are first used to measure the degree of accessing raw data; subsequently we can then measure the specification dimensions.

In contrast to the assessment of raw data, assessing the quality of information products is primarily a human-based procedure. An ideal procedure for assessing information products can follow subsequent steps. An information product is first assessed by acquisition dimensions. Then it is assessed by context dimensions. If the information products are completely inaccessible or accessible but irrelevant, we can conclude the assessment at this step. Otherwise specification and expectation dimensions are employed in assessing the quality of the information products. We summarise the above procedures of assessing the quality of raw data and information products in Figure 24.

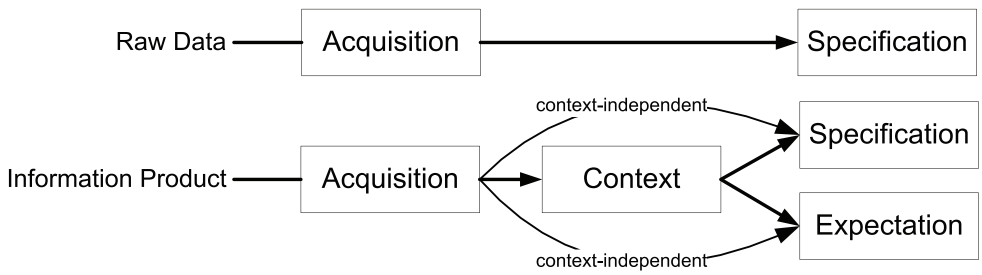


Figure 24: Information quality assessment methodology

In order to further refine the assessment, we developed an approach for assessing raw data as well as information products. First we discuss the assessment approach for raw data and subsequently the assessment approach for information products. The assessment approach for raw data is organised into following 5 steps.

(a) Creating specifications. The standard values need to be specified for each field in the database.

(b) Identifying information quality problems. Based on the specifications, we recognise the problems which violated the specifications.

(c) Relating information quality problems to the dimensions. If one information quality problem is connected to multiple dimensions, it may cause dependencies amongst these dimensions. In order to clarify the distinct function of each dimension, we link one single problem to only one single dimension; however one dimension can be connected to multiple information quality problems. Any dimension which is not linked to any information quality problems is dropped from the assessment.

(d) Assessing the quality of raw data. This is an automatic procedure performed by software.

(e) Generating information quality report. This report provides the basis for the information quality analysis and improvement.

The above approach is summarised in Figure 25.

Figure 25: Approach for assessing raw data

In order to facilitate the assessment, the assessing procedure is always operated by a piece of software. The software reads one data record from the database and compares it to the specification. If the value conforms to the specification, the software will check the next data record. Otherwise the software maps the inconformity to information quality problems, and then links to information quality dimensions. The value of the dimension is calculated by dividing the number of information quality problems by the total number of data records. This assessing procedure can be used to determine the value of *completeness* (D_k) in equation 4-5 and *consistency* (D_k) in equation 4-9. After completing the entire assessment, the software can generate an assessment report. The whole procedure is illustrated in Figure 26.

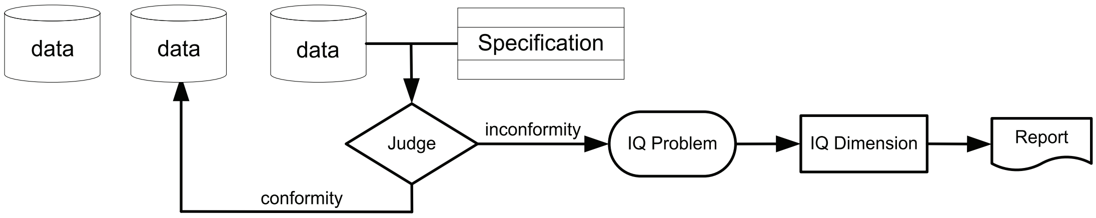


Figure 26: Procedure of assessing raw data

In order to assess the quality of information products, we also developed a 5-step approach. The 5 steps are described as follows and illustrated in Figure 27.

(a) Understanding the intended use. As most of the information quality assessments are context related, users firstly have to understand the purpose of information utility.

(b) Listing the information product inventory, according to the context, information products need to be identified.

(c) Understanding information quality dimensions. Before the evaluation, users need to comprehend the definitions and subscales of each dimension.

(d) Assessing the quality of the information product. When completing the above three steps, users employ information quality dimensions to evaluate the listed information products.

(e) Finally each user provides an information quality assessment report.

Figure 27: Approach for assessing information products

According to users' expectations, users evaluate the extent to which information products are fit for the intended use. Since subjective standards and expectations vary from person to person, each user will generate an individual report. This procedure is illustrated in Figure 28.

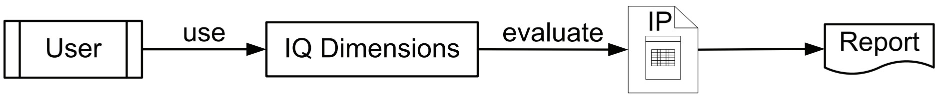


Figure 28: Procedure of assessing information products

In order to recapture our assessment approach for single databases, we position the measurement approach in the context of information quality management. Therefore we map the assessment approaches and procedures to the TDQM cycle (Wang 1998). The first three strategic steps of assessing raw data and information products are in the defining phase. The other two steps are in the measuring phase. The components of the above assessment procedure are the detailed outcomes of strategic steps. Figure 29 outlines the relationships between TDQM cycle, the approach for assessing the quality of raw data and components of the assessment procedure.

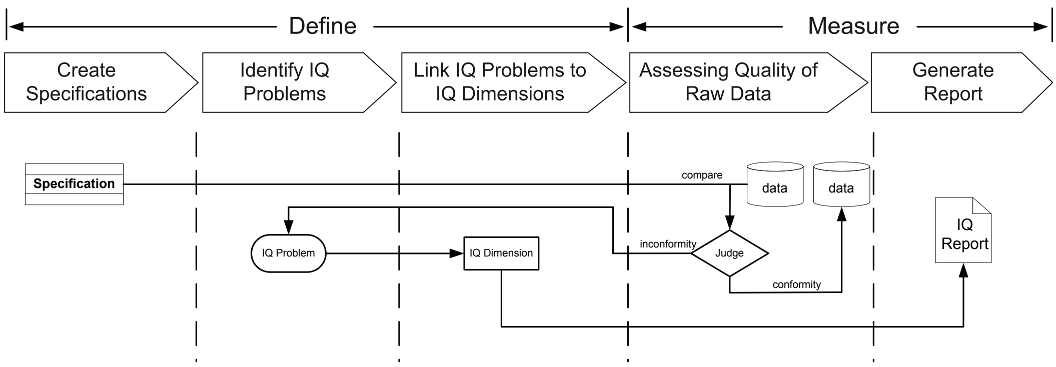


Figure 29: Relationship between TDQM and assessing raw data

As the approach for assessing the quality of raw data, we describe relationships between TDQM cycle, the approach for assessing the quality of information products and the components of the assessing procedure in Figure 30.

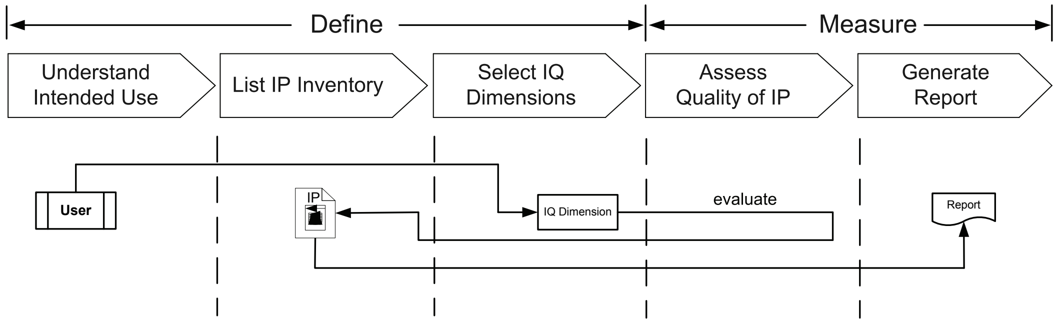


Figure 30: Relationship between TDQM and assessing information products

5.4.2 Assess Information Quality in Multiple Databases

In the previous section, we presented and discussed an assessment approach for single-database systems. However, in certain information systems, data is usually shared through an information chain. We consider information chain as an instance of multi-database system. In this thesis, we define information chain as a system that distributes the shared information in sequence amongst the entities of a chain. The shared information is then respectively used in each entity for business operations, such as decision-making and transactions. At regular intervals, information alignments are carried out to refresh the shared information among the entities. However during these intervals, data operations may be undertaken in an entity. These data operations may generate information quality problems, which in turn affect business performance. Therefore assessment and improvement of information quality are crucial in the information chain.

Our approach for assessing information quality in multiple databases builds on three earlier contributions. In Chapter 2 these contributions were included in our literature review. However due to the significance of the works, we decided to summarise these contributions in details. Considering these contributions, we then present our assessment approach.

First, Ballou and Pazer (1985) propose a model to assess information quality in multi-input and multi-output information systems. This pioneering work posits that output information quality is affected by input information quality and information processing procedure. According to a "garbage in garbage out" principle, input information quality is directly related to output information quality. Moreover, information processing can augment input information quality, diminish input information quality or leave it unchanged. Hence their model is represented by Output Information Quality = Input Information Quality + Information Processing. This model is also confirmed by Ballou et al. (1998) in the environment of information manufacturing systems.

Second, Wand and Wang (1996) use ontology to assess information quality. Their method focuses on the discrepancies between information systems and real-world systems. They proposed that information quality is determined by how well the real-world system is presented by information systems. For example, problems of incomplete information occurred when the elements of a real-world system should be presented in the information system but were not. Thus information quality is the result of the relationship between real-world systems and information systems. It can be presented by information quality = f (real-world system, information system). In practical applications, the real-world system can be considered as a reference system.

Third, Cappiello et al. (2003) propose a model to determine information quality in multi-channel information systems. The typical structure of multi-channel information systems comprises one reference database and one operational database. The reference database is used to refresh the operational database for information alignment. In the database chain of multi-channel information systems, one database may be simultaneously used as an operational database and considered as a reference database by another downstream operational database. During the information alignment period, the reference database may execute certain operations such as deleting or modifying information. These operations could result in low information quality in the operational database as the corresponding information in its reference database has been changed. Based on accepted information quality dimensions, Cappiello et al. (2003) focused on data currency, accuracy and completeness and investigated how data operations within the reference database affect the operational database. However in many applications, data operations usually take place in both the reference database and the operational database. Recognising this limitation, our work attempts to assess information quality in operational databases when both reference database and operational database are operating information in the refresh period. This work can be used when our experiment is extended to a multi-user context.

By reviewing the three significant works above, we summarise their research foci and our research objective by Figure 31.

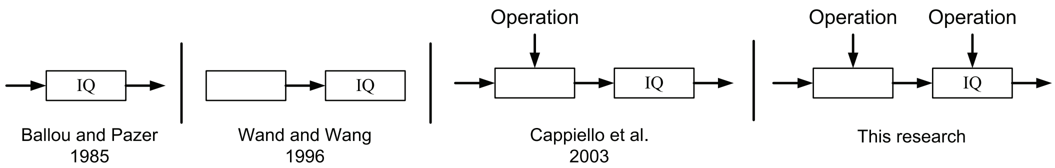


Figure 31: Summary of three related works

As indicated in Figure 31, we can observe that Ballou and Pazer (1985) focused on information quality in information flow among different data processing blocks. Wand and Wang (1996) concentrated on information quality as determined by the relationships between reference and operational databases. Cappiello et al. (2003) centred their research on the effects of operating a reference database on an operational database. Extending these three studies, our work investigates operational database information quality by considering data operations on both the reference and operational databases.

5.4.2.1 Assessment between Databases

In order to derive the algorithms of information quality dimensions, we need to further refine the measurement of the dimensions. As we discussed in Figure 7 of Chapter 2, we can map information quality dimensions onto information quality problems. To identify the related information quality problems, we need to determine which data operation generates which kind of information quality problems. This relationship is expressed by Figure 32.

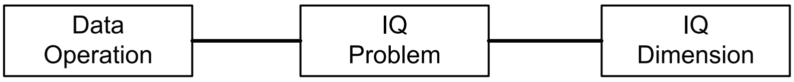


Figure 32: Relationship between data operation, information quality problems and information quality dimensions

In order to describe the data operations in a formal way, we first discuss the databases in an information chain. In the information chain, different databases are sharing information and the shared information is periodically refreshed for data alignment. Consider DB_{m-1} and DB_m in Figure 33 as adjacent databases in an information chain. As the information flow indicated in this figure, DB_{m-1} and DB_m are respectively considered as reference database and operational database.

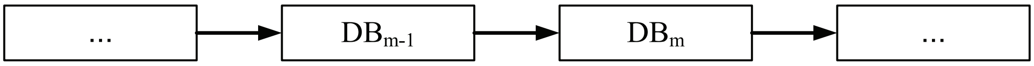


Figure 33: Typical information chain

As the reference database, DB_{m-1} periodically refreshes the shared information in DB_m. During each refresh period, data operations may take place in both DB_{m-1} and DB_m. The typical data operations are “create”, “update” and “delete”. In the following, we analyse the three data operations during the refresh period.

The operation “create” means new data are introduced to the database during the refresh period. If the data are only created in the operational database, there is no corresponding data in the reference database. Therefore from the perspective of reference database, data are created redundantly in the operational database. We term this operation “Redundantly Create”. If the same data are created in both operational database and reference database or the corresponding data exist in the reference database when we create new data in operational database, we term this operation “Completely Create”. If the data are only created in the reference database, we term this operation “Incompletely Create”. In the above data operations, “Redundantly Create” generates redundant data problems and “Incompletely Create” generates missing data problems. Although "Completely Create" does not generate any data problems, it does affect the total number of the data, which in turn affects the value of information quality dimensions. The three forms of creating data in reference database and operational database are described by Figure 34.

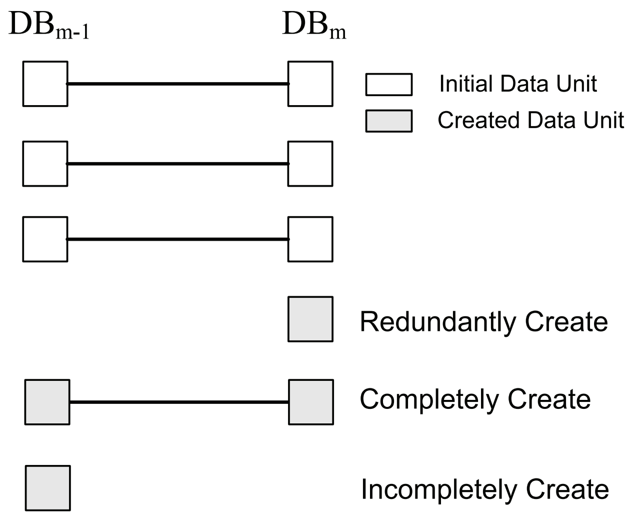


Figure 34: Data operation "create" in reference database and operational database

The data operation "update" means the existing data in databases are modified during the refresh period. After modifying the data in the reference database, the corresponding data in the operational database are no longer consistent with the modified data in the reference database. Therefore this operation makes the data in operational database incorrect. We term this operation "Inconsistently Update". This operation consists of three situations: (1) data are only updated in operational database; (2) data are only updated in reference database; (3) the same data, in both the operational and reference databases, are updated to different values. The three situations above all result in the generation of incorrect data in the operational database.

When we modify the data in both the reference database and the operational database to the same value, no data relationship is affected. Therefore we term this operation "Completely Update". This operation does not affect any information quality dimension. The above two forms of updating data in reference databases and operational databases are described by Figure 35.

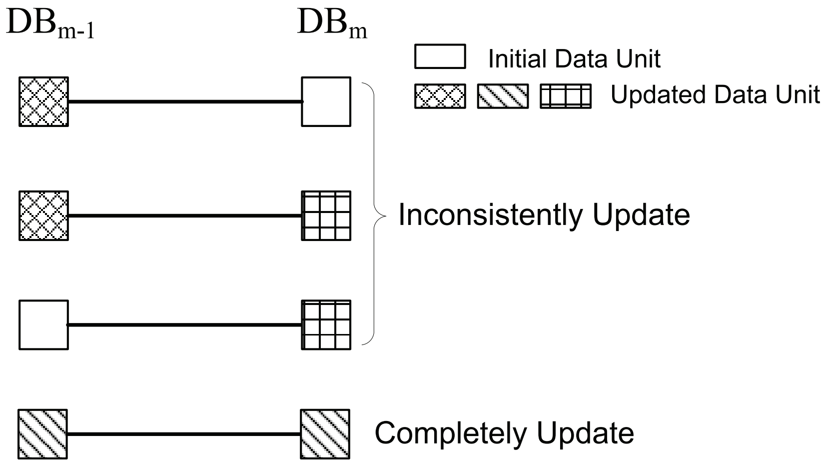


Figure 35: Data operation “update” in reference database and operational database

The data operation “delete” means the existing data in databases are deleted during the refresh period. If we delete the data only in the reference database, as a result no reference data will exist for the data in the operational database. This means the data in the operational database exist redundantly. We term this operation “Redundantly Delete”. If we delete the corresponding data in both the reference database and the operational database, we term this operation “Completely Delete”. This operation does not generate any data problems but it does influence information quality dimensions. When we delete the data only in the operational database, this results in missing data problems and we term the operation “Incompletely Delete”. We describe these three forms of deleting data in Figure 36.

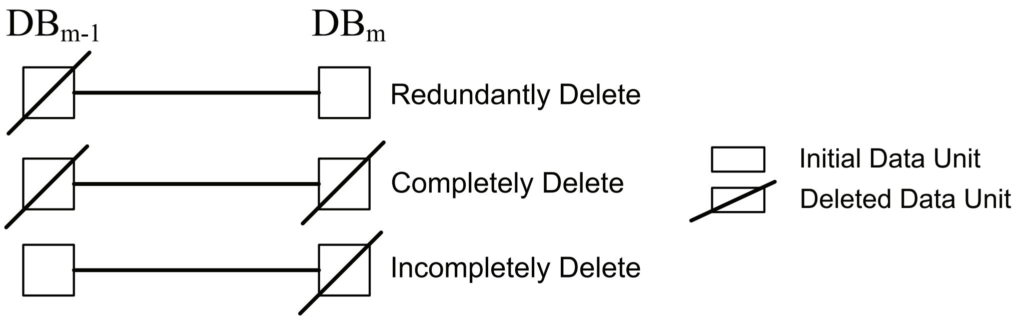


Figure 36: Data operation "delete" in reference database and operational database

From the discussion above, we can observe that data operations mainly cause three information quality problems: missing data, redundant data and incorrect data. Missing data exist when we incompletely create data in the reference database or delete data in the operational database. Redundant data are the results of data which are redundantly created in operational database or deleted in reference database. Incorrect data are generated by redundantly or incompletely updating data. Except in the above data operations, completely creating and deleting data does not generate any information quality problems; however, this still affects the value of information quality dimensions because it influences the number of total data units. Therefore the only data operation which does not affect information quality is updating the data completely.

According to the definition of accuracy in Chapter 4, the incorrect data, which inaccurately represent the actual value, are obviously mapped to accuracy. Redundant data are usually inaccurately used since their reference data cannot be found. This means that from the user's perspective, these redundant data are incorrect. We therefore map redundant data onto accuracy.

According to the definition of completeness in Chapter 4, data which, by their absence, fail to represent the data in the reference database are mapped to completeness. We summarise the mappings between data operations, information quality problems and information quality dimensions in Figure 37.

