

**LOOKING FOR A GOOD DOCTOR (OR REALTOR OR MECHANIC):
CONSTRUING QUALITY WITH CREDENCE SERVICES**

A Dissertation

by

ANN MARIE MIRABITO

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

August 2008

Major Subject: Marketing

**LOOKING FOR A GOOD DOCTOR (OR REALTOR OR MECHANIC):
CONSTRUING QUALITY WITH CREDENCE SERVICES**

A Dissertation

by

ANN MARIE MIRABITO

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee, Leonard L. Berry
Committee Members, Charles Samuelson
Venkatesh Shankar
Jelena Spanjol
Bruce Thompson
Head of Department, Jeffrey Conant

August 2008

Major Subject: Marketing

ABSTRACT

Looking for a Good Doctor (or Realtor or Mechanic):

Construing Quality with Credence Services. (August 2008)

Ann Marie Mirabito, B.A., Duke University; M.B.A., Stanford University

Chair of Advisory Committee: Dr. Leonard L. Berry

Little is known about how people evaluate credence attributes, that is, those attributes which the consumer often cannot fully evaluate even after purchasing and consuming the product. And yet consumers struggle to evaluate quality in several important product categories dominated by credence attributes such as food safety, medical services, legal services, and pharmaceuticals, among others. The dissertation explores the processes by which people form quality evaluations of services high in credence attributes and the consequences of those evaluations. Drawing on the service quality, dual-process social information processing, expert-novice and risk literatures, I develop a conceptual model to illustrate how skill and motivation moderate the ways people seek and integrate observable information to infer unobservable quality. The influence of quality evaluations on outcome, satisfaction, value, and loyalty is mapped. The model is tested in the context of a classic credence service, health care services with two large datasets using structural equation modeling.

Study 1 draws on an existing patient satisfaction database (6,280 records) to measure the sources and consequences of quality evaluations. Study 2 validates Study 1

findings and extends those findings to show the moderating roles of product expertise and perceived risk on quality evaluation processes. The second study is tested with 1,379 consumers (patients) drawn from an online consumer panel.

The research suggests service quality in this context refers narrowly to the attributes of the core product (here, the physician's medical competence); interpersonal and organizational quality are associated with value, satisfaction and loyalty, rather than overall quality. Two paths to quality evaluations appear to exist. In the first, consumers integrate evidence of the physician's capabilities, practices, and prior outcomes to reach evaluations of technical quality. In the second path, consumers rely on a trust heuristic in which observed interpersonal and organizational quality signals are used to build trust in the physician; that trust, in turn, influences perceptions of technical quality. The trust heuristic appears to be used when the stakes are low and, counterintuitively, when the stakes are high, just when superior evaluations are most needed.

TABLE OF CONTENTS

		Page
ABSTRACT		iii
TABLE OF CONTENTS		v
LIST OF FIGURES		vii
LIST OF TABLES		viii
CHAPTER		
I	INTRODUCTION.....	1
	Study 1: Customer Perceptions of the Quality of Credence Services	3
	Study 2: The Influence of Expertise and Perceived Risk on Perceptions of the Quality of Credence Services	3
II	THEORY AND HYPOTHESES	7
	Overview of the Conceptual Model	7
	Classifying Products Based on Attribute Evaluability	9
	What Are the Indicators of Credence Service Quality?	18
	How Is Credence Service Quality Assessed?.....	44
	How Does Expertise Influence Credence Service Quality Evaluations?	52
	How Does Perceived Risk Influence Credence Service Quality Evaluations?	56
	What Are the Consequences of Credence Quality Evaluation?.....	59
III	CREDENCE SERVICE QUALITY EVALUTIONS (STUDY 1).....	70
	Pilot Study	70
	Initial Exploratory Analysis (Study 1)	71
	Summary and Discussion	87
IV	PRODUCT EXPERTISE AND RISK (STUDY 2)	89
	Model Formulation and Estimation.....	89

CHAPTER	Page
Rival Models	105
Role of Product Expertise in Quality Evaluations	106
Role of Risk in Quality Evaluations.....	109
Summary and Discussion	112
V SUMMARY AND CONCLUSIONS.....	114
Summary of Research Findings	117
Implications for Credence Service Providers and Policymakers ...	126
Limitations and Direction for Future Research.....	128
REFERENCES.....	131
APPENDIX A: STUDY 1 SURVEY ITEMS.....	145
APPENDIX B: STUDY 2 SURVEY ITEMS	147
VITA	150

LIST OF FIGURES

FIGURE		Page
1	Role of Product Expertise in Interpreting Unobservable Attributes	8
2	Dimensions of Credence Service Quality	31
3	Influence of Risk on Quality Evaluation Process.....	59
4	Sources and Consequences of Credence Service Evaluations	67
5	Study 1: Structural Relationships.....	85
6	Study 1: Rival Structural Model.....	86
7	Study 2: Structural Model	102
8	Influence of Risk and Expertise on Evaluation	126

LIST OF TABLES

TABLE		Page
1	Properties of Search, Experience and Credence Quality Attributes	12
2	Service Quality Dimensions: Selected Factor Analytic Studies	21
3	Sources and Consequences of Credence Service Quality: Selected Empirical Studies	23
4	Sources and Consequences of Patient Perceptions of Physician Service Quality: A Review of the Literature.....	26
5	Summary of Hypotheses	68
6	Study 1: Profile of Sample, by Gender and by Age	76
7	Study 1: Overview of Missing Data.....	78
8	Study 1: Measurement Model Results	80
9	Study 1: Correlations Among Latent Constructs	82
10	Study 2: Measurement Model Results	96
11	Study 2: Correlations Among Latent Constructs	97
12	Formative and Reflective Indicator Models.....	100
13	Study 2: Standardized Coefficients and Fit Statistics for the Proposed Model, the Revised Model, and the Replication Analysis	103
14	Study 2: Standardized Coefficients and Fit Statistics for the Revised Model and Rivals	104
15	Invariance Testing - Experts and Novices.....	107
16	Role of Product Expertise in Quality Evaluations	108
17	Invariance Testing – Perceived Risk	111

TABLE	Page
18 Role of Risk Perceptions in Quality Evaluations	112
19 Summary of Research Findings	121

CHAPTER I

INTRODUCTION

How do people decide just how good their doctor is? How do parents conclude that a particular college will prepare their child for a satisfying and successful career? Is genetically-engineered beef safe? Does a failed business initiative reflect poor strategic work by the management consulting firm or poor execution of sound strategic advice? Medical services, education, food safety, legal representation, auto repair, and software installation are services high in credence attributes, that is, attributes which people often cannot fully judge even after purchasing and consuming the product. We know little about how buyers evaluate credence attributes. And yet, as the earlier examples suggest, products with credence attributes are often important to both consumers and industrial buyers.

My dissertation explores the process by which people form quality evaluations of services high in credence attributes and the consequences of those evaluations.¹ Drawing on the service quality, dual-process social information processing, expert-novice and risk literatures, I develop a conceptual model to illustrate how skill and motivation influence the ways consumers seek and integrate information to reach overall quality conclusions. I use the model to shed light on the following research questions:

This dissertation follows the style of *Journal of Marketing*.

¹ For convenience, “service quality evaluations of services high in credence attributes” is referred to as “credence service quality” hereafter.

- Is overall quality of services high in credence attributes assessed differently than overall quality for other services and if so, how?
- How do buyers integrate a service's credence, experience and search attributes to assess overall quality evaluations and what are the consequences of those evaluations?
- How does expertise moderate these relationships?
- How does perceived risk moderate these relationships?

I develop and test the model in the context of a typical credence service, physician services. The context provides an ideal backdrop for this study as it enables me to study the way patients (sometimes complete novices, sometimes knowledgeable) evaluate physician quality under varying levels of risk (e.g., routine primary care, general surgery, and oncology). The context is also apt because the task of evaluating physician quality is both real and important. The uneven quality of health care services has been widely documented and has been blamed for medical errors and for wasteful health care spending (Kaiser Family Foundation Institute of Medicine 2001). But judging physician quality is a complex task (Brook et al. 1996; Donabedian 2005). As a physician interviewed in an early stage of this study commented:

Judging another physician's competency is one of the things we physicians do least well. And if I can't figure out another physician's competency, how can patients do it? It's easy to assess competence for the extremes, docs who are totally incompetent, don't have licenses, make numerous mistakes. But it's much more difficult for the subgroup, the gray zone.

-- Gastroenterologist, Mayo Clinic

Two studies form the dissertation.

Study 1: Customer Perceptions of the Quality of Credence Services

I conceptualize credence service quality as a third-order hierarchical construct. The second-order dimensions include *technical* quality of the primary service provider, *interpersonal* quality of the primary service provider and *organizational quality* incorporating the support staff and systems, convenience, and the physical surroundings. The first study provides a platform for exploratory analysis of the measurement model and structural model. I draw on an existing large database (over 6,000 records) of patient satisfaction with physician services in a large multispecialty clinic. The items available in the preexisting database are similar to the items conceptualized to influence quality evaluations (technical, interpersonal and organizational performance measurements); however, several important constructs are not measured (trust, perceived risk, expertise) or are measured with only single indicators (overall quality, value, loyalty); both limitations are addressed in the second study. The dataset's scope offers a key advantage, permitting tests of the model across a wide range of medical contexts including primary care and specialty clinics. The dataset's size permits replication analysis as a guide to the internal validity of the findings.

Study 2: The Influence of Expertise and Perceived Risk on Perceptions of the Quality of Credence Services

Study 2 tests a conceptual framework incorporating two paths for integrating technical, interpersonal, and organizational quality attributes to construe overall credence service quality. People with sufficient skill and motivation are hypothesized to

follow a direct path in which technical quality indicators are assessed directly and combined with interpersonal and organizational quality evaluations to form an overall evaluation of the credence service. People lacking the skill or motivation to assess technical quality directly focus instead on interpersonal and organizational quality indicators. They first develop trust in the service provider based on their evaluation of interpersonal attributes and the organization and then use that trust to form their perceptions of the service provider's technical quality and overall quality. The mental models used to evaluate technical indicators are hypothesized and tested. Finally, the influence of credence service quality evaluations on outcome, value, satisfaction, and loyalty is mapped.

The role of perceived risk on quality evaluations is also explored. The conceptual framework for the study stems from the premise that risk operates on an evaluator's motivation and ability to assess quality. When the stakes are high, the *motivation* to assess technical quality increases but paradoxically the *ability* may decline due to stress, vulnerability and cognitive overload. I show that risk has an inverted U-shaped influence on the use of technical quality indicators to judge overall quality. That is, consumers emphasize interpersonal and organizational quality indicators in low and, counterintuitively, in high risk situations but emphasize technical quality indicators in moderate risk situations.

The model is tested with an online panel of 1,379 consumer-patients. The measurement model created in Study 1 is validated and extended in the second study using confirmatory factor analysis. Using respondent self-ratings of product knowledge,

the distinct evaluation approaches of novices and experts are evaluated using multigroup structural equation modeling. Similarly, using respondents' self-ratings of perceived risk, the influence of risk on quality evaluations is investigated.

The relationships identified in the research better elucidate the way people use available search and experience indicators to assess credence attributes, an area with limited theoretical work and very sparse empirical work. Many, arguably most, products have at least some credence quality attributes (Hahn 2004). This study focuses on credence services such as medical, legal, and repair services, all of which are often personally important, high consequence purchases. Understanding how people make those evaluations will enable service providers to better educate prospective buyers about the service providers' credence performance attributes. The research findings also better equip policymakers to protect consumers. For example, physician quality has become an important issue, fueled initially by awareness of variations in quality and more recently by concerns over cost control. Some policymakers advocate using market forces to improve quality, proposing that patients with the financial incentives to become 'better' consumers of health care will choose health care providers who offer high quality, efficiency, and overall value. This research, however, suggests patients lack the capability to accurately identify quality under many circumstances, including when they are most vulnerable.

The conceptual framework may also apply to experience services which cannot be evaluated until they are actually needed. For example, the OnStar system is an experience service because it can be evaluated after it is used, but it has the characteristic

of a credence service because the buyer may not be sure of the system's responsiveness until a serious accident happens. The study may also provide the foundation for better understanding the evaluation of credence goods such as food safety, pharmaceuticals, and nutraceuticals.

The next chapter sets forth the theory and hypotheses of my conceptual model of the antecedents and consequences of credence service quality. Chapters III and IV address the methodology and findings of Study 1 and Study 2, respectively. Implications are discussed in the final chapter.

CHAPTER II

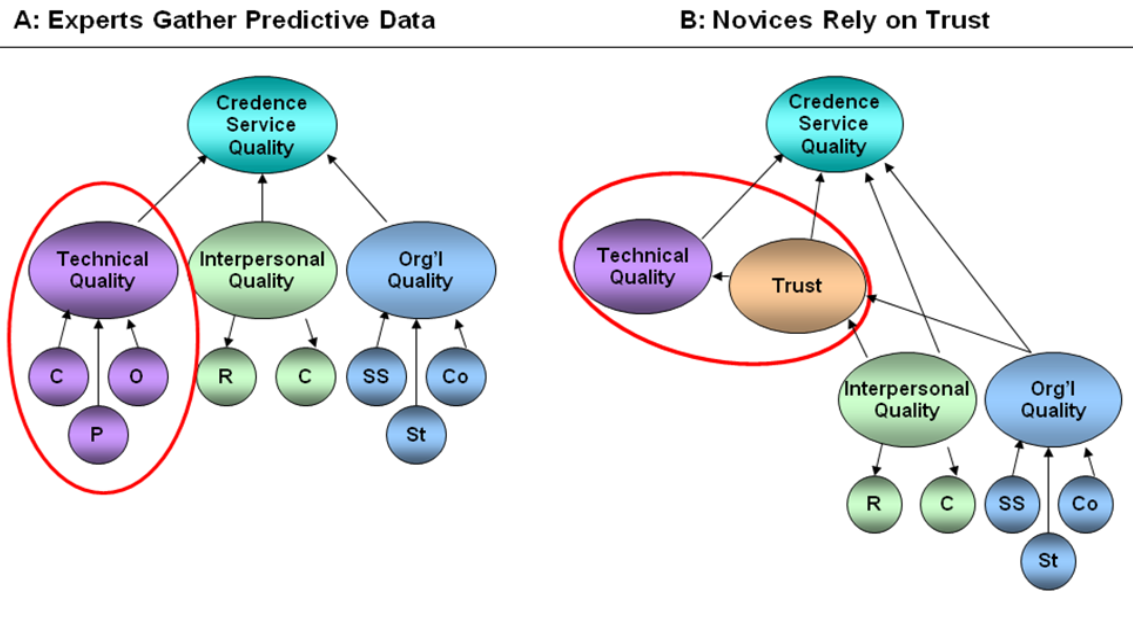
THEORY AND HYPOTHESES

Overview of the Conceptual Model

My conceptual framework integrates theories of service quality evaluations and dual-process models of attitude formation to show two paths consumers use to assess credence service quality and the consequences of those evaluations. In my framework, highly motivated product experts directly evaluate the three components of credence service quality – technical, interpersonal, and organizational – to reach overall judgments of credence service quality (see Figure 1A). People lacking product expertise or the motivation to form highly accurate assessments take an indirect path in forming credence service quality judgments (see Figure 1B). First, they form trust judgments based on evaluations of interpersonal and organizational quality. The trust judgments mediate evaluations of technical quality. The three components are then summed to determine overall quality evaluation.

Several features distinguish my model from previous work. First, in the technical quality section of the model, I explicitly model the formation of credence attribute judgments. Previous empirical work (cf, Brady and Cronin 2001; Parasuraman et al. 1991) has provided for a global measure of technical quality (e.g. reliability) without showing how the judgment is reached, a crucial issue because technical quality is unobservable with credence services. Second, I take into consideration the different

FIGURE 1
Role of Product Expertise in Interpreting Unobservable Attributes



mental structures of experts and novices to show how the two approach the evaluation of technical quality differently. I propose that experts rely on systematic cue processing to reach evaluations about credence quality; novices rely on trust. Third, I show that trust can be an antecedent of quality evaluations. The signaling role of trust with ambiguous, consequential exchanges has been elegantly elaborated by Singh, Sirdeshmukh and Sabon (2000; 2002). This research provides preliminary evidence of the role of trust in that capacity. Fourth, I show how perceived risk counterintuitively influences the way evaluations are made. Specifically, I show that people rely on weakly related signals in low risk situations and, counterintuitively, also in high risk situations, just when more ecological cues may be most important. Fifth, I show how credence quality evaluations influence perceived outcome, satisfaction, value, and loyalty.

In this section, I first describe credence services and the theory and associated hypotheses for a proposed measurement model for credence service quality. Then I develop a conceptual model for integrating the quality dimensions to reach overall credence service quality evaluations. Finally, I propose the consequences of credence service quality evaluation.

Classifying Products Based on Attribute Evaluability

Consumers are constantly evaluating products and making product choices based on those evaluations. Their ability to evaluate is influenced by the amount of information available and the evaluability of that information. Sometimes they have a wealth of information to evaluate products, other times very little information is available. Products are bundles of attributes, attributes which vary in their evaluability. A tripartite quality classification into search, experience and credence organizes attributes by their evaluability (Darby and Karni 1973; Nelson 1970; Zeithaml 1981).

Search attributes such as price, size, texture, style, color and fit can be evaluated accurately and efficiently prior to purchase. Search values can typically be measured and expressed in objective language, such as miles per gallons, wattage, or weight. As such, consumers can delegate the purchase process to others, with confidence that their requirements will be met (Mittal 2004).

The assessment of *experience* attributes requires either a purchase or a trial because the product must be used before evaluation. A restaurant meal can be evaluated after eating, a car's driving feel can be assessed after a test drive, and paint durability

can be determined after it has been put on the wall. Experience attributes often have a sensory or emotional component, which cannot be reduced to language (e.g. the fun associated with a water park) (Mittal 2004). Because experiences cannot be objectively measured and codified (just how exhilarating is a convertible ride?), consumers typically must personally experience the service in order to evaluate it and cannot delegate the evaluation to another. Service providers may offer a free trial in order to reduce the perceived risk associated with a purchase, a tactic that can be effective unless the benefit can be appreciated only after prolonged use. Sellers may also try to transform an experience attribute into a search attribute by providing information about search characteristics of the product (Mittal 2004). Wine merchants, for example, may seek to inform buyers' evaluations of taste (an experience attribute) with descriptions of a wine's varietal composition, density and color, and tannin (search attributes).

Credence attributes are often costly or impossible for a consumer to judge accurately even after purchase and use. Products high in credence attributes are often highly complex. Gathering and processing the diagnostic information about quality is time consuming and cognitively demanding, and so is infeasible for most consumers (Maute and Forrester 1991), forcing the buyer to rely on other people's evaluations or to use other attributes as indicators of product quality. While performance outcomes may offer a clue to quality, they are not perfectly diagnostic because of the way that credence attributes arise (Darby and Karni 1973). First, credence attributes can arise when an input of uncertain quality is combined with other inputs. For examples, some services incorporate both diagnosis (a less evaluable input) and repair (a more evaluable input). A

surgeon may successfully perform an appendectomy (the repair) but the patient may remain uncertain that the organ was diseased (diagnosis), and thus be unsure of the credence service quality. Second, credence attributes can arise when an outcome can have a random component. A client wrongly convicted of a crime may be uncertain as to whether the outcome reflects poor lawyering or the predispositions of members of the jury.

Table 1 summarizes the distinctions among search, experience and credence attributes. Most products contain a mixture of attributes, with some attributes more important than others. Restaurant patrons can search some attributes (price, hours, décor, type of cuisine), experience others (food flavor and texture, waiter's demeanor, wait time), while recognizing some qualities as credence (food safety). Moreover, the line between experience and credence attributes may blur if substantial time must elapse before the quality of an experience attribute can be discerned (Darby and Karni 1973). An airbag has credence characteristics because the auto owner may never need to deploy it. Similarly, a term life insurance policy holder will never know how well the issuer handles the claim.

Evaluating Product Quality

To reduce the risk associated with product purchase, buyers ordinarily undertake information search, attend to quality signals, and engage in product trial. In this section, I review the usefulness of each of these strategies for credence service buyers. As will be

TABLE 1**Properties of Search, Experience, and Credence Quality Attributes**

	Search	Experience	Credence
Examples	Often goods. Products high in tangibles e.g. clothing, jewelry, furniture	Vacations, restaurant meals. Many consumer packaged goods where quality cannot be ascertained prior to purchase, e.g., canned tuna.	Medical services, legal services, pharmaceuticals, nutraceuticals
When can attribute quality can be verified?	Prior to purchase (Darby and Karni 1973)	During or after consumption (Darby and Karni 1973)	Often impossible to evaluate costlessly (Darby and Karni 1973)
Defining characteristics	<p>Quality attributes are:</p> <ul style="list-style-type: none"> • Cognitive/linguistic i.e. can be coded in language and can be interpreted objectively (Mittal 2004) • Can be communicated to a surrogate buyer (Wright and Lynch Jr 1995) • Always accessible before purchase 	<p>Quality attributes are (Mittal 2004):</p> <ul style="list-style-type: none"> • Sensory (taste, smell, brightness, softness) • Must be searched personally • Can only be evaluated after purchase or through a free trial 	<p>Weak association between the attribute and the benefit (Mittal 2004) because:</p> <ul style="list-style-type: none"> • an input of uncertain quality is combined with other inputs (Darby and Karni 1973), or • output is stochastic (Darby and Karni 1973)

TABLE 1 (CONTINUED)

	Search	Experience	Credence
How do consumers evaluate quality?			
Preferred source of quality information	<ul style="list-style-type: none"> • Nonpersonal sources (Zeithaml 1981) 	<ul style="list-style-type: none"> • Prepurchase trial (Mittal 2004) • WOM (vicarious experience) (Mittal 2004) 	<ul style="list-style-type: none"> • Personal sources (Zeithaml 1981) • Advice from experts (Mittal 2004)
What cues do buyers rely on?	<ul style="list-style-type: none"> • Advertising claims for search attributes are better trusted than claims for experience attributes (Ford et al. 1990), consistent with the economics of information theory that consumers can verify search claims prior to purchase 	<ul style="list-style-type: none"> • Price and physical facilities (Zeithaml 1981) • Source credibility (Jain and Posavac 2001) 	<ul style="list-style-type: none"> • Seller's reputation (Mittal 2004; Zeithaml 1981) • Trust in the seller (Mittal 2004) • Credibility improves with highly credible sources (Jain and Posavac 2001) • Not: trial offers or advertising signals (Kirmani and Rao 2000; Zhao 2000)
How extensive is the quality evaluation process?	<ul style="list-style-type: none"> • Theoretically, more overall search because search is less costly (Nagle 1984; Nelson 1970) 	<ul style="list-style-type: none"> • Least extensive, based on experimental findings (Mittra et al. 1999) 	<ul style="list-style-type: none"> • More postpurchase than prepurchase evaluation (Zeithaml 1981) • Conflicting findings: More extensive than search or experience, based on experimental findings (Mittra et al. 1999); Less extensive due to high costs of gathering and interpreting information (Maute and Forrester 1991)

TABLE 1 (CONTINUED)

	Search	Experience	Credence
Size of evoked set	Larger (Zeithaml 1981)	Smaller because typically only a single brand is carried, limited competition in a geographical area, and difficulty in obtaining sufficient prepurchase information (Nelson 1970; Zeithaml 1981)	
Influence of price		Less price sensitive because price is considered a proxy for quality	Less price sensitive, especially for high criticality purchases (Ostrom and Iacobucci 1995)
Can quality be signaled through advertising or other tools?	Yes	Yes	Yes, but only to experts (Hahn 2004)
Perceived risk	Smallest (Mitra et al. 1999; Zeithaml 1981) because quality can be assessed prior to purchase	Purchase risk can be eliminated by prepurchase trial if the attribute benefits are: <ul style="list-style-type: none"> • Conveyed through the experience (e.g. sensory benefits like a car's 'feel'), or • Are unrelated to the experience attribute but can be observed after a single use (e.g. Viagra) (Mittal 2004) 	Greatest perceived risk (Mitra et al. 1999)
Brand loyalty	Smallest (Zeithaml 1981)		Highest (Zeithaml 1981) due to high switching costs and importance of building a relationship

shown, while useful theoretical insights have been developed around each strategy, empirical evidence about buyers' behaviors is sparse and sometimes conflicting.

Information search. Buyers engage in information search to reduce the risk associated with purchase. The economics of information literature proposes buyers continue searching until the costs of acquiring additional information exceed the benefits in terms of reduced purchase risk (Calfee and Ford 1988; Stigler 1961). Credence services are believed to be riskier than search or experience services and because risk is often reduced through information, it is plausible that credence services would be associated with higher levels of information search. However, credence information is always costly and often impossible to gather and to interpret, suggesting information search will be attenuated (Nagle 1984). In one study, services high in credence attributes (therapist, market research firm) were more thoroughly searched than services high in search (checking account, mail service) or experience attributes (hair cut, waiter), apparently because the credence services were riskiest and least familiar to the subjects (Mitra et al. 1999). But in a study of information search behavior, respondents dedicated less time and effort to search related to a bank's credence attributes (e.g. accurate recordkeeping, safety, trained employees) than experience attributes (e.g. helpful employees, individual attention) (Maute and Forrester 1991). Respondents claimed that the credence attributes were most critical for effective performance, but presumably were aware that, with their limited banking knowledge, the costs associated with gathering and interpreting information about complex attributes exceeded benefits in terms of reduced purchase

risk. The latter finding is consistent with research suggesting services buyers rely on internal search more than direct observation and trial (Murray 1991).

Similarly, advertising represents a low cost information source, but buyers may be skeptical of the quality of that information. Early researchers theorized that buyers would be most skeptical of unverifiable claims, reasoning that sellers would be less likely to risk losing their reputations by advertising verifiably false information (Darby and Karni 1973; Nelson 1970). Ford, Smith and Swasy (1990) found some support for that hypothesis, noting that buyers are more skeptical of experience attribute claims than search claims and of subjective claims than objective claims. However, they found no support for additional skepticism of credence claims relative to experience claims.

Signaling product quality information. Sellers frequently use warranties, price strategies or investments in advertising, branding, reputation building or location to signal quality to rational buyers (Nelson 1974). Good signals give buyers insight into quality, reduce information costs, and lower perceived risk (Erdem and Swait 1998). Signals are often self-enforcing when sellers who send misleading signals are penalized by market forces (Kirmani and Rao 2000). A low-quality seller offering a warranty will be saddled with costly warranty expenses once the true product quality is revealed. Similarly, a low quality seller who makes a large investment in advertising or branding will be unable to count on repeat purchases in order to recoup initial advertising expenses with future purchases.

Signals are particularly useful to inform buyers about the initially unobservable quality of a newly introduced experience product; they are less useful with search or

credence products (Kirmani and Rao 2000). This is because, first, signals are useful if buyers lack information prior to purchase, a characteristic of experience products. With search products, buyers can evaluate quality prior to purchase and so there is little need to decipher signals although signals may provide some value by decreasing buyers' information costs (Erdem and Swait 1998). Second, effective signals require clear information about quality after purchase. Without being certain of quality post-purchase, buyers are hobbled in their efforts to enforce penalties (Ippolito 1990). Experience products reveal quality after purchase and so are good candidates for signaling. The quality of credence products, however, may be impossible to evaluate even after purchase and use. Clearly negligent behavior can be observed (e.g., the doctor amputates the wrong foot) and penalized. But because the outcome may not reflect performance quality, it is difficult for credence service buyers to enforce the bond implicit with signals.

While signals may not be self-enforcing for credence products, credence buyers nonetheless are believed to use signals to inform quality evaluations. The seller's reputation is believed to signal product quality for credence products (Zeithaml 1981). Reputation is important for services (Berry 2000) and the additional uncertainty associated with credence attributes suggests brand equity is particularly important in this domain. Advice from credible personal sources has also been theorized to influence quality evaluations (Jain and Posavac 2001; Mittal 2004; Zeithaml 1981). Buyers relying on word-of-mouth (WOM) are relying on another consumer's evaluation of the service,

although that consumer also had limited trials and may not have full insight into the missing credence attribute (Mittal 2004).

Product trial. Free trial offers promote the evaluation of search and experience products by enabling the buyer to observe the product quality without risk. But free and reduced-cost trial offers are usually ineffective to induce trial of credence products because buyers recognize that they are unable to evaluate the product's underlying quality by simply experiencing the service. Other promotional offers may provide credence buyers with even less information. A free promotional dinner at an expensive restaurant is unlikely to give a prospective investment client the one piece of information he needs: the stockbroker's future stock-picking prowess. Instead, buyers view accepting a free trial offer as taking on an obligation to eventually buy the product without gaining definitive insight into the quality of the service (Laochumnanvanit and Bednall).

In summary, the existing literature suggests a limited role for information search, signaling and product trial with credence attribute evaluation. While those tools are informative for search and experience attributes, they are not diagnostic of credence attributes. With those limitations identified, I turn now to the measurement of credence service quality.

What Are the Indicators of Credence Service Quality?

Perceived quality is a customer's subjective evaluation of a product's superiority or excellence (Zeithaml 1988). The service quality construct is generally believed to be multidimensional, although the number and shape of the dimensions is unresolved.

Building on Swan and Combs' (1976) bifurcation of perceived goods quality into instrumental and expressive components, Gronroos (1984) proposed two fundamental dimensions of service quality. He labeled the core benefit of the service 'technical quality' and the manner in which the product is delivered 'functional quality.' Working in the same vein, Rust and Oliver (1994) proposed three components: service product, service delivery and service environment. The first two dimensions parallel Gronroos' technical and functional quality. Service environment incorporates both the internal processes that enable the delivery of good service and the physical surroundings.

Parasuraman, Zeithaml, and Berry (1991; 1985; 1988) identified five dimensions of service quality: reliability, responsiveness, empathy, assurances, and tangibles.

Validation of the instrument in insurance, banking and telephone industries revealed that the first four dimensions, representing assessments of the intangible aspects of the service, were most closely associated with overall service quality evaluations. More recently, Brady and Cronin (2001) sought to integrate the various approaches with a third-order hierarchical conceptualization with three broad dimensions: interaction, physical environment, and outcome. Interaction quality reflects attitude, behavior and expertise; physical environment quality reflects ambient conditions, design and other customers; and outcome quality reflects waiting time, tangibles and valence. Each of the subdimensions is measured with reliability, responsiveness and empathy indicators.

Table 2 includes a summary of selected service quality factor analytic studies. Fewer studies have focused on credence services; a selection of those studies are reviewed in

Table 3. Studies examining the sources and consequences of patient perceptions of physician service quality, an important credence service, are described in Table 4.

While these models provide rich insights into the way consumers evaluate service quality, I propose they are more suited to the evaluation of experience-laden services than to credence-laden services. Each of these models rests on consumer evaluation of performance outcome quality, a difficult task for credence buyers. The performance of the credence attribute is encapsulated in “technical quality” in the Gronroos model, “service product” in the Rust and Oliver model, “reliability” in SERVQUAL and “outcome valence” in Brady and Cronin. The SERVQUAL indicators “performs the service right the first time” and “employees have the knowledge to answer your questions” (Parasuraman et al. 1991 p. 448) suggest the buyer is able to determine the technical quality results and the accuracy of employees’ knowledge. Similarly, Brady and Cronin ask buyers to think about “whether the outcome of your experience was good or bad” (Brady and Cronin 2001, p. 46). Service quality research in contexts with credence attributes such as health care, libraries, real estate and insurance measure evaluations of the unobservable credence attribute without addressing the underlying processes leading to the evaluations (cf. Bowers et al. 1994; Brown and Swartz 1989; Dabholkar and Overby 2005; McAlexander et al. 1994; Thompson et al. 2003).

TABLE 2

Service Quality Dimensions: Selected Factor Analytic Studies

Researcher	Factor Structure	Industry	Instrument
(Babakus et al. 2004)	Search attributes Interest rates on deposits/loans Overdraft privileges Fees charged Banking hours New bank services Credence attributes Confidentiality Integrity Adoption of technology Management Competence Experience attributes Teller helpfulness Timely provision of services Attentiveness Willingness to listen Friendliness Employee expertise Speed of decisions	Consumer retail banking	New instrument
(Brady and Cronin 2001)	Interaction quality Attitude Behavior Expertise Physical environment quality Ambient conditions Design Social factors (other customers) Outcome quality Social factors Waiting time Tangibles Valence	Fast food Photograph developing Amusement parks Dry cleaning	New instrument
(Carman 1990)	5 to 6 dimensions, depending on setting Tangibles Reliability Security Courtesy Access Convenience Cost	Tire center Placement center Dental clinic Hospital (see Table 4)	SERVQUAL with additions, deletions, and modifications
(Cronin Jr and Taylor 1992)	One dimension	Banking Pest control Dry cleaning	SERVQUAL 22

TABLE 2 (CONTINUED)

Researcher	Factor Structure	Industry	Instrument
(Dabholkar and Overby 2005)	Communication Effort Guidance Professionalism	Fast food Real estate agents	Qualitative
(Dabholkar et al. 1996)	Physical aspects Appearance Convenience Reliability Promises Doing it right Personal interaction Inspiring confidence Courteous/helpful Problem solving Policy	Retail	28 item scale
(Thompson and Cook 2002)	Service affect Library as place Information content	Libraries	LIBQUAL
(Parasuraman et al. 1991)	Tangibles Reliability Responsiveness	Telephone repair, retail banking, insurance, credit card company	SERVQUAL
(Parasuraman et al. 1988)	Assurance Empathy	Banking, appliance repair and maintenance, long distance telephone	
(Taylor and Cronin Jr 1994)	6 dimensions for expectations scale 4 dimensions for the performance scale	Hospital services	SERVQUAL

TABLE 3**Sources and Consequences of Credence Service Quality: Selected Empirical Studies**

Researcher	Sources and Consequences	Industry	Instrument
(Bowers et al. 1994)	Sources: Tangibles Communications Competence Access Courtesy Understanding or knowing customer Responsiveness Reliability Security Credibility Outcomes Caring	Hospital services	New instrument
(Brown and Swartz 1989)	Sources: Physician interactions Staff interactions Diagnostic Professional competence Time convenience Location convenience	Physician offices	Gap model approach.
(Carman 1990)	Sources: Dental clinic and placement center Tangibles Reliability Assurance Convenience Responsiveness (Placement center only) Access (Placement Center only) Cost (Dental clinic) Acute care hospital Admission service Tangible accommodations Tangible food Tangible privacy Nursing care Explanations Visitors access and courtesy Discharge planning Billing	Dental clinic, Placement office, Acute care hospital	New instrument
(Cutler et al. 2003)	Sources: Dependability Experience and reputation Availability Courtesy	Law firms	New instrument
(Dabholkar and Overby 2005)	Sources: Service process	Real estate	Qualitative

TABLE 3 (CONTINUED)

Researcher	Sources and Consequences	Industry	Instrument
	Service outcome		
(McAlexander et al. 1994)	<p>Consequences Service process → service quality evaluations Service outcome → satisfaction</p> <p>Sources: Reliability Responsiveness Competence Access Courtesy Communication Credibility Security Understanding Tangibles</p>	Dental practices	SERVPERF (original SERVQUAL, using performance perceptions rather than expectations)
Physician offices (Safran et al. 2006)	<p>Consequences: Quality → Satisfaction Quality, Satisfaction → Purchase intentions</p> <p>Organizational features of care Access Continuity Integration Clinical team Office staff</p> <p>Physician-patient interactions Communication, Whole-person orientation Health promotion Interpersonal treatment Patient trust Relationship duration</p>		ACES (Ambulatory Care Experiences Survey) ▪ n = 12,916
(Solomon et al. 2005)	<p>Sources: Access 1: Getting care quickly/staff helpful Get care as soon as wanted Get appointment for regular care Get advice needed Office staff respect Staff helpful</p> <p>Access 2: Getting needed care Problem to get care Referral problem Help to decide on specialist</p> <p>Communication Doctors listen Doctors explain results Respect for what you had to say Spend enough time</p>		CAHPS (Clinician and Group Survey) 4.0

TABLE 3 (CONTINUED)

Researcher	Sources and Consequences	Industry	Instrument
	<p>Informed Preventive counseling Foods you eat Exercises Global rating</p> <p>Consequences: Overall satisfaction Loyalty</p>		
(Taylor and Cronin Jr 1994)	<p>Sources: Tangibles Reliability Responsiveness Assurance Empathy</p> <p>Consequences: Satisfaction (nonrecursive)</p>	Hospital services	SERVQUAL, measured as perceptions of performance
(Thompson et al. 2003)	<p>Sources: Service affect – e.g. helpfulness Library as place – e.g. haven for quiet and solitude Personal control - e.g. web site enabling me to locate information on my own Information access – e.g. convenient business hours</p>	Academic research libraries	LIBQUAL+ n = 60,027 patrons from 146 universities in the US and Canada

TABLE 4

Sources and Consequences of Patient Perceptions of Physician Service Quality: A Review of the Literature

Researcher	Sources/Consequences	Instrument/ Analysis
(Brown and Swartz 1989)	Sources: Physician interactions Staff interactions Diagnostic Professional competence Time convenience Location convenience	Gap model approach.
(Safran et al. 2006)	Sources factor structure Organizational features of care - access, continuity, integration, clinical team, office staff Physician-patient interactions - communication, whole-person orientation, health promotion, interpersonal treatment, patient trust, relationship duration	ACES (Ambulatory Care Experiences Survey) <ul style="list-style-type: none"> ▪ Massachusetts, statewide demonstration project conducted in 2002-2003 ▪ Managed care ▪ Commercially insured and Medicaid ▪ 12,916 completed surveys (30% response rate) ▪ Mail/telephone
(Solomon et al. 2005)	Sources factor structure Access 1: Getting care quickly/staff helpful – care as soon as wanted; get appointment for regular care; get advice needed; office staff respect; staff helpful Access 2: Getting needed care - Problem to get care; referral problem; help decide specialist Communication - doctors listen; doctors explain results; respect for what you had to say; spend enough time; informed. Preventive counseling - foods you eat; exercises Specialist rating Personal MD or nurse rating Rating of all care Consequences Overall satisfaction Loyalty	CAHPS – Clinician and Group Survey 4.0
(Fung et al. 2005)	Choice. When forced to trade-off technical quality for interpersonal quality <ul style="list-style-type: none"> ▪ 2/3 chose the physician who was higher in technical quality at least 3 out of 5 times Age, gender and ethnicity were not predictors of choice	Lab experiment
(Bigley et al. 2003)	Healthy consumers prefer information on benefits, office location and hours of operation and access. Patients want information on the clinician's ability to treat	Literature review of studies based on expressed preferences for information

TABLE 4 (CONTINUED)

Researcher	Sources/Consequences	Instrument/ Analysis
	people with similar conditions.	about physician quality
(Edgman- Levitan and Cleary 1996)	Sources: Access to care Coordination of care Information, communication, education Respect for patients' values, preferences, and expressed needs Emotional support and alleviation of fear and anxiety Processes of care: waiting times in the office, assistance from office staff, tests and procedures, follow up care and info	Focus groups
Hibbard & Jewett 1996 Mechanic 1989 Scotti, Bonner and Winman 1986	Sources Physician's ability to communicate and listen Knowledge and ability Thoroughness Explanations and answers Friendliness and courtesy Certification Licensure Training Diagnostic experience Proficiency of treatment	
(Kaiser Family Foundation 2000)	Most important factor in determining quality (open-ended responses): Doctors' qualification (22%) Ability to chose own doctor (7%) Patient/provider relationship (7%) Insurance coverage of care and procedures (6%) Affordability/cost (5%) Availability of appointments (5%) % saying each would tell them "a lot" about the quality of a doctor: Number of malpractice suits file (70%) Doctor's experience with a specific procedure (65%) Board certified (63%) Patient surveys on doctor's communication (57%) Admission privileges to a local hospital (37%) Attended a well known medical school or training program (36%) Highly rated by a government or independent agency (36%) Rated "the best" by a local newspaper or magazine (26%) Charges more than others do (19%)	Survey

TABLE 4 (CONTINUED)

Researcher	Sources/Consequences	Instrument/ Analysis
Robinson & Brodie 1997 [Reported in (Bigley et al. 2003)	Physician's ability to communicate and show a caring attitude (84%) Board certification status (71%) Health insurance accepted (49%) Admitting privileges to a particular hospital (38%)	Kaiser/AHRQ national survey]
(Lubalin and Harris-Kojetin 1999)	Dimensions (providers and plans) Access to chosen doctor, to specialists, length of time to get an appointment, ability to get care when needed, telephone access Amount of paperwork Benefits Choice of provider Communication/interpersonal skills/caring of provider Convenience (of choosing doctor, getting care, location) Coordination of care Costs Courtesy and manner of physicians and staff Hospital ratings Good value for the money Plan administrative hassles Quality of care overall, of particular types of care, of providers	Literature review
(NCQA 2002)	Most important topic area for evaluating a physician Patient centered care (58%) Safe care (12%) Care that works (30%) 90% agreed that it is very important or important to be treated by a physician recognized by the ADA/NCQA DPRP Consequences: Willingness to pay an additional \$10-\$30 from a physician who performs near the top of the list of physician related to each of the performance topic areas ▪ Patient centered care (68%) ▪ Safe care (60%) Care that works (68%)	Commonwealth Fund Consumer Information Project ▪ Convenience sample, visitors to American Diabetes Association website ▪ Sample characteristics: diabetic, good health, female, educated, Caucasian, have health insurance
(Pillittere et al. 2002)	<ul style="list-style-type: none"> • Patient-doctor relationship • Safety and effectiveness (after prompts) • Moderators <ul style="list-style-type: none"> ○ SES, education ○ Medicare population tended to equate good physician care with the patient-doctor relationship and to a less extent with the doctor's expertise. ○ Medicare population appears to be more forgiving of physician errors. 	Focus groups Goal: To understand the extent to which consumers can understand and value technical quality

TABLE 4 (CONTINUED)

Researcher	Sources/Consequences	Instrument/ Analysis
Rand/ California Health Care Foundation (2000) [Reported in (Bigley et al. 2003)]	Experience and training (69.9% ranked as important) Wait time for appointments (58.4%) Malpractice suits or complaints (57.7%) Ease of working with office staff (54%) Amount of time dr spends with patient (53.2%) Language spoken by dr and staff (50.2%) Hospital doctor uses (49.3%) Friend/relative recommendations (43.6%) Rating of dr by experts (38.7%) Nurse/staff recommendations (35.3%) Patient survey results (34.4%) Availability of transportation to dr's office (28.9%) Doctor's gender (17%) Doctor's race/ethnicity (11.5%)	Survey
VHA 2000 [Reported in (Bigley et al. 2003)]	Clinical quality Reputation for service quality Whether the physicians in the health plan network Referral from another physician Whether the physician is affiliated with a preferred hospital Office location is close to work/home Experience Recommendation of family/friends Medical schools attended Number of doctors working in the practice Gender	

Measurement Model of Credence Service Quality

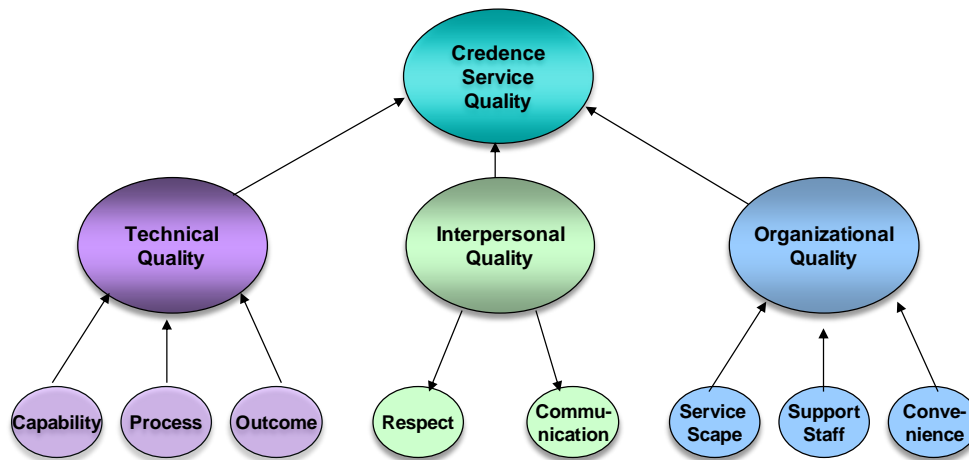
In summary, existing service quality models implicitly assume the consumer has observed and evaluated the reliability and accuracy with which the service has been performed. But existing models do not address *how* attribute judgments are made when the attribute quality is unobservable. This dissertation is designed to explore that critical gap.

Building on the service quality research tradition, I conceptualize the perceived quality of credence services as a hierarchical construct. I propose three primary dimensions: *technical quality*, *interpersonal quality*, and *organizational quality*. Technical quality refers to the provision of the core benefits of the service. Interpersonal quality refers to the interactive manner of the principal service provider. In many credence services (engineering, law, medicine, real estate), that individual has direct communications with the client. In some credence services (auto repair), the client may not actually meet the principal service provider (mechanic) and so the dimension is unavailable for evaluation. Organizational quality addresses the support staff, systems and structures facilitating the provision of the service. Figure 2 illustrates the measurement model.

H₁: Higher levels of a) technical, b) interpersonal, and c) organizational quality are associated with greater overall service quality.

The three dimensions are similar to, but distinct from, the three dimensions used by Brady and Cronin (2001). A primary difference is that the proposed model offers a

FIGURE 2
Dimensions of Credence Service Quality



finer grained examination of technical quality, the credence attributes. Another important difference is that, unlike Brady and Cronin, I hypothesize that technical and organizational quality are formative constructs, as discussed below.

Technical quality. Technical quality refers to the provision of the core benefits of the service. Its influence on overall service quality evaluations has been demonstrated extensively (Berry et al. 2006; Brady and Cronin 2001; Gronroos 1984; Parasuraman et al. 1988; Rust and Oliver 1994). Indeed, in studies of 13 industries, Berry, Parasuraman and Zeithaml (2003) consistently found that performing the promised service dependably and accurately is the most important dimension of service quality. “Little else matters to customers when a service firm is unreliable” (p. 65).

Credence service buyers often cannot directly observe overall technical quality. The bundling of diagnosis and treatment results in a credence attribute because buyers remain uncertain of their need for the service. Thus although the outcome is observable, the buyer cannot evaluate a critical diagnostic input to that outcome. The car owner may be pleased with the outcome, that is, the performance of a rebuilt transmission, but continue to wonder if flushing the fluid would have achieved the same results. In other cases, credence attributes arise when a product outcome varies depending on situational conditions. While a simple measure of valenced outcomes may capture the technical quality performance of deterministic efforts, the noise associated with the situational boundaries may make valenced outcomes an inappropriate measure for credence attributes. Buyers may recognize the possibility that outcome is beyond the service provider's control and accordingly not hold the provider responsible for outcome ("You did all that could be done").

The industrial organization (IO) literature has grappled with a similar issue: how to predict the competitive performance of firms within industries. Executive management is a credence service. Outstanding performance may reflect superior management decisions or it may reflect luck (Denrell 2004). The classic Bain/Mason IO model represents an important effort to gain insight into competitive performance. According to the paradigm, industry structure influences firm conduct; the conduct, in turn, influences firm performance (Bain 1951; Porter 1981). In the IO context, industry structure refers to the industrial environment and includes the number and sizes of competitors, consumer preferences, product mix and barriers to entry. The industry

structure informs the strategic conduct or behavior of firms. Firm conduct refers to the firm's decisions about price, advertising, and product quality. Those choices, in concert with the competitive landscape, influence firm performance. A strength of the descriptive model is its ability to show how a firm's resources and environment systematically influence its choices and ultimately its performance.

An analogous structure-process-outcome framework has been proposed as a normative tool for evaluating the technical quality of physicians (Donabedian 1982; Donabedian 1980). In that model, a physician's training and equipment (structure) influence the choice of medical processes followed and both influence medical outcomes. Building on these approaches, I use a capabilities-process-outcome framework to shed insight into the evaluation of technical quality for credence services. Quality remains probabilistic; these cues improve the odds of an accurate assessment.

Capabilities reflect the service provider's capacity for delivering the core benefit of the service. Specialized training, equipment and systems equip the technician with the tools to do the job. GM's Goodwrench service stations feature GM-trained mechanics, advanced diagnostic equipment, and access to the latest GM vehicle technical information. Certifications and membership in exclusive professional associations serve as an endorsement of competency by highly qualified peers. A brand may represent the endorsement of a product's capabilities to fulfill credence promises. The credentials and affiliations are treated search attributes related to the unobservable technical quality (Mittal 2004).

Structure may influence the *process* of service delivery and, ultimately, service quality. It involves both the appropriateness of the service and the skill with which that service is delivered (Blumenthal 1996). Strong processes are indicated by the service provider's thoroughness, judgment, and adherence to evidence-based practices, that is, practices supported by the industry's best evidence. An in-house attorney evaluating the quality of an opinion letter drafted by external counsel will examine the process used to draft the opinion. Does the opinion letter demonstrate knowledge of all the relevant facts and familiarity with relevant authorities? Did the drafting attorney consider both the narrow focal issue and other relevant issues? Was the drafting attorney able to cogently answer questions during the course of the consultation? Buyers lacking the specific expertise to evaluate the proper use of processes tend to validate process performance by comparing the service provider's advice and behavior with what the consumer knows from general reading (Mittal 2004).

Perceptions of technical quality are also shaped by *prior outcomes* the consumer experienced personally and outcomes reported by others shape. Credence outcomes may not offer perfect insight into quality. Outcomes are sometimes probabilistic; excellent providers may deliver unfavorable outcomes for reasons beyond their control. Outcomes may also mask uncertain input performance; the car runs great with the new transmission, but a fluid change might have worked as well. Nonetheless, outcomes appear to influence evaluations of the quality of the service provider. A patient may consider not only how well her physician treated her prior ailments but also how quickly her friend healed under the physician's care and the physician's malpractice record.

Outcomes may be communicated by word of mouth but also through advertising and other communication mechanisms. Investment bankers, providers of a credence service, place so-called tombstone ads in the financial press announcing successful equity placements. Regional magazines often conduct “best of” surveys to identify their readers’ favorite service providers. Plaques trumpeting a firm’s “best of” designation tell prospective customers that the service firm has delivered successful outcomes for many other customers.

In summary, I propose that an examination of a physician’s capabilities, processes and prior outcomes provide insight into the unobservable technical quality. Moreover, I propose that technical quality is a formative indicator model. With formative models, the direction of causality is from the observed measures to the factor. Technical quality meets the conceptual test for formative constructs (MacKenzie et al. 2005): indicators are not necessarily correlated, dropping an indicator may change the meaning of the construct; and increases/decreases in one indicator may be expected to influence the construct. Such a specification seems plausible. For example, a physician’s adherence to best practices might offset weak credentialing.

H₂: Higher levels of the service provider’s a) technical capabilities, b) technical processes, and c) prior outcomes are more likely to be associated with greater technical quality.

H₃: Technical quality is a formative construct.

Interpersonal quality. Interpersonal quality refers to the manner in which the principal service provider interacts with the buyer. Interpersonal quality has consistently been demonstrated to influence overall service quality perceptions (Berry et al. 2006; Bigley et al. 2003; Brady and Cronin 2001; Parasuraman et al. 1991). The quality of the client-provider relationship emerges as a critical quality evaluation criterion in other credence quality studies (Pillittere et al. 2002; Solomon et al. 2005; Taylor and Cronin Jr 1994).

The organizational justice literature provides theoretical support for interpersonal quality as a dimension of overall credence service quality. People like fairness in outcomes, in processes, and in interactions. Interactional justice is the social aspect of organizational justice, dealing with one-to-one exchanges between people with asymmetrical power (Cropanzano et al. 2002; Greenberg 1990). Concerns for interactional and procedural justice arise from people's fundamental need to establish and maintain social bonds both during a particular encounter and thereafter (Tyler 1994). Drawing on social exchange theory, the interactional justice literature suggests that people look at their standing in a group in order to validate their self-identity, self-esteem and self-respect. They expect interpersonal relationships to be marked by honesty, courtesy, respect, and ethical treatment (Greenberg 1990). Fair interactional relationships lead to satisfaction, trust and commitment (Cohen-Charash and Spector 2001).

Most of the empirical research on interactional justice is in the workplace, focusing on relationships between supervisors and employees, although there is evidence

that the construct extends to buyer-seller relationships. Customers want to be treated fairly. Vulnerable customers who are treated unfairly downgrade their quality ratings (Seiders and Berry 1998). Fairness is particularly important in credence service settings where buyer-seller relationships are characterized by asymmetrical informational power. The service provider often has better insight into the appropriateness of the service and the skill with which it is performed.²

Building on this stream, I propose that interpersonal quality is manifest by the respect and communication skills of the service provider. These dimensions are consistent with the interpersonal and informational dimensions of the interactional justice construct (Colquitt et al. 2001). *Respect* refers to the sensitivity and dignity with which the buyer is treated. It is manifest by courtesy, sympathy, and appreciation for the client's values and abilities to understand the issues. Empirically, similar attributes have been identified as indicators of service quality, including attitude (Brady and Cronin 2001), courtesy (Carman 1990), inspiring confidence (Dabholkar 1996), and aspects of responsiveness, empathy and assurance (Parasuraman et al. 1991). In the physician-patient context, respect for patients' values, preferences, and expressed needs (Bendapudi et al. 2006; Edgman-Levitan and Cleary 1996; Solomon et al. 2005) and caring attitude (Bendapudi et al. 2006) have been identified as important indicators of

² Service providers are not always omniscient, of course. An examination of general practice physicians in New Zealand found little relationship ($r \leq .2$) between the physicians' self-assessed and actual knowledge about common illnesses likely to be encountered in general practice. Physicians who rated their knowledge highly in the self-assessment survey tended to perform poorly on the knowledge test, and physicians who didn't think they knew as much performed well (Tracey et al. 1997). But while service providers may not be fully knowledgeable, they generally have greater insights into the technical provision of the service than the customers.

satisfaction. A patient in the pilot phase of this project commented on the best physician he had ever known by saying, “He remembers me, talks to me as if I’m his friend.”

Communication flows out of the informational justice construct, referring to the manner and content of explanations (Colquitt 2001). Willingness to listen suggests a neutrality and openness on the part of the service provider. The client’s problems and values have not been prejudged or dismissed as unimportant. Clarity of communications is particularly important in credence services because the technical domain is obscure for most consumers. Clear explanations give the buyer insight into the capabilities and processes that indicate technical quality.

For services that require coproduction by the buyer and the seller, clear instructions to the buyer facilitate better outcomes. Honesty is an important element of interactional justice and honesty can be discerned through clear, candid explanations. Communication performance has appeared in other services research as willingness to listen (Babakus et al. 2004), communication (Dabholkar and Overby 2005), and elements of responsiveness (Parasuraman et al. 1991).

H₄: Higher levels of the service provider’s a) respectfulness and b)

communication performance are more likely to be associated with greater interpersonal quality.

Organizational quality. A service provider’s operational infrastructure and physical setting can influence buyers’ perceptions of overall service quality (Berry et al. 2006; Bigley et al. 2003; Brady and Cronin 2001; Parasuraman et al. 1991; Rust and Oliver 1994). Strong infrastructures promote coordinated delivery of services, short

waiting times, and accurate administrative transactions. Organizational quality has a direct influence on perceptions of quality – a service delivered conveniently in attractive, comfortable surroundings by attentive personnel is likely to be more enjoyable. It also offers an indirect influence, particularly in nonroutine services and services in which intrinsic quality cues may be limited and ambiguous. The buyer's impression of organizational quality is formed by evaluations of the physical surroundings, the convenience of accessing and using the service, and the interpersonal manner of the auxiliary staff.

SERVQUAL items associated with organizational quality appear in tangibles (modern-looking equipment, attractive physical facilities) and reliability (error free records), as well as several items related to front line employee behaviors (Parasuraman et al. 1991). Rust and Oliver (1994) reasoned that service environment contributed to quality perceptions. Similarly, Brady and Cronin (2001) found that ambient conditions, design of the service facility, and the numbers and behaviors of other customers influenced overall service quality perceptions.

The *physical setting* can make the service experience more pleasant as well as provide clues to the overall performance quality (Baker et al. 2002). Environmental psychologists have demonstrated that people respond cognitively, emotionally and physiologically to their physical surroundings. Good design can encourage customers to visit a facility, to linger and browse, to interact with others. It can also enable customers to achieve the primary benefit of the service. Well-designed hospitals accelerate patient healing, reduce injuries, and lower costs (Berry et al. 2004). With services, the physical

surroundings may affect customers not only directly but also indirectly, through the motivational influence on employees who will serve the customer. Attractive workplaces foster employees' productivity and organizational commitment. The surroundings can also facilitate or hinder interactions between customers and employees (Bitner 1992).

Physical surroundings are more likely to influence customers' perceived quality in service settings in which the customer spends an extended period of time or which deliver hedonic experiences (Reimer and Kuehn 2005; Wakefield and Blodgett 1999). Customers have a greater opportunity to observe and to be impacted by the servicescape on a cruise ship, in a hotel or in a hospital. Attractive facilities stimulate the arousal and pleasure patrons are looking for in sporting events, theaters, or vacation locations. While the servicescape may be less important in utilitarian contexts, it can influence perceptions of merchandise quality (Baker et al. 2002). A dry cleaning patron might be more concerned about technical quality (How clean are the clothes? Was the dry cleaning fluid fresh and clean?) when retrieving clothing from a cramped, dirty facility than from a clean, spacious facility. The servicescape includes the mix of ambient conditions, spatial layout and functionality, and design (Bitner 1992).