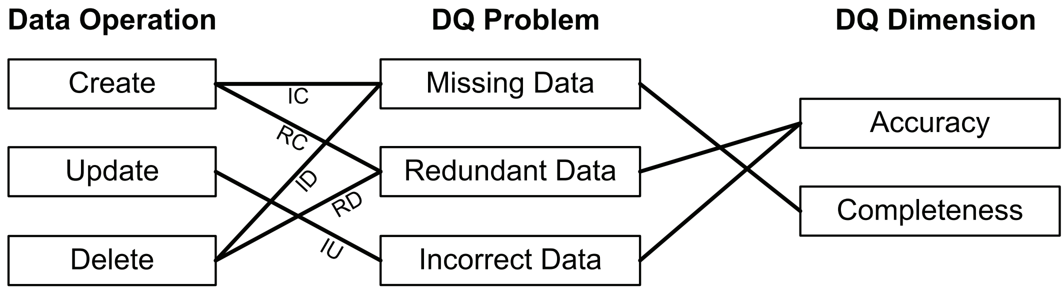


Figure 37: Mappings between data operations, information quality problems and information quality dimensions

Two assumptions are associated with the above mappings:

(1) Users can access all the data in the database. This assumption states that we are able to access the data before we determine information quality dimensions. It is also the premise for data operations.

(2) Data in operational databases become aligned after data refresh. In this study, it is hypothesised that data in reference databases are considered to be a specification.

Based on these two assumptions, we can derive the equations of accuracy and completeness. Firstly, we provide the notations used in this chapter (Table 27).

Variable	Symbol	Description
Database	DB_m	The m[th] database in the information chain
Create	$C_m(t)$	This function is used to calculate the number of data created in DB_m at time t of one refresh period.
Redundantly Create	$RC_m(t)$	This function is used to calculate the number of data redundantly created in DB_m at time t of one refresh period.
Completely Create	$CC_m(t)$	This function is used to calculate the number of data completely created in DB_m at time t of one refresh period.
Incompletely Create	$IC_m(t)$	This function is used to calculate the number of data incompletely created in DB_m at time t of one refresh period.
Inconsistently Update	$IU_m(t)$	This function is used to calculate the number of data inconsistently updated in DB_m at time t of one refresh period.
Delete	$D_m(t)$	This function is used to calculate the number of data deleted in DB_m at time t of one refresh period.
Redundantly Delete	$RD_m(t)$	This function is used to calculate the number of data redundantly deleted in DB_m at time t of one refresh period.
Completely Delete	$CD_m(t)$	This function is used to calculate the number of data completely deleted in DB_m at time t of one refresh period.
Incompletely Delete	$ID_m(t)$	This function is used to calculate the number of data incompletely deleted in DB_m at time t of one refresh period.

Table 27: Notations used in chapter 5

Let's still consider DB_{m-1} is the reference database and DB_m is the operational database. Data are shared between DB_{m-1} and DB_m. The shared dataset $DB_{m-1,m}$ can be expressed by.

$$DB_{m-1,m} = DB_{m-1} \cap DB_m \neq \phi \tag{5-1}$$

In the information chain, the shared data in DB_m are aligned with DB_{m-1} with every refresh period. During the refresh period, we determine accuracy and completeness in operational database DB_m.

5.4.2.2 Measurement of Accuracy

In the database community of information quality research, accuracy is calculated by the ratio of the number of accurate data divided by the quantity of total data. That is:

$$\text{Accuracy} = \frac{\text{Number of Accurate Data}}{\text{Number of Total Data}} \tag{5-2}$$

The number of accurate data can also be expressed by subtracting the number of inaccurate data from the number of total data. Therefore equation 5-2 can be re-formed as:

$$\text{Accuracy} = \frac{\text{Number of Total Data - Number of Inaccurate Data}}{\text{Number of Total Data}} \tag{5-3}$$

In one refresh period, the number of total data is determined by the number of initial data, deleted data and created data in DB_m. Specifically, at time t the number of total data is

$$N_t = |DB_{m-1,m}| - D_m(t) + C_m(t) \tag{5-4}$$

Where N_t is the number of total data in DB_m. $|DB_{m-1,m}|$ is the cardinality of $DB_{m-1,m}$, which is the initial number of shared data in $DB_{m-1} \cap DB_m$. $D_m(t)$ is determined by $CD_m(t)$ and $ID_m(t)$ which both delete data in DB_m. $C_m(t)$ is determined by $RC_m(t)$ and $CC_m(t)$ which represent the data created in DB_m. Therefore,

$$D_m(t) = CD_m(t) + ID_m(t) \tag{5-5}$$

$$C_m(t) = RC_m(t) + CC_m(t) \tag{5-6}$$

$$N_t = |DB_{m-1,m}| - CD_m(t) - ID_m(t) + RC_m(t) + CC_m(t) \tag{5-7}$$

According to the mapping between accuracy, information quality problems and data operation, inaccuracy is associated with redundant data and incorrect data, and in turn mapped to $RC_m(t)$, $RD_m(t)$, and $IU_m(t)$, which all generate redundant or incorrect data. Therefore,

$$N_{ia} = RC_m(t) + RD_m(t) + IU_m(t) \quad (5\text{-}8)$$

where N_{ia} is the number of inaccurate data in DB_m. According to equation 5-3, we can obtain:

$$\text{Accuracy} = \frac{N_t - N_i}{N_t} \quad (5\text{-}9)$$

$$\text{Accuracy} = \frac{|DB_{m\text{-}1,m}| - CD_m(t) - ID_m(t) + CC_m(t) - RD_m(t) - IU_m(t)}{|DB_{m\text{-}1,m}| - CD_m(t) - ID_m(t) + RC_m(t) + CC_m(t)} \quad (5\text{-}10)$$

In data operations, the combination of redundant and incomplete operations can appear to be the complete operation. For example, while we redundantly delete one data unit in DB_{m-1}, we also incompletely delete its corresponding data unit in DB_m. As a consequence of the two operations, data are completely deleted and it is computed in $CD_m(t)$.

5.4.2.3 Measurement of Completeness

Completeness is calculated by the ratio of the number of data that represent the reference data, divided by the number of total reference data. That is:

$$\text{Completeness} = \frac{\text{Number of Data That Represent Referencing Data}}{\text{Number of Total Referencing Data}} \quad (5\text{-}11)$$

The data that represent the reference data can also be expressed by subtracting the number of reference data, which are not represented, from the number of total reference data. This means subtracting the data (missing data) that fail to represent their reference data. Therefore equation 5-10 can be re-formed into:

$$\text{Completeness} = \frac{\text{Number of Total Referencing Data - Number of Missing Data}}{\text{Number of Total Referencing Data}} \quad (5\text{-}12)$$

In our case, the total number of reference data is the number of shared data in DB_{m-1} at time t., which is determined by the quantity of initial data, deleted data and created data in DB_{m-1}.

$$N_r = |DB_{m-1}| - D_{m-1}(t) + C_{m-1}(t) \quad (5\text{-}13)$$

$$D_{m-1}(t) = RD_m(t) + CD_m(t) \quad (5\text{-}14)$$

$$C_{m-1}(t) = CC_m(t) + IC_m(t) \quad (5\text{-}15)$$

$$N_r = |DB_{m-1}| - RD_m(t) - CD_m(t) + CC_m(t) + IC_m(t) \quad (5\text{-}16)$$

Where N_r is the number of total reference data in DB_{m-1}. $|DB_{m-1}|$ is the cardinality of DB_{m-1}, which is the initial number of shared data in DB_{m-1}. According to the mapping between completeness, information quality problems and data operations, we can observe that incompleteness is only related to $IC_m(t)$ and $ID_m(t)$. That indicates that although the data in DB_m may not correctly represent the corresponding data in DB_{m-1}, they do not influence completeness because they are still representing reference data, merely in the incorrect way. Therefore

$$N_{ic} = IC_m(t) + ID_m(t) \quad (5\text{-}17)$$

Where N_{ic} is the number of data in DB_{m-1} without presentations in DB_m. According to equation 5-12, we can obtain:

$$\text{Completeness} = \frac{N_r - N_{ic}}{N_r} \quad (5\text{-}18)$$

$$\text{Completeness} = \frac{|DB_{m-1}| - RD_m(t) - CD_m(t) + CC_m(t) - ID_m(t)}{|DB_{m-1}| - RD_m(t) - CD_m(t) + CC_m(t) + IC_m(t)} \quad (5\text{-}19)$$

In the assessment of accuracy and completeness, functions of data operations are not determined by those operations executed before time t. It is at time t that we assess the data alternations and link them to data operations, whereby the output of data operation functions can be calculated.

In order to validate the assessment methodologies above, in the next sections we shall both develop a real-world application to validate the assessment in a single database and a simulation to validate the assessment in multiple databases

5.5 Validation of the Assessment in a Single Database

In order to validate the assessment methodology in a single database, we used both software-oriented and user-oriented approaches to assess information quality. We selected a public available dataset, the Sam’s Club database (ua_samsclub), offered by Walton College. As the data in this database are collected from the real world, it is reliable to validate our assessment approaches based on such a practical database. This database consists of 6 tables and 57 attributes. It contains retail sales information gathered from sales at Sam's Club stores, which is a division of Wal-Mart Stores Inc. In the example dataset, we assume that both raw data and information products are accessible. Thus acquisition information quality is not considered in the current study. According to our validated dimensions in Table 26, the raw data is measured by the selected 4 information quality dimensions: accuracy, completeness, consistency and timeliness. The information product is evaluated by the selected 7 information quality dimensions: relevancy, integrity, timeliness, understandability, reliability, objectivity and ease of manipulation.

5.5.1 Software-oriented Assessment

As discussed in Section 5.4.1, the first step in assessing the quality of raw data is the creation of specifications. However most attributes in Sam’s Club database are without specifications. Only some attributes provided specifications centralised on the table member_index and store_visits. Therefore for the illustration of the use of our framework, we selected the table member_index as the measuring object for software-oriented information quality assessment. Using our approach for assessing the quality of raw data, we summarised the procedure of assessing this dataset in Table 28.

	Field	Specification
Make Specifications	BUS_CR_TYP_STAT_CD	1-5,7,9
	CMPLMNTRY_CARD_CNT	0,1,2
	ELITE_STAT_CODE	0,2,3,4
	MEMBER_STATUS_CD	A,D,E,T
	MEMBER_TYPE	1,A,E,G,V,W,X
	QUALIFY_ORG_CODE	null, 0015-3001

Identify information quality Problems	**P1**. The data value is null except in the field of QUALIFY_ORG_CODE.
	P2. Figures are expressed by English such as describing 0 as zero.
	P3. Spelling errors such as case sensitivity.
	P4. Except the above situations, data value does not conform to the specification.

Link information quality Problems to information quality Dimensions		Accuracy	Completeness	Consistency	Timeliness
	P1		√		
	P2			√	
	P3	√			
	P4	√			

Assess Quality of Raw Data	Automatic Procedure

Generate Report	Database: UA_SAMSCLUB, Table: MEMBER_INDEX, Records: **5668375**			
		Accuracy	Completeness	Consistency
	BUS_CR_TYP_STAT_CD	99.983% (933)	99.999% (17)	100%
	CMPLMNTRY_CARD_CNT	99.999% (19)	100%	100%
	ELITE_STAT_CODE	99.806% (10954)	100%	100%
	MEMBER_STATUS_CD	100%	100%	100%
	MEMBER_TYPE	99.833% (9418)	100%	100%
	QUALIFY_ORG_CODE	98.263% (98424)	100%	100%

Table 28: Assessing the quality of raw data

Based on the given specifications, 4 information quality problems were identified and linked to different information quality dimensions. Timeliness was eliminated because no information quality problem was connected to this dimension. After the automatic assessment procedure, a simplified information quality report was generated. This report stated that 5668375 records were assessed and 119765 records contained information quality problems. As Sam's Club store was a membership-based store, information about its members was crucial for their business. This result shows that information quality deficiencies did exist in the database.

5.5.2 User-oriented Assessment

To assess the quality of information products, we developed an online survey (Figure 38) to facilitate the evaluating procedure. This survey consists of three major functions: The first function is to provide introductory information. This information includes a description of the scenario, a procedure for identifying information products and an introduction to information quality dimensions. The second function is the provision of an evaluation tool. Users can evaluate the quality of information products by adjusting slider bars for each information quality dimension. The slide bar is scaled from 0 % to 100 %. 0 % represents "not at all" and 100 % represents "completely". If the information quality dimension is not applicable to the current evaluation, users can label this dimension as "N/A". The third function is to collect demographic information and evaluation results. Based on the collected information, the online survey generated an assessment report for each user.

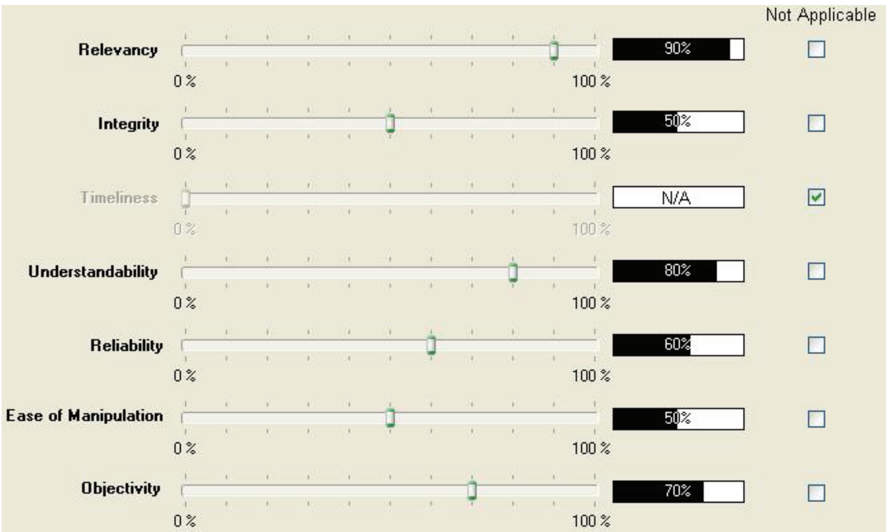


Figure 38: Online survey of assessing information products

A pilot study was conducted. We distributed the information to 10 researchers who are directors or board members in Teradata University Network. Seven researchers responded. The respondents provided 3 major comments. Firstly, most respondents pointed out that the database needed to be further specified because the database system contained different databases. Secondly, one respondent stated that concrete types of information products should be illustrated, such as the customer information or teaching information. Thirdly, two respondents suggested that system quality and information quality should have been differentiated, since users may have considered whether the system worked well instead of simply considering information quality.

Subsequently, we improved the online survey according to these comments, and carried out the evaluation with 30 postgraduate students. These students were registered users of Teradata system and all had used the system for different purposes such as data integration and data analysis. In the online survey, we provided a customer service scenario to participants. Associated with this scenario, we listed an inventory of information products manufactured from ua_samsclub dataset. After users understood information quality dimensions and the intended use of these information products, they were asked to evaluate information products using the given information quality dimensions.

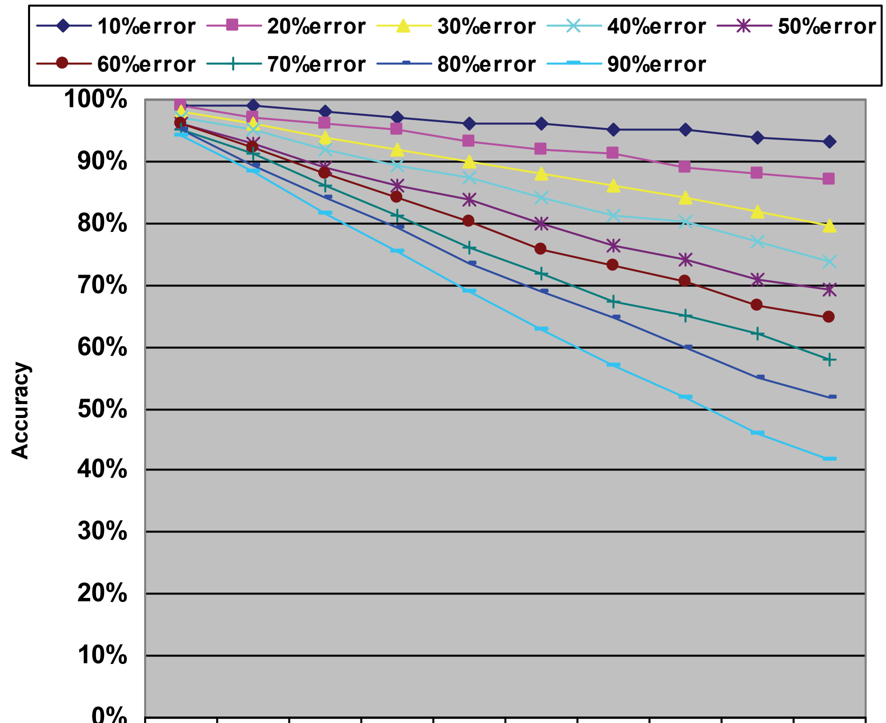
Evaluation of the results shows that integrity is rated as the lowest mean value (54%), with 90% of users assigning integrity their lowest value. We observe that users are generally not satisfied with the accuracy, completeness and consistency of information products. 30% of users designate ease of manipulation “not applicable”. The reason could possibly be that some users believe this dimension is only used when integrating data. Understandability and reliability exhibited the greatest changing range. This indicates that users may possess different opinions on the understandability and reliability of the information products.

Using our framework, we assessed the quality of raw data in the database and the quality of information products in the contexts. Both results have confirmed that information quality deficiencies existed in the example dataset. These deficiencies may negatively affect business growth and organisational competitiveness. The assessment also shows that our framework is valid for practical applications.

5.6 Validation of the Assessment in Multiple Databases

Based on the algorithms in Section 5.4.2, we developed a system to simulate the assessment of accuracy and completeness in the domain of the supply chain management. In the supply chain, retailer and distributor considered two adjacent entities whose databases share customer dataset. Since a retailer is closer to customers than a distributor in the supply chain, the database of the retailer is considered as the reference database and the database of the distributor is considered as the operational database. The shared dataset contains 100 data records. The period of data alignment is 100 seconds. In each second, we select at random one data operation to operate the randomly selected data. Within the refresh period, we assess accuracy and completeness every 10 seconds.