Service convenience refers to the buyer's perception of time and effort expenditures related to buying or using a service (Berry et al. 2002). Like money, time and effort are expended to acquire products. Unlike money, the valuation of time and effort are ambiguous for an individual. Money is tangible, fungible and can be stored for future use. Its opportunity cost (that is, its next best use) is more easily estimated and recalled. The ambiguity in the value of time arises because its opportunity cost is

malleable (Okada and Hoch 2004). People decide the next best use for their time based on the context. An individual with free time on her hands may be less aggravated by time delays than an individual with pressing demands. Similarly, cognitive, physical and emotional effort can be difficult to measure. And like time, effort is perishable (yet often renewable), again making its opportunity cost difficult to estimate. Yet like money, people generally prefer to minimize the unnecessary expenditure of time and effort. By making services convenient, sellers can improve the quality of the buyer's experience.

Convenience is salient at multiple points during the process of buying and using a service (Berry et al. 2002; Seiders et al. 2007). Buyers dedicate time and effort to decisions about whether to go without a service, to make it or buy it. The make-buy decision is particularly salient for labor-intensive services the buyer has the capability of performing. It is less relevant with credence services as buyers typically lack the capability to perform the service. But decision convenience also encompasses the evaluation and selection of prospective service providers, an area that is particularly germane to credence service transactions because of its complexity. Most buyers have little understanding of the predictors of good performance with credence products. Service providers can facilitate the selection process by educating buyers about meaningful evaluation criteria and then showcasing their performance on these criteria.

Buyers also expend time and effort in arranging for access to the service (Berry et al. 2002; Seiders et al. 2007). Access convenience is enhanced by the ease of setting up an appointment as soon as desired and at a time convenient for the buyer. Location and parking availability are important for services conducted at the seller's premises.

Benefit convenience involves the expenditures of time and effort once the service begins (Berry et al. 2002; Seiders et al. 2007).. Services that are performed efficiently and involve little unnecessary waiting drain less time from buyers. Buyers often prefer facilities designed such that buyers can avoid unnecessary physical exertion. Transaction convenience refers to expenditures of time and effort associated with payment for the service (Berry et al. 2002; Seiders et al. 2007). Buyers particularly object to spending time waiting to pay or sorting out billing mishaps.

The interpersonal manner and actions of the *frontline support staff* can influence customers' perceptions of the organizational service quality. Whereas the interpersonal quality construct discussed earlier dealt with interpersonal interactions with the principal service provider, this construct refers to interactions with the support staff. Customers seek respect and inclusion from front line employees, as they do from the principal service provider. Customers favor frontline employees who establish appropriate rapport and provide personal attention (Baker et al. 2002; Singh 2000). In addition, when frontline employees discharge their duties effectively and efficiently, buyers may experience shorter wait times and more efficient delivery of services. Physical surroundings can support employees in their interactions with customers (Bitner 1992).

The three facets of organizational quality – physical surroundings, convenience, support staff – have an important influence on perceptions of quality. But credence services are often highly valued and involving for buyers. People are likely to be more tolerant of poor organizational quality with consequential services (Berry et al. 2002), although inexperienced buyers may rely on organizational quality signals when there are

few intrinsic clues to judge quality (Bitner 1992). Thus, while organizational quality is desired by buyers, it is unlikely to be the dominant driver of quality evaluations.

As with the technical quality construct, I propose that organizational quality is a formative indicator model. With formative models, the direction of causality is from the observed measures to the factor. Organizational quality meets the conceptual test for formative constructs (MacKenzie et al. 2005): indicators are not necessarily correlated, dropping an indicator may change the meaning of the construct; and increases/decreases in one indicator may be expected to influence the construct. For example, a friendly and professional support staff may practice in an unattractive or an inconveniently located facility.

H₅: Higher levels of the a) perceived physical surroundings, b) perceived convenience, and c) perceived support staff performance are likely to be associated with greater organizational quality.

H₆: Organizational quality is a formative construct.

To summarize, credence service quality is hypothesized to be a hierarchical construct with technical, interpersonal and organizational quality as the primary dimensions. I turn next to a discussion of the processes used to integrate these three dimensions to infer the quality of the credence attribute and to reach overall quality evaluations.

How Is Credence Service Quality Assessed?

While it may be cognitively difficult or impossible, consumers presumably try, however imperfectly, to infer quality of unobservable attributes. The inferential process is complicated because the information is not just unknown – as it is with search and experience products -- but often unknowable. The evaluation involves judging and then integrating the three technical, interpersonal and organizational quality dimensions, dimensions that vary in the amount of effort and skill required for evaluation.

Interpersonal and organizational quality are often relatively accessible for most consumers who can draw on a lifetime of experience evaluating interpersonal skills and organizational qualities in many contexts. Their preferences are likely to be fairly well-formed and stable. And because interpersonal and organizational quality are experience attributes, people are likely to spontaneously form attitudes about those dimensions (Smith and Swinyard 1988). Assessing technical quality is much more complex. The diagnostic attributes (capabilities, processes, prior outcomes) are unfamiliar to most people. Even if they know the diagnostic attributes, they may not know a particular service provider's ratings on those attributes. Moreover, to the extent that cues about technical quality are gathered through search rather than experience, attitudes are less likely to be spontaneously formed.

Steenkamp (1990) used a lens model to show how inferential beliefs about quality attributes are formed. To begin, quality cues are distinguished from quality attributes. Cues are visible; they are useful to the extent that they provide insight into the hidden quality attribute. With a lens model, the true state of the quality attributes is

inferred through the filter of the quality cues. Individual cues do not provide perfect insight into true quality. Rather the cues must be integrated in a way that reveals – or, more accurately, increases the probability of revealing – the true state. With interpersonal and organizational quality, the cues provide simple, clear insight into the underlying quality of the dimension. The auto mechanic treats you respectfully or he doesn't. The service is convenient or it is not. The observer's challenge is to weight the cues properly; that is, putting an inordinate weight on a single thoughtless comment or a single scheduling mishap may yield an inaccurate evaluation of the dimension. With technical quality, the individual cues provide less ecological validity because the link between the cue and the hidden dimension is more obscure.

Dual-Path Models of Attitude Formation

The heuristic-systematic model of attitude formation offers a useful framework for exploring how people infer quality under such circumstances. The model posits two, intertwined paths to persuasion: a systematic, cognitive-based path in which attitudes are formed using argument-based thinking and a heuristic, rule-based path in which attitudes are formed using psychological mechanisms (Chaiken and Maheswaran 1994; Eagly and Chaiken 1993; Olson and Zanna 1993; Wood 2000). When both paths yield similar conclusions, the results are likely to be additive. But when the argumentation has ambiguous interpretations, heuristics can bias cognitive processing by creating expectancies about the attitude object or the cognitive arguments. For example, if source credibility is suspect, then ambiguous messages may be interpreted with suspicion even by people highly motivated to process the argument systematically.

Systematic evaluation process. The cognitively-demanding systematic path is useful under conditions of high motivation and ability when a sufficient level of confidence in the judgment must be reached (Eagly and Chaiken 1993). I propose that it is an appropriate mechanism for motivated, able evaluators to assess technical quality. Under this mechanism, information dealing with the capabilities, processes and outcomes of the service provider is gathered and then evaluated to form a technical quality evaluation. The technical quality evaluation is integrated with experience- and search-based evaluations of interpersonal and organizational quality to form an overall evaluation. Evaluations made following this path tend to be more persistent but may require substantial time, effort and cognitive skill (Eagly and Chaiken 1993).

Heuristic evaluation process. In many cases buyers want information about credence attributes, but are reluctant to undertake an extensive information search because of the high search costs and the complexity in interpreting information (Maute and Forrester 1991). Following the dual-process models of attitude formation (Eagly and Chaiken 1993; Petty et al. 1983), I propose that people lacking the motivation or skill to gather diagnostic information and then make systematic, argument-based evaluations of technical quality instead rely on heuristic processes to assess technical quality. Attitudes formed through heuristic processing tend to remain latent until elicited (Kardes et al. 2004), as I found in my pilot study. When asked to evaluate the medical skills and knowledge (i.e., the technical quality) of their physician, the most common reply from patients was: “Gee, that’s a hard question”

While multiple heuristics may be used, I propose two as being most likely. With a correlation heuristic (Kardes et al. 2004), missing information is assumed to be correlated with available information. With a trust heuristic (Hastie 1983), available information is used to form trust in the service provider, and that trust is used to infer quality.

Correlation as a heuristic to infer technical quality. Missing attributes are sometime inferred by examining the values of available attributes and treating the missing attribute as if it were correlated with the observable attributes. Expectations or prior beliefs about the relationships between the missing and observed attributes shape the inference process (Kardes et al. 2004). The correlation heuristic has been implicated in inferences about unobserved quality drawn from prices or warranties.³ In the context of this study, consumers using a correlation heuristic infer that unobserved technical quality is correlated with observed interpersonal and organizational quality. In short, buyers substitute “style for substance” (Mittal 2004 p. 453).

Evidence that buyers substitute style for substance may be a demonstration of inference-making consistent with accessibility-diagnostics theory (Dick et al. 1990; Feldman and Lynch 1988; Kardes et al. 2004). When missing information must be inferred, diagnostic cues are given preference over non-diagnostic cues. When diagnostic cues are unavailable, however, then overall evaluations guide the evaluations of missing attributes. Following the accessibility-diagnostics framework, then, the diagnostic cues from capabilities, processes, and prior outcomes may be expected to be

³ Signaling theory offers more nuanced explanations of the price-quality and warranty-quality correlations.

used to form inferences about technical quality. Those cues are often unavailable, however, suggesting that overall evaluations formed from interpersonal and organizational cues will be used to infer technical quality.

Trust as a heuristic for evaluations of technical quality. Trust is an individual's "willingness to rely on an exchange partner in whom one has confidence" (Moorman et al. 1993 p. 82). Trust is riskier than confidence. Confidence does not imply vulnerability. With trust, the consumer actively relies on the other party, making oneself vulnerable. Confidence and reliability are at the heart of trust (Garbarino and Johnson 1999) although other researchers have used slightly different terminology. Morgan and Hunt (1994) focus on reliability and integrity; Moorman Deshpande and Zaltman (1993) emphasize expertise, reliability and intentionality; Sirdeshmukh, Singh, and Sabol (2002) emphasize dependability and reliably delivering on promises. Mayer, Davis and Schoorman (1995) emphasize ability, integrity and benevolence as the underlying dimensions. Ability refers to the trusted's expertise in a particular domain of interest. Benevolence refers to the trusted's willingness to put the other's interests ahead. It implies a personal interest on the part of the trusted, an interest motivated by something other than economic profit. Integrity refers to the trusted's adherence to a set of principles that the truster finds acceptable. The parties values' are congruent. Benevolence and integrity and distinct constructs. A customer may be reluctant to trust a service provider who adheres to a set of ethical standards (high integrity) but lacks a personal interest and attachment to the customer (low benevolence) to go the extra mile to assure good outcomes.

Trust can be a decision heuristic enabling trusters to conserve cognitive resources or to cope when cognitive resources are scarce (McEvily et al. 2003). There is no need to verify the performance, capabilities or processes of trusted individuals. Singh and Sirdeshmukh (2000) reason that consumer trust is particularly important in buyer-seller relationships marked by performance ambiguity, consequentiality and interdependence – characteristics of many credence services. The information asymmetry inherent in many credence services renders the consumer vulnerable to exploitation and heightens the value of well-founded trust.

While most research finds the trust is a consequence of quality (Garbarino and Johnson 1999), I propose that trust acts as a heuristic to facilitate credence quality evaluations. Observable interpersonal and organizational attributes serve as cues for cognitive and affective evaluations that are the foundation of trust (McAllister 1995). Trust, in turn, I propose, influences technical and overall quality evaluations. Cognitive trust reflects beliefs molded from contextual cues and preexisting beliefs. Some beliefs must be in place to form trust. Those beliefs may arise from several sources (Kramer 1999):

- Third party endorsements in the shape of credentials, licenses, and brands; with physicians, licensure and specializations may be salient.
- Shared membership in social or organizational categories as people tend to attribute positive characteristics to other ingroup members and negative characteristics to outgroup members.

- Awareness of a particular role the trusted has, based on the presumption is that the trusted would not be in that role if she were not competent and willing to fulfill the role's duties. With physicians, it is plausible people may assume all physicians are competent.
- A shared understanding of the rules guiding behavior and the consequences of failure to comply. The medical code of ethics may contribute to patients' trust in physicians.

While cognitive beliefs are the foundation for trust, affect is a powerful motivator for trust. Affective trust reflects the emotional bond between the truster and the trusted; the deeper the emotional bonds, the stronger the trust. The respectfulness and communication displayed by the principal service provider form a strong basis for affective trust. Ongoing interactions strengthen affective trust.

Trust has a dynamic quality. Its formation is contingent on expectations about the other's behavior. If the other's behavior were known with certainty, there is no need for trust. And the expectations are updated based on future supporting or refuting experiences. Diagnostic cues are asymmetric; negative cues are more powerful than positive cues (Sirdeshmukh et al. 2002).

Marketplace trust can arise at both the interpersonal and the organizational level. The performance of a principal service provider is pivotal to the effectiveness of many credence services. Many people entrust their car to a particular mechanic. Surgery is often performed by an individual surgeon. In these situations, the client is likely to rely on the performance of the focal service provider and a trusting relationship with that

individual will be critical to the development of a trust heuristic. At the same time, the organization the principal service provider is associated with may influence trust relationships, particularly if the principal service provider is not personally known. A car owner may trust Midas for muffler work without having a relationship with a particular mechanic.

A trust heuristic has been identified in the management, psychology and political science literatures. In a study of interfirm relationships, Uzzi (1999) found evidence of a trust heuristic acting as a form of relationship governance. Firms with embedded ties did not engage in extensive analyses of the terms of transactions nor did they resort to reading contracts when mishaps arose. Rather they relied on their partners to operate in ways that would protect their best interests. Using trust as a heuristic does not imply the trusting party blindly goes along with the other party. If the grounds for trust are eroded, then the trust is eroded. In Uzzi's study, the firms loosely monitored the way trust was used in their relationships. If they found their trust was repeatedly abused, they cut off the relationship with the other firm. There is also evidence of trust as a heuristic in political science. Affective trust in government forms when policies are perceived as effective and consistent with citizens' expectations (Rudolph and Evans 2005). Performance of the economy, crime levels, and political scandals all serve to influence the formation of trust. Political trust is used as a decision heuristic when citizens decide whether to support redistributive policies. Citizens who trust the government to "do what is right" are willing to sacrifice their personal interests for those of political minorities.

To summarize, the quality of the credence attribute – technical quality – is complex to evaluate and ambiguous. The heuristic-systematic model of attitude formation (Eagly and Chaiken 1993) offers a framework for understanding how people make evaluations under those circumstances. According to this model, buyers with sufficient skill and motivation will undertake a systematic evaluation of technical quality by examining the capabilities, processes and prior outcomes of the principal service provider. Buyers lacking either the skill or the motivation rely on heuristic processes, specifically, I propose, the correlation heuristic and the trust heuristic.

H_{7a}: Perceptions of high interpersonal quality promote trust.

H_{7b}: Perceptions of high organizational quality promote trust.

H_{7c}: Trust positively influences perceptions of technical quality.

Next, I show how product expertise and perceived risk inform, respectively, the skill and motivation of the buyer making the quality evaluation.

How Does Expertise Influence Credence Service Quality Evaluations?

Both novice and expert buyers infer missing information about the credence attribute when forming quality judgments. Expert evaluators are distinguished by the domain-specific knowledge acquired through experience and training (Spence and Brucks 1997). A physician referring a patient to a specialist is an expert; a chief information officer reviewing a software installation is an expert. Consumers making a first time purchase in a category may be novices, depending on the complexity of the product category. With the experience gained from multiple transactions and extensive

research, they may develop some expertise. Experts and novices are likely to differ in way they approach the evaluation of credence attributes. Specifically, I propose, experts are better equipped to use systematic processing to reach credence quality evaluations while novices are forced to rely on heuristic processing.

Experts' judgments are not necessarily more accurate than novices' (Camerer and Johnson 1991) but their evaluation process is different. When solving a problem, experts require less information than novices, can find needed information more easily, and are more apt to agree among themselves on what information is relevant (Spence and Brucks 1997). Experts use a deductive process when forming evaluations about the missing information. That is, they apply general principles to specific examples.

Experts have better intuition about the underlying concepts relevant to a problem and are more likely to use top down strategies to deduce the right solution (Chi et al. 1981; Larkin et al. 1980). Experts outperform novices with ill-structured but structurable problems (Spence and Brucks 1997). Structure refers to the knowledge required to solve the problem (Smith 1988). Structured problems are those that have a clear solution strategy. For example, finding the most conveniently-located physician to treat a medical problem might involve ranking physicians based on the travel time to each physician's office. Both novices and experts excel in these problems. Ill-structured but structurable problems are those that are unstructured as formulated but can be structured with a structuring methodology. Assessing technical quality, the credence attribute, is an ill-structure but structurable problem. Experts outperform novices with such problems (Spence and Brucks 1997). In this case, experts are likely to recognize

that the physician's capabilities, processes and prior outcomes can provide valuable insight into technical quality (Donabedian 1980).

Thus it is reasonable to anticipate that experts approach quality evaluation using systematic processing. Experts are more likely to understand the probabilistic determinants of technical quality and are more likely to have the resources to gather the information efficiently. Information they acquire may also be more accurate. Darby and Karni (1973) demonstrated that credence sellers have little incentive to mislead customers with the capability of discerning quality and customers who offer substantial future profits. Expert buyers often fit both criteria. In particular, experts who serve as gatekeepers (e.g., primary care physicians referring patients to specialists) may have the knowledge to evaluate the work and the ability to send or withhold future referrals. In addition, experts may feel particularly accountable for their evaluations, another hallmark of systematic processing (Eagly and Chaiken 1993). Thus I hypothesize experts are more likely to recognize that technical indicators are predictive of the competence of the service provider and are more likely to seek out and use information about those indicators.

In contrast, novices are likely to be at a loss as to how to evaluate technical quality. Even when novices think of a predictor of technical quality, they are vulnerable to counterargumentation. For example, a decision rule advocating "choose the doctor with the best medical education" can be counterargued with "I'd rather have the best doctor from an unknown school than the worst doctor from Harvard." More generally, novices examine problems on the basis of surface features and use bottom up strategies

to solve the problem (Chi et al. 1981). They may generalize from other experiences which may or may not be relevant. Thus, I hypothesize that novices will resort to heuristic-based processing using a correlation heuristic or the trust heuristic.

H₈: Compared to experts, novices are more likely to evaluate technical quality with a) a correlation heuristic in which technical quality evaluations are correlated with interpersonal and organizational quality evaluations or b) a trust heuristic in which trust mediates (i) interpersonal and (ii) organizational quality to form technical quality evaluations.

Confidence in an evaluation is higher when the information is plentiful, credible, and consistent (Smith and Swinyard 1988). Some dimensions of credence service quality are more accessible than others. Interpersonal and organizational quality, for example, are characteristics that consumers routinely observe and evaluate in many settings whereas accurate technical quality evaluations require some domain competency. Thus I hypothesize that consumers will tend to have more confidence in their evaluations of interpersonal and organizational quality than in their assessment of technical quality.

Experts are also more confident than novices in domain-related judgments (Spence and Brucks 1997), and so relative to novices are likely to have more confidence in their judgments of technical quality. While this hypothesis is hardly surprising, I am including it because it may help to explain the observed labile quality of novice's preferences for credence service quality attributes.

H_{9a}: Compared to their evaluations of technical quality, novices are more confident of their evaluations of interpersonal and organization quality.

H_{9b}: Compared to novices, experts are more confident of their judgments of technical quality.

How Does Perceived Risk Influence Credence Service Quality Evaluations?

Buyers face risk “when a decision or action produces social and economic consequences that cannot be estimated with certainty” (Conchar et al. 2004). The assessment of risk is subjective, with individual characteristics serving as a lens for the appraisal of the situation. The uncertainty inherent in credence services sets the stage for the full range of potential losses: performance, financial, physical, psychological, social and linked-decision losses.

Both the classical information processing model and the newer risk-as-feelings model (Loewenstein et al. 2001) recognize the roles of cognitive and affective appraisals in forming risk evaluations. With the information processing model, a risk is framed through the lens of risk importance (how important is it to achieve a target outcome?) and inherent uncertainty (what is the likelihood of a target outcome?) Risk assessments can be refined and sometimes reduced through information search. External search may come from market information such as salespeople, advertising, media and word of mouth. Internal search may examine learning from prior experiences. In this framework, emotions may be an outcome of the cognitive evaluation (Conchar et al. 2004).

According to the risk-as-feelings hypothesis, emotions play a more central role in risk assessment and decision making (Loewenstein et al. 2001). This model suggests that emotional reactions and cognitive evaluations are intertwined during a risk appraisal phase. Outcome and probabilities are inputs for both paths. But emotions are also swayed by vividness, associations, and preparedness for unacceptable outcomes. When emotions and cognitions point in the same direction, decision making is facilitated. But when emotions diverge from cognitive appraisals, emotions are often more influential.

Perceived risk affects an evaluator's motivation and ability to assess quality. Risk triggers involvement, that is, it increases personal relevance and perceived importance (Mittal 1995). High involvement promotes a greater interest in making an accurate judgment (Petty et al. 1983). Under conditions of high involvement, arguments are scrutinized more carefully, attributes differences are made clearer, strong arguments are accepted and weak arguments are rejected (Johnson and Eagly 1989). In other words, involvement promotes systematic processing, and thus when risk increases, it is reasonable to expect systematic processing of evaluation criteria.

At very high levels, however, perceived risk promotes stress, vulnerability, and cognitive overload, conditions that inhibit systematic processing and promote the use of heuristics (Eagly and Chaiken 1993). In a demonstration of this effect, participants selecting choice rules for high stakes decisions rejected compensatory decision rules based on multiattribute utility theory in favor of no rules or simple decisions heuristics (Kahn and Baron 1995). At very high levels of risk, therefore, I propose that heuristic processing will prevail.

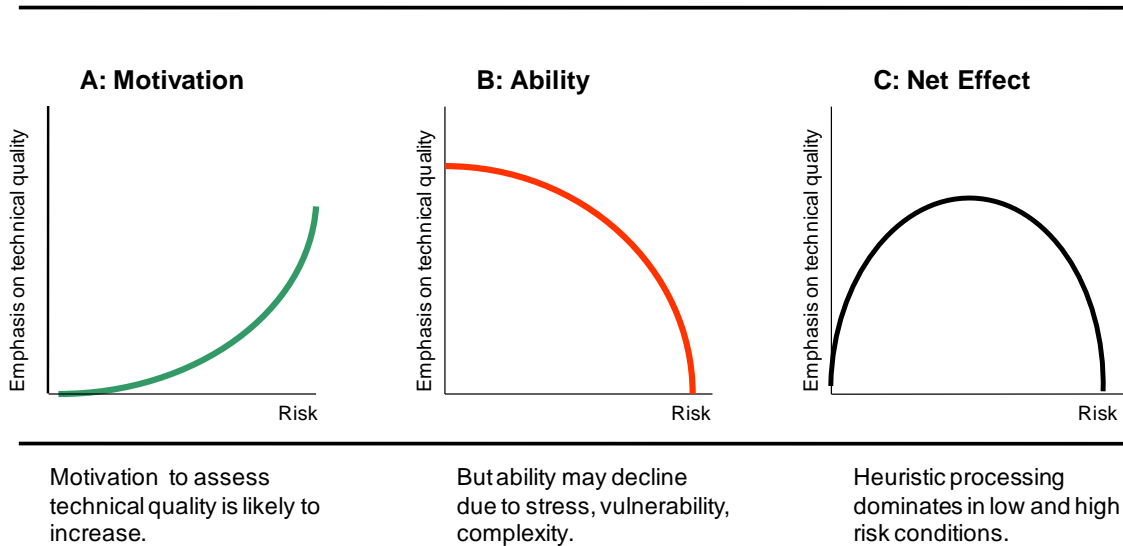
Thus I hypothesize an inverted U shaped relationship between risk and evaluation style, as illustrated in Figure 3. When risk is low, buyers have relatively little motivation to seek to evaluate the credence attribute. The incremental benefit is unlikely to exceed the costs associated with gathering additional information, and heuristic processing is useful. As the stakes increase, the motivation to assess technical quality increases (Figure 3A) promoting systematic processing of evaluation. At the same time, the ability to evaluate technical quality decreases (Figure 3B), promoting reliance on evaluation heuristics. The interaction of motivation and ability suggests that the systematic evaluation of technical quality is lowest when risk is low and, paradoxically, when risk is very high (Figure 3C).

H_{10a}: Compared to people with low perceptions of risk, those with moderate perceptions of risk are more likely to rely on systematic processing.

H_{10b}: Compared to people with moderate perceptions of risk, those with high perceptions of risk are more likely to rely on heuristic processing.

The response to risk is likely to have a temporal component. As risk is adapted to, the cognitive load may decline, permitting more systematic evaluations.

FIGURE 3
Influence of Risk on Quality Evaluation Process



What Are the Consequences of Credence Quality Evaluation?

The relationships between perceived service quality, satisfaction, value and loyalty are complex and depend, in part, on whether the constructs are viewed at the global or transactional level. In this research, I follow the prevailing quality-satisfaction-loyalty paradigm and focus on global evaluations. Loyalty researchers observe that a single transaction, whether favorable or unfavorable, is unlikely to influence loyalty. Rather the transaction is placed in the context of the full set of experiences suggesting a global focus is more appropriate (cf. Agustin and Singh 2005 for a perspective on transactional satisfaction; Anderson et al. 1994 for a perspective on cumulative satisfaction).

Determinants of Outcome

Outcome quality is a determinant of service quality evaluations for many services (Parasuraman et al. 1991). Extreme outcomes are likely to be particularly vivid and thus more easily recalled and emphasized, particularly if the results are negative (Camerer 2005; Kahneman and Tversky 1979; Kahnemann and Tversky 1974). For credence services, however, the direction of the relationship between service quality and outcome is less clear. Credence outcomes do not necessarily reflect the underlying quality of the service. It is more accurate to consider quality as a predictor of outcome.

Given the difficulty in evaluating credence quality, however, it is plausible that consumers use outcome as a measuring stick for the quality of the credence attribute. Research into the hindsight bias shows a tendency to judge decision makers as incompetent “for not having foreseen what was so clearly evident in hindsight” (Agans and Shaffer 1994 p. 440). In one study, participants judged a physician’s decision to recommend surgery favorably when a successful outcome occurred and unfavorably when an unsuccessful outcome occurred (Baron and Hershey 1988). Evaluations made in foresight reflect information associated with multiple possible outcomes. In hindsight, however, evaluations focus on information associated only with the known outcome (Hawkins and Hastie 1990 p. 315). Foresight evaluators of the surgeon’s quality considered the likely benefits and the minimal risks associated with the surgery. But when considered in hindsight, outcome knowledge creeps into the evaluator’s mental representation of the situation and influences the evaluation.

Because of the dueling conceptual arguments, I do not predict the direction of the relationship between quality and outcome.

H₁₁: Outcome is positively associated with a) technical quality, b) interpersonal quality, c) organizational quality, and d) overall service quality.

Determinants of Satisfaction

Satisfaction is the sense of pleasure arising from the way a need or goal was fulfilled (Oliver 1999). It is a psychological state, an attitude arising from the affective and cognitive evaluations of perceived performance set against the backdrop of expectations. Satisfaction influences word-of-mouth referrals and, within a range, repeat purchases. Dissatisfaction or weak satisfaction tends to knock a product out of the consideration set and squash repeat purchases. Higher levels of satisfaction promote higher levels of repeat purchases. But after a satisfaction threshold has been met, additional satisfaction is unlikely to be translated into additional purchases (Oliva et al. 1992).

Satisfaction has an affective component, distinguishing it from quality (Oliver 1993; 1994). Satisfaction stems from an interplay of cognitive appraisals of the service performance and of feelings created by the experience. Quality evaluations reflect a cognitive appraisal of the excellence of a product. Moreover satisfaction reflects a post-purchase appraisal whereas, at least with search goods, quality can be evaluated prior to purchase.

Service quality has been found to influence customer satisfaction in several contexts (Anderson et al. 1994; Gotlieb et al. 1994; Lam et al. 2004; Spreng and Mackoy

1996). With credence services, I expect quality to influence satisfaction in a more complex way depending on the evaluability of service attributes. Interpersonal and organizational quality attributes are generally more accessible and easier to evaluate than technical quality, and so, I propose, are more likely to influence satisfaction. When people cannot assess outcomes directly, perceived procedural fairness influences satisfaction (Van den Bos et al. 1997). Service failure research shows that, in the event of a service failure, perceptions of procedural and interactive justice help salvage satisfaction (Smith et al. 1999). By analogy, then, when technical quality is difficult to evaluate, I propose that satisfaction appraisals will rely heavily on perceptions of procedural fairness gleaned from interpersonal and organizational experiences. A service provider who willfully disregards a customer's need for convenience violates principles of procedural justice.⁴

In addition, appraisal of the perceived value associated with a product can lead to satisfaction feelings and evaluations (Woodruff 1997). The value construct reflects the net benefit of attributes and costs associated with a product and customers have a preference for attributes that help them to meet their goals. The influence of value on satisfaction has been demonstrated empirically (Tam 2004; Yang and Peterson 2004).