The objective of this simulation is to test the relationships between the error rate of data operations, assessment time and information quality dimensions. We divide the error rate of data operations into 9 levels from 10% error to 90% error. To define the error rate, we classify data operations as correct operations or incorrect operations. Correct operations include “completely create”, “completely update” and “completely delete”. Incorrect operations consist of “redundantly create”, “incompletely create”, “redundantly update”, “incompletely update”, “redundantly delete” and “incompletely delete”. The error rate is determined by the number of incorrect data operations divided by the total number of data operations. These data operations are selected at random in the simulation. For example, within the first assessment time (10s), 10 data operations are taken. If the error rate is 30%, that means 3 data operations are randomly selected from the set of incorrect operations and 7 data operations are randomly selected from the set of correct operations. After the simulation, we have obtained the following results:



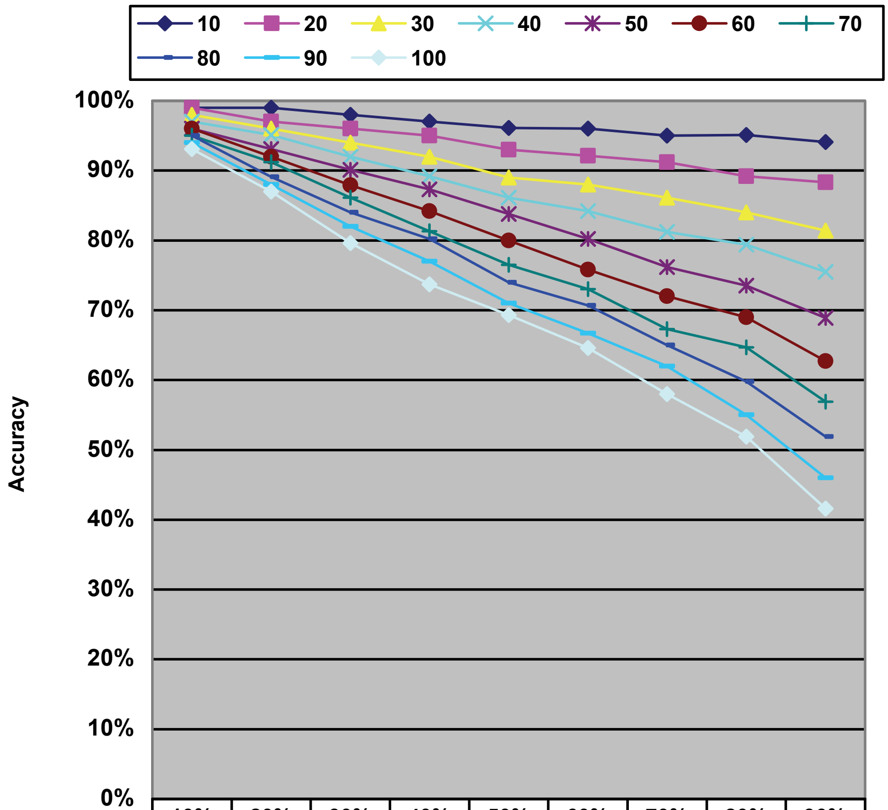
	10	20	30	40	50	60	70	80	90	100
10%error	0.99	0.99	0.98	0.97	0.96	0.96	0.95	0.95	0.94	0.931
20%error	0.99	0.97	0.96	0.951	0.931	0.92	0.912	0.891	0.88	0.87
30%error	0.98	0.96	0.94	0.92	0.901	0.879	0.861	0.84	0.82	0.796
40%error	0.97	0.95	0.92	0.892	0.873	0.842	0.813	0.802	0.77	0.737
50%error	0.961	0.93	0.89	0.861	0.838	0.8	0.765	0.74	0.71	0.693
60%error	0.96	0.921	0.88	0.842	0.802	0.758	0.73	0.707	0.667	0.646
70%error	0.95	0.912	0.861	0.812	0.762	0.72	0.673	0.65	0.62	0.58
80%error	0.951	0.892	0.84	0.794	0.735	0.69	0.647	0.598	0.55	0.519
90%error	0.941	0.883	0.814	0.755	0.689	0.627	0.569	0.519	0.46	0.416

Figure 39: Simulation result for accuracy and operation error rate

When the error rate of data operation is fixed, assessment time is negatively related to accuracy; it can also be obtained from equation 5-10. As time is closing to the refresh point, erroneous data operations such as RU(t) and IU(t) are increasing. Therefore accuracy decreases as the assessment time increases.

When we know the current error rate of data operation and the tolerable level of accuracy, we can determine the correct time period for using the data. The tolerable level of information quality dimensions means people can recognise and correct within a range of data errors, or that these errors do not significantly influence the intended use. For example, in our simulation (Figure 39), if our current error rate of data operation is 50% and we require the accuracy to be above 80%, we are able to use data between 0 and 60 seconds to meet the requirement of accuracy.

Furthermore, when we require accuracy to achieve a certain level, we can determine the optimal error rate of data operation by considering the cost of improvement. As an example, if we require the accuracy to be above 80% in the whole refresh period and our current error rate is 60 %, from the simulation result (Figure 38), we can observe that the error rate of data operation which meets 80% accuracy requirement can vary from 0% to 30%. Therefore we only need to reduce the error rate of data operation to 30% in order to meet the requirement of accuracy, rather than to a value of less than 30% which would be more costly. Consequently, it can reduce the costs of improving data operations to an optimal value.



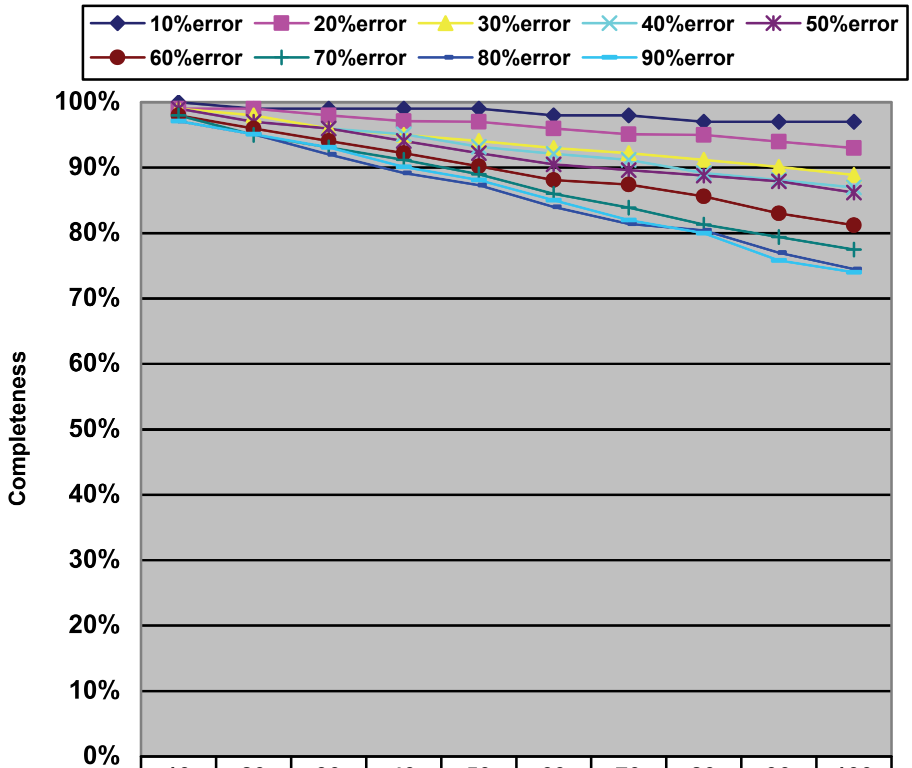
	10%	20%	30%	40%	50%	60%	70%	80%	90%
10	0.99	0.99	0.98	0.97	0.961	0.96	0.95	0.951	0.941
20	0.99	0.97	0.96	0.95	0.93	0.921	0.912	0.892	0.883
30	0.98	0.96	0.94	0.92	0.89	0.88	0.861	0.84	0.814
40	0.97	0.951	0.92	0.892	0.861	0.842	0.812	0.794	0.755
50	0.96	0.931	0.901	0.873	0.838	0.802	0.762	0.735	0.689
60	0.96	0.92	0.879	0.842	0.8	0.758	0.72	0.69	0.627
70	0.95	0.912	0.861	0.813	0.765	0.73	0.673	0.647	0.569
80	0.95	0.891	0.84	0.802	0.74	0.707	0.65	0.598	0.519
90	0.94	0.88	0.82	0.77	0.71	0.667	0.62	0.55	0.46
100	0.931	0.87	0.796	0.737	0.693	0.646	0.58	0.519	0.416

Figure 40: Simulation result for accuracy and assessment time

When the assessment time is fixed, the error rate of data operation is negatively related to accuracy. That means during the same period of time, the more erroneous data operations take place, the more accuracy decreases. We also can observe this result from equation 5-10.

When we have obtained the time we use the data and the level of accuracy we can tolerate, we can determine the optimal error rate of data operation. For example, in our simulation (Figure 40), if we use data at 50s in each refresh period and we can tolerate the accuracy at no less than 80%, the greatest error rate of data operation which can meet this requirement is 60%. Therefore if our current error rate of data operation is 80%, we only need to reduce the error rate to 60%, in turn to achieve the optimum cost-effectiveness.

Moreover, when we require accuracy to achieve a certain level without regarding the error rate of data operations, we can determine the correct time for using data. For example, in our simulation if we require that accuracy is above 80% in the whole refresh period, we can use from 0 to 30 seconds without affecting the error rate.



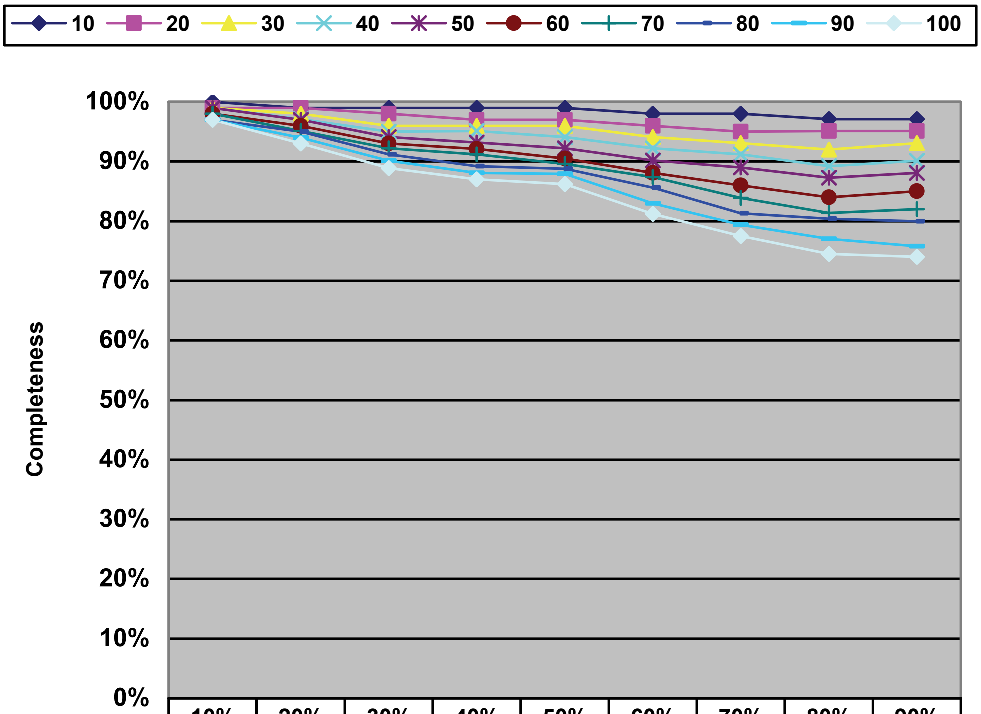
	10	20	30	40	50	60	70	80	90	100
10%error	1	0.99	0.99	0.99	0.99	0.98	0.98	0.97	0.97	0.97
20%error	0.99	0.99	0.98	0.971	0.97	0.96	0.951	0.95	0.94	0.93
30%error	0.99	0.98	0.96	0.95	0.941	0.93	0.922	0.912	0.901	0.889
40%error	0.99	0.97	0.96	0.951	0.932	0.921	0.912	0.892	0.881	0.87
50%error	0.99	0.97	0.96	0.941	0.922	0.905	0.896	0.888	0.879	0.862
60%error	0.98	0.96	0.941	0.922	0.902	0.881	0.874	0.856	0.83	0.812
70%error	0.98	0.95	0.931	0.912	0.89	0.86	0.839	0.813	0.794	0.775
80%error	0.971	0.951	0.92	0.892	0.873	0.84	0.814	0.804	0.77	0.745
90%error	0.971	0.951	0.931	0.901	0.881	0.85	0.82	0.8	0.758	0.74

Figure 41: Simulation result for completeness and operation error rate

We can obtain the following conclusions with reference to completeness: When the error rate of data operation is fixed, assessment time is negatively related to completeness. As the time increases, erroneous data operations that are related to completeness also increase. Therefore completeness is negatively associated with assessment time.

When we have obtained the current error rate of data operation and the tolerable level of completeness, we can determine the optimal error rate of data operation. For example, in our simulation, if our current error rate of data operation is 50% and we require the completeness to be above 90%, from the simulation result (Figure 41) we can observe that we could use the data between 0 to 60 seconds to meet the requirement of completeness.

When we require completeness to achieve a certain level, we can determine the optimal error rate of data operation by considering the cost of improvement. For example, in our simulation, if we require that completeness is above 90% in the whole refresh period, the error rate of data operation can vary from 0% to 20%. Consequently we only need to improve the error rate to 20% to achieve the minimum cost of improving data operations.



	10%	20%	30%	40%	50%	60%	70%	80%	90%
10	1	0.99	0.99	0.99	0.99	0.98	0.98	0.971	0.971
20	0.99	0.99	0.98	0.97	0.97	0.96	0.95	0.951	0.951
30	0.99	0.98	0.96	0.96	0.96	0.941	0.931	0.92	0.931
40	0.99	0.971	0.95	0.951	0.941	0.922	0.912	0.892	0.901
50	0.99	0.97	0.941	0.932	0.922	0.902	0.89	0.873	0.881
60	0.98	0.96	0.93	0.921	0.905	0.881	0.86	0.84	0.85
70	0.98	0.951	0.922	0.912	0.896	0.874	0.839	0.814	0.82
80	0.97	0.95	0.912	0.892	0.888	0.856	0.813	0.804	0.8
90	0.97	0.94	0.901	0.881	0.879	0.83	0.794	0.77	0.758
100	0.97	0.93	0.889	0.87	0.862	0.812	0.775	0.745	0.74

Figure 42: Simulation result for completeness and assessment time

When the assessment time is fixed, the error rate of data operation is negatively related to completeness. This means completeness decreases as erroneous data operations increase. Therefore erroneous data operations contribute negatively towards completeness.

When we have obtained the time we need to use the data and the level of completeness we can tolerate, we can determine the optimal error rate of data operation. For example in our simulation (Figure 42), if we use data at 50s in each refresh period and we can tolerate completeness of no less than 90%, we only need to reduce the error rate to 60% in order to achieve the optimal cost.

When we require completeness to achieve a certain level without regarding the error rate of data operations, we can determine the correct time for using data. For example in our simulation, if we require the accuracy to be above 90%, we then can use the data from 0 to 30 seconds without influencing the error rate of data operations.

From the comparison between accuracy and completeness, accuracy decreases faster than completeness in the refresh period. Therefore when users operate data for intended use, the impact of accuracy problems may be more significant than that of completeness problems. It also indicates that when we improve information quality, the improvement of accuracy contributes more than the improvement of completeness.

5.7 Summary

This chapter provides a suitable and validated information quality assessment framework for our experiments. This framework consists of two major components: validated dimensions and assessment methodologies. From the relevant information quality literature, we adopt 17 information quality dimensions and classify them into four categories, which are acquisition, context, specification and expectation. By using confirmatory factor analysis and computing Cronbach alpha, we tailor 17 dimensions into 9 dimensions. Analysis of the results supports strong reliability and the validity of the dimensions. Based on the validated dimensions, we consequently propose methodologies to assess information quality in the scenarios of a single database and of multiple databases.

In the scenario of a single database, we propose two assessment approaches that respectively measure the quality of raw data and information products. To validate the assessment, we apply the two assessment approaches in the real-world application.

In this application, a software prototype and an online survey are developed to facilitate the assessment. The research findings have shown that information quality deficiencies existed in a real-world dataset. These deficiencies are confirmed from both software and user-oriented scenarios. Information quality issues may result in significant losses for organisations, for example losing customers, missing opportunities and making incorrect decisions. The application indicates that our assessment framework provides a feasible prototype for practical information quality assessment. Using this framework, we can control the levels of information quality in our experiments.

In the scenario of multiple databases, we propose a model to assess information quality in the context of an information chain. The model consists of three major components: data operations, information quality problems and information quality dimensions. We focus on three frequently used data operations: create, update and delete. In each operation, we divide the operation into redundant operation, complete operation and incomplete operation. During the refresh time in an information chain, these data operations may generate information quality problems. Based on the identification of information quality problems, we map these information quality problems to information quality dimensions.

As information quality is a multi-dimensional concept, we focus on the two most-cited information quality dimensions: accuracy and completeness. To derive the algorithms, accuracy is mapped to incorrect and redundant data problems, and completeness is mapped to missing data problems. Each problem is connected to different data operations. In the light of the relationship between data operations, information quality problems and information quality dimensions, we propose algorithms to determine accuracy and completeness. Using the algorithms, a simulation is carried out in the context of information chain. The simulation results have shown that when the error rate of data operation is fixed, accuracy and completeness are negatively related to assessment time. Based on random selections of data operations, we can determine the optimal time of using data. In addition, accuracy decreases faster than completeness in the same refresh period. Since our experiments are currently carried out only in the scenario of a single database, the algorithms of multi-database assessment are not used in our experiments. However, as we emphasize in Chapter 7, the experiment can be extended to a game that multiple users can play in the same supply chain. Thus the algorithms can be used to control the level of information quality in the future multi-user experiment.

Chapter 6: Data Analysis

In Chapters 4 and 5, we presented our research hypotheses, experimental design and the measurement of information quality. After carrying out the experiments, we collected the experimental results through the software we developed. Using the collected data, we analyse the effects of information quality categories and dimensions on decision quality. Additionally, we develop a workshop to further confirm the analysis results. In summary, this chapter mainly discusses the results and the indications derived from the data analysis.

6.1 Data Analysis for Experiment 1

A total of 80 college students enrolled in a master's degree were involved in experiment 1. All the subjects are selected from the school of computing and the school of business at the Dublin City University. 58% of the subjects were male and the remaining 42% were female. The average age of subjects was 27. Each subject made 10 decisions for the JIT scenario and 10 decisions for the EOQ scenario. From the 10 decisions, we can calculate a value for decision quality by considering the correct decisions in the 10 decisions. Thus each subject contributed one decision quality score for each scenario. The data was examined by histogram and confirmed to be normally distributed. After the experiment, subjects received a monetary reward for their participation in the experiment. The subject who achieved the highest score in the experiment was awarded an additional monetary reward. The descriptive statistics of the decision quality and the subjects are summarised in Table 29.

In Table 29, IIQ represents intrinsic information quality, CIQ represents contextual information quality and RIQ represents representational information quality. As discussed in Chapter 4, high- or low-quality level in each category is achieved by setting all the dimensions in this category to high- or low-quality level. For example, IIQ consists of accuracy, believability, objectivity and reputation. High IIQ means all the four dimensions are considered to be in the high-quality level. Table 29 also shows the mean value and standard deviation of decision quality, sample size, average age of the participants and the gender ratio in each treatment.

	High IIQ				Low IIQ			
	High CIQ		Low CIQ		High CIQ		Low CIQ	
	High RIQ	Low RIQ	High RIQ	Low RIQ	High RIQ	Low RIQ	High RIQ	Low RIQ
Scenario 1 (JIT Decision Quality)								
Mean	.79	.75	.41	.44	.50	.35	.14	.19
Std. Dev.	.09944	.11785	.11972	.09661	.13333	.14337	.08433	.14491
Sample	10	10	10	10	10	10	10	10
Age	24	25	25	27	27	25	25	23
M/F	6/4	5/5	7/3	6/4	6/4	5/5	5/5	6/4
Scenario 2 (EOQ Decision Quality)								
Mean	.54	.49	.29	.31	.21	.24	.08	.02
Std. Dev.	.08433	.11972	.09944	.11005	.09944	.05164	.06325	.04216
Sample	10	10	10	10	10	10	10	10
Age	25	26	27	25	26	23	24	25
M/F	6/3	5/5	4/6	6/4	6/4	5/5	7/3	6/4

Table 29: Descriptive statistics of JIT and EOQ experimental treatments

In order to clarify the effects of information quality categories on decision quality, ANOVA was executed to analyse the data collected in the experiments. JIT and EOQ decision quality were tested by two three-way ANOVAs. The tests were at a 95% confidence level, and SPSS 14.0 was used for data analysis. The result of between-subjects using ANOVA for JIT scenario is shown in Table 30.

Source	Type III Sum of Squares	Df	Mean Square	F	Sig.
Corrected Model	3.834(a)	7	.548	38.472	.000
Intercept	15.931	1	15.931	1119.064	.000
IIQ	1.830	1	1.830	128.555	.000
CIQ	1.682	1	1.682	127.210	.000
RIQ	.015	1	.015	1.062	.306
IIQ * CIQ	.036	1	.036	2.538	.116
IIQ * RIQ	.010	1	.010	.711	.402
CIQ * RIQ	.091	1	.091	6.401	.014
IIQ * CIQ * RIQ	.021	1	.021	1.484	.227
Error	1.025	72	.014		
Total	20.790	80			
Corrected Total	4.859	79			

Table 30: Data analysis for the effects of information quality categories on JIT decision quality

The ANOVA result of information quality effects on JIT decision quality showed that the main effects of intrinsic information quality ($F(1, 72) = 128.555$, $p = < 0.001$) and contextual information quality ($F(1, 72) = 127.210$, $p < 0.001$) were significant, indicating that the two variables are crucial factors in the making of decisions. This also means that intrinsic and contextual information quality are positively related to decision quality. Therefore, H1 and H2 are supported. The effect of representational information quality ($F(1, 72) = 1.062$, $p = 0.306$) was found to be non-significant. As a result, H3 was rejected. However, the interaction effect of contextual information quality and representational information quality was significant. It indicated that contextual and representational information quality jointly affect JIT decision quality. This finding implied that contextual and representational information quality were mutually intensified. Therefore, an alternative approach to the improvement of contextual information quality is through enhancing representational information quality. In addition, no significant interaction effect was found between intrinsic and contextual information quality ($F(1, 72) = 2.538$, $p = 0.116$) as well as intrinsic and representational information quality ($F(1, 72) = 0.711$, $p = 0.402$). Finally, intrinsic information quality, contextual information quality and representational information quality (F (1, 72) = 1.484, p = 0.227) did not jointly affect JIT decision quality.

In order to support the external validity (refer to Section 3.4) of the experiment, we perform a three-way ANOVA (Table 31) in the scenario of EOQ to confirm the results in JIT scenario.

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	2.503(a)	7	.358	52.431	.000
Intercept	6.216	1	6.216	911.530	.000
IIQ	1.596	1	1.596	234.055	.000
CIQ	.861	1	.861	126.275	.000
RIQ	.000	1	.000	.018	.893
IIQ * CIQ	.021	1	.021	3.098	.083
IIQ * RIQ	.003	1	.003	.458	.501
CIQ * RIQ	.006	1	.006	.898	.346
IIQ * CIQ * RIQ	.015	1	.015	2.218	.141
Error	.491	72	.007		
Total	9.210	80			
Corrected Total	2.994	79			

Table 31: Data analysis for the effects of information quality categories on EOQ decision quality

The main effects of intrinsic information quality (F (1, 72) = 234.055, p < 0.001) and contextual information quality (F (1, 72) = 126.275, p < 0.001) were significant, indicating that intrinsic and contextual information quality significantly contributed to inventory decision quality. The main effect of representational information quality (F (1, 72) = 0.018, p = 0.893) was found to be non-significant. This means representational information quality is not positively related to inventory decision quality.

The three main effects confirmed the result of the JIT scenario. Surprisingly, no significant interaction effect of contextual information quality and representational information quality was found. This is not consistent with the result of JIT scenario. One possible explanation could be the variety of scenarios and the distinction between experimental goals. This means the interaction effects of information quality categories varied according to specific organisational contexts. Except this particular interaction effect, the remainder of the effects conformed to the results of JIT scenario.

6.2 Data Analysis for Experiment 2

80 postgraduate students participated in experiment 2. One third of the subjects were research students in computer science. Others were involved in computing or business Master's programmes. All the subjects are from Dublin City University. 62% of the subjects were male and 38% were female. The average age of subjects was 27.5. In the experiment, each subject contributed an interval data to decision cost. After the data collection, the data set was described by histogram and examined by Levene's test. The results showed that the collected data were normally distributed and exhibited homogeneity of variance (Levene's sig. = 0.324). The descriptive analysis is summarised in Table 32. In the following table, Comp, Accu and Cons respectively represent completeness, accuracy and consistency.

Comp	Accu	Cons	Mean	Std. Deviation	N
.00	.00	.00	882.6000	68.05096	10
		1.00	902.0000	76.34862	10
		Total	892.3000	71.09008	20
	1.00	.00	779.1000	38.54709	10
		1.00	754.8000	40.46617	10
		Total	766.9500	40.43380	20
	Total	.00	830.8500	75.60721	20
		1.00	828.4000	96.11911	20
		Total	829.6250	85.36669	40
1.00	.00	.00	664.5000	42.10635	10
		1.00	753.9000	57.17702	10
		Total	709.2000	67.01971	20
	1.00	.00	560.1000	50.78812	10
		1.00	522.5000	62.39703	10
		Total	541.3000	58.63545	20
	Total	.00	612.3000	70.21328	20
		1.00	638.2000	132.22652	20
		Total	625.2500	105.31631	40

Table 32: Descriptive analysis for experiment 2

Based on the descriptive analysis, a three-way ANOVA was executed for inferential analysis. The ANOVA was carried out at a 95% confidence level and SPSS 14.0 was used for this analysis. The between-group result for decision quality is summarised in Table 33.

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	1326277.988(a)	7	189468.284	60.392	.000
Intercept	42333225.313	1	42333225.313	13493.516	.000
Comp	835382.813	1	835382.813	266.274	.000
Accu	429977.813	1	429977.813	137.053	.000
Cons	2749.513	1	2749.513	.876	.352
Comp * Accu	9052.513	1	9052.513	2.885	.094
Comp * Cons	4018.613	1	4018.613	1.281	.261
Accu * Cons	36423.113	1	36423.113	11.610	.001
Comp * Accu * Cons	8673.613	1	8673.613	2.765	.101
Error	225885.700	72	3137.301		
Total	43885389.000	80			
Corrected Total	1552163.688	79			

Table 33: Data analysis for the effects of information quality dimensions on decision quality

The three way ANOVA results show that the main effects of completeness ($F(1, 72) = 266.274$, $p < 0.001$) and accuracy ($F(1, 72) = 137.053$, $p < 0.001$) of information on decision quality are significant. This means completeness and accuracy of information are two fundamental factors in decision-making. Increasing completeness or accuracy of information could improve decision quality. Furthermore, decision-makers could expect to make better decisions when using information with high-completeness or with high-accuracy. Therefore hypotheses H7 and H8 are supported. The main effect of consistency ($F(1, 72) = 0.876$, $p = 0.352$) of information on decision quality is non-significant, indicating that there is no significant difference in the effect of consistency of information on decision quality. Regardless of information completeness and accuracy, the improvement of information consistency could not enhance decision quality. Therefore hypothesis H9 is rejected.

Except the main effects, four interaction effects are illustrated in Table 33. Firstly, the interaction effect of completeness and consistency of information is found to be non-significant ($F (1, 72) = 1.281$, $p = 0.261$). The main cause can be attributed to the non-significance of consistency. This means with both high and low completeness of information, high consistency of information does not benefit decision quality.

Secondly, it is found that the interaction effect of completeness and accuracy is also non-significant ($F (1, 72) = 2.885$, $p = 0.094$). This means that although completeness or accuracy of information significantly affect decision quality, their combined effects on decision quality are non-significant.

Third, the interaction effect of accuracy and consistency of information is found to be significant ($F (1, 72) = 11.610$, $p = 0.001$), indicating that accuracy and consistency of information jointly affect decision quality. This finding implies that accuracy and consistency of information are mutually intensified. The positive effect of increasing accuracy with high consistency is greater than that with low consistency. Although the effect of information consistency on decision quality is not significant, improving consistency can intensify the effects of increasing accuracy. To further analyse this significant interaction effect, we carried out two one-way ANOVAs for testing accuracy and consistency.

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	429977.813	1	429977.813	29.887	.000
Within Groups	1122185.875	78	14386.998		
Total	1552163.688	79			

Table 34: One way ANOVA of accuracy and decision quality

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	2749.513	1	2749.513	.138	.711
Within Groups	1549414.175	78	19864.284		
Total	1552163.688	79			

Table 35: One way ANOVA of consistency and decision quality

The results in Table 34 and Table 35 show that the effect of accuracy is significant ($F(1, 78) = 2.885$, $p < 0.001$) and the accuracy of information is positively related to decision quality. The effect of information consistency on decision quality is not significant ($F(1, 78) = 2.885$, $p < 0.001$). Both results have confirmed the results of the three-way ANOVA presented above in Table 33.

Finally, the interaction effect between completeness, accuracy and consistency is non-significant. This may result from the non-significant effects of consistency, completeness and accuracy, as well as completeness and consistency. To further refine this analysis, we control two variables as a testing condition and analyse how the other variable responds in such a condition. For example, when information is incomplete and accurate, we investigate the effect of information consistency on decision quality. Thus 12 more in-depth data analyses are performed and presented in Table 36.

Hypotheses	df	Mean Square	F	P-value	Result
H7a	1	267729.800	74.758	<.001	Supported
H7b	1	54496.800	25.042	<.001	Supported
H7c	1	108339.200	29.020	<.001	Supported
H7d	1	53561.250	17.513	.001	Supported
H8a	1	269816.450	97.567	<.001	Supported
H8b	1	239805.000	117.976	<.001	Supported
H8c	1	109668.050	24.107	<.001	Supported
H8d	1	237838.050	74.279	<.001	Supported
H9a	1	7068.800	2.184	.157	Rejected
H9b	1	39961.800	15.851	.001	Rejected
H9c	1	2952.450	1.891	.186	Rejected
H9d	1	1881.800	.360	.556	Rejected

Table 36: In-depth data analysis for experiment 2

Four situations are used to study the effects of accuracy: high completeness and high consistency, high completeness and low consistency, low completeness and high consistency, and low completeness and low consistency. Under the above four situations, the effects of accuracy of information on decision quality are all significant. That means if the quality of information used for decision-making belongs to any particular one of the four situations, increasing accuracy of information could significantly contribute to decision quality. Therefore hypotheses H7a, H7b, H7c and H7d are supported.

Analogous to the above four situations, effects of completeness are investigated by four conditions: high accuracy and high consistency, high accuracy and low consistency, low accuracy and high consistency, and low accuracy and low consistency. The results show that under each condition, the effect of completeness on decision quality is found to be significant. Therefore hypotheses H8a, H8b, H8c and H8d are also supported.

When information is complete but inaccurate, the effect of information consistency is found to be significant. From the descriptive analysis, we can observe that consistency of information is negatively related to decision quality. That means when the information is assessed as complete but inaccurate, the decision-maker using inconsistent information performs better than the one using consistent information. A potential explanation is that information inconsistency may force decision-makers to consider given information more carefully, to try to correct the inaccurate values. Therefore H9b is rejected. Under the other three conditions - high completeness and high accuracy, low completeness and high accuracy, and low completeness and low accuracy - the effects of consistency are non-significant. Consequently hypotheses H9a, H9c, H9d are rejected.

From the data analysis of experiment 1, we found that the information quality categories, intrinsic and contextual, are positively related to decision quality. The two categories respectively contain accuracy and completeness as their dimension (Figure 16), which are also found to be positively related to decision quality in experiment 2. Thus information accuracy and completeness are observed to be the key elements for improving decision quality. Therefore to further investigate how accuracy and completeness affect decision quality. We develop more quality levels for accuracy and completeness and conduct the following workshop.

6.3 Workshop

6.3.1 Workshop Overview

The objective of this workshop is to show the effects of information quality from user perspective. The effects of information quality are shown through our beer game experiment. Participants were invited to attend this experiment and provide their comments when completing their decisions. After the experiment, we introduced the research on information quality management, which includes TDQM cycle, information quality assessment and information quality improvement. This workshop is also designed to show the importance of information quality and help participants to manage information quality in their organisations.

The participants were invited from both academia and industry. These participants are mainly postgraduate students from computer science or management science, academic researchers who are interested in information system research and organisational staff whose job involves dealing with data analysis or information management. The workshop was held in the University of Oxford on 10 January 2009. A total of 30 participants attended this workshop. These attendees are mainly from Dublin City University, University of Oxford, Singapore Nanyang Technological University, University of Dundee, Microsoft Research Asian, Hibernia Atlantic Ltd., Avaya Ireland, and J.P. Morgan UK.

The participants are expected to understand the impact of information quality dimensions on decision quality, learn how to manage information quality, and contribute to an experimental result that is aligned with the goal of the workshop.

6.3.2 Workshop Layout

Before introducing the concept of information quality, we first presented several case studies to demonstrate the impact of information quality. These case studies cover the information quality issues arising in our daily life and disasters caused by poor quality information. For example, we frequently encounter misspelling or outdated information in certain documents. We may not pay full attention to these information quality problems but they could potentially cause inconveniences to our lives. In addition to cases drawn from everyday life, we also present information quality disasters on a much larger scale: for instance, in July 1988 the U.S. Navy Cruiser USS Vincennes shot down an Iranian commercial passenger jet and killed all 290 people on board (refer to the details in Chapter 1). It was found that information quality was a major influencing factor in the USS Vincennes decision-making process. After the case studies, we introduce the definition of information quality and present a review of information quality definitions.

Following on from the case studies and the concept of information quality, we invited the workshop attendees to participate in an information quality experiment. This experiment is based on the proposed experiments in Chapter 4. However, instead of controlling information quality in the predicting information, this experiment controls information quality in the ordering information. As discussed in the end of Section 6.2, we further focus on how accuracy and completeness affect decision quality in more details. By using a practical approach, decision quality is indicated by the decision cost. Before conducting the experiment we give an oral introduction to it. The introduction is instructed as follows.

“This is a supply chain simulation game, which involves the manufacturer, distributor and customer. The participants are asked to play the distributor and the other two roles are taken over by the computer. According to the inventory information and customer ordering history, participants will order products from the manufacturer and supply products to the customer. There will be two sources of cost associated with the game: (1) If the distributor cannot supply the customer’s order, a cost will be occurred (1 euro per unfilled item). (2) When the items are stored in the inventory, a cost will be incurred (0.5 euro per stored item per *week*). The goal of the game is to minimise the cost to the distributor's inventory management. Decision quality is indicated by decision cost. The higher the cost of the decision, the lower the decision quality is.”

Followed by the experiment description above, we directed each participant to one computer and showed them the experiment website. This website contains the text description of this experiment and the background knowledge of information quality. The snapshot of this website is captured as follows (Figure 43):

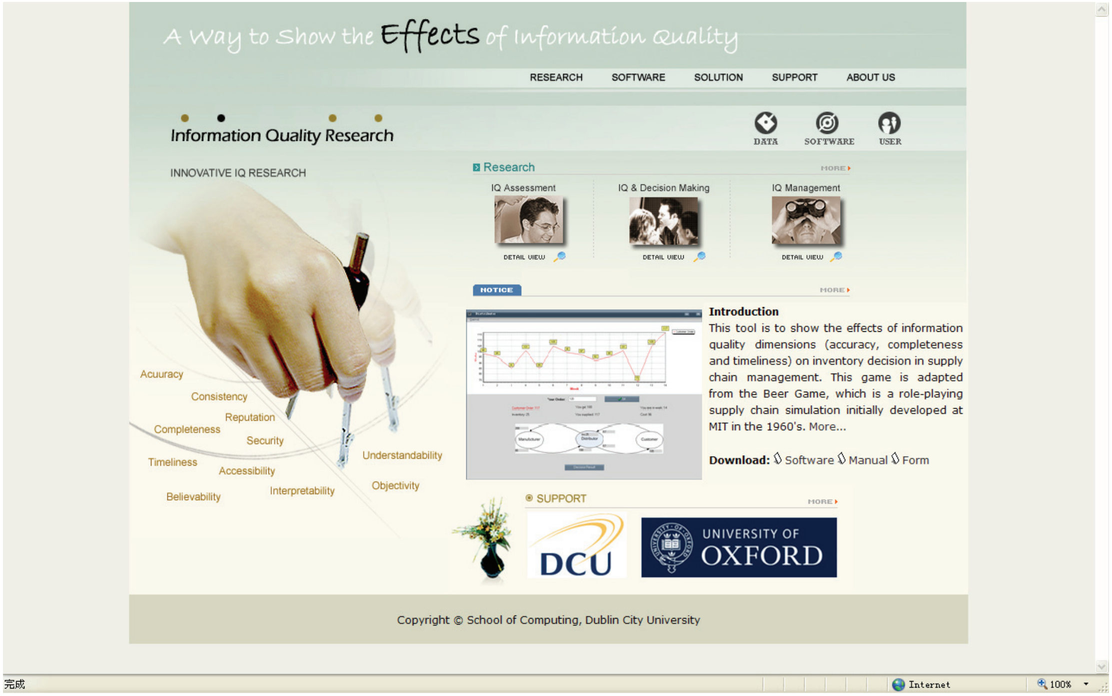


Figure 43: Screenshot of the website

When participants log on to the website, they have 20 minutes to browse the introduction of the experiment and review the case studies. After that, we instructed participants to download the software and run this software on their computers. We could observe the initial interface of this software in Figure 44. In the beginning of the experiment, participants are given previous customer orders for the last 10 *weeks*. For example, Figure 44 showed that customer order was 99 items in *week* 1 and 106 in *week* 2. Participants will play the game in *week* 10. In the beginning week, the inventory is 100 items. This figure can be found at the inventory label or inside the distributor oval.

In order to start the game, participants place the order in the input box and click the *OK button*. This order will be sent to the manufacturer. The procurement delay from manufacturer takes 3 *weeks*. That means it takes 3 *weeks* from making the order and receiving the ordered items. Note that the manufacturer may not fill the order if this order exceeds the inventory of the manufacturer. For example, in *week* 10, participants place an order to the manufacturer. This order will arrive to the manufacturer in the beginning of *week* 11. The manufacturer needs 1 *week* to process the order. In the beginning of *week* 12, the manufacturer ships the ordered items to the distributor and the shipment takes 1 *week*. Therefore if participants make an order in *week* 10, the order will arrive in *week* 13.