H_{12a}: Perceived credence service quality positively influences satisfaction.

⁴ Related to this prediction, Dabholkar and Overby (2005) found that outcome valence from a real estate transaction – a credence service -- predicted satisfaction. The real estate agent's processes led to service quality evaluations, but seemed unrelated to satisfaction. Using a "what counts is the outcome" argument, the researchers hypothesized that customers delighted by the outcome were satisfied regardless of the quality of other quality dimensions.

H_{12b}: With credence services, perceived interpersonal quality positively influences satisfaction.

H_{12c}: With credence services, perceived organizational quality positively influences satisfaction.

H_{12d}: Value positively influences satisfaction.

Determinants of Value

Consumers are looking for value from exchange relationships. Value refers to the consumer's overall assessment of the tradeoff between benefits received and sacrifices made to maintain the relationship (Agustin and Singh 2005; Zeithaml 1988). The notion of value is rooted in equity theory, which holds that people expect benefits to be proportional to costs.

Perceived quality, a contributor to benefits, has been found to mediate the influence of extrinsic cues such as product brand, store reputation, and servicescape on value perceptions (Baker et al. 2002; Teas and Agarwal 2000). With credence services, I propose that outcome, a key benefit of quality, will directly influence value. In a service relationship, the relationship itself also provides benefits to the buyer (Gwinner et al. 1998). While price is commonly a focal sacrifice, time, effort and stress are nonmonetary costs the buyer bears to maintain the relationship. Trust reduces the nonmonetary costs associated with a relationship. Trusting relationships free the buyer from continually monitoring and interpreting indicators of quality. The relational benefits associated with trust also contribute value. (Sirdeshmukh et al. 2002)

H_{13a}: Perceived outcome positively influences value.

H_{13b}: Perceived interpersonal quality positively influences value.

H_{13c}: Trust positively influences value.

Determinants of Loyalty

While consumers seek value, firms want loyalty. Loyal customers contribute disproportionately to the firm's financial performance (Oliver 1999). They forgive errors, tolerate higher prices, resist competitive efforts, and attract other customers to the firm. Loyalty is a deeply held, enduring commitment to continue to buy a product, leading to extended repurchases despite situational influences that may promote switching (Oliver 1999). There are degrees of loyalty. At its simplest, loyalty is represented through favorable beliefs about a brand, beliefs which may evolve into liking for the product. The positive affect may then evolve into loyalty intentions which may ultimately result in loyalty purchase behaviors. That highest level of loyalty – action loyalty -- is critical to the firm, because it leads to financial performance. It may be measured by future intentions (Zeithaml et al. 1996), by “share of choice” (Neal 1999), and by willingness to recommend. In this research, I focus on attitudinal loyalty: does the buyer freely⁵ and without reservation choose to remain with the service provider and recommend the service provider to friends and family?

Loyalty can arise from a hybrid of cognitive and emotional factors. Cognitively driven loyalty can arise through commitment, defined as “the decisions or cognitions

⁵ Attitudinal loyalty cannot always be inferred from actions. High switching costs can prevent attitudinally disloyal customers from defecting. In the case of medical care, insurance restrictions may influence consumers' choices. In this research, however, I controlled for insurance restrictions by measuring switching costs.

that bind an individual to a behavioral disposition” (Pritchard et al. 1999). Commitment attitudes are more stable when: evaluations are reached through the kind of systematic processing described earlier; cues are consistent individually and when considered together; and consumers are confident in their judgment. Emotional response to a service can also influence future purchase intentions (Berry 2000). Customers are loyal to service providers with whom they feel an emotional bond, a bond that is especially strong when the consumer believes the provider shares her core values.

The influence of customer satisfaction on both attitudinal and behavioral loyalty is well established (Lam et al. 2004; Shankar et al. 2003). Service quality has been found to influence behavioral intentions both directly (Zeithaml et al. 1996) and when mediated through satisfaction (Dabholkar et al. 2000; Gotlieb et al. 1994).

Perceived value has been found to mediate the relationship between quality and loyalty (Bolton and Drew 1991; Tam 2004; Yang and Peterson 2004). In a retail setting, merchandise value only partially mediated the influence of quality attributes on loyalty, with interpersonal service quality and convenience also playing direct roles (Baker et al. 2002).

Trust has also been found to influence loyalty. Sirdeshmukh, Singh, and Sabol (2002) found that value partially mediates the influence of trust on loyalty. In transactional exchanges, satisfaction mediates trust’s influence on loyalty whereas in relational exchanges, trust mediates satisfaction’s influence on loyalty (Garbarino and Johnson 1999). Seeking to integrate the findings, Agustin and Singh (2005) proposed and tested an elegant solution in which transactional satisfaction, trust, and relational

value have a curvilinear relationship with loyalty. In their formulation, transactional satisfaction and value fulfill lower-order hygiene needs related to the consumer's overall goal. To a point, higher levels of transactional satisfaction and value help the consumer achieve her ultimate goal, but after that point, increasing levels of satisfaction and value have less of an impact on achieving the overall goal. [Note that by using transactional rather than cumulative satisfaction, the researchers expected to capture an economic evaluation and minimize the influence of social bonds.] Trust, however, represents a higher-order, growth need: You can't have too much. With credence services, I propose that trust is particularly important. Specifically, I propose that trust plays the dual role of shedding light on past quality and promising successful future transactions with the service provider.

H_{14a}: Outcome valence has a positive influence on loyalty intentions.

H_{14b}: Satisfaction has a positive influence on loyalty intentions.

H_{14c}: Value has a positive influence on loyalty intentions.

H_{14d}: Trust has a positive influence on loyalty intentions.

Switching Costs

Switching costs moderate the influence of customer satisfaction and value on loyalty intentions (Hennig-Thurau et al. 2002; Lam et al. 2004; Yang and Peterson 2004). Switching costs refer to the monetary and nonmonetary costs associated with changing suppliers. Jones, Mothesbaugh and Beatty (2002) identified three classes of switching costs. Continuity costs refer to the actual or potential loss of performance benefits associated with staying with the current provider. Frequent flyer points and

preferential treatment are more obvious examples of continuity costs. A less obvious, but potentially more powerful continuity cost, is the uncertainty associated with the performance of a prospective service provider relative to the current provider. Learning costs are a switching cost arising from the process of searching and evaluating new service providers and then adapting to the new provider's systems. Sunk costs, perhaps the least obvious switching cost, arise from the economically irrelevant but psychologically powerful prior investments of money, time and effort in a relationship that will be abandoned when the relationship is terminated. I expect high switching costs to moderate the influence of value on loyalty.

H₁₅: Switching costs moderate the influence of value on loyalty.

The conceptualized sources and consequences of credence service quality evaluations are illustrated in Figure 4. Hypotheses are summarized in Table 5.

FIGURE 4
Sources and Consequences of Credence Service Evaluations
(Conceptualized)

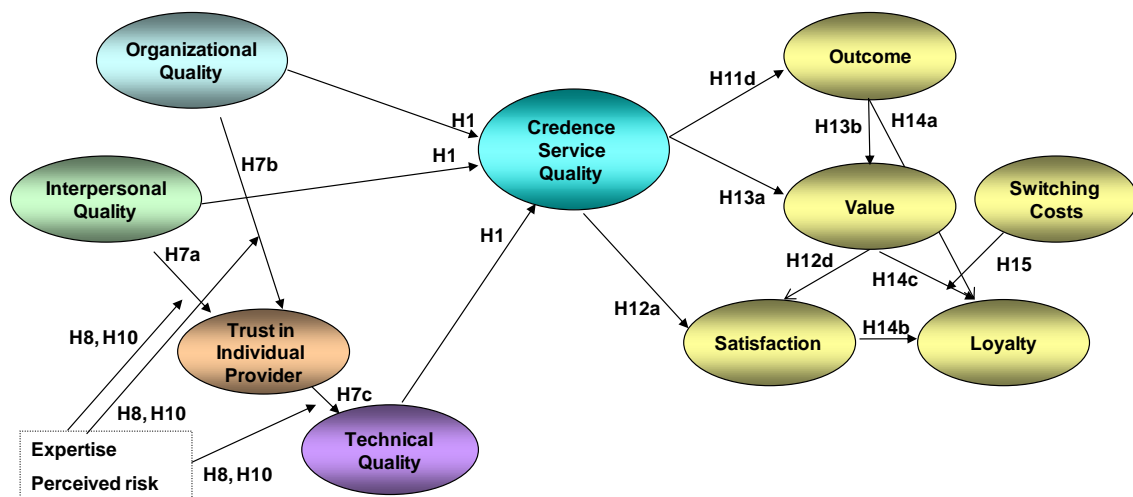


TABLE 5
Summary of Hypotheses

Domain	Hypotheses
Overall Service Quality	H ₁ : Higher levels of a) technical b) interpersonal and c) organizational quality are associated with greater overall service quality.
Technical Quality	H ₂ : Higher levels of the service provider's a) technical capabilities, b) technical processes, and c) prior outcomes are associated with greater technical quality. H ₃ : Technical quality is a formative construct.
Interpersonal Quality	H ₄ : Higher levels of the service provider's a) respectfulness and b) communication performance are associated with greater interpersonal quality.
Organizational Quality	H ₅ : Higher levels of the a) physical surroundings, b) convenience, and c) support staff performance are associated with greater organizational quality. H ₆ : Organizational quality is a formative construct.
Trust	H _{7a} : Perceptions of high interpersonal quality promote trust. H _{7b} : Perceptions of high organizational quality promote trust. H _{7c} : Trust positively influences perceptions of technical quality.
Expertise	H ₈ : Compared to experts, novices are more likely to evaluate technical quality with a) a correlation heuristic, or b) a trust heuristic
Confidence	H _{9a} : Compared to their evaluations of technical quality, novices are more confident of their evaluations of interpersonal and organization quality. H _{9b} : Compared to novices, experts are more confident of their judgments of technical quality.

TABLE 5 (CONTINUED)

Domain	Hypotheses
Risk	<p>H_{10a}: Compared to people with low perceptions of risk, those with moderate perceptions of risk are more likely to rely on systematic processing.</p> <p>H_{10b}: Compared to people with moderate perceptions of risk, those with high perceptions of risk are more likely to rely on heuristic processing.</p>
Outcome	H ₁₁ : Outcome is positively associated with overall service quality.
Satisfaction	<p>H_{12a}: Perceived credence service quality positively influences satisfaction.</p> <p>H_{12b}: With credence services, perceived interpersonal quality positively influences satisfaction.</p> <p>H_{12c}: With credence services, perceived organizational quality positively influences satisfaction.</p> <p>H_{12d}: Value positively influences satisfaction.</p>
Value	<p>H_{13a}: Perceived outcome positively influences value.</p> <p>H_{13b}: Interpersonal quality positively influences value.</p> <p>H_{13c}: Trust positively influences value.</p>
Loyalty	<p>H_{14a}: Outcome valence has a positive influence on loyalty intentions.</p> <p>H_{14b}: Satisfaction has a positive influence on loyalty intentions.</p> <p>H_{14c}: Value has a positive influence on loyalty intentions.</p> <p>H_{14d}: Trust has a positive influence on loyalty intentions.</p>
Switching Costs	H ₁₅ : Switching costs moderate the influence of value on loyalty.

CHAPTER III

CREDENCE SERVICE QUALITY EVALUTIONS (STUDY 1)

Pilot Study

A qualitative pilot study was conducted to refine the conceptual model. The study included 34 individual, in-depth interviews conducted in spring 2006.

Respondents included 23 patients and 11 practicing physicians. I continued interviewing participants in each group until the marginal value of additional interviews was minimal.

To recruit patient participants, I began by identifying people (approximately one-fifth of the total) who I believed would be interested in the topic. The rest of the sample was then selected using the snowball sampling method in which interview participants referred me to others, an appropriate technique when the goal is hypothesis-generation (Stake 2000). Typically, prospective participants were sent an email describing the study and requesting participation. Patient participation rates were very high; all but two of the people invited to participate agreed to do so. The patients interviewed represented a broad range of demographics (age, income, education) and varying health status. Patients described: the best and the worst doctors they knew; how they judge physician competence; and how they integrate perceptions of technical quality and experience quality to form overall service quality evaluations.

Physicians were selected based on their reputations for medical excellence. Physicians received an email describing the study and asking them to participate.

Approximately 50% of the physicians invited chose to participate in the study. The physicians interviewed included specialists and generalists in large and small groups. Physicians described their perceptions of: the best doctor they knew; how they judge technical quality; how patients judge technical quality; and why patients visit technically incompetent physicians.

All interviews were conducted by telephone, by appointment. Patient interviews typically lasted 45 minutes and physician interviews typically lasted 25 minutes. I took detailed notes and later transcribed and coded the notes. I analyzed the interview transcripts, looking for novel insights into the ways people evaluate quality.

I refined the model based on the pilot data, resulting in the conceptual model described above. I also studied the transcripts to understand the terminology people use for describing medical quality in order to improve the question format used in the survey-based studies.

Initial Exploratory Analysis (Study 1)

Drawing on an existing large dataset (6,280 records) of patient satisfaction with physician services in a large multispecialty clinic, the first study provides a preliminary test of the conceptual model. A key advantage of the dataset is its size and its scope, that is, the inclusion of primary care and a wide range of specialty care medical conditions. Study 1 has two primary purposes. First, the dataset provides a unique opportunity to begin to identify the measurement of the three conceptualized domains of credence service quality: technical, interpersonal and organizational quality. Second, a

preliminary structural model of the relationship among the focal constructs of service quality, outcome, and perceived value is estimated. A limitation to this study is that the items included in the existing patient satisfaction database are not as extensive as those conceptualized. That limitation is remedied in Study 2 in which a full complement of the indicators of the conceptualized constructs is included.

Model Formulation and Estimation

My analysis includes four tasks: 1) SEM analysis of an overall model linking service provider quality characteristics with evaluations of outcome and perceived value; 2) mediation analysis of quality on outcome and perceived value; 3) robustness checks on the measurement and structural estimates; and 4) analysis of consumer heterogeneity. To estimate the relationships between service quality and its consequences, I estimated a measurement model that involved the antecedents and consequences of service quality and an SEM that associates service provider performance with outcome, value and loyalty. For the mediation analysis, I performed partial and full mediation tests of quality on outcome and value using the methodology outlined by Baron and Kenny (1986). I evaluated the robustness of the estimates by randomly assigning each record in the full sample to one of three buckets. The first bucket (n=2,155) was used to estimate the measurement model. The second bucket (n=2,108) was used to validate the initial estimates using methodology recommended by Cudeck and Browne (1983) in which the restricted covariance matrix created from the measurement model of the calibration sample is compared to the unrestricted covariance matrix for the validation sample. After demonstrating satisfactory fit, the first two buckets were combined (n=4,173) to estimate

the structural model which was later validated with the third bucket (n=6,280). All reported estimates reflect the full sample (n=6,280).

I performed the consumer heterogeneity analyses using a priori segmentation. Biased estimates may result from SEMs conducted with an aggregate sample. To ferret out consumer heterogeneity, I conducted multigroup SEM on known demographic (age, gender, health status), psychographic (health self-efficacy), and contextual (was there a doctor in charge?) variables.

Data sample. The sample is a cross-sectional dataset drawn from an existing patient satisfaction dataset collected by a large multispecialty, multi-location outpatient clinic. The patient satisfaction data are collected as part of the organization's ongoing quality improvement process. Patients are alerted by mail of the survey one to two weeks after their outpatient visit and a telephone survey is conducted one week later. The survey is conducted continually. A random sample is drawn to achieve a quota of 75 completed interviews per quarter for each of the 108 outpatient reporting units (i.e. primary care or specialty care, by geography). The survey is conducted by telephone in order to speed data collection and processing, to achieve high response rates and to reduce response bias. Additionally, telephone surveying accommodates low literacy populations and non-English readers. The sample excludes patients who are deceased, have no phone number, live outside the U.S., require an interpreter, obstetric patients under 18, outpatients with appointments in the next 60 days, those surveyed in the past 52 weeks, and those who have asked not to be contacted.

The data were collected between October 2005 and March 2006. Of the 21,028 patients who were eligible for contact, 9,101 (43.3%) met the sampling quota, were reached by telephone and completed the survey. Pediatric patients (2,143) and patients who received health care from a nurse or other non-physician provider (678) were removed from the sample for this study. The remaining 6,280 patients were included in the study. IRB approval was obtained for use of the secondary data in this study.

Measurement items. Perceived performance is measured for physician behaviors (listening, explaining, thoroughness, involving patients in decisions), organizational behaviors (teamwork, respect for privacy, courtesy, efficiency, appointment convenience, and facilities), value, and loyalty intentions. Control variables include a self-assessed health rating, perceived health self-efficacy, race, age and gender. Patient records identify the department performing the service, but not the physician.

Appendix A identifies the items collected. Performance items are measured on a five-point scale (1=excellent, very good, good, fair, 5=poor). The distribution of the performance variables departs substantially from multivariate normality. Performance variables are negatively skewed and platykurtic (peaked). Variance for the variables included in the model averages 0.6, just 15% of the maximum possible variance with a 5-point rating scale (Thompson 2006). While that pattern is common in patient satisfaction studies, the effect may be amplified in this sample because of the strong quality orientation of the focal clinic and because the measurement scale was limited to five points. A descriptive analysis of the study sample shows that the sample mirrors

demographics of U.S. patient visits to physician office (see Table 6), supporting generalizability of study findings.

Model estimation. The unit of analysis is the physician. The analysis was conducted with structural equation modeling in LISREL (version 8.80) and Mplus (version 5).⁶ Because of the violations of multivariate normality, an asymptotically distribution free (ADF) robust weighted least squares estimator (LISREL's diagonal WLS and Mplus's WLSMV) was used (Flora and Curran 2004)⁷ as it generates unbiased parameter estimates, standard errors and χ^2 goodness of fits statistics. As a robustness check, normal theory ML was used with the Satorra-Bentler adjustment of the χ^2 and the standard errors generated from normal theory ML.⁸ Adjusted ML estimates were consistent with the robust WLS findings reported here.

The dataset contained a limited number of missing values (3.4% of the sample); see Table 7 for a description of missing data patterns. Missing data were imputed using

⁶ Analyses were replicated using multiple techniques in both software packages. Mplus was used for certain techniques not available in LISREL, e.g. latent class analysis and formative factor analysis.

⁷The high skewness and kurtosis present in the sample violate key assumptions underlying the more commonly used estimation techniques such as maximum likelihood (ML) and normal theory generalized least squares (GLS) (West, Finch and Curran 1995). GLS, for example, assumes that the distribution is fully described by the mean and variance. With nonnormal data, measures of skewness and kurtosis are also required in order to describe the distribution. Violation of the multivariate normality requirement will still yield parameters that are unbiased (neither overestimating nor underestimating the population parameter) and consistent (as the sample size increases, converging to the true value of the population parameter). However parameter estimates are not efficient (variance can be substantial even in large samples). With nonnormal data, ML and GLS produce inflated χ^2 values and underestimated fit indexes such as the Comparative Fit Index (CFI), both working to increase Type I errors. At the same time, with nonnormal data, ML and GLS underestimate standard errors of parameter estimates, thereby increasing Type II errors.

⁸ The Satorra-Bentler adjustment involves scaling the normal theory χ^2 by a function of the residual weight matrix, the observed multivariate kurtosis and the model degrees of freedom and making a similar adjustment to the standard errors (West et al. 1995).

TABLE 6
Study 1: Profile of Sample, by Gender and by Age

	Gender			Age			
	Sample	U.S. ¹	Difference	Sample	U.S. ¹	Difference	
Female	55%	58%	-3.0 pts.	18-44	32% ²	33%	-0.4 pts.
Male	45%	42%	3.0 pts.	45-64	31% ³	36%	-5.1 pts.
Total	100%	100%	0.0 pts.	65-74	21%	15%	6.0 pts.
				75/Over	16%	16%	-0.5 pts.
				Total	100%	100%	0.0 pts.

¹ Visits to Physician Offices in 2004. Source: National Center for Health Statistics 2006, p. 324.

² Ages 18-49 .

³ Ages 50-64.

the multiple imputation (MI) procedure (Schafer and Graham 2002; Schafer and Olsen 1998) as implemented in LISREL. In the technique, each missing value is replaced by a set of m simulated values based on the observed values and random noise. MI is robust to violations of normality (Schafer and Graham 2002).

All items are treated as ordinal because the data were measured on 5-point Likert-like scales. Accordingly, covariance matrices and the asymptotic covariance matrix required for estimation were based on polychoric correlations.⁹

Measurement findings. Following Anderson and Gerbing (1988), the measurement model was evaluated first. Seven first order factors were identified from the measured attributes: service quality, physician quality, support staff quality, access quality, facilities quality, Outcome, and Value. Loyalty showed poor discrimination from service quality and was dropped from further analyses. Sample statistics and

⁹ I also fit the model with the data defined as continuous. Model fit was substantially weaker. So for empirical and theoretical reasons, the ordinal treatment is reported.

parameter estimates for each indicator are noted in Table 8. The measurement model is similar to the conceptualized model, within the constraints of the indicators available in the preexisting dataset. Specifically, Trust is not measured and Outcome, Value and Loyalty are measured with single indicators, adjusted to account for measurement error.

A key difference between the conceptual and empirical models, however, is that the physician quality construct incorporates measures of both the physician's technical and interpersonal performance. While those attributes were conceptualized as distinct dimensions, they were found to lack discriminant validity and have been collapsed to form a single dimension. The support staff, access, and facilities constructs parallel the hypothesized subdimensions of the organizational quality construct, with the exception that access reflects a narrower range of the convenience construct.¹⁰ All constructs were fit with reflective indicators in which the direction of causality flows from the construct to the indicators. Formative indicator models in which the direction of causality is from the indicators to the construct could not be identified with the extant dataset. Specifically, formative indicator models require at least two paths emanating from the construct to other indicators or to other constructs. The sparse dataset did not provide for those additional measures. The model fit is good. The chi-square statistic is 835.99 with 184 degrees of freedom ($p < 0.01$). The CFI (1.00) and RMSEA (.024) are acceptable, with values better than the recommended cutoffs of .95 and .06 (Hu and Bentler 1999).

¹⁰ In Mplus, support staff quality, access quality, facilities quality are found to reflect a higher order construct, organizational quality. Second order factor analysis in LISREL did not yield a solution.

TABLE 7

Study 1: Overview of Missing Data

Number of Missing Values per Variable

Pr3Sp4	Attempts	q01qlty	q02MDchg	q05thoro	q06enftm	q07listn	q08ezwrđ
-----	-----	-----	-----	-----	-----	-----	-----
0	3	12	134	80	12	31	19

Number of Missing Values per Variable

q09invdc	q10commu	q13cocar	q14medsk	q15ontim	q16instr	q17expla	q18MDQ
-----	-----	-----	-----	-----	-----	-----	-----
116	8	13	60	62	98	70	22

Number of Missing Values per Variable

q23RN	q24recep	q25tmwrk	q26pinfm	q27expla	q28priva	q29court	q30time
-----	-----	-----	-----	-----	-----	-----	-----
118	33	90	306	129	57	12	84

Number of Missing Values per Variable

q31effic	q34adays	q35apacc	q36apcrt	q37ainfo	q38qaccs	q39fcour	q40atmos
-----	-----	-----	-----	-----	-----	-----	-----
14	894	238	383	446	1037	538	30

Number of Missing Values per Variable

q41clean	q42prkg	q43bill	q44value	q45outco	q46safet	q49rcmd	q1bqlty
-----	-----	-----	-----	-----	-----	-----	-----
8	1438	1109	632	441	5	41	5

Number of Missing Values per Variable

q52compa	q53healt	q54hse	q55race	q56OKfu	AgeGroup	gender	Employee
-----	-----	-----	-----	-----	-----	-----	-----
624	123	319	75	0	0	0	0

TABLE 7 (CONTINUED)

Number of Missing Values per Variable

q33balg

420

Distribution of Missing Values

Total Sample Size = 6280

Number of Missing Values	0	1	2	3	4	5	6	7	8	9	10
Number of Cases	2093	1677	1012	608	381	202	113	72	45	35	16
Number of Missing Values	11	12	13	14	15	16	17				
Number of Cases	16	3	2								

Notes: Items in boxes are included in the final model.

TABLE 8
Study 1: Measurement Model Results (N=6,280)

ITEM	MEAN	STANDARD DEVIATION	SERVICE QUALITY	MD	STAFF	ACCESS	SETTING	OUTCOME	VALUE
Q01b Overall quality of care	1.39	0.71	0.92						
Q01 Overall quality of care	1.36	0.69	0.84						
Q05 Thoroughness	1.55	0.83		0.91					
Q14 Medical skills	1.44	0.73		0.91					
Q09 Involved dec making	1.57	0.85		0.92					
Q07 Listening	1.54	0.84		0.91					
Q13 Courtesy/caring	1.43	0.75		0.90					
Q06 MD spends enough time	1.56	0.87		0.88					
Q10 Communication	1.56	0.85		0.92					
Q16 Instructions	1.60	0.86		0.89					
Q17 Explanations	1.57	0.85		0.92					
Q25 Staff teamwork	1.52	0.76			0.90				
Q26 Prompt information	1.65	0.86			0.85				
Q28 Privacy	1.42	0.66			0.85				
Q29 Courtesy	1.43	0.67			0.82				
Q23 RN	1.49	0.70			0.84				
Q35 ApptAccess	1.65	0.85				0.880			
Q38 QuestAccess	1.76	0.92				0.870			
Q40 Atmosphere	1.50	0.70					0.920		
Q41 Cleanliness	1.34	0.60					0.870		
Q45 Outcome	1.75	0.99						0.940	
Q44 Value	1.79	0.96							0.940
Composite reliability			0.874	0.972	0.930	0.867	0.890	0.884	0.884
Variance extracted			0.776	0.822	0.727	0.766	0.802	0.884	0.884
Cronbach's alpha			0.794	0.958	0.882	0.789	0.784		

Note: Table entries indicate completely standardized factor loadings. All indicators are significant at $p < .001$

The convergent validity and the reliability of each of the measures is strong. All the factor loadings were highly significant ($p < .001$) and substantively large (.84-.94), indicating convergent validity. Three reliability indexes were constructed to ascertain how well the constructs measured their indicators. Cronbach's alpha ranged from .78 to .96, exceeding the .7 benchmark (Nunnally 1978). The composite reliability of each construct was between .87 and .97. Between 74% and 88% of the variance in the items was captured by the respective measures, well above the 50% average variance extracted guideline (Fornell and Larcker 1981).

Discriminant validity was investigated using two approaches. First, using the approach suggested by Bagozzi, Yi, and Phillips (1991) and MacKenzie, Podsakoff and Jarvis (2005), each pair of constructs was tested to determine if "1" falls within the confidence intervals of the correlation estimates, implying perfect correlation and unidimensionality. In LISREL, the test was conducted by comparing the chi-square statistic for a measurement model in which the correlation was constrained to unity with an unrestricted measurement model. The correlation was found to be statistically different from 1 for all pairs, a predictable finding in light of the very small standard errors and, thus, narrow confidence intervals, observed throughout the model estimates. For the second approach, the average variance extracted (AVE) of each construct was compared with the squared correlation between construct pairs (Fornell and Larcker 1981). The AVE for each contributing factor should exceed the squared correlations. As Table 9 shows, the AVEs exceed the squared correlations for all but three pairs of

constructs. Because at least one procedure supports discriminant validity, I conclude that the scales reflect distinct constructs.

Two robustness checks of overall model fit were conducted. First, the measurement model was initially fit to an estimation sample (n=2,155) and then validated with a second sample (n=2,018) using methodology recommended by Cudeck and Browne (1983) in which the restricted covariance matrix created from the measurement model of the calibration sample is compared to the unrestricted covariance matrix for the validation sample. Fit was excellent (CVI <2.0). The first two samples were combined (n=4,173) to estimate the structural model which was later validated with the third bucket (n=6,280).

In addition, to guard against the possibility that model fit may be an artifact of outliers, the model was refit without outliers. Following Thompson (2004), the Mahalanobis distances were calculated for each case to identify outliers. Approximately 9% of the sample were selected as outliers (564 of 6,280). Model parameter estimates

TABLE 9
Study 1: Correlations Among Latent Constructs (N=6,280)

Construct	AVE	SERVICE QUALITY	MD	STAFF	ACCESS	SETTING	OUTCOME	VALUE
SVC_QLTY	0.79		0.69	0.76	0.66	0.56	0.76	0.74
MD	0.81	0.83		0.69	0.62	0.44	0.59	0.55
STAFF	0.74	0.87	0.83		0.79	0.74	0.64	0.66
ACCESS	0.77	0.81	0.79	0.89		0.61	0.64	0.64
SETTING	0.80	0.75	0.66	0.86	0.78		0.48	0.53
OUTCOME	0.86	0.87	0.77	0.80	0.80	0.69		0.72
VALUE	0.88	0.86	0.74	0.81	0.80	0.73	0.85	

Notes: Factor correlations appear in the lower triangle; squared correlations appear in the upper triangle.