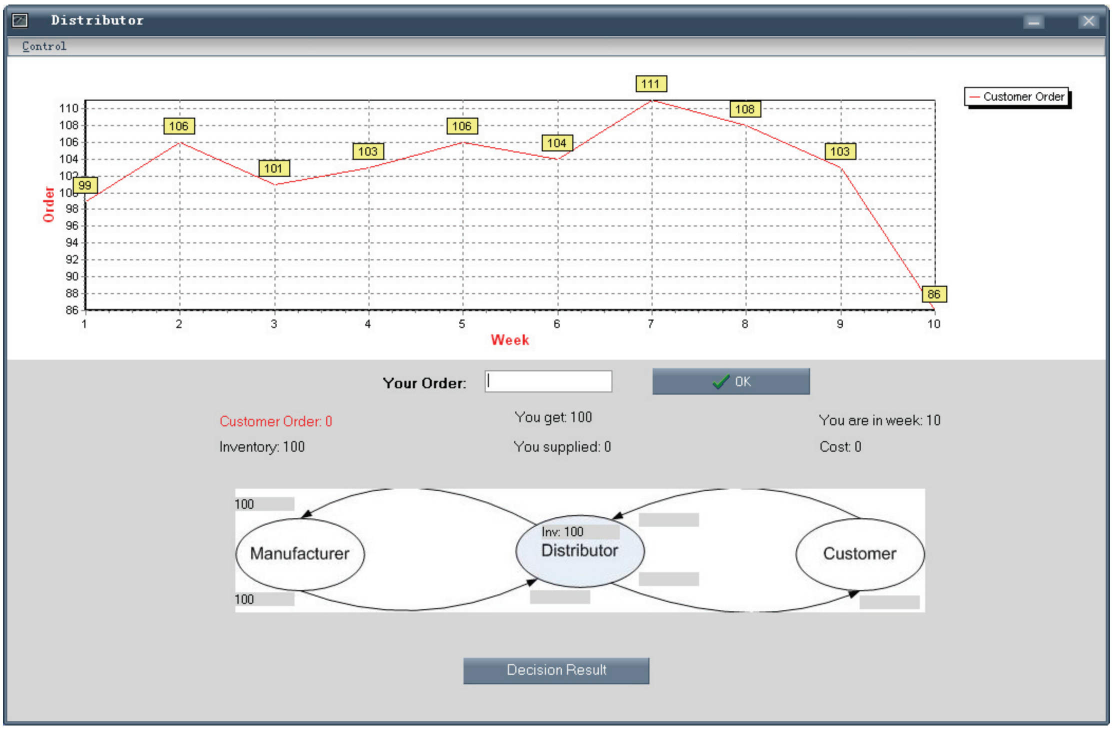


Figure 44: Initial software interface in the workshop

When playing the game, participants can obtain the following information from the software interface: customer order, current inventory, how many ordered items are arrived, how many items are shipped to the customer, current week, and the total cost. Participants also can observe the delays between the manufacturer, the distributor and the customer. One round of the game contains 10 *weeks*. After played for 10 *weeks*, the *OK button* becomes *another 10-week play button*. Participants could either play another round of the game by clicking this button or check the decision result by clicking *decision result button*. Figure 45 displays a screenshot of the software at the end of one round of the experiment.

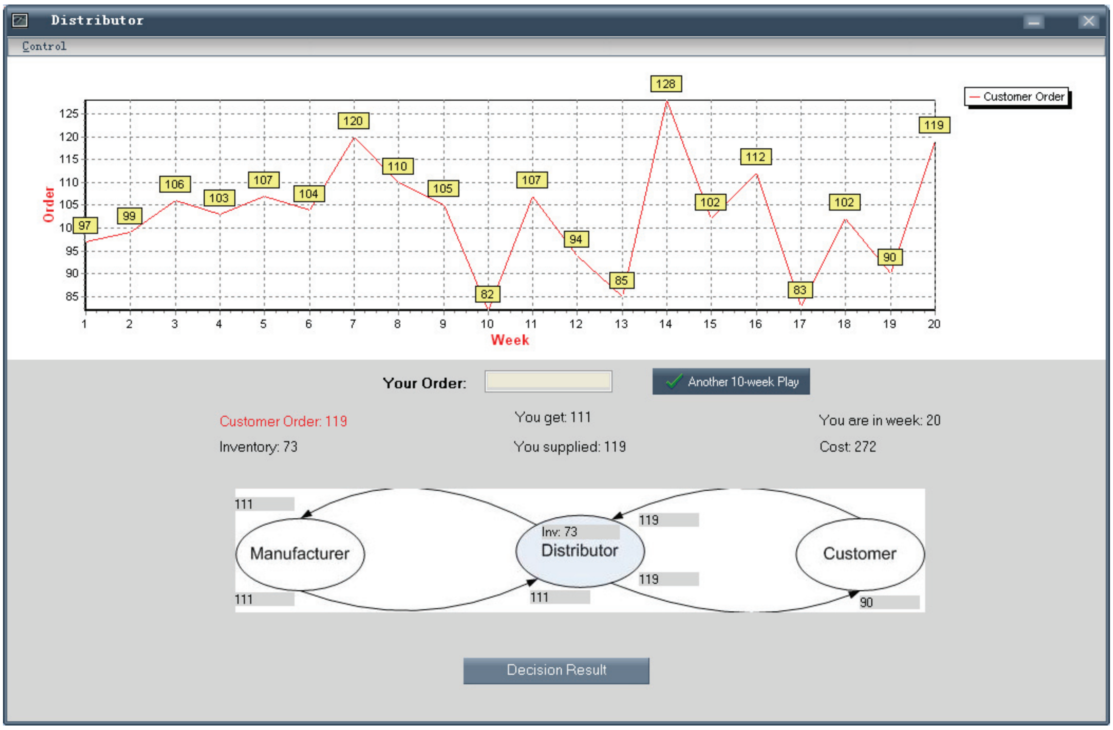


Figure 45: Screenshot of one round of the workshop experiment

When participants click the *decision result button*, the decision results of last 10 weeks are showed as Figure 46.

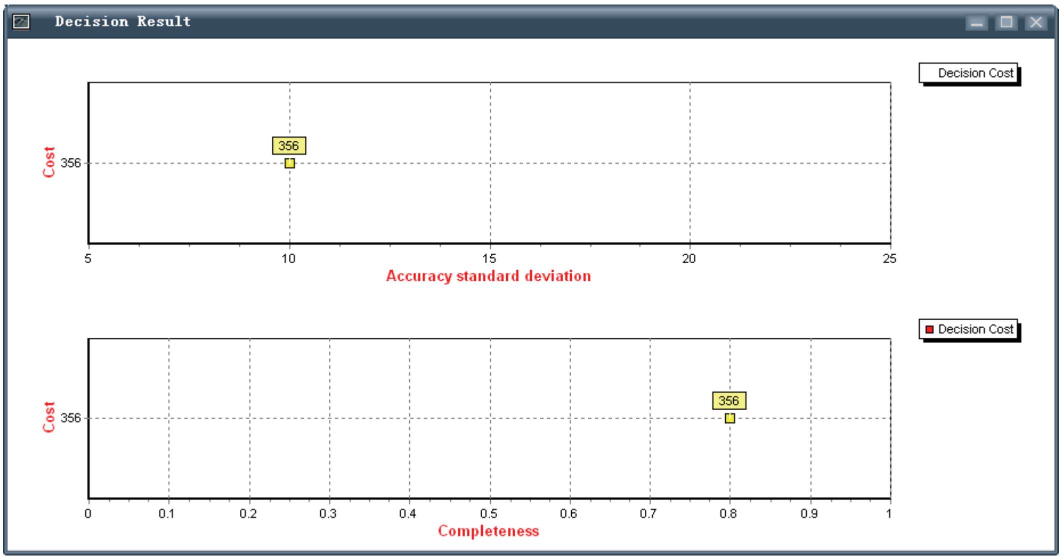


Figure 46: Decision result of the workshop experiment

In Figure 46, we can observe that when the standard deviation of information accuracy is 10 and the information completeness is 80%, the participant made 10 ordering decisions and the total cost incurred is 356 euro.

The measurements of completeness and accuracy are organised as follows. The levels of completeness are expressed by percentages. The percentages can be determined by the equation 4-5. In the experiment, 5 completeness levels are used: 20%, 40%, 60%, 80% and 100%. For example, 60% completeness means only 60% of the information is provided.

Accuracy is divided into 5 levels by different standard deviations. These 5 standard deviation levels are 5, 10, 15, 20, and 25. The greater the standard deviation, the greater the degree of inaccuracy is. For example, the orders generated by standard deviation 20 is always more inaccurate than the ones generated by standard deviation 5. The reason we employ the standard deviation is that we can design the inaccuracy distributed in every order. The accuracy in each order can be calculated by the equation 4-6.

In order to examine the effect of information accuracy on decision quality, we set the information completeness level to 100% and vary the 5 standard deviations of information accuracy. When we examine the effect of information completeness on decision quality, we set the standard deviation of information accuracy to 0 and vary the 5 levels of information completeness. In the experiment, each participant is asked to make two decisions. One is based on inaccurate data and the other is based on incomplete data.

6.3.3 Workshop Result

A total of 30 workshop participants were involved in the experiment. 65% of subjects were male and the other 35% female. The average age of subjects is 32. The data is collected through a software-based system as shown in Figure 45. All the data are checked and confirmed as valid and usable for data analysis. Based on the collected data, we carry out a regression analysis to examine the effects of accuracy and completeness on decision quality. The confidence interval of this regression analysis is 95%. In table 37, it shows the result of regression analysis between accuracy and decision cost.

Dependent Variable: Cost. The independent variable is Accuracy.

Equation	Model Summary					Parameter Estimates	
	R Square	F	df1	df2	Sig.	Constant	b1
Linear	.797	110.159	1	28	.000	180.283	18.977

Table 37: Regression analysis between accuracy and decision cost

The linear regression between accuracy and decision cost is found to be significant (Sig. < 0.001), indicating that information accuracy is well correlated with decision cost. 79.7% (R Square = 0.797) of the variation in decision cost is explained by the standard deviation of accuracy. That means decision cost is 79.7% accounted for or predicated by information accuracy. As the sign on the coefficient (b1=18.977) is positive, increasing the standard deviation of accuracy is expected to increase decision cost. That indicates decision quality is negatively related to the standard deviation of accuracy. A decision-maker could expect higher decision quality through increasing the level of accuracy. We show the curve estimation between accuracy and decision cost in Figure 47.

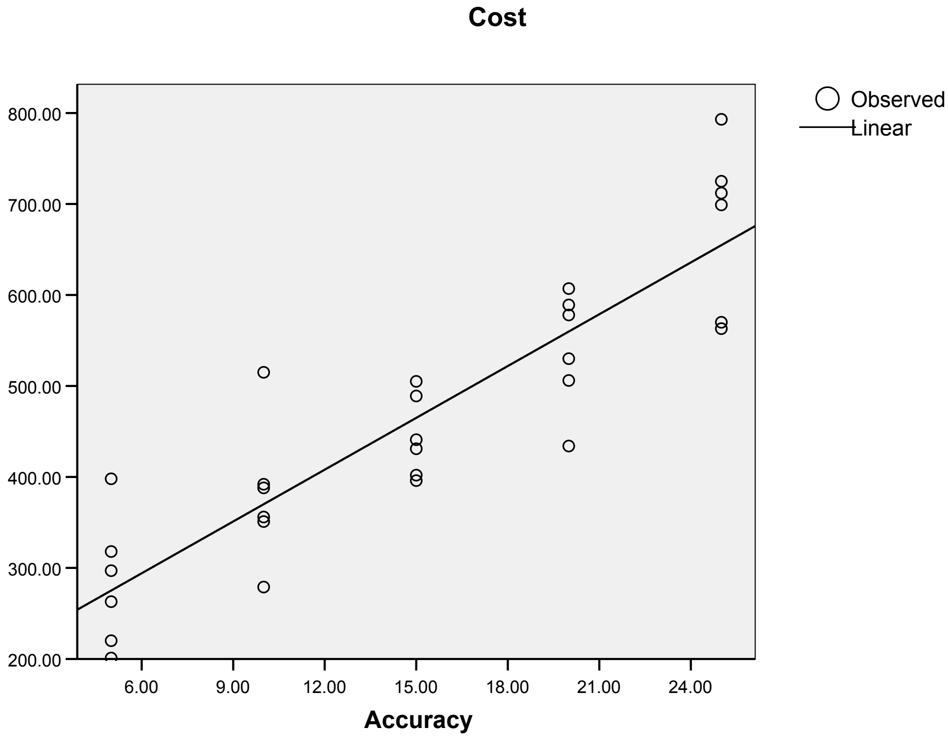


Figure 47: Curve estimation between accuracy and decision cost

As with the analysis above, we list the result of regression analysis between completeness and decision cost in Table 38.

Dependent Variable: Cost. The independent variable is Completeness.

Equation	Model Summary					Parameter Estimates	
	R Square	F	df1	df2	Sig.	Constant	b1
Linear	.878	202.330	1	28	.000	1023.500	-727.167

Table 38: Regression analysis between completeness and decision cost

The linear regression between completeness and decision cost is also found to be significant (Sig. <0.001), indicating that information completeness is well correlated with decision cost. 87.8% (R Square = 0.878) of the variation in decision cost is explained by the level of completeness. That means decision cost is 87.8% accounted for or predicated by information completeness. As the sign on the coefficient (b1=-727.167) is negative, increasing the level of completeness is expected to decrease decision cost. That indicates decision quality is positively related to the level of completeness. A decision-maker can expect higher decision quality through increasing information completeness. We show the curve estimation between completeness and decision cost in Figure 48.

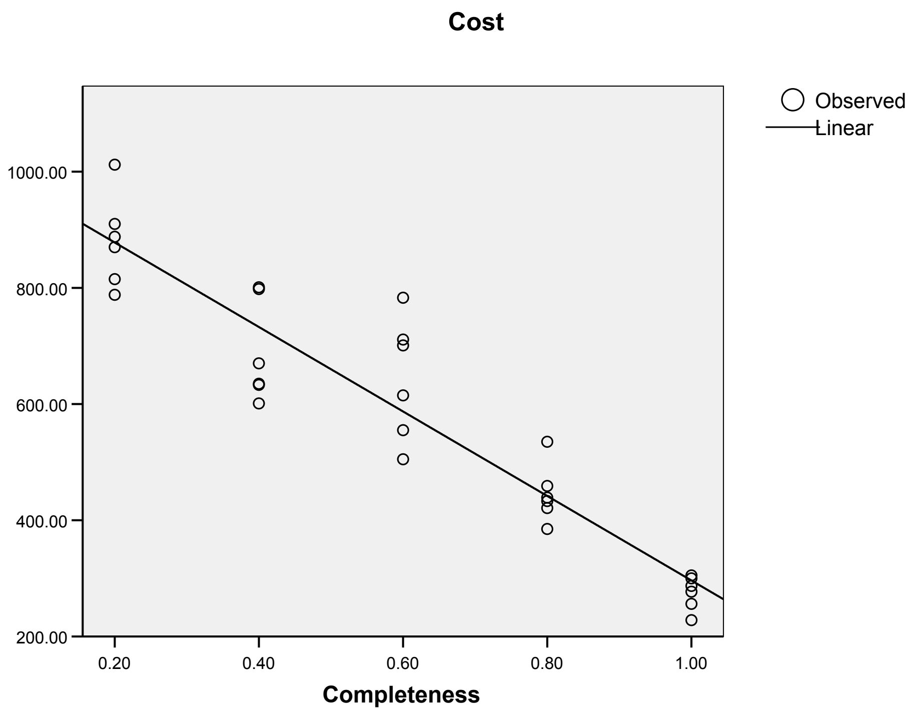


Figure 48: Curve estimation between completeness and decision cost

This regression analysis illustrated that decision quality increases as the level of accuracy or completeness is rising. This result confirmed our three-way ANOVA analysis conclusion that accuracy and completeness are positively related to decision quality.

In order to further confirm our experimental result, we carried out a structured interview to the participants. This interview consists of 4 structured questions: (1) How is your work or study related to information processing? (2) Do you encounter any information quality problems in your work or life? (3) What do you think is the significance of information accuracy and completeness in decision-making? (4) What are the possible root causes of the information quality problems? After collating the answers, we organise the feedbacks as follows.

- All the interviewees' works or studies are related to information processing. This information can be presented by text, audio, video, conversation etc. The most common information processing concerned checking emails and collecting business information. Some interviewees' work concluded with a report. This report can be considered as a result of gathering and processing different information. Therefore information quality is directly related to the quality of their work. Some interviewees' work is centred on data analysis and data management. Information quality is vital since poor quality information may generate erroneous analysis results, which could incur a variety of business losses.

- Most of the interviewees have met information quality problems in their work or daily life. Interviewees provided us with various examples of such information quality problems. Compiling their answers, we find that information quality problem is highly pervasive in work and everyday life. Hence it is valuable to improve information quality in people's work and daily life. However, although all the interviewees have met different kinds of information quality problems, only a few people had considered information quality improvement. Some interviewees have grown accustomed to information quality problems and were not aware of the possibility of improving information quality. Some student interviewees complained that information quality problems always happened within their life and study, especially with regards to information from the internet. Some industrial interviewees had considered improving information quality in their company. However due to unknown budgets and the inexistence of mature information quality management models, high management always denied proposals for systematic information quality improvement. The answers to this question showed that although many information quality problems exist, the operation of information quality improvement procedures is very rare.

- To confirm our quantitative analysis, we collect the interviewees' opinions concerning information accuracy and completeness. The interviewees stated that both accuracy and completeness of information are crucial in decision-making. Some interviewees emphasise the difficulty of improving information completeness since in certain circumstances it might be not easy to recognise complete information. However, all the interviewees confirm the importance of information accuracy and completeness in decision-making.

- Following the work of Lee (2004), we collect the possible reasons of triggering an information quality problem. Except some typical causes such as typing errors or delayed input, interviewees are mainly concerned with two causes: system design and information processing. Some interviewees emphasized that it is important to prevent information quality problems when designing the system. For example, the lack of constraint checking in the system can result in poor-quality data in the database. Some interviewees considered that information quality problems can be generated in the information processing or transferring procedure. Therefore it is critical to increase information quality awareness when we process or transfer the information.

6.4 Summary

This chapter first investigates the effects of information quality categories on decision quality. Three information quality categories are used as independent variables: intrinsic information quality, contextual information quality and representational information quality. Each information quality category is divided into two levels: high and low information quality. The design of the two levels is according to the measurement of the dimensions in the category. The dependent variable is inventory decision quality. As JIT and EOQ policies can be used to establish the best decision, the two inventory control policies are used to determine the decision quality. Accordingly, two scenarios are designed and implemented. After the experiment, two three-way ANOVAs are used to analyse the experimental results.

This research found that intrinsic information quality and contextual information quality are confirmed to be positively related to inventory decision quality in both scenarios. Therefore, intrinsic and contextual information quality are the main concerns in improving inventory decision quality. One interesting result we found is that in the JIT scenario, contextual and representational information quality jointly affect inventory decision quality, but in the EOQ scenario this joint effect appears to be non-significant. This nonconformity indicates that contextual information quality and representational information quality are positively related to inventory decision quality under certain circumstances. Therefore, if the financial cost of improving contextual information quality is extremely high, decision-makers could try to enhance representational information quality in order to improve contextual information quality. In addition, representational information quality was found to be non-significant in affecting inventory decision quality in both experiments. Thus, inventory decision quality may not be improved when decision-makers enhance representational information quality. A review of the findings on our hypothesis 1-6 is summarised in Table 39.

Hypotheses		p-value	Scenario 1
H1	Intrinsic information quality is positively associated with JIT decision quality.	<.001	Supported
H2	Contextual information quality is positively associated with JIT decision quality.	<.001	Supported
H3	Representational information quality is positively associated with JIT decision quality.	.306	Rejected
Hypotheses		p-value	Scenario 2
H4	Intrinsic information quality is positively associated with EOQ decision quality.	<.001	Supported
H5	Contextual information quality is positively associated with EOQ decision quality.	<.001	Supported
H6	Representational information quality is positively associated with EOQ decision quality.	.893	Rejected

Table 39: Summary of the experiment results

Instead of controlling information quality in the category level, we further control information quality by different dimensions in experiment 2. This experiment investigated the effects of accuracy, completeness and consistency of information on decision quality. Three significant effects are found: the main effect of accuracy, the main effect of completeness, and the interaction effect of accuracy and consistency. Thus there are at least three methods to improve decision quality. The first is increasing information accuracy. This requires the correction of inaccurate values and their representation in tangible form. The second method of improving decision quality is to enhance information completeness. This can be achieved by filling NULL values and including full meanings. The third method is to trade-off the improvement of accuracy and consistency. For example, when the information is complete and accurate, increasing consistency of information may improve the decision quality. However, when the information is complete but inaccurate, increasing consistency of information provides no benefit for decision-making. Another important finding is that the effect of consistency of information on decision quality is found to be non-significant, indicating that increasing consistency of information does not significantly affect decision quality. Decision-makers cannot expect to improve the decision quality only through enhancing consistency of information. We model the major findings in Figure 49.

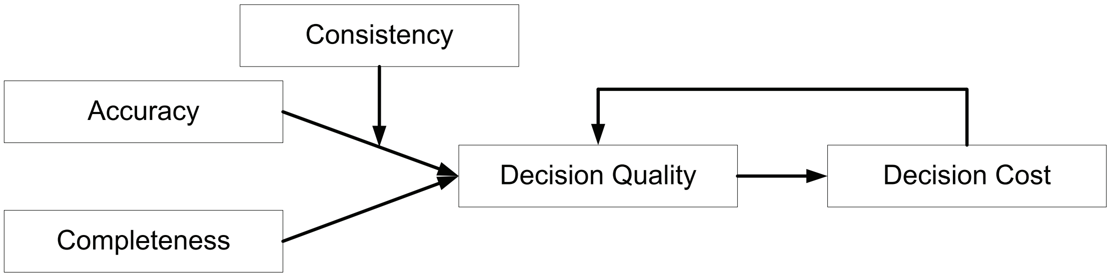


Figure 49: Effects of accuracy, completeness and consistency on decision quality

Figure 49 indicates that information accuracy and completeness are two crucial determinants in decision quality. This finding is also confirmed by experiment and interview in our workshop. The effects of accuracy can be intensified by the effect of consistency. As a surrogate measure of decision quality, less decision cost indicates higher decision quality.

After the data analysis, we carried out a workshop to confirm the results and further investigate how accuracy and completeness affect decision quality. In this workshop, we develop an experiment in which accuracy and completeness are measured by five levels. After collecting the data in the experiment, we performed a regression analysis between the two dimensions and the decision cost. As decision quality is indicated by the decision cost in the experiment, the analysis shows that we are able to improve decision quality when we increase accuracy or completeness. Besides the regression analysis, we carried out an interview to confirm the result and locate the root causes of information quality problems.

Chapter 7: Conclusion

In this chapter, we present our conclusions relating to each aspect of our work. In addition, we discuss the limitations of our finished work and the extensions for future research.

7.1 Summary

In this thesis, we investigated the relationship between information quality and decision-making. In detail, we refined this relationship to "information quality categories and decision quality" and "information quality dimensions and decision quality". The experimental methodology is used to investigate the relationships.

For the relationship of information quality categories and decision quality, we selected intrinsic, contextual and representational information quality as the categories. In our supply chain experiment, since the inventory policy JIT and EOQ can be used to determine the value of decision quality, we analysed the effects of information quality categories on the JIT and the EOQ decision quality. For the relationship between information quality dimensions and decision quality, we selected accuracy, completeness and consistency as the dimensions. Extending the experiment of information quality category, we use the decision cost to indicate the value of decision quality, as the decision cost can be used in more supply chain scenarios. Thus we analysed the effects of accuracy, completeness and consistency on inventory decision quality.

By refining information quality to categories and dimensions, we further investigated the relationship between information quality and decision-making. Our results clarified how different aspects of information quality affect decision-making. To confirm our results, we carried out a workshop to perform a further analysis and an interview regarding the analysis results.

Since information quality assessment is a crucial factor in investigating the relationship between information quality and decision-making, we proposed an information quality assessment framework that consists of two major components: validated dimensions and measurement methodology. To develop validated dimensions, we adopted 17 widely accepted information quality dimensions from previous literature. Based on these dimensions, we carried out a confirmatory factor analysis and a Cronbach calculation, by which we obtained a valid and reliable set of information quality dimensions. Using our validated dimensions, we proposed methodologies to assess information quality in the scenario of a single database and multiple databases.

In the scenario of a single database, the assessment approaches were proposed to respectively assess the quality of raw data and the quality of information products. Using the proposed approaches, we applied the assessment to an example dataset. This application not only validated the assessment approaches but also demonstrated information quality deficiencies in the assessed dataset. This result allows us to measure information quality in the experiments.

In the scenario of multiple databases, we proposed algorithms to determine the value of accuracy and completeness. The algorithms are derived by the relationships between data operations, information quality problems and information quality dimensions. To validate the algorithms, we carried out a simulation in the context of an information chain. By analysing the simulation results, we can control the level of accuracy and completeness by data operations. This result can be used in the experiment when multiple subjects play in the same supply chain.

To facilitate our investigations, we programmed 3 software-based experiments, 1 software prototype and 1 online survey for the assessment in a single database, and 1 simulation for the assessment between multiple databases. These research products have validated our models and provided us expected results. They also can be used in the future research and for practical applications.

7.2 Conclusion

In this section, we provide our results and conclusions from six aspects. The first result focuses on the contribution from studying the relationship between information quality and decision-making. The following three results are related to information quality assessment and measurement. Then we will discuss the findings from our literature review. Finally we will discuss the experiments we developed and their future usage.

The first result is that we investigated the effects of information quality categories and dimensions on decision quality. Since different organisations may consider information quality from different aspects, such as category or dimension, they expect to employ the result that is suitable for their specific concerns. Therefore we have analysed the effects of information quality from both category and dimension aspects. Along with this research objective, a set of hypotheses was proposed. To test these hypotheses, two experiments were developed in the context of supply chain management. Analysing the experimental results, we have obtained the following conclusions.

For the effects of information quality categories, the effects of intrinsic information quality and contextual information quality on decision quality were found to be significant, indicating that decision quality can be improved by increasing intrinsic and contextual information quality. When information used for decision-making was assessed as of high intrinsic and contextual information quality, decision-makers could expect a high quality decision and vice versa. The effect of representational information quality on decision quality appeared to be non-significant. That means improving representational information quality does not significantly affect decision quality.

Therefore, intrinsic information quality and contextual information quality are the main concerns when we seek to improve information quality. However, we also found that the interaction effect of contextual information quality and representational information quality was significant, indicating that contextual information quality and representational information quality may be mutually intensified. This means that if under some circumstances, such as high financial cost, we are not able to improve contextual information quality, we could try to improve representational information quality, and so through this, to improve contextual information quality.

For the effects of information quality dimensions, it was found that the effects of accuracy and completeness on decision quality are significant. That means increasing accuracy and completeness of information could significantly improve the decision quality. Thus decision-makers could expect to make better decisions when using high accuracy and high completeness information. Furthermore, there was no significant difference in low consistency and high consistency on decision quality. It implies that, regardless of information completeness and accuracy, the improvement of information consistency could not significantly enhance decision quality. Therefore, accuracy and completeness are the main concerns when we improve information quality. However, the interaction effect of accuracy and consistency of information was found to be significant. That means accuracy and consistency of information jointly affect decision quality. Although consistency alone does not significantly affect decision quality, it may improve decision quality by combining accuracy and consistency. Specifically, when information is assessed to be complete but inaccurate, a decision-maker using inconsistent information performed better than one using consistent information. A potential explanation is that information inconsistency may force decision-makers to consider the given information more carefully in order to try to compensate for the inaccurate values. We organise the results of main effects in Table 40.

Hypothesis	Result
Intrinsic information quality is positively related to decision quality.	Supported
Contextual information quality is positively related to decision quality.	Supported
Representational information quality is positively related to decision quality.	Rejected
Accuracy is positively related to decision quality.	Supported
Completeness is positively related to decision quality.	Supported
Consistency is positively related to decision quality.	Rejected

Table 40: Results of the effects of information quality on decision quality

Relating our research to previous studies such as Chengalur-Smith et al. (1999) and Fisher et al. (2003), our result supported their research findings and further analyses the effects of different information quality aspects on decision-making.

For information system practitioners, it is important to enhance the awareness of information quality and recognise information quality as a fundamental element in organisational decision-making. Poor information quality may lead to various organisational losses such as losing customers and missing opportunities (Keller and Staelin 1987, Ballou and Pazer 1990, Raghunathan 1999, Chengalur-Smith et al. 1999, Fisher et al 2003). Therefore to enhance the quality of decision-making, one feasible solution is to improve the quality of information that is used for decision-making.

From our results, top management or decision-makers are able to observe the relationship between information quality and decision quality. Therefore if an organisation is considering whether to implement information quality management, our result can demonstrate the importance of information quality and how information quality affects decision-making.

In previous information quality improvement, top management or decision-makers would improve every aspect of information quality. However, our results indicate that not all the aspects of information quality are equally effective for improving decision quality. In the context of supply chain management, the representational information quality and information consistency are found to be not as effective as other aspects of information quality. We are therefore able to prioritise the procedure of information quality improvement. For example, when organisations attempt to improve decision quality through improving information quality, decision-makers can decide to pay little or even no attention to the improvement of representational information quality or information consistency.

In certain database, it may require human to perform or assist information quality improvement. Therefore information quality improvement can be very costly and time-consuming. If we can reduce some aspects of information quality we need improve, it will directly reduce the cost and the time for information quality improvement. Software designers can also benefit from our result, which indicates that it is valuable to allow users to prioritise the improvement of different aspects of information quality.

The second result is that we further confirmed a set of valid and reliable information quality dimensions. Based on the literature review, mixed results were found in deriving information quality dimensions. Two prominent works were those of Wang and Strong (1996) and Lee et al. (2002). In Wang and Strong (1996)'s work, they conducted an exploratory factor analysis on 118 information quality metrics and derived 15 information quality dimensions. The loading of each dimension was greater than 0.5. Their results supported the construct validity, which includes convergent and discriminant validity. On the other hand Lee et al. (2002) re-examined the correlations of the 15 information quality dimensions. They found strong correlations among the information quality dimensions, indicating the weak form of discriminant validity. To further confirm the correlations between information quality dimensions, we carried out a confirmatory factor analysis and a Cronbach calculation with 316 subjects. Our results generated 9 constructs for measuring information quality: accessibility, security, relevancy, integrity, timeliness, understandability, reliability, ease of manipulation and objectivity. In the above dimensions, we considered integrity as a category, including accuracy, completeness and consistency. The results of our factor analysis and Cronbach calculation are summarised in Table 41.

Dimension	Number of Items	factor analysis	Cronbach Alpha
Accessibility	4	> 0.600	0.91
Security	3	> 0.770	0.88
Relevancy	4	> 0.685	0.81
Integrity	10	> 0.621	0.89
Timeliness	3	> 0.795	0.92
Understandability	5	> 0.692	0.86
Reliability	6	> 0.627	0.91
Ease of Manipulation	3	> 0.604	0.83
Objectivity	4	> 0.692	0.77

Table 41: A valid and reliable set of information quality dimensions

According to our review, there is so far no related works that use the confirmatory factor analysis to validate information quality dimensions. Based on previous works, our work has rigorously confirmed and validated the dependency of information quality dimensions. Using the validated dimensions, following researchers can more confidently develop the measurement of each dimension. Consequently our work increases the confidence of using such information quality dimensions. Moreover, instead of spending time on validating the dimensions, researchers and practitioners can directly employ the validated dimensions to carry out their assessments. Therefore our work has clarified the dimension-validation part for information quality assessment.

The third result is that we proposed a practical methodology for assessing information quality in the scenario of a single database. Based on the review of objective and subjective assessment, we separate the measuring objects into raw data stored in the database and information products delivered to information consumers. Driven by the idea of "who uses which dimensions to measure what", we proposed two measuring approaches, one for raw data and the other for information products. Using these two approaches, we applied our methodology to a database in the Teradata system. After measuring the quality of raw data in ua_samsclub dataset, we found 119765 out of 5668375 data records contained information quality problems. Once these poor quality data records are manufactured to information products, the quality of those information products would also be poor. When measuring information products, we carried out the assessment with an online survey. The results also indicated pervasive information quality deficiencies. Therefore information quality problems are confirmed from both a database and contextual scenario. In addition, this application validated our assessment methodology from a practical point of view.

In previous information quality assessment, most methodologies only focus on a general approach to assess information quality. However, in practice organisations always meet assessment difficulties when they are dealing with different measuring objects such as raw data and information products. In order to solve this problem, our result concludes a practical framework that allows organisations to carry out assessment with different measuring objects. This framework has eliminated the confusion when simultaneously implementing the assessment to raw data and information products. It therefore clarified the measuring target and corresponding assessment approach. Furthermore, compare to other frameworks that require software engineers to understand information quality theory, our framework allows software engineers to only follow a step-by-step procedure to build an application of information quality assessment. This will directly reduce the working load of software engineers. In turn, it will save the time and budget for the project.

The fourth result is that we proposed algorithms to assess information quality dimensions in the scenario of multiple databases. In the context of an information chain, algorithms for measuring accuracy and completeness are derived by the relationship between data operations, information quality problems and information quality dimensions. “Create”, “update” and “delete” are used as the data operations. Each operation was connected to the possible information quality problems. For example creating new data in the reference database will generate missing data problems in the operational database. Based on the definitions of information quality dimensions, we mapped the information quality problems onto information quality dimensions. For example, missing data problems were mapped to completeness. Using the above procedure, we are able to use data operations to determine the values of information quality dimensions.

In the end, we applied the algorithms to an information chain simulation. The simulation results have shown that when the error rate of data operation is fixed, accuracy and completeness are negatively related to assessment time. Based on random selections of data operations, we can determine the optimal error rate of data operations and the correct time for using data. In addition, accuracy decreases faster than completeness in the same refresh period. Therefore to improve information quality in the information chain, we need to focus more attention on the improvement of accuracy.

When managing an information chain, organisations may need to deal with multiple databases. These databases are connected and sharing data with each other. If we only consider the assessment in a single database, we may obtain the erroneous assessment result. For example, database M and N are the two connected databases in an information chain. They are sharing certain amount of data. When database M has created a new data in the shared data and the data alignment is not carried out yet, the data in Database N now becomes incomplete. However if we only focus on database N and ignore database M, we would mistakenly consider that the data in Database N are complete. Therefore it is crucial to consider multiple databases when we carry out information quality assessment across databases.

As little research has been done to solve the assessment issue in multiple databases, our result has provided a novel algorithm that allows organisations to assess information quality by considering other databases. This algorithm solves the assessment problems when the data operations simultaneously occurred in related databases. It also helps database engineers to locate the information quality problems amongst different databases. Based on our algorithms, practitioners can implement a piece of software to supervise the changing of information quality in the information chain. It also can be used to decide when we can use the information within an acceptable quality level. In addition, under a required information quality level, the algorithms can help database or software designers to determine how long the data alignment should be. Furthermore, the algorithms may be commercialised through information quality software.

The fifth result is our literature review. In our review, we primarily analysed three aspects of information quality research: contextual information quality, information quality management and information quality assessment. In the contextual information quality section, we analysed two typical contexts in information quality research, which are information systems and decision-making. First, we positioned information quality development and showed the relationship between information quality research and information system research. Then we analysed the effects of information quality on decision-making. Our review showed that different factors need to be considered when investigating the relationship between information quality and decision-making.

In the information quality management section, our review showed that information quality management has been merging with information management, quality management and knowledge management. Since information quality management research remains its infant stage, further knowledge can be adopted from other fields.

Therefore, our review indicated the future research trend for information quality management.

In the information quality assessment section, we reviewed information quality problems, dimensions and assessment methodologies. The review facilitated the understanding of information quality assessment and concluded with open questions for future research. These open questions are listed in Table 42 of Section 7.4

From our review, new researchers, who intend to perform information quality research, can obtain an insight view of this research area. The review has also provided a research agenda (Table 42), which highlights the open research questions and research steams for future research. With this agenda, researchers can easily locate the research gaps and contribute knowledge to the research domain of information quality.

The final results are the software-based experiments we developed in this research. The software was used to examine the relationship between information quality and decision quality. Our result successfully showed the relationship between information quality and decision quality. Based on our software, more information quality dimensions and influencing factors can be introduced into the experiment. This means that our software can be reused to carry out more experiments for future research. In addition, as it is a practical demonstration, our software can be used as a teaching tool to show the impact of information quality.

Reviewing the works in our research, this study has reviewed the state-of-art of information quality research. Based on the review, we proposed our major research question that is “how does information quality affect decision-making”. Using our proposed experiments, we have clarified the relationship between information quality and decision-making from information quality category and dimension perspective. We also bridged the gaps in information quality assessment and proposed a validated and practical assessment framework. In summary, this research has presented feasible guidance for information quality management, which can be applied to both academia and industry.

7.3 Critical Review and Limitation

In this thesis, we have investigated the effects of information quality on decision-making. When designing the experiment, one limitation is information complexity. In our experiment, we provide relatively simplified information in the decision-making process. Thus we limit our decision-relevant information to a low complexity level. As the information complexity increases in the experiment, representational information quality and consistency of information may become more important, which in turn may significantly influence decision quality.

Since our experiment is conducted in the domain of supply chain management, we limit our conclusions mainly in this domain. However, since supply chain management is a typical business domain, we may apply our results to other business domains. Besides, our results have been confirmed by different scenarios, different experiments and qualitative studies. The confirmation also increases the external validity to support that the results can be applied in other domains.

When our experiment is carried out by the subjects who are experienced in the domain of supply chain management, subjects might bring a biased view towards the information before the information is even presented. As discussed in Chapter 4, we control the experience of the subjects in a novice level.

Another limitation of this study is the correlation between information quality dimensions. Since information quality is a relative rather than an absolute concept (Ballou and Pazer 1985), the dimensions of information quality may be defined differently in different organisational contexts. Therefore, the correlations of information quality dimensions may be higher than the correlations in this study. For example, Olson (2003) considers consistency as a part of accuracy. Thus when we consider consistency from format perspective, increasing accuracy may not improve decision quality. When we consider consistency from value perspective, increasing consistency may significantly improve decision quality.

Furthermore, the measurement of information quality is also considered as a limitation. In order to adapt our experiment to factorial design, we limit the measurement of information quality categories and dimensions to two levels. Although more levels or percentages can be used as the method for measuring information quality, current information quality research is still confronting measurement difficulties such as how to divide equal interval levels and how to determine the weight of information in the percentage measurement. Therefore only high level and low level of information quality are used in experiment 1 and 2.

In experiment 2, when designing the low level of information consistency, we introduced another piece of information to increase the inconsistency. However, this inconsistent information may indirectly provide more information to the decision-makers. Therefore we consider the design of consistency as a limitation.

In the validation of information quality assessment in a single database, the identification of information quality problems is dependent on the specific business rules of an organisation. In our assessment, we merely intend to demonstrate an assessment example rather than trying to cover all the information quality problems in the Teradata database. Therefore other information quality problems may be also found in this database.

In the validation of information quality assessment in multiple databases, mapping among data operations, information quality problems and information quality dimensions may vary between organisations. Using our mappings, we propose two algorithms to measure accuracy and completeness. The algorithms are limited to the definition of information quality dimensions. If the definitions of accuracy and completeness are different from this research, the linkage between information quality problems and information quality dimensions would be also different. Therefore, in practical application, organisations could create their own mappings according to their business rules and specific organisational context.

From the methodology perspective, we address the following limitations for the survey research and experimental research in our study. In our assessment framework, we use survey research methodology to validate information quality dimensions. Limitations of survey research are whether the respondents accurately reflect their viewpoints and whether their viewpoints accurately reflect the real world. Based on previous literature, this study moderates this limitation by adopting the validated survey and providing reliability and validity analysis of our survey.

Limitations of the laboratory experiment are mainly focused on the existence of the identified relationship due to the over-simplified experimental scenario and the effects of exogenous variables in the real world. In order to alleviate the over-simplified experimental scenario, we adapt the Beer Game as our experimental context, which is already considered to be a mature game in management research. In order to lighten the effects of exogenous variables, we perform a systematic review on the exogenous variables that may influence the relationship between information quality and decision-making. Based on our review, we propose a framework to control the effects of exogenous variables. This framework is then applied in our experiments.