All correlations are significant ($p < .01$).

The AVE for each contributing factor should exceed 1) the squared correlations and 2) 0.5 (Fornell and Larcker 1981).

Values in boxes do not meet that requirement but do meet the Bagozzi, Yi and Phillips (1991) test for discriminant validity.

and overall fit without the outliers were consistent with the full sample, suggesting outliers did not unduly influence model fit.

Discussion of measurement model findings. The measurement model shows limited support for the construction of service quality evaluations. Technical quality and interpersonal quality were construed as distinct dimensions, a hypothesis not supported by the data. However, the combined MD construct incorporated many of the subdimensions originally associated with technical and interpersonal quality. The measurement of MD supports H_{2a} (capabilities) and H_{2b} (process). The dataset did not include measures for H_{2c} (prior outcomes). H_{3a} (respectfulness) and H_{3b} (communication) were also supported.

The model also provided some support for the organizational quality construct. Physical surroundings (H_{4a}), convenience (H_{4b}), and support staff performance (H_{4c}) are hypothesized to be related to overall organizational quality. However, the convenience measure is narrower than hypothesized and incorporates only access convenience (appointment availability, getting questions answered). Time convenience (physician is on time for the appointment, the patient's time is well spent, the organization is efficient) does not fit into the model.

Estimating and Testing Models with Direct, Indirect and Moderating Effects

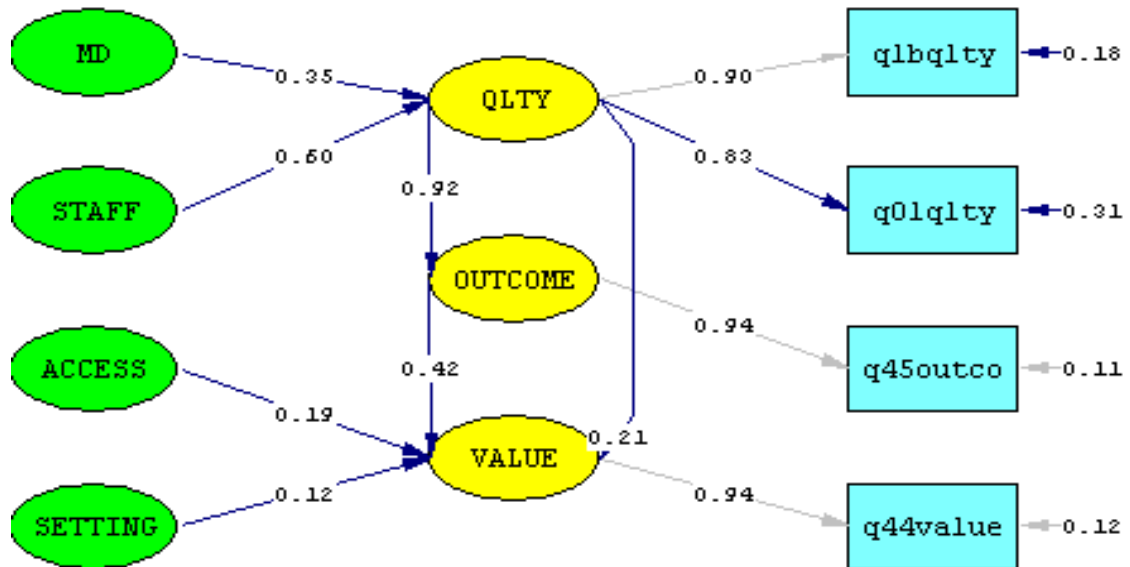
Parameter estimates and goodness of fit indicators for the structural model of service quality are strong. The model estimated with the full sample of 6,280 shows a χ^2 of 946.27, $df = 192$, CFI=1.00, and RMSEA=.025. The physician's technical and interpersonal attributes (MD) and auxiliary staff performance (STAFF) influenced

physician service quality (QLTY), as predicated by H_{1a} and H_{1b}. Service quality fully mediated the influence of those factors on outcome and on value, as predicted by H_{11d} and H_{13b}. Access convenience (ACCESS) and physical surroundings (SETTING), together with service quality and outcome, influenced value perceptions. The influence of service quality and outcome on value was predicted (H_{13a} and H_{13b}); the influence of organizational attributes on value was not anticipated. Following the procedures recommended by Baron and Kenny (1986), quality and outcome were found to fully mediate the influence of the core product attributes MD and support staff, on value. See Figure 5.

Discussion of structural model findings. The analysis provides some support for the measurement of overall service quality (QLTY). MD performance and auxiliary staff performance were found to influence overall service quality, as predicted in H_{1a}, H_{1b}, and H_{1c}. However, the introduction of access convenience and physical surroundings, also predicted by H_{1c}, led to sign reversals or reduced the loadings of the core attributes on quality. Instead, those two dimensions were better associated with value evaluations.

An augmented product level paradigm may provide insight into these findings. That paradigm distinguishes between core and augmented or peripheral features of a good or service (Kotler 2000). The core product incorporates the product and service features required to solve the buyer's fundamental problem. In this context, performance of the principal service provider and auxiliary staff may be most critical to providing patient care. The augmented features of a product, while not mission critical, provide an

FIGURE 5
Study 1: Structural Relationships

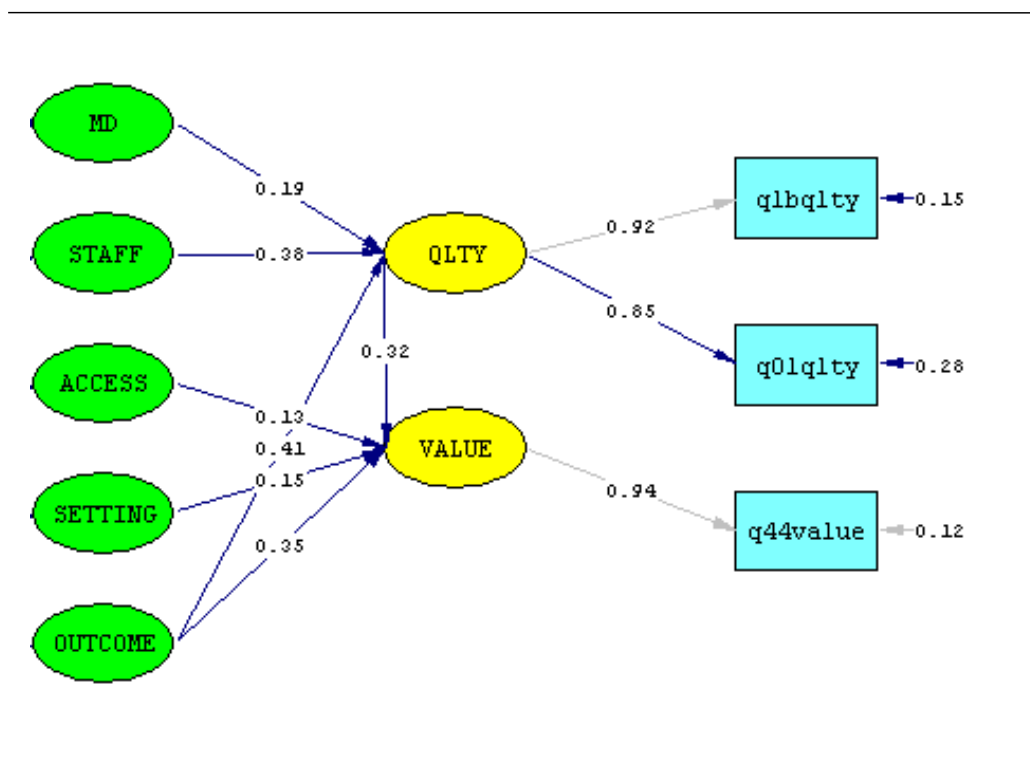


additional source of differentiation. In this case, access convenience and physical surroundings may be providing that role.

While both core and augmented attributes generally contribute to the overall evaluation of service quality (cf. Ozment and Morash 1994; Parasuraman et al. 1991), the relative influence of individual attributes on quality evaluation outcomes may depend on the criticality of the service encounter. When the outcome is highly consequential, the

performance of the core product is of paramount importance. Augmented benefits are “nice to have.” In low stakes situations, augmented or experiential components may become more influential in overall evaluations perhaps because buyers can afford the risk of poor core product performance (Ostrom and Iacobucci 1995). The present findings contribute to the existing literature, suggesting that for a highly consequential service, only core attributes contribute to quality evaluations, with augmented attributes influencing value.

FIGURE 6
Study 1: Rival Structural Model



Rival models. The estimated model shows outcome as a consequence of quality. It is plausible, however, that patients view quality through the lens of the outcome. Consumers construct preferences on the fly in difficult to evaluate arenas (Slovic 1995) and it is plausible that consumers also construct evaluations on the fly. In that case, available cues may be the hook consumers look for. Perceived outcome may be an available cue for many consumers. I tested a model in which outcome served as a predictor of quality and of value. That model provides a superior fit ($\chi^2 = 848.4$, $df = 188$, $p = 0$, $CFI = 1.0$, $RMSEA = .024$) (Figure 6). To the extent that this model better reflects consumer behavior, it suggests consumers use outcome as an indicator of quality.

Summary and Discussion

The exploratory empirical study provides some support for the conceptualized model. Based on the findings, it appears that in this highly consequential context, core product attributes are closely associated with overall service quality while augmented product attributes are associated with value. The technical and interpersonal components of the physician's behavior were unidimensional, contrary to expectations. Interestingly, the performance of the support staff have about twice the impact on overall quality evaluations as the principal service provider. That finding may be an artifact of the context, as the focal clinic is known for overall superior quality.

A rival model provides particularly intriguing insights, suggesting that quality evaluations may be made through the lens of outcome. Such a construction is consistent

with existing service quality models including SERVQUAL (Parasuraman et al. 1991) and Brady and Cronin (2001). The model is surprising with credence services because credence outcomes are not always indicative of quality. These findings suggest that people may prefer to ignore the complexity associated with accurate credence service quality evaluations and focus instead on outcomes.

A limitation to this study, as noted earlier, is that the database does not include several constructs included in the full conceptual model. Those constructs include: subdimensions of the conceptualized technical, interpersonal and organizational quality dimensions; trust; satisfaction; perceived switching costs. In addition, the multispecialty clinic has a reputation for quality, suggesting a full range of perceived quality is unlikely to be observed and limiting external generalizability. The second study is designed to address those limitations.

CHAPTER IV

PRODUCT EXPERTISE AND RISK (STUDY 2)

Study 2 tests a conceptual framework incorporating two paths for integrating technical, interpersonal, and organizational quality attributes to construe overall service quality in services high in credence attributes.

Model Formulation and Estimation

As with Study 1, my analysis includes four tasks: 1) SEM analysis of an overall model linking service provider performance characteristics with evaluations of outcome and perceived value; 2) mediation analysis of quality on outcome and perceived value and loyalty intentions; 3) robustness checks on the measurement and structural estimates; and 4) analysis of consumer heterogeneity. To estimate the relationships between service provider quality and its consequences, I estimated a measurement model that involved the antecedents and consequences of service quality and an SEM that associates service provider quality with outcome, value, satisfaction, trust and loyalty. For the mediation analysis, I performed partial and full mediation tests of quality on outcome, value, satisfaction, trust and loyalty using the methodology outlined by Baron and Kenny (1986). I evaluated the robustness of the estimates using two approaches. First, following Cudeck and Browne (1983), the restricted covariance matrix created from the measurement model of the calibration sample is compared to the unrestricted covariance matrix for the validation sample. Second, following Thompson (2004), the

final model was reestimated without outliers to determine the influence of outliers on model parameters and overall fit. All reported estimates reflect the full sample (n=1,379).

I performed the consumer heterogeneity analyses using a priori and post hoc segmentation. Biased estimates may result from SEMs conducted with an aggregate sample. To ferret out consumer heterogeneity, I first conducted multigroup SEM on known demographic (age, gender, health status) and psychographic (health self-efficacy), variables. Then I used finite mixture SEM analysis to search for unobserved consumer subgroups.

Data Sample

A stratified random sample was drawn from participants in a leading national online consumer research panel. The use of an online panel was chosen because of the superior opportunity for external generalizability. An attractive feature of the online panel is its national scope, which precludes the risk of local geographical biases if the panel had instead been drawn from a particular community. Also, because of the national scope, respondents are unlikely to evaluate the same physician or physicians working from the same clinic, providing the opportunity for more variability in responses. Online research databases are increasingly utilized for marketing research (cf., Bart et al. 2005). Online surveys offer several important advantages over mail surveys including speed of completion, flexibility in achieving sample quotas, and potentially higher response rates.

The sampling plan proceeded as follows. To increase the likelihood of familiarity with the topic, panelists were selected based on self-reports that they or someone in their

household had a chronic ailment. Two groups were selected. The first group included people with a self-reported history of a serious medical condition (primarily cancer and heart disease). The second group included people with mild to moderate chronic ailments, mainly allergies, bone or joint conditions, high cholesterol, hypertension, and stomach problems. To help ensure that memories were fresh, respondents were screened for a recent visit to a doctor. The low to moderate severity group reported visiting a physician in the past 4 weeks; the high severity group, past 3 months. The recency requirement was extended for the high risk group in anticipation that those patients would have had longer and more memorable relationships with the focal physician.

Invitations to participate in the survey were emailed to a nationally representative sample of 13,178 participants in the national consumer panel. The initial response rate was 28% including 43% for the high severity and 22% for the low to moderate severity group. Of the respondents, 38% met the recency qualification and completed the survey. The survey was closed once the quotas had been met. The final sample total 1,379 respondents including 990 with minor to moderate severity and 389 with high severity. Data were collected online in July 2007. IRB approval for the study was obtained.

Measurement

The measurement instrument including ratings of the technical, interpersonal and organizational components of quality as well as value, outcome, satisfaction, loyalty, and switching costs. Covariates included self-ratings of illness severity, self-expertise, health self-efficacy, and health status, along with demographic characteristics. See Appendix B for survey items. Data drawn from physician satisfaction surveys is typically highly

skewed and has limited variance. In order to promote measurement variance, performance ratings were measured with an 11-item Likert scale anchored by “one of the worst in the country” and “one of the best in the country” with the midpoint labeled “about average.” For all items, the negative rating (e.g. “worst in the country” or “disagree completely”) was shown first in order to attenuate respondents’ reluctance to provide negative ratings (Fowler 1995).

Respondents were required to answer all questions. Specifically, respondents were not permitted to skip questions or to answer “no opinion/don’t know.” While it is reasonable to assume that respondents would be able to rate the provider’s interpersonal relations (e.g. the provider’s listening skills, courtesy, and caring) and organizational performance (e.g. convenience, facility atmosphere), respondents were expected to struggle with the hypothesized indicators of technical quality (e.g. professional qualifications and adherence to best-practice procedures). One solution would be to offer a “no opinion” response choice for those items. A disadvantage of that solution is that it provides no insight into respondents’ latent evaluations about the provider’s performance in those hard-to-evaluate arenas. Instead, I chose to measure respondents’ assumptions about performance on technical quality, by asking about “*your impression of the doctor’s medical education and training*” (emphasis added).¹¹

The survey was designed to maximize breadth of domains at the expense of fine-grained analysis of some of the covered constructs. The implication is that reliability scores may be lower than they would be if more items were included about each

¹¹ Later in the survey, in order to gain some insight into the face validity of evaluations of technical competence, respondents’ specific knowledge about physician training and credentials was probed.

construct. However, as Feldman and Lynch (1988) observe, multiple measures of a construct create a different measurement problem because the thoughts activated in response to early judgments will crowd out new inputs. Later judgments will systematically reflect inputs to earlier judgments, creating an illusion of measurement reliability.¹²

Estimation. The analysis was conducted with structural equation modeling using Mplus (version 5). All items are treated as continuous, an appropriate assumption given the 11-point scale. As with Study 1, the data are highly skewed and platykurtic (peaked). Variances for each variable included in the model average 4.3, just 17% of the maximum possible variance from an 11-point measurement scale (Thompson 2006). Because of the violations of multivariate normality, normal theory ML was used with the Satorra-Bentler mean- and variance-adjustment of the χ^2 and the standard errors generated from normal theory ML (MPlus MLMV) (Flora and Curran 2004). The model was also fit with asymptotically distribution free (ADF) robust weighted least squares using a diagonal weight matrix with standard errors and mean- and variance-adjusted chi-square test statistic using a full weight matrix (Mplus's WLSMV). Parameter estimates were similar; however, the model failed to converge when subsamples were estimated, a common occurrence when an asymptotic covariance matrix is created from a small

¹² “[T]he problem of artifactual influence of early judgments on later judgments should be greatest when multiple rating scales are used to measure the same construct, as in (a) the assessment of reliability, (b) some applications of multitrait, multimethod analyses, and (c) causal modeling when each construct is measured by multiple indicators. In these cases, even if the multiple indicators are interspersed throughout a longer questionnaire, it is likely that responses to later items will be based on answers to earlier ones. In such cases, high levels of internal consistency can scarcely be interpreted to mean that measurement error accounts for only a small proportion of variance in the multi-item scale. The memory dynamics have produced an effect similar to that obtained by sampling only from a small part of the domain of potential items.” (p. 427)

sample. Following the recommendations of West, Finch and Curran (1995) and Mosteller and Tukey (1977) for dealing with negatively skewed data, a power transformation was applied to all measurement items. The transformation did not appreciably improve model performance, perhaps because of the extremeness of the skew. Results are reported for the untransformed data.

Measurement findings. Following Anderson and Gerbing (1988), the measurement model was evaluated first. Nine first order factors were identified from the measured attributes: service quality, Technical quality, Interpersonal quality, Organizational quality, Outcome, Satisfaction, Value, Trust and Loyalty. A tenth factor, Switching Costs,¹³ did not contribute to model fit and was dropped from analyses. All factors met tests for convergent and discriminant validity.

Factor composition and sample statistics are included in Table 10. Overall service quality was measured in multiple ways in an effort to better discriminate the construct from technical quality and the hypothesized consequences of quality evaluations. These include: i) two indicator measures (Q1 “overall quality of care” and MDQ “this doctor is one of the best in the country”); ii) constructed measure (average of the global measures of technical, interpersonal and organizational quality TQ1, TQ2, IQ1, IQ2, OQ1, OQ); iii) single indicator (Q1) without measurement error; and iii) single indicator (Q1) with measurement error. The last approach (single indicator with measurement error) provided the best discrimination and was retained. While using

¹³ Measured switching costs included time, money, and certainty of improvement. Construct reliability (.61) and average variance extracted (.2) were low, indicating lack of convergent reliability. The failure of the Switching Cost items to contribute to model fit suggests switching costs, while highly discussed in this domain, are less important than other constructs.

single indicators with measurement error is useful in structural equation modeling (Brown 2006), it is inappropriate to consider the adjusted item a factor with surplus meaning. The model fit is good. The chi-square statistic is 268.2 (df = 80, $p < 0.01$).¹⁴ The CFI (.971), RMSEA (.041) and SRMR (.024) are acceptable, with values better than the recommended cutoffs of .95 and .06 (Hu and Bentler 1999).

The convergent validity and the reliability of each of the measures is strong. All the factor loadings were highly significant ($p < .001$) and substantively large (.77-.95), indicating convergent validity. Three reliability indexes were constructed to ascertain how well the constructs measured their indicators. Cronbach's alpha ranged from .86 to .96, exceeding the .7 benchmark (Nunnally 1978). The composite reliability of each construct was between .87 and .96. Between 75% and 91% of the variance in the items was captured by the respective measures, well above the 50% average variance extracted guideline (Fornell and Larcker 1981).

Discriminant validity was investigated using two approaches. First, using the approach suggested by Bagozzi, Yi, and Phillips (1991) and MacKenzie, Podsakoff and Jarvis (2005), each pair of constructs was tested to determine if "1" falls within the confidence intervals of the correlation estimates, implying perfect correlation. The test was conducted in Mplus by constraining the correlation of each pair of constructs to unity and then performing a Wald chi-square test of the restriction. For each pair of constructs, the constraint was rejected, indicating discriminant validity. For the second

¹⁴ Mplus's estimators that perform the Satorra-Bentler correction for non-normality solve for "p." The reported χ^2 statistics cannot be used for χ^2 difference tests because the difference in SB corrected χ^2 s are not distributed χ^2 . I report χ^2 for completeness. But when making χ^2 comparisons, I follow the appropriate procedure recommended by the software.

TABLE 10
Study 2: Measurement Model Results (N=1,379)

ITEM	MEAN	STANDARD DEVIATION	QUALITY	TECHNICAL QUALITY	INTER- PERSONAL QUALITY	ORGANIZA- TIONAL QUALITY	SATIS- FACTION	OUTCOME VALUE	LOYALTY	TRUST	
QI	9.00	1.71	0.95								
MDEXP	9.11	1.90		0.87							
MDEDUC	9.22	1.77		0.88							
THORO	8.98	2.01		0.91							
REPUTE	9.06	1.66		0.80							
MDCOURT	9.51	1.79			0.90						
CARE	9.27	1.98			0.92						
INVDM	9.20	1.91			0.90						
LISTEN	9.22	2.01			0.95						
EXPLAIN	9.34	1.82			0.91						
EMPCOURT	8.94	2.02				0.84					
TMWRK	8.90	1.98				0.88					
APPTSOON	8.36	2.30				0.70					
ANSSOON	8.19	2.30				0.80					
COMFORT	8.72	1.91				0.79					
CLEAN	9.28	1.77				0.81					
EQPMT	8.92	1.88				0.83					
SAT	9.12	1.96					0.97				
SATMD	9.26	1.91					0.94				
OUTPHYS	8.83	1.96						0.89			
OUTEMOT	8.60	2.21						0.89			
VALUTIME	9.02	2.11							0.97		
VALUCASH	9.00	2.19							0.90		
RCMD	8.80	2.61								0.95	
LOYLOC	7.78	2.94								0.81	
TELLALL	8.87	2.24								0.85	
HELPS	8.98	2.15								0.97	
TRUSTMD	8.95	2.20								0.92	
Composite reliability			0.90	0.92	0.96	0.93	0.95	0.89	0.94	0.88	0.94
Variance extracted			0.90	0.75	0.84	0.65	0.91	0.80	0.88	0.78	0.83
Cronbach's alpha				0.93	0.96	0.93	0.95	0.88	0.94	0.87	0.94

Note: Table entries indicate completely standardized factor loadings. All indicators are significant at $p < .001$

approach, the average variance extracted (AVE) of each construct was compared with the squared correlation between construct pairs (Fornell and Larcker 1981). The AVE for each contributing factor should exceed the squared correlations. As Table 11 shows, the AVEs exceed the squared correlations for 25 of the 36 pairs of constructs. As in Study 1, the approaches yield different findings because the first approach leverages the large sample size and the small standard errors. Because at least one procedure supports discriminant validity, I conclude that the scales reflect distinct constructs.

Two robustness checks of overall model fit were conducted. First, the measurement model was initially fit to an estimation sample (n=699) and then validated with a second sample (n=680) using methodology recommended by Cudeck and Browne (1983) in which the restricted covariance matrix created from the measurement model of the calibration sample is compared to the unrestricted covariance matrix for the validation sample. Fit was excellent (CVI <3.2).

TABLE 11
Study 2: Correlations Among Latent Constructs (N=1,379)

Construct	AVE	QUALITY	TECHNICAL QUALITY	INTER- PERSONAL QUALITY	ORGANIZA- TIONAL QUALITY	SATIS- FACTION	OUTCOME	VALUE	LOYALTY	TRUST
QUALITY	0.90		0.89	0.70	0.56	0.69	0.68	0.61	0.57	0.64
TECHNICAL	0.75	0.94		0.82	0.65	0.82	0.80	0.73	0.66	0.75
INTERPERSONAL	0.84	0.84	0.91		0.60	0.80	0.78	0.71	0.62	0.73
ORGANIZATIONAL	0.65	0.75	0.81	0.78		0.71	0.70	0.65	0.48	0.53
SATISFACTION	0.91	0.83	0.91	0.89	0.84		0.95	0.91	0.68	0.75
OUTCOME	0.80	0.83	0.89	0.88	0.84	0.97		0.83	0.61	0.76
VALUE	0.88	0.78	0.85	0.84	0.80	0.95	0.91		0.62	0.69
LOYALTY	0.78	0.76	0.82	0.79	0.69	0.83	0.78	0.79		0.65
TRUST	0.83	0.80	0.87	0.86	0.73	0.87	0.87	0.83	0.81	

Notes: Factor correlations appear in the lower triangle; squared correlations appear in the upper triangle.

All correlations are significant ($p < .01$).

The AVE for each contributing factor should exceed 1) the squared correlations and 2) 0.5 (Fornell and Larcker 1981).

Values in boxes do not meet that requirement but do meet the Bagozzi, Yi and Phillips (1991) test for discriminant validity.

In addition, to guard against the possibility that model fit may be an artifact of outliers, the model was refit without outliers. Following Thompson (2004), the Mahalanobis distances were calculated for each case to identify outliers. Approximately 11% of the sample were selected as outliers (150 of 1,379). Model parameter estimates and overall fit without the outliers were considerably weaker than with the full sample, suggesting outliers have influence model fit. Multigroup analyses is used to shed light on the outliers, as discussed below.

Measurement model findings. As predicted, technical quality is measured through a combination of perceptions of capabilities (education, experience), processes (thoroughness) and prior results (reputation), providing support for H₂. As predicted by H₃, interpersonal quality is measured through two facets: respectfulness (courtesy, caring, involved decision making) and communication (listening, explaining). Consistent with H₄, organizational quality incorporates support staff performance (employee courtesy, teamwork), convenience (getting an appointment and getting answers to telephone questions as soon as needed), and physical surroundings (comfort, cleanliness, and condition of equipment). However, convenience is measured more narrowly than expected. Temporal convenience (avoiding long waits) and locational convenience were not associated with organizational quality.

Formative vs. reflective indicator models. Formative indicator models were fit for technical quality and organizational quality as rival models to reflective indicator models. With formative models, the direction of causality is from the observed measures to the factor. Both constructs meet the conceptual test for formative constructs (MacKenzie et

al. 2005): indicators are not necessarily correlated, dropping an indicator may change the meaning of the construct; and increases/decreases in one indicator may be expected to influence the construct. Such a specification seems plausible on both counts. For example, for technical quality, a physician's adherence to best practices might offset weak credentialing. For organizational quality, a friendly and professional support staff may practice in an inconveniently located or unattractive facility. To meet model identification requirements, factors were measured by two global measures (TQ1/TQ2 and OQ1/OQ2, respectively).

For technical quality, the formative model fit the data poorly, not supporting H₃. See Table 12. Similar results were observed when the model was fit only with experts. Poor fit may be evidence of a poorly specified model. In this case, however, it is more plausible to interpret this as evidence that technical quality is a reflective construct. Classical factor analysis assumes that factors are attitudes, a construal that seems appropriate for a difficult-to-evaluate construct. An analysis of answers to related questions support this reasoning. Note that respondents rated the provider's medical training on an 11-point scale anchored by best in the country/worst in the country. Later, as a validity check, respondents were asked if they knew where the provider was trained. It seems reasonable that respondents who did not know where the physician's received medical training would have offered neutral ratings on that item. Instead, the physician's medical training ratings correlated highly with overall technical quality ratings, regardless of whether the respondent knew where the provider was trained ($r=.83$ for respondents who knew where the physician was trained and $r=.84$ for those who did not

TABLE 12
Formative and Reflective Indicator Models

A: TECHNICAL QUALITY				
Indicators	Full panel n=1,379		Experts n=446	
	Reflective	Formative	Reflective	Formative
MDEDUC	0.87	0.42	0.87	0.47
MDEXP	0.88		0.85	
THORO	0.91	0.25	0.90	0.23
REPUTE	0.80	0.41	0.80	0.40
χ^2	268.20	417.00	72.70	112.50
df	80.00	93.00	37.00	46.00
p	0.00	0.00	0.00	0.00
CFI	0.97	0.96	0.95	0.93
RMSEA	0.04	0.05	0.05	0.06
SRMR	0.02	0.03	0.04	0.04
B: ORGANIZATIONAL QUALITY				
Indicators	Full panel n=1,379			
	Reflective	Formative		
EMPCOURT	0.84	0.27		
TMWRK	0.88	0.39		
APPTSOON	0.70	0.08		
ANSSOON	0.80			
COMFORT	0.79	0.14		
CLEAN	0.81			
EQPMT	0.83	0.16		
χ^2	268.20	208.80		
df	80.00	62.00		
p	0.00	0.00		
CFI	0.97	0.97		
RMSEA	0.04	0.04		
SRMR	0.02	0.04		

Notes: All loadings are significant $p < .001$.

know). This suggests credential and training ratings are inferred from global attitudes, rather than contributing to global attitudes.

For organizational quality, the formative model provided a reasonable fit for the data with a slight improvement in chi-square based goodness of fit statistics but a slight worsening in SRMR, supporting H₄. Indicator coefficients were smaller, possibly reflecting the high inter-indicator correlations. Note that, because the reflective model is less cumbersome to work with, the reflective model was retained for other analyses.

Direct, Indirect and Moderating Effects

The structural model was evaluated for model identification, fit (CFI, RMSEA, SRMR) and plausible rival models. The model was initially fit with a calibration sample and adjusted. The revised model was validated with a holdout sample (n=699). The model fit the validation sample well and then the model was applied to the entire sample. The model fit is good. The chi-square statistic is 393.1 (df=84, p<0.001). The CFI (.953), RMSEA (.052) and SRMR (.053) are acceptable, with values better than the recommended cutoffs of .95 and .06 (Hu and Bentler 1999). Figure 7 illustrates the relationships and Table 13 reports standardized coefficients and fit statistics. As predicted by the trust heuristic, interpersonal and organizational quality influenced trust (H_{5a} and H_{5b}). Trust influenced technical quality as expected in H_{5c}. Technical quality is closely associated with overall quality, supporting H_{1a}. Interpersonal and organizational quality were not associated with overall quality, contrary to the predictions of H_{1b} and H_{1c}. This finding is similar to that in Study 1, further suggesting that, in highly consequential services, reliability dominates all other considerations. Outcome is associated with

evaluations of quality quality, supporting H₁₁. Surprisingly, quality does not mediate the relationship of technical quality on satisfaction, as predicted by H_{12a}. Instead satisfaction is influenced by quality mediated through outcome, and also by interpersonal quality (H_{12b}) and value (H_{12d}).

Organizational quality had a statistically significant effect on satisfaction, but the effect size is small (H_{12c}). Value is influenced by outcome and trust, supporting H_{13a} and H_{13c}, but not interpersonal quality (H_{13b}). Loyalty is influenced by trust (H_{14d}) and satisfaction (H_{14b}), but not by outcome (H_{14a}) or value (H_{14c}). Switching costs were not influential, contrary to expectations (H15).

FIGURE 7
Study 2: Structural Model (Revised, n=1329)

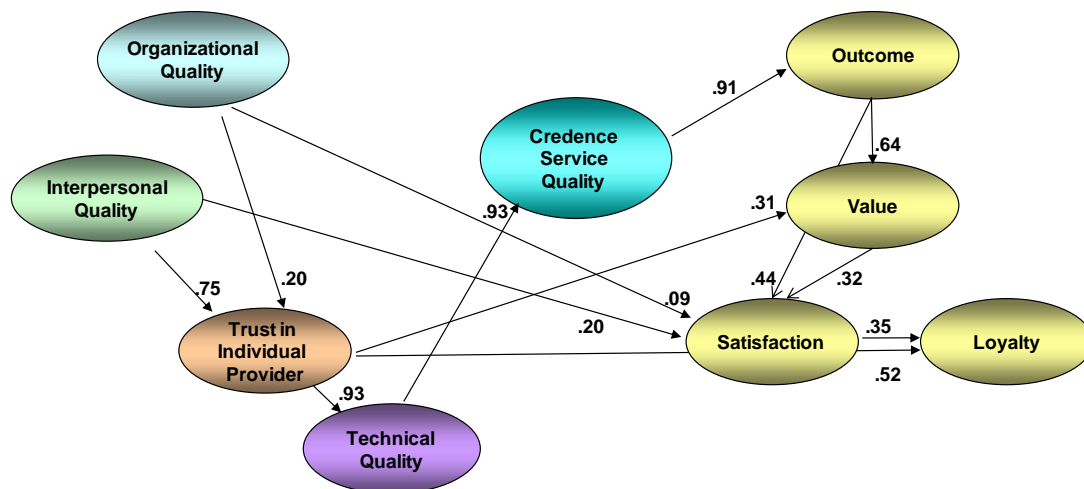


TABLE 13

Study 2: Standardized Coefficients and Fit Statistics for the Proposed Model, the Revised Model, and the Replication Analysis

			Expected Sign	Proposed Model	Revised Model	Replication Analysis	Full Sample
H1a	Technical quality	--> Overall Quality	+	0.793 ***	0.912 ***	0.947 ***	0.929 ***
H1b	Interpersonal Quality	--> Overall Quality	+	0.066			
H1c	Organizational Quality	--> Overall Quality	+	0.141 ***			
H7a	Interpersonal Quality	--> Trust	+	0.738 ***	0.733 ***	0.771 ***	0.752 ***
H7b	Organizational Quality	--> Trust	+	0.192 ***	0.212 ***	0.176 ***	0.195 ***
H7c	Trust	--> Technical Quality	+	0.898 ***	0.925 ***	0.923 ***	0.925 ***
H11	Quality	--> Outcome	+	0.881 ***	0.908 ***	0.915 ***	0.911 ***
H12a	Quality	--> Satisfaction	+	0.162 ***			
	Outcome	--> Satisfaction	+		0.457 ***	0.400 ***	0.444 ***
H12b	Interpersonal Quality	--> Satisfaction	+	0.246 ***	0.195 ***	0.198 ***	0.204 ***
H12c	Organizational Quality	--> Satisfaction	+	0.154 ***	0.114 ***	0.087 ***	0.094 ***
H12d	Value	--> Satisfaction	+	0.493 ***	0.297 ***	0.375 ***	0.318 ***
H13a	Outcome	--> Value	+	0.579 ***	0.624 ***	0.684 ***	0.643 ***
H13b	Interpersonal Quality	--> Value	+	0.134 ***			
H13c	Trust	--> Value	+	0.243 ***	0.301 ***	0.291 ***	0.306 ***
H14a	Outcome	--> Loyalty	+	-0.105			
H14b	Satisfaction	--> Loyalty	+	0.286 **	0.263 ***	0.435 ***	0.345 ***
H14c	Value	--> Loyalty	+	0.121			
H14d	Trust	--> Loyalty	+	0.554 ***	0.593 ***	0.441 ***	0.521 ***
H15	Switching costs	--> Loyalty	+	0.000			
	df			91	91	54	84
	χ^2			324.5	292.0	155.0	393.1
	p			0.000	0.000	0.000	0.000
	CFI --> Loyalty			0.934	0.944	0.956	0.953
	RMSEA			0.061	0.057	0.052	0.052
	SRMR			0.060	0.053	0.058	0.053
	N			680	680	699	1379

NOTES: *** p<.001 ** p<.01 * p<.05

TABLE 14
Study 2: Standardized Coefficients and Fit Statistics for the Revised Model and Rivals

			Expected Sign	Trust Heuristic Trust as Antecedent	Rival 1 Trust as Consequence	Rival 2 Trust as Consequence -- No Overall Quality
H1a	Technical quality	--> Overall Quality	+	0.929 ***	0.929 ***	
H5a	Interpersonal quality	--> Trust	+	0.752 ***	0.393 ***	0.393 ***
H5b	Organizational quality	--> Trust	+	0.195 ***		
	Outcome	--> Trust	+		0.542 ***	0.542 ***
H5c	Trust	--> Technical Quality	+	0.925 ***		
H8	Overall Quality	--> Outcome	+	0.911 ***		
	Technical quality	--> Outcome	+		0.924 ***	0.593 ***
	Interpersonal quality	--> Outcome	+			0.361 ***
	Outcome	--> Satisfaction	+	0.444 ***		0.931 ***
H9b	Interpersonal quality	--> Satisfaction	+	0.204 ***	0.304 ***	
H9c	Organizational quality	--> Satisfaction	+	0.094 ***		0.078 ***
H9d	Value	--> Satisfaction	+	0.318 ***	0.715 ***	
H10a	Outcome	--> Value	+	0.643 ***	0.802 ***	0.941 ***
	Organizational quality	--> Value	+		0.188 ***	
H10c	Trust	--> Value	+	0.306 ***		
	Overall Quality	--> Loyalty	+		0.188 ***	
	Technical quality	--> Loyalty	+			0.375 ***
H11b	Satisfaction	--> Loyalty	+	0.345 ***	0.372 ***	
H11c	Value	--> Loyalty	+			0.174 ***
H11d	Trust	--> Loyalty	+	0.521 ***	0.331 ***	0.341 ***
H11e	Switching costs	--> Loyalty	+			
			df	84	85	78
			χ^2	393.1	331.6	286.7
			p	0.000	0.000	0.000
			CFI	0.953	0.962	0.966
			RMSEA	0.052	0.046	0.044
			SRMR	0.053	0.029	0.030
			N	1379	1379	1379

NOTES: *** p<.001 ** p<.01 * p<.05

Rival Models

Two rival models provide intriguing insights (see Table 14). First, a trust-as-consequence to quality model was constructed in contrast to the trust heuristic model in which trust is an antecedent of quality. Model fit is acceptable ($\chi^2 = 346.05$, $df=78$, $p<0.001$, $CFI = .957$, $RMSEA=.050$, $SRMR=.044$) and marginally better than the trust heuristic model.

The second rival involved dropping overall quality from the model. The rationale for doing so is as follows. First, quality is playing a more limited role in the model than expected. In Study 1, the quality construct mediated only the principal service provider's performance, not the performance of other organizational attributes as is common in the service quality literature (Parasuraman et al. 1991). Second, similarly, in Study 2, only the principal service provider's technical skills influence quality. Third, because quality is measured with a single indicator, the item cannot be construed to be a factor with surplus meaning. Accordingly, I dropped overall quality and refit the model. Fit is comparable ($\chi^2= 277.1$, $df=78$, $p<0.001$, $CFI =.968$, $RMSEA=.043$, $SRMR=.028$). Because the model is more parsimonious, I used it for the investigations of the roles of product expertise and risk in quality evaluations described below.

Latent class analysis of hidden consumer heterogeneity was unsuccessful, perhaps because the models are empirically unidentified. Latent class models are identified if the known groups models are identified and if the unknown groups are distributed multivariate normal, an unlikely condition with this dataset. Instead, I restrict the investigation of consumer heterogeneity to evaluations of known groups.

Role of Product Expertise in Quality Evaluations

Multigroup analysis was used to identify the influence of expertise. Expertise was measured based on self-reports (“how knowledgeable do you feel about the condition or problem this doctor has evaluated you for?”)¹⁵ The high expertise group (self rating 7 out of 7; n= 446) was contrasted with the low expertise group (self rating \leq 4, n = 237).

The measurement model fit both groups well, demonstrating (partial) strong factorial invariance. Specifically, the factor structure and unstandardized factor loadings are invariant across groups and indicator intercepts differed for only three indicators (Tables 15). However, factor variances and latent means were substantively different (Table 16). Factor means were higher for the experts. Experts appear more certain of their ratings, with factor variances about half that of novices, supporting H_{9b}. Experts also stated that they were confident of their ratings, further supporting H_{9b}.

I had expected trust to drive the formation of technical quality inferences for novices (H₈). That hypothesis was tested in multiple ways. First, both the trust heuristic model and the trust-as-a-consequence-of-quality model were fit to both novices and to experts. The trust-as-a-consequence-of-quality of quality model slightly outperformed the trust heuristic model for both groups by approximately the same amounts, providing no support for the hypothesis. Additionally, novices’ and experts’ path coefficients were

¹⁵ Other measures that could shed light on expertise are duration (“how long have you been bothered by the problem or condition this doctor has treated you for”) and MDSSEEN (“how many doctors have you seen for the same condition or problem that you have seen this doctor for?”). Pairwise correlations among all items is statistically significant ($p < .01$). Self-reported expertise appeared most influential for because it is most highly correlated with confidence in evaluating the physician’s technical skills. Moreover, the other measures (duration and number of doctors seen) are measures of familiarity rather than expertise.

TABLE 15
Invariance Testing - Experts and Novices

	χ^2 SB	df	p >.05	CFI >.95	RMSEA <.06	SRMR <.08	n	Observations
MEASUREMENT INVARIANCE	1 Does model fit all groups?							
	Low (Expertise <=4)	110.1	60	0.000	0.95	0.06	237	No evidence of local strain
	Med (Expertise = 5)	123.4	67	0.000	0.96	0.05	323	No evidence of local strain
	High (Expertise =7)	72.5	36	0.000	0.95	0.05	446	No evidence of local strain
	2 Are the FACTOR STRUCTURES EQUAL across groups? (Equal form, configural invariance)							
		256.1	134	0.00	0.95	0.05	0.04	Yes, equal form Good fit
	3 Are the unstandardized FACTOR LOADINGS EQUAL across groups? (Metric invariance; weak factorial invariance)							
	<i>Model 3 - Model 2</i>	32.0	25	0.00 0.16	0.96	0.05	0.04	Yes, equal loadings DIFFTEST
	4 Are the INDICATOR INTERCEPTS EQUAL across groups? (Scalar invariance; strong factorial invariance)							
	Partial invariance <i>Model 4 - Model 3</i>	44.30	32	0.00 0.07	0.96	0.05	0.04	No, indicator intercepts are not the same Intercepts are different for: reputation, explaining (High only), emotional outcome DIFFTEST
POPULATION HETEROGENEITY	6 Are the FACTOR VARIANCES equal across groups?							
	<i>Model 6- Model 4</i>	3.9	1	0.00 0.05	0.96	0.04	0.04	No, unequal factor variances for: all factors except OQ Implication: comparisons of factor covariances not feasible
	7 Are the LATENT MEANS equal across groups?							
<i>Model 6- Model 5</i>			0.95 0.00	0.05	0.05	0.04	No, unequal latent means	

* Mplus's DIFFTEST procedure used to obtain a correct chi-square difference test. Required because the difference in chi-square values for models estimated with mean- and variance-adjusted chi-square test statistics are not distributed as chi-square.

TABLE 16
Role of Product Expertise in Quality Evaluations

A: Latent Factor Means and Variances						
Latent Constructs	Factor Means		Factor Variances			
	Novice	Expert	Novice	Expert	Δ	
TQ	0.0	2.30	3.60	1.80	1.80	
IQ	0.0	2.00	3.80	1.40	2.40	
OQ	0.0	1.70	3.30	2.40	0.90	
Trust	0.0	2.20	4.10	2.40	1.70	
Outcome	0.0	2.30	3.90	1.70	2.20	
Satisfaction	0.0	2.30	5.10	2.00	3.10	
Value	0.0	2.40	6.00	2.90	3.10	
Loyalty	0.0	2.80	8.30	3.80	4.50	
			←-----			

B: Path Coefficients				C: Overall Model Fit (Rival Models)			
	Novice		Expert	Δ	Novice		Expert
	Outcome					n	237
TQ	0.57	0.46	0.11		Trust-as-a-consequence of quality		
IQ	0.37	0.47	-0.10		χ^2 (df)*	115.1 (61)	76 (37)
Trust					p	0.000	0.000
Outcome	0.75	0.58	0.17		CFI	0.948	0.944
IQ	0.16	0.32	-0.16		RMSEA	0.061	0.049
Satisfaction					SRMR	0.038	0.042
Outcome	0.63	0.86	-0.23		Trust heuristic		
OQ	0.23	0.15	0.08		χ^2 (df)*	125.1 (62)	84.3 (37)
IQ	0.17	0.00	0.17		p	0.000	0.000
Value					CFI	0.939	0.932
Satisfaction	0.94	0.89	0.05		RMSEA	0.066	0.054
Loyalty					SRMR	0.046	0.056
Value	0.36	0.15	0.21				
Trust	0.34	0.20					
TQ	0.23	0.45	-0.22				

Notes: Coefficients in bold are significant $p < .01$

examined to identify the importance placed on technical quality in the evaluation of outcomes, satisfaction and loyalty. Both groups determine outcome valence from technical and interpersonal quality evaluations; surprisingly, novices place greater weight on technical quality and less on interpersonal quality than do experts, again in contrast with the prediction of H₈. However, novices rely more heavily on peripheral cues – interpersonal quality and organization quality – when judging satisfaction; experts depend on outcomes, providing some support for H₈. Both groups look to value, trust, and technical quality to form loyalty intentions, with novices favoring value and experts paying particular attention to technical quality, again supporting H₈.

In a more direct test of the hypothesis, I conducted a multigroup analysis in which TQ was modeled formatively. If experts pay more attention to TQ indicators, then experts' coefficients on TQ formative indicators should be higher than novices' coefficients. Coefficients for both groups were approximately the same, providing no support for the hypothesis (see Table 12).

Role of Risk in Quality Evaluations

Multigroup analysis was used to identify the influence of perceived risk associated with the service. Risk was measured based on self-reports (“thinking about the most important problem or condition you have seen this doctor for, would you say the condition is [not too serious/life threatening]?”)¹⁶ The moderate risk group (rating 5

¹⁶ Another measure that could shed light on risk is MDIMP (“thinking about the most important problem or condition you have seen this doctor for, how important is it to you to choose a very highly qualified doctor”). The selected measure appears more informative because there is more variance.

or 6 out of 11, $n = 288$) was contrasted with the high risk group (rating = 10 or 11 out of 11, $n=305$) and the low risk group (rating ≤ 4 , $n = 157$). A common factor structure fit across the three groups. Factor loadings were largely invariant with the exception of outcome indicators in the high risk group. Indicator intercepts were invariant except for two technical quality indicators in the high risk group (see Table 17). Factor variances differed notably with the smallest variances in the high risk group and the largest in the low group, suggesting less certainty in the low risk group (see Table 18, panel A).

H_{10} proposes an inverted U-shaped relationship between perceived risk and the use of systematic processing. An examination of path coefficients shows technical quality evaluations influence outcome more heavily for the mid group than the low or high risk groups (see Table 18, panel B). Similarly, interpersonal quality evaluations are less important for the mid group relative to the others, again suggesting a heavier focus on cognitive reasoning. Other evidence of heightened cognitive reasoning among the mid group includes the heavy emphasis placed on outcome in reaching satisfaction evaluations. Loyalty evaluations are also sensitive to risk. Trust and value influence loyalty for the low risk group; technical quality and trust for the high risk group; and all three factors are important for the mid group. The trust-as-an-antecedent of quality model had a reasonably good fit for all three groups, but weaker than the trust-as-a-consequence of quality model (see Table 18, panel C). In summary, the results support H_{10} .

TABLE 17
Invariance Testing - Perceived Risk

	χ^2 SB	df	p >.05	CFI >.95	RMSEA <.06	SRMR <.08	n	Observations
MEASUREMENT INVARIANCE	1 Does model fit all groups?							<i>Yes, model fits all groups</i>
	Low (Severity <= 3)	58.9	41	0.035	0.97	0.05	157	No evidence of local strain
	Mid (Severity = 5 or 6)	89.7	54	0.002	0.96	0.05	288	No evidence of local strain
	High (Severity >=10)	52.5	32	0.013	0.96	0.05	305	No evidence of local strain
	2 Are the FACTOR STRUCTURES EQUAL across groups? (Equal form, configural invariance)							<i>Yes, equal form</i>
		181.1	114	0.00	0.96	0.05	0.03	<i>Good fit</i>
	3 Are the unstandardized FACTOR LOADINGS EQUAL across groups? (Metric invariance; weak factorial invariance)							
	<i>Model 3 - Model 2</i>	27.5	23	0.00	0.96	0.04	0.04	<i>Different factor loadings for High Risk Outcome indicators DIFFTEST</i>
	4 Are the INDICATOR INTERCEPTS EQUAL across groups? (Scalar invariance; strong factorial invariance)							
	Partial invariance			0.00	0.97	0.05	0.04	<i>Intercepts are different for High Risk group: REPUTE and</i>
<i>Model 4 - Model 3</i>	48.10	35	0.07				<i>DIFFTEST</i>	
POPULATION HETEROGENEITY	6 Are the FACTOR VARIANCES equal across groups?							
				0.00	0.97	0.05	0.10	<i>Different factor variances for: Mid (TQ), High (TQ, Trust,OUT,Satn, Value)</i>
	<i>Model 6- Model 4</i>	53.7	40	0.07				<i>Implication: comparisons of factor covariances not DIFFTEST</i>
	7 Are the LATENT MEANS equal across groups?							
<i>Model 6- Model 4</i>	93.2	41.0	0.00	0.95	0.05	0.05	0.04	<i>No, unequal latent means DIFFTEST</i>

* Mplus's DIFFTEST procedure used to obtain a correct chi-square difference test. Required because the difference in chi-square values for models estimated with mean- and variance-adjusted chi-square test statistics are not distributed as chi-square.

TABLE 18
Role of Perceived Risk in Quality Evaluations

A: Latent Factor Means and Variances							
Latent Factors	Factor Means			Factor Variances			
	LOW	MID	HIGH	LOW	MID	HIGH	
TQ	0.00	-0.03	1.00	3.78	2.56	1.70	
IQ	0.00	-0.02	0.71	2.89	2.44	2.02	
OQ	0.00	-0.02	0.74	3.64	2.48	2.52	
Trust	0.00	0.03	1.31	4.87	3.50	1.75	
Outcome	0.00	-0.14	0.99	3.77	3.41	1.90	
Satisfaction	0.00	-0.24	0.77	4.19	3.63	2.50	
Value	0.00	-0.06	0.97	5.16	3.99	3.07	
Loyalty	0.00	-0.07	0.99	6.29	6.86	5.17	

←-----

B: Path Coefficients				C: Overall Fit in Rival Models			
	Low	Mid	High	Low	Mid	High	
Outcome				Trust-as-a-consequence of quality			
TQ	0.67	0.96	0.53	χ^2 (df)*	61.30 (41)	94.40 (56)	53.40 (32)
IQ	0.22	0.00	0.43	p	0.00	0.00	0.00
				CFI	0.96	0.97	0.96
Trust				RMSEA	0.06	0.05	0.05
Outcome	0.44	0.31	0.39	SRMR	0.04	0.03	0.04
IQ	0.48	0.60	0.54				
				Trust-as-an-antecedent of quality			
Satisfaction				χ^2 (df)*	70.10 (42)	113.20 (56)	0.62 (33)
Outcome	0.72	0.88	0.75	p	0.00	0.00	0.00
OQ	0.11	0.13	0.13	CFI	0.95	0.95	0.94
IQ	0.18	0.00	0.16	RMSEA	0.07	0.06	0.05
				SRMR	0.06	0.04	0.07
Value							
Satisfaction	0.96	0.96	0.91				
Loyalty							
Value	0.30	0.31	0.07				
Trust	0.59	0.26	0.31				
TQ	0.07	0.29	0.42				

Notes: Coefficients in bold are significant $p < .01$

Risk groups are based on self-reported severity. 11 pt scale anchored by "not too serious" and "life threatening."
HI risk group = 10 or 11. Med risk group = 5 or 6. Low risk group <4.

Summary and Discussion

The results shed light on the processes people use to evaluate unobservable technical quality and the consequences of those evaluations. The data show support for both the trust heuristic model and the rival trust-as-a-consequence-of-quality model, suggesting there may be two paths to processing evaluations, as hypothesized. There was

little support for the hypothesis that experts rely on systematic processing while novices use heuristics. This may be a result of the sample. Experts are defined as individuals with domain-specific training and experience (Spence and Brucks 1997). For this research, experts were identified based on a self-rating. It is plausible that the self-rating reflects familiarity rather than expertise. Finally, there is support for the risk hypothesis, which suggests that people facing moderate illness severity seek out predictive indicators of technical quality while those not at risk and, counterintuitively, those facing high risk rely on heuristic processing.

Overall quality played a limited role in both the trust heuristic model and the trust-as-a-consequence-of-quality model. This may suggest that the concept of overall quality is less meaningful in a highly consequential credence service.

CHAPTER V

SUMMARY AND CONCLUSIONS

Evaluating credence services is a complex judgment process. Evaluators must combine multiple, elusive, fallible, and often conflicting clues to reach an overall quality evaluation. The task is a challenge for people from all walks of life. Parents grapple with choosing the right schools for their children. Sophisticated hedge fund investors fall prey to allegedly fraudulent statements by Wall Street investment bankers. The goal of this dissertation is to begin to understand how people judge the quality of credence services and the consequences of those evaluations. I chose to begin my investigation in the context of physicians' services, a classic and highly important credence service.

I created a conceptual framework with credence service quality conceptualized as a third-order hierarchical construct with three second-order dimensions: the *technical* quality of the primary service provider, the *interpersonal* quality of the primary service provider and the *organizational quality* incorporating the support staff, convenience, and the physical surroundings. Drawing on dual-process social information processing theory, I expected skill and motivation to provide boundary conditions. Accordingly, the influence of product expertise (novice-expert) and risk were examined.

The model was refined through a qualitative pilot study and then tested in two large scale empirical studies. Study 1 examines consumer perceptions of the quality of credence services. I draw on an existing large database (over 6,000 records) of patient

satisfaction with physician services in a large multispecialty clinic to measure credence service quality and estimate the consequences of service quality evaluations. The same relationships were examined in Study 2 with a panel of 1,379 consumers, providing a test of the generalizability of the initial findings. Study 2 also provided a test of the hypothesis of two paths for integrating technical, interpersonal, and organizational quality attributes to construe overall credence service quality.

Both datasets consisted of patient ratings of a recently visited physician. In a demonstration of the Lake Wobegon effect,¹⁷ nearly all of the physicians evaluated were exceptional in the eyes of their reporting patients. Respondents in the first study rated their physicians' overall quality as 4.5 out of 5-points (s.d. = .7); respondents in the second study delivered an average physician quality rating of 9 out of 11-points (s.d.= 1.7). The outstanding physician performance reported by patients in the first study may be a reflection of the study context. The evaluated physicians are associated with a world-class medical clinic which has a reputation for employing highly qualified physicians. Moreover, in that study, patients rated physicians on a 5-point scale anchored by "excellent" and "poor". The limited number of ratings options and the anchor terminology (perhaps all the physicians *are* excellent) may have dampened variability.

The sampling plan for the second study was designed to address those limitations and to maximize the opportunity for variability in patient's ratings of physician performance. Data were collected through an online national consumer research panel,

¹⁷ Lake Wobegon, created by writer and radio presenter Garrison Keillor, is a fictional community in rural Minnesota where "all the women are strong, all the men are good-looking, and all the children are above average."

minimizing the likelihood that only patients with superior physicians would be selected for the study. Moreover, the survey instrument used an eleven-point rating scale anchored by “one of the best in the country” and “one of the worst in the country.” The best/worst anchors were chosen over other terminology, such as excellent/poor, to encourage participants to compare their physician against the universe of physicians. Yet once again, respondents rated their physicians very highly.

What explains this apparent ratings bias? Respondents agreed that it was important to find a highly qualified physician to treat the particular problem or disorder they saw this physician for (9.4 out of 11 points, s.d.=1.9). And they do not think “all doctors are the same” (agreement 4.0 out of 11 points, s.d. = 2.7). It is possible, then, that the high ratings reflect respondents’ attempts to avoid the cognitive dissonance associated with giving only moderate performance ratings to a personally important service provider.

Alternatively, the strong ratings may reflect a leniency bias or a halo bias. These biases are correlation heuristics offering people an economical solution to a complex evaluation while providing desirable cognitive consistency. With a leniency bias, respondents rate people they like more favorably (Podsakoff et al. 2003). The data provide some support for this hypothesis. In the second study, respondents’ agreement with the statement “I like this doctor” explained 49% of the variance in overall quality ratings.

With a halo bias, raters’ overall impressions spill over to individual attributes or raters’ evaluations of one attribute activate impressions of another attribute (Murphy et

al. 1993). In marketing, halo errors are believed to occur when consumers use observable attributes to infer unobservable attribute qualities. Halo errors are fostered when unreasonable demands are placed on the rater, such as asking raters to make fine-grained evaluations of unfamiliar attributes (Murphy et al. 1993). These conditions existed in the study, as patients were asked to make fine-grained evaluations of unobservable technical quality, among other items.¹⁸

While study respondents showed only limited variability in their ratings, respondents appeared to be interested in the study topic. The overall response rate to participating in the studies is high. The marketing research firm that collected the data for Study 1 reports that 90% of the patients they are able to contact by telephone choose to participate in the study. Response rates were also high for Study 2. Invitations to participate in the survey were emailed to 13,178 participants in the national online consumer panel. Within 48 hours, twenty-eight percent of those invited agreed to participate. The response rate would likely have been even higher had the quota not been filled so rapidly.

Summary of Research Findings

The dissertation was designed to address four questions:

¹⁸ While halo biases have been considered to reflect suboptimal thinking and decision error, recent research suggests halo biases lower the rater's evaluation risk (that is, the variance). Specifically, when important attributes are unobservable, the heuristic narrows the gap between a consumer's belief and the true value of the product compared to trying to infer the missing information directly (Boatwright et al. 2008). More formally, when there are at least three unrelated attributes, the estimation risk is reduced, that is the expected value of the estimation loss created by the gap between a consumer's belief and the true value of a product attribute. The effect becomes more reliable as the number of attributes grows and, of course, as the variance among attributes diminishes.

Research Question 1. Is overall quality of credence services evaluated differently than quality for other services and, if so, how?

The service quality literature suggests overall quality evaluations center around three to five dimensions. Berry, Wall and Carbone (2006) observe that customers focus on technical performance, the behaviors and appearance of service providers, and the tangibles associated with service. Brady and Cronin (2001) point to outcome quality, interaction quality, and physical environment quality. Rust and Oliver (1994) proposed three components: service product, service delivery and service environment. SERVQUAL points to reliability, responsiveness, empathy, assurances, and tangibles (Parasuraman et al. 1991).