7.4 Future Work

This research has investigated the effects of intrinsic, contextual and representational information quality on decision-making. In future research, the measurement of each information quality category could be further improved. For example, it would be valuable to develop a comprehensive and validated framework to measure information quality categories. In addition to the above three categories, studying the effect of accessibility information quality on decision-making is also considered as a stimulus for future work. However, researchers firstly need to differentiate between perceived system accessibility and perceived information accessibility.

After investigating the effects of accuracy, completeness and consistency on decision-making, in the future research more information quality dimensions can be included as the variables that influence decision-making. In order to determine the focus of information quality improvement, it is valuable to further compare the importance of the information quality dimensions in various decision contexts. For example, in real-time decision-making, timeliness may be considered as the most important determinant in information quality improvement. Once the importance of information quality dimensions has been clarified, it would then be possible to develop an effective strategy for information quality improvement. For example, under a fixed financial budget, we would be able to find out the key information quality dimensions we need to focus on.

The further improvement for the experiment can focus on a group competition. That means each supply chain is played by a group of participants. Each component of the supply chain such as distributor or manufacturer is played by one participant. Thus different groups are able to compete for their performances by comparing the benefit in a determined period. To facilitate this extension, we have provided the measurement of information quality in Chapter 5. In addition, we can extend the improvement by including more information quality dimensions. It would enable us to observe how different dimensions perform in a practical application. In practice, organisations can extend our experiment to a training tool that can demonstrate the effects of information quality in an organisational context.

In order to assess information quality, this research has contributed a practical methodology to assess the quality of raw data and information products. However, in some information manufacturing systems, raw data are firstly assembled into component data and finally transformed into information products. These component data are often operated and delivered amongst different processes. Between these processes, the quality of component data can be augmented, diminished or kept unchanged. The shifting quality of component data will directly influence the quality of information products. Therefore, future research can focus on assessing the quality of component data. Through controlling the quality of component data, we will be able to control the quality of information products.

In order to assess information quality in multiple databases, this research has proposed algorithms to assess information accuracy and completeness. As information quality is also measured by other dimensions, future research could focus on deriving the assessment algorithms of other dimensions such as timeliness and consistency. Once the algorithms of information quality dimensions are determined, researchers could further investigate the dependencies between these dimensions. Based on the cost and dependency analysis, it would be possible to find an optimal strategy for information quality improvement.

After information quality assessment, information quality analysis and improvement are the important following phases in information quality management. Therefore future research could focus on how to analyse the assessment results and improve current information quality. In information quality analysis, researchers can investigate how to coordinate the objective and subjective assessment results and how to locate the sources of information quality deficiencies. Using the analysis results, it will be valuable to further develop a practical framework for information quality improvement.

Finally, based on our review, we propose a research agenda including 15 questions for future research. These research questions are listed in Table 42.

Research Question 1: How to assess information quality effectively?
Research Question 2: How to identify potential information quality problems?
Research Question 3: How to define and select information quality dimensions?
Research Question 4: What is the relationship between information quality problems and information quality dimensions?
Research Question 5: How are information quality dimensions mutually dependent? And how to deal with the dependencies of information quality dimensions?
Research Question 6: How to choose the most suitable information quality assessment methodology for organisations?
Research Question 7: How to manage information quality in organisations?
Research Question 8: How to analyse cost and benefit of information quality management?
Research Question 9: How to evaluate the maturity of information quality management?
Research Question 10: How to deploy information quality management in organisations?
Research Question 11: How to build information quality awareness and information quality culture in organisations?
Research Question 12: What is the relationship between information quality and application contexts?
Research Question 13: How does information quality impact application contexts?
Research Question 14: What is the relationship between information quality research and information system research?
Research Question 15: How to control extraneous variables in information quality experiment?

Table 42: Research agenda for future information quality research

This research agenda has addressed the open issues related to information quality assessment and management. As the research questions stem from a critical review, researchers in the future are able to focus on significant issues and contribute knowledge to the information quality community.

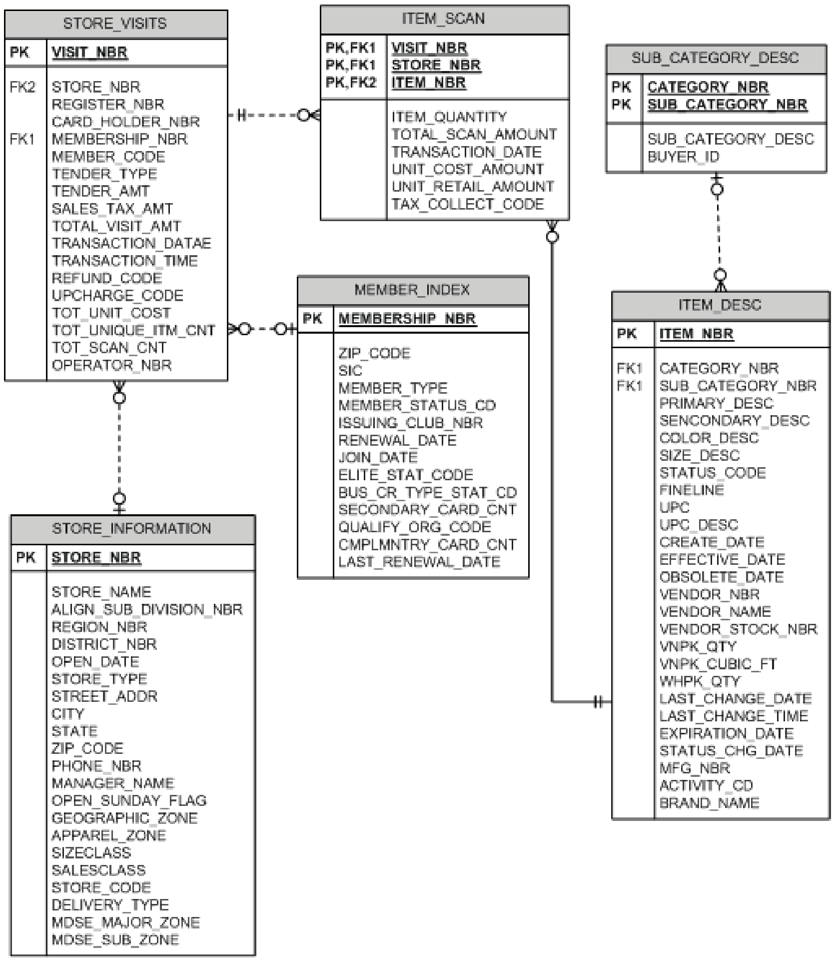
Appendix

Survey for Validating Information Quality Dimensions

	Not Important At All					Average					Extremely Important
1. The information that is **accessible** is	0	1	2	3	4	5	6	7	8	9	10
2. The information that is **secure** is	0	1	2	3	4	5	6	7	8	9	10
3. The information that is **useful** is	0	1	2	3	4	5	6	7	8	9	10
4. The information that is **correct** is	0	1	2	3	4	5	6	7	8	9	10
5. The information that is **sufficient** is	0	1	2	3	4	5	6	7	8	9	10
6. The information that is **current** is	0	1	2	3	4	5	6	7	8	9	10
7. The information that is **impartial** is	0	1	2	3	4	5	6	7	8	9	10
8. The information that is **concise** is	0	1	2	3	4	5	6	7	8	9	10
9. The information that is **precise** is	0	1	2	3	4	5	6	7	8	9	10
10. The information that is **believable** is	0	1	2	3	4	5	6	7	8	9	10
11. The information that is **timely** is	0	1	2	3	4	5	6	7	8	9	10
12. The information that is **relevant** is	0	1	2	3	4	5	6	7	8	9	10
13. The information that is **accurate** is	0	1	2	3	4	5	6	7	8	9	10
14. The information that is **complete** is	0	1	2	3	4	5	6	7	8	9	10
15. The information that is **up to date** is	0	1	2	3	4	5	6	7	8	9	10
16. The information that is **available** is	0	1	2	3	4	5	6	7	8	9	10
17. The information that is **obtainable** is	0	1	2	3	4	5	6	7	8	9	10
18. The information that is **protected** is	0	1	2	3	4	5	6	7	8	9	10
19. The information that is **free of error** is	0	1	2	3	4	5	6	7	8	9	10
20. The information that is **reliable** is	0	1	2	3	4	5	6	7	8	9	10
21. The information that is **detailed** is	0	1	2	3	4	5	6	7	8	9	10
22. The information that is **unbiased** is	0	1	2	3	4	5	6	7	8	9	10
23. The information that is **compact** is	0	1	2	3	4	5	6	7	8	9	10
24. The information that is **dependable** is	0	1	2	3	4	5	6	7	8	9	10
25. The information that is **trustworthy** is	0	1	2	3	4	5	6	7	8	9	10
26. The information that is **readable** is	0	1	2	3	4	5	6	7	8	9	10
27. The information that is **credible** is	0	1	2	3	4	5	6	7	8	9	10
28. The information that is **objective** is	0	1	2	3	4	5	6	7	8	9	10
29. The information that is **helpful** is	0	1	2	3	4	5	6	7	8	9	10
30. The information that is **retrievable** is	0	1	2	3	4	5	6	7	8	9	10
31. The information that is **applicable** is	0	1	2	3	4	5	6	7	8	9	10

32. The information that contains **consistent meaning** is	0	1	2	3	4	5	6	7	8	9	10
33. The information that is **interpretable** is	0	1	2	3	4	5	6	7	8	9	10
34. The information that is **from good sources** is	0	1	2	3	4	5	6	7	8	9	10
35. The information that is **easy to manipulate** is	0	1	2	3	4	5	6	7	8	9	10
36. The information that is **easy to understand** is	0	1	2	3	4	5	6	7	8	9	10
37. The information that is **presented in the same format** is	0	1	2	3	4	5	6	7	8	9	10
38. The information that is **well referenced** is	0	1	2	3	4	5	6	7	8	9	10
39. The information that is **easy to identify the key point** is	0	1	2	3	4	5	6	7	8	9	10
40. The information that **includes all necessary values** is	0	1	2	3	4	5	6	7	8	9	10
41. The information that contains **consistent structure** is	0	1	2	3	4	5	6	7	8	9	10
42. The information that is **without inappropriate language and symbol** is	0	1	2	3	4	5	6	7	8	9	10
43. The information that is of **good reputation** is	0	1	2	3	4	5	6	7	8	9	10
44. The information that is **easy to aggregate** is	0	1	2	3	4	5	6	7	8	9	10
45. The information that is **easy to comprehend** is	0	1	2	3	4	5	6	7	8	9	10
46. The information that is **authorized to access** is	0	1	2	3	4	5	6	7	8	9	10
47. The information that is **comprehensive** is	0	1	2	3	4	5	6	7	8	9	10
48. The information that is **delivered on time** is	0	1	2	3	4	5	6	7	8	9	10
49. The information that was **easy to combine** was	0	1	2	3	4	5	6	7	8	9	10
50. The information that is **based on facts** is	0	1	2	3	4	5	6	7	8	9	10

ERD Diagram of Sam's Club Database in Teradata System



Reference

Abelson, R.P. and Levi, A. (1985), Decision making and decision theory. *In Lindzey G. and Aronson E. (Eds.), The Handbook of Social Psychology*, Random House, New York, USA, pp. 231-309.

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

Ahituv, N. (1980), A systematic approach toward assessing the value of an information system, *MIS Quarterly*, 4(4), pp. 61-75.

Ahituv, N., Igbaria, M. and Stella, A. (1998), The effects of time pressure and completeness of information on decision making, *Journal of Management Information Systems*, 15(2), pp. 153-172.

Almeida, A. and Marreiros, G. (2005), A scheduling model based on group decision support, *IEEE International Conference on Systems, Man and Cybernetics*, pp. 1180 - 1187.

Amicis, F.D., Baron, E.D., and Batini, C. (2006), An analytical framework to analyze dependencies among data quality dimensions, *11th International Conference on Information Quality,* Boston, Massachusetts, USA.

Amicis, F.D. and Batini, C. (2004), A methodology for data quality assessment on financial data, *Studies in Communication Sciences*, 4(2), pp. 115-137.

Amason, A.C. (1996), Distinguishing the effects of functional and dysfunctional conflict on strategic decision making: resolving a paradox for top management teams, *Academy Management Journal*, 39(1), pp. 123–148.

Anastasi, A. (1988), *Psychological Testing*. 6th edition. Macmillan publication, New York.

Bailey, K. (1978), *Methods of Social Research,* 3rd edition, The Free Press, New

York.

Ballinger, T.P., Palumbo, M.G. and Wilcox, N.T. (2003), Precautionary saving and social learning across generations: an experiment, *The Economic Journal*, 113(490), pp. 920-947.

Ballou, D.P. and Pazer, H.L. (1990), A framework for the analysis of error conjunctive, multi-criteria, satisfying decision processes, *Decision Sciences*, 21(4), pp. 752-770.

Ballou, D.P. and Pazer, H.L. (1985), Modeling data and process quality in multi-input, multi-output information systems, *Management Science*, 31(2), pp. 150-162.

Ballou, D.P. and Pazer, H.L. (1995), Designing information systems to optimize the accuracy-timeliness tradeoff, *Information Systems Research*, 6(1), pp. 51-72.

Ballou, D.P. and Tayi, G.K. (1989), Methodology for allocating resources for data quality enhancement, *Communications of the ACM*, 32(3), pp. 320-329.

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

Ballou, D.P. and Pazer, H.L. (2003), Modeling completeness versus consistency tradeoffs in information decision contexts. *IEEE Transactions on Knowledge and Data Engineering*, 15(1), pp. 240-243.

Batini, C. and Scannapieco, M. (2006), *Data Quality, Concepts, Methodologies and Techniques*, Publisher: Springer, Berlin, Germany.

Beauclair, R. (1987), An experimental study of the effects of specific GDSS applications on small group decision making, *Working Paper, University of Louisville*.

Belardo, S. and Pazer, H. (1995), A framework for analyzing the information monitoring and decision support system investment tradeoff dilemma: an application to crisis management, *IEEE Transactions on Engineering Management*, 42(4), pp. 352-358.

Benbasat, I.G., Goldstein, D.K. and Mead, M. (1987), The case research strategy in studies of information systems, *MIS Quarterly*, 11(3), pp. 369-386.

Berndt, D.J., Fisher, J.W., Hevner, A.R. and Studnicki, J. (2001), Healthcare data warehousing and quality assurance, *Computer*, 36(12), pp. 56-65.

Blalock, H.M. (1963), Correlated independent variables: the problem of multicollinearity, *Social Forces*, 42(2), pp. 233-237.

Bovee, M., Srivastava, R.P., Mak B. (2003), A conceptual framework and belief-function approach to assessing overall information quality, *International journal of intelligent systems*, 18(1), pp. 51-74.

Brien J.A. (1991), *Introduction to information systems in business management*, 6th edition. Boston, Massachusetts, USA.

Brodie, M. L. (1980), Data quality in information systems. *Information and Management*, 3(6), pp. 245-258.

Bronner, R. (1982), *Decision making under time pressure*, Lexington Books, Lexington, Massachusetts, USA.

Bruggen G. V., Smidts, A. and Wierenga, B. (1998), Improving decision making by means of a marketing decision support system, *Management Science*, 44(5), pp. 645-658.

Buckland, M.K. and Florian, D. (1991), Expertise, task complexity, and the role of intelligent information systems, *Journal of the American Society for Information Science*, 42(9), pp. 635-643.

Burleson, B.R., Levine, B.J. and Samter, W. (1984), Decision making procedure and decision quality. *Human Communication Research*, 10, pp. 557–574.

Caldwell, D.F. and O'Reilly, C.A. (1985), The impact of information on job choices and turnover, *The Academy of Management Journal*, 28(4), pp. 934-943.

Campbell, D.J. (1988), Task complexity: a review and analysis, *Academy of Management Review*, 13(1), pp. 40-52.

Campbell, D.T. and Stanley, J. (1963), *Experimental and quasi-experimental designs for research*. Houghton-Mifflin, Boston, Massachusetts, USA.

Cappiello, C., Francalanci, C. and Pernici, B. (2004), Time-related factors of data quality in multi-channel information systems, *Journal of Management Information Systems*, 20(3), pp. 71-91.

Carmines, E.G. and Zeller, R.A. (1991), *Reliability and validity assessment*. Sage Publications, Newbury Park, California, USA.

Carter, L.F. (1947), An experiment on the design of tables and graphs used for presenting numerical data, *Journal of Applied Psychology*, 31, pp. 640-650.

Chengalur-Smith, I.N., Ballou, D.P. and Pazer, H.L. (1999), The impact of data quality information on decision making: an exploratory analysis. *IEEE Transactions on Knowledge and Data Engineering*, 11(6), pp. 853-864.

Chewning, E.C., Jr., and Harrell, A.M. (1990), The effect of information load on decision makers' cue utilization levels and decision quality in a financial distress decision task. *Accounting, Organizations and Society*, 15(6), pp. 527-542.

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

Citera, M. (1998), Distributed teamwork: the impact of communication media on influence and decision quality, *Journal of American Society for Information Science*, 49(9), pp. 792-800.

Cushing, B.E. (1974), A mathematical approach to the analysis and design of internal control systems, *The Accounting Review*, 49(1), pp. 24-41.

David, C. (2004), *The cheating culture*, Harvest Books, Publisher: Harcourt, New York.

Davis, G.D. (1974), *Management information systems*: *conceptual foundations, structure and development*, McGraw Hill, New York.

DeLone, W.H. and McLean, E.R. (1992), Information system success: the quest for dependent variables, *Information System Research*, 3(1), pp. 60-96.

Deming, W.E. (1982), *Out of the crisis,* MIT Press, Boston, Massachusetts, USA.

Dickson, G., Senn, J. and Chervany, N. (1977), Research in management information systems: the Minnesota experiments, *Management Science*, 23(9), pp. 913-923.

English, L.P. (1999), *Improving data warehouse and business information quality: methods for reducing costs and increasing profits,* Publisher: Wiley, New York.

Eppler, M. J., (2006), *Managing Information Quality*, 2nd edition, Publisher: Springer. Berlin, Germany.

Eppler, M. and Helfert, M. (2004), A classification and analysis of data quality costs, *9th International Conference on Information Quality,* Boston, Massachusetts, USA.

Eppler, M.J. and Mengis, J. (2003), The concept of information overload: a review of literature from organization science, accounting, marketing, MIS, and related disciplines, *The Information Society. An International Journal*, 20(5), pp. 1-20.

Fazel, F. (1997), A comparative analysis of inventory cost of JIT and EOQ, *International Journal of Physical Distribution and Logistics Management*, 27(8), pp. 496–505.

Feliciano, G.D., Powers, R.D. and Bryant, E.K. (1963), The presentation of statistical information, *Audio visual Communication Review*, 11(13), pp. 32-39.

Field, A, and Hole, G. (2003), *How to design and report experiments*, Sage publication. London, UK.

Fisher, C.W., Chengalur-Smith I. and Ballou, D.P. (2003), The impact of experience and time on the use of data quality information in decision making, *Information Systems Research*, 14(2), pp.170-188.

Fisher, C.W. and Kingma, B.R. (2001), Criticality of data quality as exemplified in two disasters, *Information & Management*, 39(2), pp 109-116.

Fisher, C.W., Lauría, E., Chengalur-Smith, I. and Wang, R. (2006), *Introduction to information quality*, MITIQ Publication. Boston, Massachusetts, USA.