For this study I found evidence of three drivers of quality and satisfaction: technical, interpersonal and organizational quality. The technical quality dimension and interpersonal quality dimension were highly correlated. Indeed in Study 1, the two dimensions could not be separated. While this is contrary to expectation, it is consistent with comments from participants in the pilot study who rejected the term “competence” to describe technical skills, noting that a physician without communication skills would be unlikely to elicit useful information from patients.

Findings from both Studies 1 and 2 suggest that quality may connote a narrower construct for consumers of highly consequential, credence services. In Study 1, only the performances of the principal service provider and the auxiliary staff contributed to overall quality evaluations; organizational attributes did not contribute to quality but, instead, influenced value perceptions. In Study 2, only the physician’s technical quality

influenced overall quality. The physician's interpersonal quality and the organizational quality influenced satisfaction and value, respectively. At this point, it is not clear whether the narrower interpretation of quality seen in these studies reflects the evaluation processes for all credence services or whether it reflects the consequentiality of the focal service, health care. It is plausible that, with highly consequential services, buyers focus on the performance of the core product in making quality evaluations. Performance of the augmented product features (here, organizational quality) influences satisfaction and value.

The research also explored whether two of the key constructs, technical quality and organizational quality, are more appropriately modeled as reflective or formative constructs. Both constructs meet the tests for formative constructs: indicators are not necessarily correlated, dropping an indicator may change the meaning of the construct; and increases/decreases in one indicator may be expected to influence the construct (MacKenzie et al. 2005). To explore this issue, the technical and organizational quality dimensions were measured as both formative and reflective constructs in Study 2. I found that a reflective model outperforms a formative model with technical quality. That finding, coupled with evidence that ratings of the indicators of technical quality correlated with overall technical ratings, rather than the respondent's actual knowledge of the indicator, suggests technical quality evaluations are an overall attitude and should be modeled as a reflective construct. For organizational quality, however, the formative structure slightly outperforms the reflective model. It will be useful for future service quality researchers to explore formative modeling with organizational quality in order to

understand the generalizability of this finding. Research findings are summarized in Table 19.

Research Question 2: How do buyers integrate a service's credence, experience and search attributes to assess overall quality evaluations and what are the consequences of those evaluations?

Buyers of credence services have a difficult challenge: how to use the signals from observable service components to infer the quality of unobservable attributes. Two paths to evaluation have been identified from this study. The role of trust distinguishes the two approaches. In the first approach, trust is a consequence of quality evaluations. Technical quality influences overall quality and outcomes; interpersonal quality influences satisfaction and trust; organizational quality and outcomes influence value; overall quality, satisfaction and trust influence loyalty. The model is consistent with other research which finds that trust is a consequence of quality (Garbarino and Johnson 1999).

Under the rival approach, trust is used as a heuristic for evaluating unobservable technical quality. Specifically, perceptions of interpersonal and organizational quality influence trust formation; trust, in turn, influences technical quality perceptions. Technical quality influences overall quality and, through overall quality, outcome. Outcome plays a central role. Outcome, interpersonal quality, and value influence satisfaction. Outcome and trust influence value. And satisfaction and trust influence loyalty.

TABLE 19
Summary of Research Findings

Domain	Hypotheses	Study 1 Findings	Study 2 Findings
Overall Service Quality	H ₁ : Higher levels of a) technical b) interpersonal and c) organizational quality are associated with greater overall service quality.	✓ Supported ✓ Supported ~ Partially supported	✓ Supported ✗ Not supported ✗ Not supported
Technical Quality	H ₂ : Higher levels of the service provider's a) technical capabilities, b) technical processes, and c) prior outcomes are associated with greater technical quality.	✓ Supported ✓ Supported	✓ Supported ✓ Supported ✓ Supported
	H ₃ : Technical quality is a formative construct.		✗ Not supported
Interpersonal Quality	H ₄ : Higher levels of the service provider's a) respectfulness and b) communication performance are associated with greater interpersonal quality.	✓ Supported ✓ Supported	✓ Supported ✓ Supported
Organizational Quality	H ₅ : Higher levels of the a) physical surroundings, b) convenience, and c) support staff performance are associated with greater organizational quality.	✓ Supported ~ Partially ✓ Supported	✓ Supported ~ Partially ✓ Supported
	H ₆ : Organizational quality is a formative construct.		✓ Supported
Trust	H _{7a} : Perceptions of high interpersonal quality promote trust.		✓ Supported
	H _{7b} : Perceptions of high organizational quality promote trust.		✓ Supported
	H _{7c} : Trust positively influences perceptions of technical quality.		✓ Supported

TABLE 19 (CONTINUED)

Domain	Hypotheses	Study 1 Findings	Study 2 Findings
Expertise	H ₈ : Compared to experts, novices are more likely to evaluate technical quality with a) a correlation heuristic, or b) a trust heuristic		✗ Not supported
Confidence	H _{9a} : Compared to their evaluations of technical quality, novices are more confident of their evaluations of interpersonal and organization quality.		✗ Not supported
	H _{9b} : Compared to novices, experts are more confident of their judgments of technical quality.		✓ Supported
Risk	H _{10a} : Compared to people with low perceptions of risk, those with moderate perceptions of risk are more likely to rely on systematic processing.		✓ Supported
	H _{10b} : Compared to people with moderate perceptions of risk, those with high perceptions of risk are more likely to rely on heuristic processing.		✓ Supported
Outcome	H ₁₁ : Outcome is positively associated with overall service quality.	✓ Supported	✓ Supported
Satisfaction	H _{12a} : Perceived credence service quality positively influences satisfaction.		✗ Supported
	H _{12b} : With credence services, perceived interpersonal quality positively influences satisfaction.		✓ Supported
	H _{12c} : With credence services, perceived organizational quality positively influences satisfaction.		✓ Supported (but small effect)
	H _{12d} : Value positively influences satisfaction.		✓ Supported

TABLE 19 (CONTINUED)

Domain	Hypotheses	Study 1 Findings	Study 2 Findings
Value	H _{13a} : Perceived outcome positively influences value.	✓ Supported	✓ Supported
	H _{13b} : Interpersonal quality positively influences value.	✗ Not supported	✗ Not supported
	H _{13c} : Trust positively influences value.		✓ Supported
Loyalty	H _{14a} : Outcome valence has a positive influence on loyalty intentions.		✗ Not supported
	H _{14b} : Satisfaction has a positive influence on loyalty intentions.		✓ Supported
	H _{14c} : Value has a positive influence on loyalty intentions.		✗ Not supported
	H _{14d} : Trust has a positive influence on loyalty intentions.		✓ Supported
Switching Costs	H ₁₅ : Switching costs moderate the influence of value on loyalty.		✗ Not supported

Both models performed equally well, suggesting the trust heuristic warrants additional attention in contexts in which peripheral cues may influence inferences about unobservable attributes. As discussed earlier, theorists believe signals have only limited usefulness with credence products because effective signals require clear information about quality after purchase (Kirmani and Rao 2000). Without being certain of quality post-purchase, buyers are hobbled in their efforts to enforce penalties (Ippolito 1990). The quality of credence products, however, may be impossible to evaluate even after purchase and use. But because the outcome may not reflect performance quality, it is difficult for credence service buyers to enforce the bond implicit with signals.

This research suggests, however, that peripheral cues such as interpersonal quality and organizational quality may work through a trust filter. That is, buyers use those signals to form trust, and the trust, in turn, mediates perceptions of the unobservable attribute.

The role of outcome in the evaluation of quality remains unclear. With search and experience products, performance reliability is an indicator of overall quality. With credence services, I expected outcome to be a consequence of other quality evaluations, as credence outcomes do not always reflect the quality of the service performance. However, a rival model in which outcome is a predictor of quality fits equally well, suggesting challenged consumers may be willing to trade-off the possible miscues from the outcome signal for the cognitive savings it offers.

Research Question 3. How does product expertise moderate these relationships?

Drawing on dual-process theory, I hypothesized that skill and motivation are likely to influence the processing paths credence service consumers follow. Skill is reflected in product expertise; perceived risk may influence moderation. Based on the findings from Study 2, it appears that novices and experts interpret quality measures in similar ways. Multigroup analysis showed partial measurement invariance between the two groups, with equal form, loadings and intercepts prevailing across the groups. There was evidence of population heterogeneity. Experts' ratings of the latent constructs were both higher and more certain than novices' ratings, supporting the expectation that novices would be less confident of their ratings.

Contrary to expectations, experts were no less likely than novices to subscribe to the trust heuristic, in which technical quality evaluations are filtered through perceptions of interpersonal and organizational quality and trust. However, important differences between the groups were identified. In making satisfaction evaluations, organizational and interpersonal quality was important to novices, whereas experts focused on outcome. In making loyalty judgments, value is relatively the most important contributor for novices whereas technical quality is most important for experts.

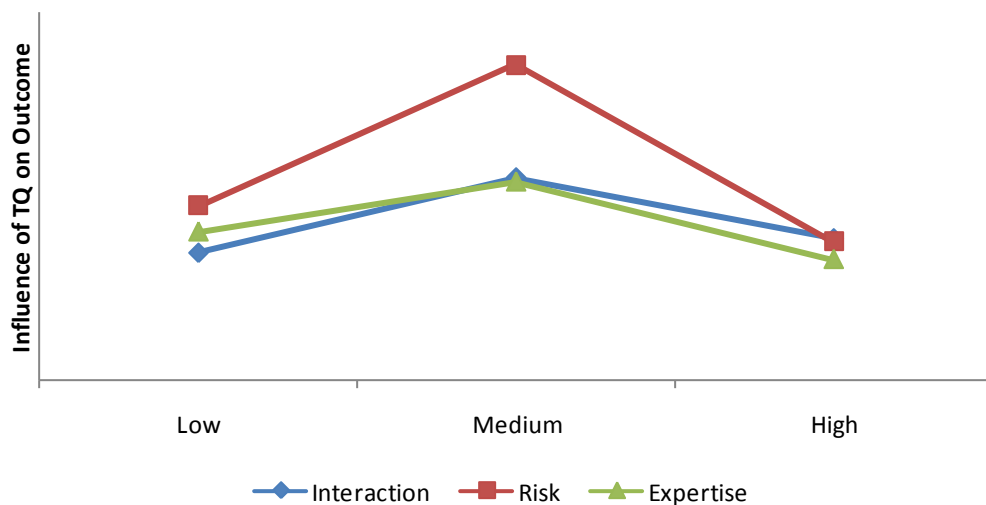
Research Question 4. How does perceived risk moderate these relationships?

Perceived risk appears to influence the way credence quality evaluations are made. I hypothesized that higher levels of risk would increase the motivation for thoughtful evaluations of technical quality, but would, at the same time, decrease the ability and increase reliance on cognitive heuristics. A counterintuitive inverted U-

shaped relationship between perceived risk and relative emphasis on technical quality was found. As shown in Figure 8, people reporting moderate levels of risk associated with their illness or disorder tended to focus on the technical quality evaluations rather than observable interpersonal cues when explaining outcomes, suggesting a heavier use of cognitive reasoning.

FIGURE 8

Influence of Risk and Expertise on Evaluation



Implications for Credence Service Providers and Policymakers

The research offers several implications for credence service providers. First, service firms should show buyers how to evaluate their technical quality. The heuristic-

based evaluations consumers tend to use facilitate quick judgments but carry the risk that important information is overlooked, resulting in flawed evaluations. Firms can guide consumers to relevant performance indicators, indicators consumers may be unaware of and be unsure how to interpret. Indeed, service providers should not assume clients know how to interpret the clues. For example, in the qualitative pilot study, most respondents were unaware that board certification involved more advanced training and examinations than medical licensure.

The research suggests that performance outcome influence satisfaction and value perceptions. Outcome valence may represent an easily accessible quality signal for consumers. But outcomes are not always within the control of credence service providers. Providers will benefit from educating consumers about the uncontrollable aspects of the service in order to immunize buyers to poor outcomes. Alternatively, service providers may consider guaranteeing performance results. That strategy is risky, to the extent that sellers do not have control over outcomes and to the extent that successful outcomes may rely on the consumer's co-creation efforts. But if outcomes are the beacon consumers look to, then the strategy may be worth consideration.

The healthcare system in the U.S. is challenged by unaffordably high costs, lack of access, and uneven quality. Some healthcare policymakers are advocating a market-based solution. With this approach, patients will act like consumers and reward healthcare providers who offer the best combination of quality, prices, and services. Innovative, patient-centric providers will be rewarded with higher fees and more patients. For the market-based approach to work, patients must have the ability to act

like consumers. This research provides clues about the conditions under which patients can be expected to act like consumers. Specifically, patients facing moderate levels of risk may have the ability and motivation to sort out quality cues. Patients facing little risk likely lack the motivation to decipher quality and will need appropriate incentives to take on the responsibility. Patients in high risk situations may have the incentive to find the best quality but may be too stressed to make systematic evaluations. Those patients will need additional resources to sort through quality clues.

Limitations and Direction for Future Research

This research contributes to the understanding of the processes consumers use to evaluate unobservable credence attributes. In particular, the role of trust in the formation of attitudes toward technical quality and overall quality is investigated. While the study provides some support for the use of a trust heuristic in evaluating unobservable attributes, further research is needed. Attitudes are formed and evolve over time. This study's cross-sectional design precludes definite statements about causality of the formation of the complex attitudes. Testing the model with a longitudinal or experimental design will improve understanding of the role of a trust heuristic in quality evaluations.

While the empirical analysis adheres to recommended procedures, there are some limitations based on the distributional properties of the data. First, although each pair of factors met at least one of two statistical tests for construct discrimination, many pairs of factors were too highly correlated. The ability of factors to meet statistical tests for

discrimination is in part an artifact of the large sample sizes and the kurtosis. Second, many factors were measured with one or two indicators. While such a practice is not uncommon (cf Brown 2006), it creates the potential for model identification problems. Investigating relationships with another tool, such as partial least squares, may be useful.

The study explored the boundary roles of skill and motivation. For this study, skill was operationalized as self-ratings of expertise. While the approach is reasonable, it is possible that the self-ratings actually measured familiarity or general self-confidence. Toward the end of the survey, respondents described their level of confidence in the global ratings they had assigned to the physician on each of the domains (technical, interpersonal, organizational). Experts were confident of their ratings across all three domains. In contrast, novices were unconfident across all three domains. While it is reasonable that novices would lack confidence in their evaluations of technical quality, there is no reason they should have lacked confidence in their ability to rate interpersonal skills and servicescapes; as consumers, they do so all the time. This area of inquiry will benefit from research that contrasts true novices from true experts.

It will also be fruitful to examine the influence of other motivators on processing styles. Perceived risk appears to influence the use of heuristics in a counterintuitive way. The influence of positive motivators should be examined.

Investigating the attributions associated with evaluations is beyond the scope of this dissertation yet it represents a fruitful area for research. “Most consumers are not rocket scientists. They simply ask why an outcome was unsatisfying, whether it will happen again, and who, if anyone, is to be blamed” (Weiner 2000 p. 387). Control and

volition are factors implicating attribution formation. It will be helpful to understand how people form attributions in the context of credence services, when controllability may be unknowable.

In addition, exploring the consequences of buyer's evaluative uncertainty is an opportunity. Attitudes are formed when needed (Feldman and Lynch 1988). Attitudes that are simple to form are created spontaneously and reside in memory. More complex attitudes, such as technical quality in the present study, are likely formed only when needed. When is judgment suspended? Are people aware that judgment is suspended? How do people cope with suspension? Some avenues may be avoidance or substitution of other attitudes. It may be fruitful to use implicit attitude techniques to tap the timing of attitude formation.

Finally, the investigation can be extended to other contexts in which credence attributes dominate. Within services, it will be useful to understand the roles of consequentiality and familiarity. Evaluation processes of credence goods is also an important part of the puzzle, particularly in light of recent concerns about food safety.

REFERENCES

- Americans as Health Care Consumers: Update on the Role of Quality Information. Highlights of a National Survey. Kaiser Family Foundation and the Agency for Healthcare Research and Quality, Rockville, MD. Available at www.ahrq.gov/qual/kffhigh00.htm. Accessed June 19, 2008.
- Agans, Robert P. and Leigh S. Shaffer (1994), "The Hindsight Bias: The Role of the Availability Heuristic and Perceived Risk," *Basic and Applied Social Psychology*, 15 (4), 439-449.
- Agustin, Clara and Jagdip Singh (2005), "Curvilinear Effects of Consumer Loyalty Determinants in Relational Exchanges," *Journal of Marketing Research*, 42 (1), 96-108.
- Anderson, Eugene W., Claes Fornell, and Donald R. Lehmann (1994), "Customer Satisfaction, Market Share, and Profitability: Findings from Sweden," *Journal of Marketing*, 58 (3), 53-66.
- Anderson, James C. and David W. Gerbing (1988), "Structural Equation Modeling in Practice: A Review and Recommended Two-Step Approach," *Psychological Bulletin*, 103 (3), 411-423.
- Babakus, Emin, Sevgin Eroglu, and Ugur Yavas (2004), "Modeling Consumers' Choice Behavior: An Application in Banking," *Journal of Services Marketing*, 18 (6), 462-470.
- Bagozzi, Richard P., Youjae Yi, and Lynn W. Phillips (1991), "Assessing Construct Validity in Organizational Research," *Administrative Science Quarterly*, 36 (3), 421-458.
- Bain, Joe S. (1951), "Relation of Profit Rate to Industry Concentration: American Manufacturing, 1936-1940," *Quarterly Journal of Economics*, 65 (3), 293-324.
- Baker, Julie, A. Parasuraman, Dhruv Grewal, and Glenn B. Voss (2002), "The Influence of Multiple Store Environment Cues on Perceived Merchandise Value and Patronage Intentions," *Journal of Marketing*, 66 (2), 120-141.
- Baron, Jonathon and John C. Hershey (1988), "Outcome Bias in Decision Evaluation," *Journal of Personality and Social Psychology*, 54 (4), 569-579.
- Baron, Reuben M. and David A. Kenny (1986), "The Moderator-Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations," *Journal of Personality and Social Psychology*, 51 (6), 1173-1182.

Bart, Yakov, Venkatesh Shankar, Fareena Sultan, and Glen L. Urban (2005), "Are the Drivers and Role of Online Trust the Same for All Web Sites and Consumers? A Large-Scale Exploratory Empirical Study," *Journal of Marketing*, 69 (4), 133-152.

Bendapudi, Neeli M., Leonard L. Berry, Keith A. Frey, Janet Turner Parish, and William L. Rayburn (2006), "Patients' Perspectives on Ideal Physician Behaviors," *Mayo Clinic Proceedings*, 81 (3), 338-344.

Berry, Leonard L. (2000), "Cultivating Service Brand Equity," *Journal of the Academy of Marketing Science*, 28 (1), 128-138.

Berry, Leonard L., A. Parasuraman, and Valarie A. Zeithaml (2003), "Ten Lessons for Improving Service Quality," MSI Report 03-104. Cambridge, MA: Marketing Science Institute.

Berry, Leonard L., Derek Parker, Russell C. Coile Jr, D. Kirk Hamilton, David D. O'Neill, and Blair L. Sadler (2004), "The Business Case for Better Buildings," *Frontiers of Health Services Management*, 21 (1), 27-34.

Berry, Leonard L., Kathleen Seiders, and Dhruv Grewal (2002), "Understanding Service Convenience," *Journal of Marketing*, 66, 1-17.

Berry, Leonard L., Eileen A. Wall, and Lewis P. Carbone (2006), "Service Clues and Customer Assessment of the Service Experience: Lessons from Marketing," *Academy of Management Perspectives*, 20 (2), 43-57.

Bigley, Mary Beth, Donna Pillittere, Judith Hibbard, and L. Gregory Pawlson (2003), "Consumer Preferences for Information on the Quality of Physician Care: A Review of the Literature," Washington, DC: NCQA.

Bitner, Mary Jo (1992), "Servicescapes: The Impact of Physical Surroundings on Customers and Employees," *Journal of Marketing*, 56 (2), 57-71.

Blumenthal, David (1996), "Quality of Care -- What Is It?- Part One of Six," *New England Journal of Medicine*, 335 (12), 891-894.

Boatwright, Peter, Ajay Kalra, and Wei Zhang (2008), "Research Note: Should Consumers Use the Halo to Form Product Evaluations?," *Management Science*, 54 (1), 217-223.

Bolton, Ruth N. and James H. Drew (1991), "A Multistage Model of Customers' Assessments of Service Quality and Value," *The Journal of Consumer Research*, 17 (4), 375-384.

Bowers, Michael R., John E. Swan, and William F. Koehler (1994), "What Attributes Determine Quality and Satisfaction with Health Care Delivery?," *Health Care Management Review*, 19 (4), 49-55.

Brady, Michael K. and J. Joseph Cronin (2001), "Some New Thoughts on Conceptualizing Perceived Service Quality: A Hierarchical Approach," *Journal of Marketing*, 65 (3), 34-49.

Brook, Robert H., Elizabeth A. McGlynn, and Paul D. Cleary (1996), "Measuring Quality of Care- Part Two of Six," *New England Journal of Medicine*, 335 (13), 966-970.

Brown, Stephen W. and Teresa A. Swartz (1989), "A Gap Analysis of Professional Service Quality," *Journal of Marketing*, 53 (2), 92-98.

Brown, Timothy A. (2006), *Confirmatory Factor Analysis for Applied Research*. New York: The Guilford Press.

Calfee, John E. and Gary Ford (1988), "Economics, Information and Consumer Behavior," *Advances in Consumer Research*, 15, 234-238.

Camerer, Colin (2005), "Three Cheers—Psychological, Theoretical, Empirical—for Loss Aversion," *Journal of Marketing Research*, 42 (2), 129-133.

Camerer, Colin F. and Eric J. Johnson (1991), "The Process-Performance Paradox in Expert Judgment: How Can Experts Know So Much and Predict So Badly," in *Toward a General Theory of Expertise: Prospects and Limits*.

Carman, James M. (1990), "Consumer Perceptions of Service Quality: An Assessment of the SERVQUAL Dimensions," *Journal of Retailing*, 66 (1), 33-56.

Chaiken, Shelly and Durairaj Maheswaran (1994), "Heuristic Processing Can Bias Systematic Processing: Effects of Source Credibility, Argument Ambiguity, and Task Importance on Attitude Judgment," *Journal of Personality and Social Psychology*, 66 (3), 460-473.

Chi, Michelene T. H., Paul J. Feltovich, and Robert Glaser (1981), "Categorization and Representation of Physics Problems by Experts and Novices," *Cognitive Science*, 5 (2), 121-152.

Cohen-Charash, Yochi and Paul E. Spector (2001), "The Role of Justice in Organizations: A Meta-Analysis," *Organizational Behavior and Human Decision Processes*, 86 (2), 278-321.

- Colquitt, Jason A. (2001), "On the Dimensionality of Organizational Justice: A Construct Validation of a Measure," *Journal of Applied Psychology*, 86 (3), 386-400.
- Colquitt, Jason A., Michael J. Wesson, Christopher O. L. H. Porter, Donald E. Conlon, and K. Yee Ng (2001), "Justice at the Millennium: A Meta-Analytic Review of 25 Years of Organizational Justice Research," *Journal of Applied Psychology*, 86 (3), 425-445.
- Conchar, Margy P., George M. Zinkhan, Cara Peters, and Sergio Olavarrieta (2004), "An Integrated Framework for the Conceptualization of Consumers' Perceived-Risk Processing," *Journal of the Academy of Marketing Science*, 32 (4), 418-436.
- Cronin Jr, J. J. and S. A. Taylor (1992), "Measuring Service Quality: A Reexamination and Extension," *Journal of Marketing*, 56 (3), 55-68.
- Cropanzano, Russell, Cynthia A. Prehar, and Peter Y. Chen (2002), "Using Social Exchange Theory to Distinguish Procedural from Interactional Justice," *Group & Organization Management*, 27 (3), 324-352.
- Cudeck, Michael W. and Robert Browne (1983), "Cross-Validation of Covariance Structures," *Multivariate Behavioral Research*, 18 (2), 147-167.
- Cutler, Bob, Rajeshkar G. Javalgi, and Kurt Schimmel (2003), "Selection of a Legal Services Firm: The Differing Perceptions of Attorneys and Consumers," *Services Marketing Quarterly*, 25 (1), 1-10.
- Dabholkar, Pratibha A. (1996), "Consumer Evaluations of New Technology-Based Self-Service Options: An Investigation of Alternative Models of Service Quality," *International Journal of Research in Marketing*, 13 (1), 29-51.
- Dabholkar, Pratibha A. and Jeffrey W. Overby (2005), "Linking Process and Outcome to Service Quality and Customer Satisfaction Evaluations," *International Journal of Service Industry Management*, 16 (1), 10-27.
- Dabholkar, Pratibha A., C. David Shepherd, and Dayle I. Thorpe (2000), "A Comprehensive Framework for Service Quality: An Investigation of Critical Conceptual and Measurement Issues through a Longitudinal Study," *Journal of Retailing*, 76 (2), 139-174.
- Dabholkar, Pratibha A., Dayle I. Thorpe, and Joseph O. Rentz (1996), "A Measure of Service Quality for Retail Stores: Scale Development and Validation," *Journal of the Academy of Marketing Science*, 24 (1), 3-17.
- Darby, Michael R. and Edi Karni (1973), "Free Competition and the Optimal Amount of Fraud," *Journal of Law and Economics*, 16 (1), 67-88.

- Denrell, Jerker (2004), "Random Walks and Sustained Competitive Advantage," *Management Science*, 50 (7), 922-934.
- Dick, Alan, Dipankar Chakravarti, and Gabriel Biehal (1990), "Memory-Based Inferences During Consumer Choice," *Journal of Consumer Research*, 17 (1), 82-93.
- Donabedian, Avedis (1982), *The Criteria and Standards of Quality*. Ann Arbor, MI: Health Administration Press.
- (1980), *The Definition of Quality and Approaches to Its Assessment*. Ann Arbor, MI: Health Administration Press.
- (2005), "Evaluating the Quality of Medical Care," *The Milbank Quarterly*, 83 (4), 691-729.
- Eagly, A. H. and S. Chaiken (1993), *The Psychology of Attitudes*. Fort Worth: Harcourt Brace Jovanovich.
- Edgman-Levitan, S. and P. D. Cleary (1996), "What Information Do Consumers Want and Need?," *Health Affairs*, 15 (4), 42-56.
- Erdem, Tulin and Joffre Swait (1998), "Brand Equity as a Signaling Phenomenon," *Journal of Consumer Psychology*, 7 (2), 131-157.
- Feldman, Jack M. and John G. Lynch, Jr (1988), "Self-Generated Validity and Other Effects of Measurement on Belief, Attitude, Intention and Behavior," *Journal of Applied Psychology*, 73 (3), 421-435.
- Flora, David B. and Patrick J. Curran (2004), "An Empirical Evaluation of Alternative Methods of Estimation for Confirmatory Factor Analysis with Ordinal Data," *Psychological Methods*, 9 (4), 466-491.
- Ford, Gary T., Darlene B. Smith, and John L. Swasy (1990), "Consumer Skepticism of Advertising Claims: Testing Hypotheses from Economics of Information," *The Journal of Consumer Research*, 16 (4), 433-441.
- Fornell, Claes and David F. Larcker (1981), "Evaluating Structural Equation Models with Unobservable Variables and Measurement Error," *Journal of Marketing Research*, 18 (1), 39-50.
- Fowler, Floyd J. (1995), *Improving Survey Questions: Design and Evaluation*. Thousand Oaks, CA: Sage Publications Inc.

- Fung, Constance H., Marc N. Elliott, Ron D. Hays, Katherine L. Kahn, David E. Kanouse, Elizabeth A. McGlynn, Mark D. Spranca, and Paul G. Shekelle (2005), "Patients' Preferences for Technical Versus Interpersonal Quality When Selecting a Primary Care Physician," *Health Services Research*, 40 (4), 957-977.
- Garbarino, Ellen and Mark S. Johnson (1999), "The Different Roles of Satisfaction, Trust, and Commitment in Customer Relationships," *Journal of Marketing*, 63 (2), 70-87.
- Gotlieb, Jerry B., Dhruv Grewal, and Stephen W. Brown (1994), "Consumer Satisfaction and Perceived Quality: Complementary or Divergent Constructs?," *Journal of Applied Psychology*, 79 (6), 875-885.
- Greenberg, Jerald (1990), "Organizational Justice: Yesterday, Today, and Tomorrow," *Journal of Management*, 16 (2), 399-433.
- Gronroos, Christian (1984), "A Service Quality Model and Its Marketing Implications," *European Journal of Marketing*, 18 (4), 36-45.
- Gwinner, Kevin P., Dwayne D. Gremler, and Mary Jo Bitner (1998), "Relational Benefits in Services Industries: The Customer's Perspective," *Journal of the Academy of Marketing Science*, 26 (2), 101-114.
- Hahn, Sunku (2004), "The Advertising of Credence Goods as a Signal of Product Quality," *Manchester School*, 72 (1), 50-59.
- Hastie, Reid (1983), "Social Inference," *Annual Review of Psychology*, 34 (1), 511-542.
- Hawkins, Scott A. and Reid Hastie (1990), "Hindsight: Biased Judgments of Past Events after the Outcomes Are Known," *Psychological Bulletin*, 107, 311-327.
- Hennig-Thurau, Thorsten, Kevin P. Gwinner, and Dwayne D. Gremler (2002), "Understanding Relationship Marketing Outcomes: An Integration of Relational Benefits and Relationship Quality," *Journal of Service Research*, 4 (3), 230-248.
- Hu, Li-tze and Peter M. Bentler (1999), "Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria Versus," *Structural Equation Modeling*, 6 (1), 1-56.
- Institute of Medicine (2001), *Crossing the Quality Chasm: A New Health System for the 21st Century*. Washington, D.C.: National Academy Press.
- Ippolito, Pauline M. (1990), "Bonding and Nonbonding Signals of Product Quality," *Journal of Business*, 63 (1), 41-60.

Jain, Shailendra P. and Steven S. Posavac (2001), "Prepurchase Attribute Verifiability, Source Credibility, and Persuasion," *Journal of Consumer Psychology*, 11 (3), 169-180.

Johnson, Blair T. and Alice H. Eagly (1989), "Effects of Involvement on Persuasion: A Meta-Analysis," *Psychological Bulletin*, 106 (2), 290-314.

Jones, Michael A., David L. Mothersbaugh, and Sharon E. Beatty (2002), "Why Customers Stay: Measuring the Underlying Dimensions of Services Switching Costs and Managing Their Differential Strategic Outcomes," *Journal of Business Research*, 55 (6), 441-450.

Kahn, Barbara E. and Jonathan Baron (1995), "An Exploratory Study of Choice Rules Favored for High-Stakes Decisions," *Journal of Consumer Psychology*, 4 (4), 305-328.

Kahneman, Daniel and Amos Tversky (1979), "Prospect Theory: An Analysis of Decision under Risk," *Econometrica*, 47 (2), 263-291.

Kahnemann, Daniel and Amos Tversky (1974), "Judgment under Uncertainty," *Science*, 185, 1124-1131.

Kardes, Frank R., Steven S. Posavac, and Maria L. Cronley (2004), "Consumer Inference: A Review of Processes, Bases, and Judgment Contexts," *Journal of Consumer Psychology*, 14 (3), 230-256.

Kirmani, Amna and Akshay R. Rao (2000), "No Pain, No Gain: A Critical Review of the Literature on Signaling Unobservable Product Quality," *Journal of Marketing*, 64 (2), 66-79.

Kotler, Philip (2000), *Marketing Management*. Upper Saddle River, NJ: Prentice-Hall.

Kramer, Roderick M. (1999), "Trust and Distrust in Organizations: Emerging Perspectives, Enduring Questions," *Annual Review of Psychology*, 569-570.

Lam, Shun Y., Venkatesh Shankar, M. Krishna Erramilli, and Bvsan Murthy (2004), "Customer Value, Satisfaction, Loyalty, and Switching Costs: An Illustration from a Business-to-Business Service Context," *Journal of the Academy of Marketing Science*, 32 (3), 293-311.

Laochumnvanit, Krongjit and David Bednall (2005), "Consumers' Evaluation of Free Service Trial Offers," *Academy of Marketing Science Review*, (11), 1-13.

Larkin, Jill, John McDermott, Dorothea P. Simon, and Herbert A. Simon (1980), "Expert and Novice Performance in Solving Physics Problems," *Science*, 208 (4450), 1335-1342.

Loewenstein, George, Elke U. Weber, Christopher K. Hsee, and Ned Welch (2001), "Risk as Feelings," *Psychological Bulletin*, 127 (2), 267-286.

Lubalin, James S. and Lauren D. Harris-Kojetin (1999), "What Do Consumers Want and Need to Know in Making Health Care Choices?," *Medical Care Research and Review*, 56 (Suppl 1), 67-102.

MacKenzie, Scott B., Philip M. Podsakoff, and Cheryl Burke Jarvis (2005), "The Problem of Measurement Model Misspecification in Behavioral and Organization Research and Some Recommended Solutions," *Journal of Applied Psychology*, 90 (4), 710-30.

Maute, Manfred F. and William R. Forrester, Jr. (1991), "The Effect of Attribute Qualities on Consumer Decision Making: A Causal Model of External Information Search," *Journal of Economic Psychology*, 12 (4), 643-667.

Mayer, Roger C., James H. Davis, and F. David Schoorman (1995), "An Integrative Model of Organizational Trust," *Academy of Management Review*, 20 (3), 709-734.

McAlexander, James H., Dennis O. Kaldenberg, and Harold F. Koenig (1994), "Service Quality Measurement," *Journal of Health Care Marketing*, 14 (3), 34-39.

McAllister, Daniel J. (1995), "Affect-and Cognition-Based Trust as Foundations for Interpersonal Cooperation in Organizations," *The Academy of Management Journal*, 38 (1), 24-59.

McEvily, Bill, Vincenzo Perrone, and Akbar Zaheer (2003), "Trust as an Organizing Principle," *Organization Science*, 14 (1), 91-103.

Mitra, Kaushik, Michelle C. Reiss, and Louis M. Capella (1999), "An Examination of Perceived Risk, Information Search and Behavioral Intentions in Search," *Journal of Services Marketing*, 13 (2/3), 208-229.

Mittal, Banwari (1995), "A Comparative Analysis of Four Scales of Consumer Involvement," *Psychology and Marketing*, 12 (7), 663-682.

---- (2004), "Lack of Attribute Searchability: Some Thoughts," *Psychology and Marketing*, 21 (6), 443-462.

Moorman, Christine, Rohit Deshpande, and Gerald Zaltman (1993), "Factors Affecting Trust in Market Research Relationships," *Journal of Marketing*, 57 (1), 81-101.

- Morgan, Robert M. and Shelby D. Hunt (1994), "The Commitment-Trust Theory of Relationship Marketing," *Journal of Marketing*, 58 (3), 20-39.
- Mosteller, Frederick and J.W. Tukey (1977), *Data Analysis and Regression*. London: Addison-Wesley.
- Murphy, Kevin R., Robert A. Jako, and Rebecca L. Anhalt (1993), "Nature and Consequences of Halo Error: A Critical Analysis," *Journal of Applied Psychology*, 78 (2), 218-225.
- Murray, Keith B. (1991), "A Test of Services Marketing Theory: Consumer Information Acquisition Activities," *Journal of Marketing*, 55 (1), 10-25.
- Nagle, Thomas (1984), "Economic Foundations for Pricing," *The Journal of Business*, 57 (1), 3-26.
- NCQA (2002), "NCQA/ADA Consumer-Patient On-Line Survey."
- Neal, William D. (1999), "Satisfaction Is Nice, but Value Drives Loyalty," *Marketing Research*, 11 (1), 20-3.
- Nelson, Phillip (1974), "Advertising as Information," *The Journal of Political Economy*, 82 (4), 729-754.
- (1970), "Information and Consumer Behavior," *Journal of Political Economy*, 78 (2), 311-330.
- Nunnally, Jum C. (1978), *Psychometric Theory* (2nd ed.). New York: McGraw Hill.
- Okada, Erika M. and Stephen J. Hoch (2004), "Spending Time Versus Spending Money," *Journal of Consumer Research*, 31 (2), 313-323.
- Oliva, Terence A., Richard L. Oliver, and Ian C. MacMillan (1992), "A Catastrophe Model for Developing Service Satisfaction Strategies," *Journal of Marketing*, 56 (3), 83-95.
- Oliver, Richard L. (1993), "Cognitive, Affective, and Attribute Bases of the Satisfaction Response," *The Journal of Consumer Research*, 20 (3), 418-430.
- (1994), "Conceptual Issues in the Structural Analysis of Consumption Emotion, Satisfaction, and Quality: Evidence in a Service Setting," *Advances in Consumer Research*, 21 (1), 16-22.
- (1999), "Whence Consumer Loyalty?," *Journal of Marketing*, 63 (4), 33-44.

Olson, James M. and Mark P. Zanna (1993), "Attitudes and Attitude Change," *Annual Review of Psychology*, 44 (1), 117-154.

Ostrom, Amy and Dawn Iacobucci (1995), "Consumer Trade-Offs and the Evaluation of Services," *Journal of Marketing*, 59 (1), 17-28.

Ozment, John and Edward A. Morash (1994), "The Augmented Service Offering for Perceived and Actual Service Quality," *Journal of the Academy of Marketing Science*, 22 (4), 352-363.

Parasuraman, A., Leonard L. Berry, and Valarie A. Zeithaml (1991), "Refinement and Reassessment of the SERVQUAL Scale," *Journal of Retailing*, 67 (4), 420-451.

Parasuraman, A., Valarie A. Zeithaml, and Leonard L. Berry (1985), "A Conceptual Model of Service Quality and Its Implications for Future Research," *Journal of Marketing*, 49 (4), 41-50.

---- (1988), "SERVQUAL: A Multiple-Item Scale for Measuring Consumer Perceptions of Service Quality," *Journal of Retailing*, 64 (1), 12-40.

Petty, Richard E., John T. Cacioppo, and David Schumann (1983), "Central and Peripheral Routes to Advertising Effectiveness: The Moderating Role of Involvement," *The Journal of Consumer Research*, 10 (2), 135-146.

Pillittere, Donna, Mary Beth Bigley, Judith Hibbard, and Greg Pawlson (2002), "Exploring Consumer Perspectives on Good Physician Care: A Summary of Focus Group Results," report # 578 available at www.cmwf.org. The Commonwealth Fund.

Podsakoff, Philip M., Scott B. MacKenzie, Jeong Yeon Lee, and Nathan. P. Podsakoff (2003), "Common Method Biases in Behavioral Research: A Critical Review of the Literature and Recommended Remedies," *Journal of Applied Psychology*, 88 (5), 879-903.

Porter, Michael E. (1981), "The Contributions of Industrial Organization to Strategic Management," *Academy of Management Review*, 6 (4), 609-620.

Pritchard, Mark P., Mark E. Havitz, and Dennis R. Howard (1999), "Analyzing the Commitment-Loyalty Link in Service Contexts," *Academy of Marketing Science Journal*, 27 (3), 333-348.

Reimer, Anja and Richard Kuehn (2005), "The Impact of Servicescape on Quality Perception," *European Journal of Marketing*, 39 (7/8), 785-808.

- Rudolph, Thomas J. and Jillian Evans (2005), "Political Trust, Ideology, and Public Support for Government Spending," *American Journal of Political Science*, 49 (3), 660-671.
- Rust, Roland T. and Richard L. Oliver (1994), "Service Quality: Insights and Managerial Implications from the Frontier," *Service Quality: New Directions in Theory and Practice*, 1-20.
- Safran, Dana Gelb, Melinda Karp, Kathryn Coltin, Hong Chang, Angela Li, John Ogren, and William H. Rogers (2006), "Measuring Patients' Experiences with Individual Primary Care Physicians. Results of a Statewide Demonstration Project," *Journal of General Internal Medicine*, 21 (1), 13-21.
- Schafer, John L. and John W. Graham (2002), "Missing Data: Our View of the State of the Art," *Psychological Methods*, 7 (2), 147-177.
- Schafer, Joseph L. and Maren K. Olsen (1998), "Multiple Imputation for Multivariate Missing-Data Problems: A Data Analyst's Perspective," *Multivariate Behavioral Research*, 33 (4), 545-571.
- Seiders, Kathleen and Leonard L. Berry (1998), "Service Fairness: What It Is and Why It Matters," *Academy of Management Executive*, 12 (2), 8-20.
- Seiders, Kathleen, Glenn B. Voss, Andrea L. Godfrey, and Dhruv Grewal (2007), "Servcon: Development and Validation of a Multidimensional Service Convenience Scale," *Journal of the Academy of Marketing Science*, 35 (1), 144-156.
- Shankar, Venkatesh, Amy K. Smith, and Arvind Rangaswamy (2003), "Customer Satisfaction and Loyalty in Online and Offline Environments," *International Journal of Research in Marketing*, 20, 153-175.
- Singh, Jagdip (2000), "Performance Productivity and Quality of Frontline Employees in Service Organizations," *Journal of Marketing*, 64 (2), 15-34.
- Singh, Jagdip and Deepak Sirdeshmukh (2000), "Agency and Trust Mechanisms in Consumer Satisfaction and Loyalty Judgments," *Journal of the Academy of Marketing Science*, 28 (1), 150-168.
- Sirdeshmukh, Deepak, Jagdip Singh, and Barry Sabol (2002), "Consumer Trust, Value, and Loyalty in Relational Exchanges," *Journal of Marketing*, 66 (1), 15-37.
- Slovic, Paul (1995), "The Construction of Preference," *American Psychologist*, 50 (5), 364-371.

- Smith, Amy K., Ruth N. Bolton, and Janet Wagner (1999), "A Model of Customer Satisfaction with Service Encounters Involving Failure and Recovery," *Journal of Marketing Research*, 36 (3), 356-372.
- Smith, Gerald F. (1988), "Towards a Heuristic Theory of Problem Structuring," *Management Science*, 34 (12), 1489-1506.
- Smith, Robert E. and William R. Swinyard (1988), "Cognitive Response to Advertising and Trial: Belief Strength, Belief Confidence and Product Curiosity," *Journal of Advertising*, 17 (3), 3-14.
- Solomon, Joel S. , Ron D. Hays, Alan M. Zaslavsky, Lin Ding, and Paul D. Cleary (2005), "Psychometric Properties of a Group-Level Consumer Assessment of Health Plans Study (CAHPS) Instrument," *Medical Care*, 43 (1), 53-60.
- Spence, Mark T. and Merrie Brucks (1997), "The Moderating Effects of Problem Characteristics on Experts' and Novices' Judgements," *Journal of Marketing Research (JMR)*, 34 (2), 233-247.
- Spreng, Richard A. and Robert D. Mackoy (1996), "An Empirical Examination of a Model of Perceived Service Quality and Satisfaction," *Journal of Retailing*, 72 (2), 201-214.
- Stake, Robert E. (2000), "Case Studies," in *Handbook of Qualitative Research*, Norman K. Denzin and Yvonna S. Lincoln, Eds. Thousand Oaks, CA: Sage Publications.
- Steenkamp, Jan-Benedict E. M. (1990), "Conceptual Model of the Quality Perception Process," *Journal of Business Research*, 21 (4), 309-333.
- Stigler, George J. (1961), "The Economics of Information," *The Journal of Political Economy*, 69 (3), 213-225.
- Swan, John E. and Linda Jones Combs (1976), "Product Performance and Consumer Satisfaction: A New Concept," *Journal of Marketing*, 40 (April), 25-33.
- Tam, Jackie L. M. (2004), "Customer Satisfaction, Service Quality and Perceived Value: An Integrative Model," *Journal of Marketing Management*, 20 (7/8), 897-917.
- Taylor, Steven A. and J. Joseph Cronin Jr (1994), "Modeling Patient Satisfaction and Service Quality," *J Health Care Mark*, 14 (1), 34-44.
- Teas, R. Kenneth and Sanjeev Agarwal (2000), "The Effects of Extrinsic Product Cues on Consumers' Perceptions of Quality, Sacrifice, and Value," *Journal of the Academy of Marketing Science*, 28 (2), 278-290.

Thompson, Bruce (2004), *Exploratory and Confirmatory Factor Analysis*. Washington, DC: American Psychological Association.

---- (2006), *Foundations of Behavioral Statistics*. New York: Guildford Press.

Thompson, Bruce and Colleen Cook (2002), "Stability of the Reliability of LibQUAL+ Scores: A Reliability Generalization Meta-Analysis Study," *Educational & Psychological Measurement*, 62 (4), 735-743.

Thompson, Bruce, Colleen Cook, and Fred Heath (2003), "Structure of Perceptions of Service Quality in Libraries: A LibQUAL+™ Study," *Structural Equation Modeling*, 10 (3), 456-464.

Tracey, Jocelyn, Bruce Arroll, Philip Barham, and David Richmond (1997), "The Validity of General Practitioners' Self Assessment of Knowledge: Cross Sectional Study," *British Medical Journal*, 315 (7120), 1426-1428.

Tyler, Tom R. (1994), "Psychological Models of the Justice Motive: Antecedents of Distributive and Procedural Justice," *Journal of Personality & Social Psychology*, 67 (5), 850-863.

Uzzi, Brian (1999), "Embeddedness in the Making of Financial Capital: How Social Relations and Networks Benefit Firms Seeking Financing," *American Sociological Review*, 64 (4), 481-505.

Van den Bos, Kees, E. Allan Lind, Riel Vermunt, and Henk A. M. Wilke (1997), "How Do I Judge My Outcome When I Do Not Know the Outcome of Others? The Psychology of the Fair Process Effect," *Journal of Personality & Social Psychology*, 72 (5), 1034-1046.

Wakefield, Kirk L. and Jeffrey G. Blodgett (1999), "Customer Response to Intangible and Tangible Service Factors," *Psychology and Marketing*, 16 (1), 51-68.

Weiner, Bernard (2000), "Attributional Thoughts About Consumer Behavior," *Journal of Consumer Research*, 27 (3), 382-387.

West, Stephen G., John F. Finch, and Patrick J. Curran (1995), "Structural Equation Models with Nonnormal Variables," in *Structural Equation Modeling*, Rick H. Hoyle, Ed. Thousand Oaks: Sage Publications.

Wood, Wendy (2000), "Attitude Change: Persuasion and Social Influence," *Annual Review of Psychology*, 51 (1), 539-570.

Woodruff, Robert B. (1997), "Customer Value: The Next Source for Competitive Advantage," *Journal of the Academy of Marketing Science*, 25 (2), 139-154.

Wright, Alice A. and John G. Lynch Jr (1995), "Communication Effects of Advertising Versus Direct Experience When Both Search and Experience Attributes Are Present," *The Journal of Consumer Research*, 21 (4), 708-718.

Yang, Zhilin and Robin T. Peterson (2004), "Customer Perceived Value, Satisfaction, and Loyalty: The Role of Switching Costs," *Psychology and Marketing*, 21 (10), 799-822.

Zeithaml, Valarie A. (1988), "Consumer Perceptions of Price, Quality and Value: A Means-End Model and Synthesis of Evidence," *Journal of Marketing*, 52 (July), 2-22.

---- (1981), "How Consumer Evaluation Processes Differ between Goods and Services," in *Marketing of Services*, J.H. Donnelly and W.R. George, Eds. Chicago: AMA.

Zeithaml, Valarie A., Leonard L. Berry, and A. Parasuraman (1996), "The Behavioral Consequences of Service Quality," *Journal of Marketing*, 60 (2), 31-46.

Zhao, Hao (2000), "Raising Awareness and Signaling Quality to Uninformed Consumers: A Price-Advertising Model," *Marketing Science*, 19 (4), 390-396.

APPENDIX A

STUDY 1 SURVEY ITEMS

Code	Item
Ptidcode	Patient ID
Year	Year
Quarter	Quarter
Month	Month
Date	
Clinic	Clinic
SiteID	Site ID
depart	Department
Pr3Sp4orig	Primary/Specialty
Attempts	Attempts
q01qlty	Overall, would you rate the quality of care provided as..
q02MDchg	Overall, did you feel there was a [clinic name] doctor in charge of your care?
q05thoro	How would you rate that provider on: Thoroughness of examination and treatment
q06enftm	How would you rate that provider on: Spending enough time with you and not seeming rushed
q07listn	How would you rate that provider on: How well he or she listened to your concerns
q08ezwrd	How would you rate that provider on: Using words and terms you could understand
q09invdc	How would you rate that provider on: Involving you in decisions about your care
q10commu	How would you rate that provider on: Communication with you
q13cocar	How would you rate that provider on: Being courteous and caring
q14medsk	How would you rate that provider on: Your impression of his or her medical skills
q15ontim	How would you rate that provider on: Being on-time for the appointment
q16instr	How would you rate that provider on: Giving clear instructions about what to do following the appointment
q17expla	How would you rate that provider on: Explaining your medical condition and treatment
q18MDQ	Overall, would you rate the quality of care that you received from the provider?
q23RN	How would you rate the: Nurses and medical assistants
q24recep	How would you rate the: Reception staff
q25tmwrk	How would you rate the: Overall teamwork between the doctors, nurses and staff
q26pinfm	How would you rate the: Staff on promptly informing you of test or exam results
q27explain	How would you rate the: Explanation of your condition and what to expect
q28priva	How would you rate the: Respect for your privacy
q29court	How would you rate the: Courtesy and friendliness shown to you by all employees
q30time	How would you rate the: Time spent waiting while at the facility
q31effic	How would you rate: Your impression of [name of clinic]'s efficiency overall
q33	Thinking about the number of days on your appointment schedule, from the first appointment to the last appointment, would you say the schedule was...
q34adays	Over how many days did your entire appointment schedule span, including the day of the first appointment to the day of the last appointment?
q35apacc	Overall, how would you rate the: Access to appointments and medical care in this department when needed
q36apcrt	Overall, how would you rate the: Courtesy and helpfulness of the appointment office staff on the phone
q37ainfo	Overall, how would you rate the: Information you were given to prepare for this visit, such as when to stop eating or taking medications
q38qaccs	Overall, how would you rate the: Ease of accessing the provider's office for medical problems or questions by phone
q39fcour	Overall, how would you rate the: Courtesy and helpfulness of the provider's office staff on the phone
q40atmos	How would you rate the atmosphere in and around the facility?
q41clean	How would you rate the cleanliness of the facility?

q42prkg	And, how would you rate the parking?
q43bill	How would you rate your interactions with the billing process and Business Office?
q44value	How would you rate the value of the care you received for the amount paid?
q45outco	How would you rate the outcome of your care, that is, how much you were helped?
q46safety	During this visit, did you ever feel that your safety was at risk?
q48	Did the doctors, nurses and other staff wash their hands or use a waterless hand sanitizer before examining you?
q49rcmd	Would you recommend [this clinic] to your friends and family? Would you say that you
q1bqity	Overall, would you rate the quality of care provided as...
q52compa	How would you compare [focal clinic] with other healthcare facilities in the area where you live?
q53healt	In general, how would you rate your overall health?
q54hse	In general, how would you rate your ability to take care of your own health without the help of a medical professional?
q55race	Which of the following best describes your race or ethnicity?
Age	Age
gender	Gender
Employee	Employee/Dependent/Retiree
St	State

APPENDIX B

STUDY 2 SURVEY ITEMS

Domain	Code	Question
		<i>Thinking about <u>all your experiences with this doctor</u>, how would you rate ...</i>
<i>Global ratings</i>	Q1	The overall quality of care
	IQ1	The way the doctor relates to you as a person
	OQ1	The convenience and professionalism of the doctor's office
	TQ1	Your impression of the doctor's medical skills and knowledge
<i>Technical quality</i>	MDPREV	Talking to you about specific things you could do to prevent illness
	ENUFTIME	Spending enough time with you and not seeming rushed
	THORO	Thoroughness of the examination and treatment
	MDEDUC	Your impression of the doctor's medical education and training
	MDEXP	Your impression of how much experience the doctor has treating patients with conditions like yours
	RESEARCH	Talking to you about the latest medical research related to your health problems
	HEALFAST	Clearing up your health problems as quickly as you hoped for
<i>Interpersonal quality</i>	LISTEN	Listening carefully to your concerns
	EXPLAIN	Explaining your medical condition and treatment in a way that is easy to understand
	CARE	Being caring
	INVDM	Involving you in decisions about your care. <i>Decisions about your health care can include choices about medicine, surgery, or other treatment.</i>
	APPEAR	The doctor's appearance
	MDCOURT	Being courteous
<i>Organizational quality</i>		<i>Thinking about <u>the doctor's office</u>, how would you rate ...</i>
	CONVLOC	The convenience of the location you usually visit
	EMPCOURT	Courtesy and friendliness shown to you by all employees
	CONVHRS	The convenience of the hours the office is open for appointments
	PRINFO	Promptly informing you of test results
	ANSSOON	Getting a prompt answer to your medical question when you call on the telephone.
	APPTSOON	Getting an appointment as soon as you want it
	RNSKILL	Your impression of the medical skills and knowledge of the nurses and medical assistants
	TMWRK	Overall teamwork between the doctors, nurses and staff
	COMFORT	The comfort and attractiveness of the facilities
	CLEAN	The cleanliness of the facility
	EQPMT	Your impression of how up to date the medical equipment is
		<i>How would you rate ...</i>
<i>Outcome</i>	OUTPHYS	How much you were helped physically
	OUTEMOT	How much you were helped emotionally to cope with your health conditions
<i>Satisfaction</i>	SAT	Satisfaction overall with the care you received
	SATMD	Satisfaction with the way the doctor and staff treated you
	SATOUT	Satisfaction with the outcome, that is, how much you were helped
<i>Value</i>	VALUTIME	The value of the care you received for the amount of time and effort it took for you to see the doctor
	VALUCASH	The value of the care you received for the amount paid

How likely are you to ...

<i>Loyalty</i>	SWITCH	Switch to another doctor because of your concerns about your medical care or the way you are treated?
	RCMD	Recommend this doctor to your friends and family?
	LOYLOC	Continue to go to this doctor even if the doctor relocated to a less convenient location?
		<i>How strongly do you disagree or agree with the following statements?</i>
<i>Trust</i>	TELLALL	You feel like you can tell this doctor anything
	HELPS	This doctor does everything possible to help you
	TRUSTMD	You completely trust this doctor's judgment about your medical care
<i>Switching costs</i>	MDSAME	All doctors have the same level of medical skills and knowledge
	OTHERCST	Other doctors are likely to cost more
	OTHERQ	It's hard to tell if another doctor would be better than this current doctor
	HASSLE	It would be a hassle switching to another doctor
<i>Organizational quality</i>	LOUD	The other patients in the doctor's waiting room are sometimes too loud or otherwise annoying
	WEALTHY	The other patients in the doctor's waiting room are probably wealthier than you are
		<i>Thinking about the most important problem or condition you have seen this doctor for ...</i>
<i>Risk</i>	SEVERITY	Seriousness of the condition (Not too serious --> life threatening)
	MDIMP	How important is it to you to choose a very highly qualified doctor?
<i>Expertise</i>	MDSSEEN	How many doctors have you seen for the same condition or problem that you have seen this doctor for?
	DURATION	How long have you been bothered by the problem or condition this doctor has treated you for?
	EXPERTIS	How knowledgeable do <i>you</i> feel about the condition or problem this doctor has treated you for?
<i>Technical quality</i>	MDSCHOOL	Do you know where this doctor went to medical school or received medical training?
	BDCERT	Is this doctor board certified?
	MDCONF	How confident does this doctor seem to be about his or her medical skills and knowledge?
<i>Reputation</i>	SOLOGRP	Does this doctor practice medicine (solo, small group, large group)
	REPUTE	<i>How would you rate the reputation for diagnosing and treating patients</i>
	REPUTMD	The reputation of the other doctors in this doctor's office
	REPUTGRP	The reputation of this doctor
<i>Global ratings</i>		<i>How would you rate</i>
	TQ2	The medical skills and knowledge of this doctor?
	IQ2	The courtesy and caring of this doctor?
		<i>How confident are you of your ratings of:</i>
<i>Confidence</i>	TQCONF	The medical skills and knowledge of this doctor?
	IQCONF	The courtesy and caring of this doctor?
	OQCONF	The efficiency and professionalism of this doctor's office?
		<i>How strongly do you disagree or agree with the following statements?</i>
<i>Affect</i>	LONGWAIT	When you have an appointment to see this doctor, the wait time is usually too long. <i>Wait time includes time spent in the waiting room and in the exam room</i>
	MDQ	This doctor is one of the best in the country
	AFFECT	You have a good feeling when you think about your experiences with this doctor.
	ANGRY	This doctor has made you angry.

Referrals

SRCMD	Recommended by another doctor or health professional
SRCFAM	Recommended by a friend or family member
SRCINS	Listing in an insurance directory
SRCADV	An advertisement or a listing in the telephone book
SRCOTHER	Other

*How did you choose this doctor?**Consumer characteristics*

HEALTH	Health
HSE	Health self-efficacy
GENDER	Gender
AGE	Age
EDUC	Income
INC	Education
GEO	Geography (part of the US)
	Race/ethnicity

VITA

Name: Ann Marie Mirabito

Address: Department of Marketing, Hankamer School of Business,
Baylor University, One Bear Place, Waco, TX

Email Address: Ann_Mirabito@Baylor.edu

Education: B.A., Economics, Duke University, 1976
M.B.A., Stanford University, 1983
Ph.D., Marketing, Texas A&M University, 2008

Professional: Product Manager, Frito Lay Inc., 1983-1987
Director of Marketing, Vie de France, 1987-1988
Vice President and Associate Publisher, Time-Life Books, 1988-1992
President, Greater Fulton Chamber of Commerce, 1992-1996
Senior Vice President, Rapidforms Inc, 1996-2000
Chief Marketing Office, Inspire Insurance Solutions, 2001-2002