Fox, C., Levitin, A. and Redman, T. (1994), The notion of data and its quality dimensions, *Information Processing and Management*, 30(1), pp. 9-19.

Galbraith, J.R. (1974). Organization design: an information processing view. *Interfaces*, 4(3), pp. 28-36.

Galliers, R.D. (1991), Choosing information systems research approaches, *in: Galliers, R.D. (Ed) Information Systems Research: Issues, Methods and Practical Guidelines, Alfred Waller: Henley-on-Thames*, pp. 144-162.

Garvin, D. A. (1988), *Managing quality*. Free Press. New York, USA.

Ge M. and Helfert M. (2009), Information quality research: current state and future research agenda, *China Science & Technology Resources Review*, 40 (1), pp. 5-16.

Ge M. and Helfert M. (2008), Effects of information quality on inventory management. *International Journal of Information Quality*, 2(2), pp. 176-191.

Ge M. and Helfert M. (2008), Data and information quality assessment in information manufacturing system, *11th International Conference on Business Information Systems*, Innsbruck, Austria, 7, pp. 380-389.

Ge M. and Helfert M. (2008), Modeling data quality in information chain, *International Conference on Business Innovation and Information Technology,* Dublin, Ireland.

Ge M. and Helfert M. (2007), Develop a research agenda: a review of information quality research, *12th International Conference on Information Quality,* Boston, Massachusetts, USA.

Ge M. and Helfert M. (2007), A framework for information quality assessment, *2nd China-Ireland International Conference on Information and Communications Technologies,* Dublin, Ireland, pp. 951-959.

Ge M. and Helfert M. (2007), A framework to address effects of information quality on decision quality, *18th Annual Conference International Information Management Association,* Beijing, China.

Ge M. and Helfert M. (2007), Information quality research in information system and decision making, *IADIS Multi Conference on Computer Science and Information Systems,* Lisbon, Portugal.

Ge M. and Helfert M. (2007), Research design to evaluate information quality effects on decision making, *8th IBIMA Conference on Information Management in the Networked Economy,* Dublin, Ireland.

Ge M. and Helfert M. (2007), A review of factors influencing decision making scenarios: applied to an information quality experiment, *12th Annual UKAIS Conference*, Manchester, UK.

Ge M. and Helfert M. (2007), A theoretical model to explain effects of information quality awareness on decision making, *9th International Conference on Enterprise Information Systems,* Funchal, Madeira, Portugal.

Ge M. and Helfert M. (2006), Assessing information quality in group decision making, *7th IBIMA Conference Internet & Information Systems in the Digital Age,* Brescia, Italy.

Ge M. and Helfert M. (2006). A framework to assess decision quality using information quality dimensions, *11th International Conference on Information Quality,* Boston, Massachusetts, USA.

George, J.F., Easton, G.K., Nunamaker, J.F. and Northcraft, G.B. (1988), A study of collaborative group work and without computer-based support, *working paper, University of Arizona*.

Gertz, M., Ozsu, T., Saake, G., and Sattler, K. (2004), Report on Dagstuhl seminar data quality on the web, *SIGMOD Report*. 33(1).

Ghani, J. A. (1981). The effects of information representation and modification of decision performance, *Ph.D. dissertation,* University of Pennsylvania, USA.

Gilmore, H.L. (1974), Product conformance cost, *Quality Progress*, 7, pp. 16-19.

Gonzalez, C. and Kasper, G.M. (1997), Animation in user interfaces designed for decision support systems: the effects of image abstraction, transition, and interactivity on decision quality, *Decision Sciences*, 28(4), pp. 793-823.

Goodhue, D.L. (1995), Understanding user evaluations of information systems. *Management Science*, 41(12), pp. 1827-1844.

Gravetter, F. J., and Wallnau, L. B. (2000). *Statistics for the behavioral sciences*, 5th edition, Wadsworth Publishing, Belmont, California, USA.

Grinnell, R. Jr. (1993), *Social work, research and evaluation,* 4th edition, Publisher: Peacock, Illinois, USA

Gronroos, C. (1983) Strategic management and marketing in the service sector. *Marketing Science Institute*. Massachusetts, USA.

Hall, R.W. (1983). *Zero Inventories*, Dow Jones-Irwin Press, Homewood, Illinois, USA.

Helfert, M. (2001), Managing and measuring data quality in data warehousing, *Proceedings of the World Multiconference on Systemics, Cybernetics and Informatics*, Orlando, Florida, USA, pp. 55-65.

Helfert, M. and Heinrich B. (2003), Analyzing data quality investments in CRM: a model-based approach, *8th International Conference on Information Quality,* Boston, Massachusetts, USA.

Hilton, R.W. (1979), The determinants of cost information value: an illustrative analysis, *Journal of Accounting Research*, 17(2), pp. 411-435.

Hogan, T.P. (2003). *Psychological testing: a practical introduction*. Wiley, New York.

Huang, K.T., Lee, Y., Wang, R.Y. (1999), *Quality information and knowledge management*, Publisher: Prentice Hall. New Jersey, USA.

Ives, B. (1982). Graphical user interfaces for business information systems. *MIS Quarterly*, 6(4), pp. 15-48.

Jacoby, J. (1977), Information load and decision quality: some contested issues, *Journal of Marketing Research*, 14(4), pp. 569-573.

Janis I. and Mann L. (1977), *Decision making: a psychological analysis of conflict, choice and commitments*, The Free Press. New York, USA.

Janson, M. (1988), Data quality: the Achilles heel of end-user computing, *OMEGA: International Journal of Management Science*, 16(5), pp. 491-502.

Jarvenpaa, S.L. (1989). The effect of task demands and graphical format on information processing strategies, *Management Science*, 35(3), pp. 285-303.

Jarvenpaa, S.L., Dickson, G.W. and DeSanctis, G. (1985), Methodological issues in experimental IS research: experiences and recommendations, *MIS Quarterly*, 9(2), pp.141-156.

Jenkins, A.M. (1985), Research methodologies and MIS research, *in Ed. Mumford et al., Research Methods in Information Systems,* Amsterdam, Holland, pp. 103-117.

Jung, W. and Olfman, L. (2005), An experimental study of the effects of contextual data quality and task complexity on decision performance, *Information Reuse and Integration Conference*, Las Vegas, Nevada, USA.

Juran, J.M., Gryna, F.M. and Bingham, R.S. (1974), *Quality Control Handbook,* 3rd edition, McGraw-Hill, New York, USA.

Kahai, S.S. and Cooper R. (2003), Exploring the core concepts of media richness theory: the impact of cue multiplicity and feedback immediacy on decision quality. *Journal of Management Information Systems*, 20(1), pp. 263-299.

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

Kahn, B.K. and Strong, D.M. (1998), Product and service performance model for information quality: an update. *4th International Conference on Information Quality,* Boston, Massachusetts, USA.

Kahneman, D. and Tversky, A. (1979), Prospect theory: an analysis of decisions under risk. *Econometrica*, 47(2), pp. 263-291.

Kaiser, M.K., Proffitt, D.R., Whelan, S.M. and Hecht, H. (1992), Influence of animation on dynamical judgments. *Journal of Experimental Psychology: Human Perception and Performance*, 18(3), pp. 669-690.

Kaminsky, P. and Simchi-Levi, D. (1998), A new computerized beer game: a tool for teaching the value of integrated supply chain management. *In Supply Chain and Technology Management, Hau Lee and Shu Ming Ng, eds., The Production and Operations Management Society,* Miami, Florida.

Kamis, A.A. and Davern, M.J. (2005), An exploratory model of decision quality and its antecedents for category novices using multiple-stage shopping engines, *e-Service Journal*, 4(1), pp.3-27.

Kaplan, B. and Duchon, D (1988), Combining qualitative and quantitative methods in information systems research: a case study, *MIS Quarterly*, 12(4), pp. 571-586.

Kaplan, D., Krishnan, R., Padman, R. and Peters, J. (1998), Assessing data quality in accounting information systems, *Communications of the ACM*, 41(2), pp. 72-78.

Kerlinger, F.N. and Lee, H.B. (2000), *Foundation of Behavioral Research*, Harcourt College Publisher, Fort Worth, Texas, USA.

Keller, K.L. and Staelin, R. (1987), Effects of quality and quantity of information on decision effectiveness, *Journal of Consumer Research*, 14(2), pp. 200-213.

Klein, B.D., Goodhue D.L., and Davis G.B. (1997), Can humans detect errors in data? Impact of base rates, incentives, and goals. *MIS Quarterly*, 21(2), pp. 169-194.

Knight, S. and Burn, J. (2005), Developing a framework for assessing information quality on the World Wide Web, Informing Science Journal, 8, pp. 160-172.

Kocher, M.G. and Sutter, M. (2006), Time is money: time pressure, incentives, and the quality of decision-making, *Journal of Economic Behavior & Organization*, 61(3), pp. 375–392.

Kumar, R. (1996), *Research Methodology,* 2nd edition, Addison Wesley Longman Australia Pty Limited, London, UK.

Laudon, K.C. (1986), Data quality and due process in large inter-organizational record systems. *Communications of the ACM*, 29(1), pp. 4-11.

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

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

Lee, Y., Pipino, L., Funk, J., and Wang R.W. (2006), *Journey to data quality*, MIT Press, Boston, Massachusetts, USA.

Lesca, H. and Lesca, E. (1995), Gestion de l'information: qualité de l'information et performances de l'entrerpise. LITEC, Les essentiels de la gestion.

Li, S. and Lin, B. (2006), Accessing information sharing and information quality in supply chain management, *Decision Support Systems*, 42(3), pp. 1641-1656.

Lillrank, P. (2003), The quality of information, *International Journal of Quality & Reliability Management*, 20(6), pp. 691-703.

Lerner, J.S. and Tetlock, P.E. (1999), Accounting for the effects of accountability, *Psychological Bulletin*, 125(2), pp. 255-275.

Letzring, T.D., Wells, S.M., and Funder, D.C. (2006), Information quantity and quality affect the realistic accuracy of personality judgment. *Journal of Personality and Social Psychology*, 91(1), pp. 111-123.

Litwin, M.S. (1995), *How to measure survey reliability and validity*, Sage Publication, California, USA.

Lucas, H.C. (1981), An experimental investigation of the use of computer based graphics in decision making, Management Science, 27(7), pp. 757-768.

Lusk, E.J., and Kersnick, M. (1979), The effect of cognitive style and report format on task performance: the MIS design consequences, *Management Science*, 25(8), pp. 787-798.

Mackay, J.M., and Elam, J.J. (1992), A comparative study of how experts and novices use a decision aid to solve problems in complex knowledge domains. *Information Systems Research*. 3(2), pp. 150-172.

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

Mingers, J. (2001), Combining IS research methods: towards a pluralist methodology. *Information Systems Research*, 12(3), pp. 240-259.

Morey, R.C. (1982), Estimating and improving the quality of information in a MIS, *Communications of the ACM*, 25(5), pp. 337-342.

Naumann, F. and Rolker, C. (2000), Assessment methods for information quality criteria, *5th International Conference on Information Quality,* Boston, Massachusetts, USA.

Neumann, S. and Hadass, M. (1980), DSS and strategic decisions. *California Management Review*, 22(2), pp. 77-84.

Nunnally, J.C. (1967). *Psychometric theory*. McGraw-Hill Publications. New York, USA.

Oates, B.J. (2006), *Researching Information Systems and Computing*, Sage Publications, California, USA.

Olson, J.E. (2003), *Data quality: the accuracy dimension*, Publisher: Morgan Kaufmann, San Francisco, California, USA.

Olson, M.H. and Lucas, H.C. (1982), The impact of office automation on the organization: some implications for research and practice, *Communications of The ACM*, 25(11), pp. 838-847.

Oliveira, P., Rodrigues, F. and Henriques, P. (2005), A formal definition of data Quality problems, *11th International Conference on Information Quality,* Boston, Massachusetts, USA.

Ordonez, L., and Benson III, L. (1997), Decisions under time pressure: how time constraints affect risky decision making. *Organization Behavior and Human Decision Process*, 71(2), pp. 121-140.

O'Reilly III, C.A (1982), Variations in decision Makers' use of information source: the impact of quality and accessibility of information, *Academy of Management Journal*, 25(4), pp. 756-771.

Paradice, D.B. and Fuerst, W.L. (1991), An MIS data quality methodology based on optimal error detection, *Journal of Information Systems*, 5(1), pp. 48-66.

Payne, J. (1976), Task complexity and contingent processing in decision making an information search and protocol analysis, *Organizational Behavior and Human Performance*, 6(2), pp. 366-387.

Payne, J.W., Bettman, J.R. and Johnson, E.J. (1988), Adaptive strategy selection in decision making, *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14(3), pp. 534-552.

Pierce, E.M. (2004), Assessing data quality with control matrices. *Communications of the ACM*, 47(2), pp. 82-86.

Pingle, M. (1992), Costly optimization: an experiment. *Journal of Economic Behavior and Organization*. 17(1), pp. 3-30.

Pipino, L., Lee, Y., and Wang, R.Y. (2002), Data quality assessment, *Communications of the ACM*, 45(4), pp. 211-218.

Pirani, C. (2004), How safe are you hospital? *The Weekend Australia.*

Raghunathan, S. (1999), Impact of information quality and decision-maker quality on decision quality: a theoretical model and simulation analysis, *Decision Support Systems*, 26(4), pp. 275-286.

Rao, A. and Scheraga, D. (1988), Moving from manufacturing resource planning to just-in-time manufacturing, *Production & Inventory Management Journal*, 29(1), pp. 44-9.

Redman, T. (1996), *Data quality for the information age*, Publisher: Artech House, Boston, Massachusetts, USA.

Redman, T. (1998), The impact of poor data quality on the typical enterprise, *Communications of the ACM*, 41(2), pp. 79-82.

Redman, T. (2001), *Data Quality: The Field Guide*, Publisher: Digital Press, Newton, Massachusetts, USA.

Remus, W. (1984), An investigation of the impact of graphical and tabular data

Presentations on decision making, *Management Science*, 30 (5), pp. 533-542.

Sage, A.P. (1991). *Decision Support Systems Engineering*. Publisher: John Wiley and Sons, New York, USA.

Sanbonmatsu, D., Kardes, M., Frank, R., and Herr, P.M. (1992), The role of prior knowledge and missing information in multiattribute evaluation. *Organizational Behavior and Human Decision Processes*, 51(1), pp. 76 - 91.

Savchenko, S. (2003), Automating objective data quality assessment, *8th International Conference on Information Quality,* Boston, Massachusetts, USA.

Schoemaker, Paul, (1982), The expected utility model: its variants, purposes, evidence, and limitations, *Journal of Economic Literature*, 20(2), pp. 529-563.

Schroder, H. Driver, M. and Streufert, S. (1967), *Human Information Processing*, Rinehart and Winston, New York, USA.

Shankaranarayan, G., Ziad, M. and Wang, R.Y. (2003). Managing data quality in dynamic decision environments: an information product approach, *Journal of Data Management*, 14(4), pp. 14-32.

Shaw, M. (1976), *Group Dynamics: The Psychology of Small Group Behavior*. McGraw-Hill, New York, USA.

Simon, H, A, (1964), The architecture of complexity, *General Systems Yearbook*, 10, pp. 63-76.

Simon, H, A, (1969), *The Science of the Artificial*, MIT Press, Cambridge, Massachusetts, USA.

Slone (2006), *Information Quality Strategy: An Empirical Investigation of the Relationship between Information Quality Improvements and Organizational Outcomes*, Ph.D. Dissertation. Capella University.

Sprague, R.H. (1980), A framework for the development of decision support systems.

MIS Quarterly, 4(4), pp. 1-26.

Stevens, J. (1996), *Applied multivariate statistics for the social sciences,* 3rd edition, Hillsdale, Erlbaum New Jersey, USA.

Streufert, S. C. (1973), Effects of information relevance on decision making in complex environments, *Memory and Cognition*, 1, pp. 389-403.

Strong, D., Lee, Y., and Wang, R.Y. (1997), Data quality in context, *Communications of the ACM*, 40(5), pp. 103-110.

Stvilia, B., Gasser, L., Twidale, M.B. and Smith, L.C. (2007), A framework for information quality assessment, *Journal of the American Society for Information Science and Technology*. 58(12), pp. 1720-1733.

Svenson, O. and Edland A. (1987), Change of preferences under time pressure: choices and judgments, *Scandinavian Journal of Psychology*, 28, pp. 322-330.

Svenson, O., Edland, A. and Slovic, P. (1990), Choices and judgments of incompletely described decision alternatives under time pressure, *Acta Psychologica*, 75, pp. 153-169.

Taylor, R.N. and Benbasat, I, (1980), Cognitive styles research and managerial information use: problems and prospects, *Joint National Meeting of the Operations Research Society of America and The Institute of Management Sciences, Colorado Springs,* Colorado, USA.

Taylor, M.S. (1981), The motivational effects of task challenge, a laboratory investigation. Organizational Behavior and Human Performance, 27, pp. 255-278.

Teflian, M. (1999), *Information Liquidity*. Cambridge, Massachusetts, USA.

Thyer, B.A.(1993), Single-systems research design *in Grinnell R.M. (ed.), Social Work, Research and Evaluation,* 4th edition, F.E. Peacock Publishers, Illinois, USA.

Tsichritzis, D.C., and Lochovsky, F.H. (1982), *Data models*, Englewood Cliffs, Prentice-Hall, New Jersey, USA.

Turban, E. (1995), *Decision support and expert systems: management support systems*. Englewood Cliffs, Prentice Hall, New Jersey, USA.

Wand, Y. and Wang, R.Y. (1996), Anchoring data quality dimensions in ontological foundations, *Communications of the ACM*, 39(11), pp. 86-95.

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

Wang, R.Y. (1998), A product perspective on total data quality management, *Communications of the ACM*, 41(2), pp. 58-65.

Wang, R.Y., Lee, Y., Pipino, L. and Strong, D.M. (1998), Manage your information as a product, *Sloan Management Review*, 39(4), pp. 95-105.

Wang, R.Y., Lee, Y.. and Ziad, M. (2001), *Data quality*, Kluwer Academic Publishers, Norwell, Massachusetts, USA.

Wang, R.Y., Storey V.C. and Firth, C.P. (1995), A framework for analysis of data quality research, *IEEE Transactions on Knowledge and Data Engineering*, 7(4), pp. 623-640.

Wetmore, W.R. and Summers, J. (2003), Group decision making: friend or foe, *IEEE International Engineering Management Conference,* pp: 405- 409.

Wright, P.L. (1976), The harassed decision maker: time pressures, distractions and the use of evidence, *Journal of Applied Psychology*, 59, pp. 555-561.

Xu, H., Horn, J., Brown, N. and Nord G.D. (2002), Data quality issues in implementing an ERP, *Industrial Management & Data Systems*, 102(1), pp. 47-59.

Xu, H. and Koronios, A. (2004), Understanding information quality in E-business, *Journal of Computer Information Systems*, 45(2), pp. 73-82.

Yu, S., and Neter, J. (1973), A stochastic model of the internal control system, *Jour-*

nal of Accounting Research, 1(3), pp. 273-295.

Yates J.F. (1990), *Judgment and Decision Making*, Prentice Hall, New Jersey, USA.

Zakay, D., and Wooler, S. (1984), Time pressure, training, and decision effectiveness, *Ergonomics*, 27(3), pp. 273-284.

Zigurs, I. Poole M.S. and DeSanctis, G. (1987), A study of influence in computer-mediated decision making, *Working paper, University of Minnesota*.

Zmud, R. (1978), Concepts, theories and techniques: an empirical investigation of the dimensionality of the concept of information. *Decision Sciences*, 9(2), pp. 187-195.