EMPIRICAL VALIDATION OF THE STRUCTURE OF AN INFORMATION QUALITY MODEL

(Completed Research: Models)

Matthew Bovee
University of Vermont School of Business
bovee@bsad.uvm.edu

Abstract: Quality of information is a critical issue in today's interconnected society. Yet while most information quality studies posit or depend on a model consisting of various information attributes and their relationships with information quality and each other, few validate the simultaneous theoretical relationships inherent in the overall model structure. Assumptions underlying the major information quality models preclude the use of structural equation modeling via tools such as LISREL and AMOS. This study addresses both issues, empirically validating the structure of an information quality model with partial least squares applied to data from professionals who process patient claims information for health care providers. The structure of the model showed reasonable fit with the data sample, and predicted 59% of the variance in perceived overall IQ for information they received. Although participants rated all model attributes as important to information quality, two well-accepted information quality attributes, Accessibility and Relevance, did not adequately contribute to its prediction in the model. The results support and illustrate a formative approach to modeling information quality and the applicability of the tested model to patient claims information used by health care providers. They also raise questions about the ability of importance ratings to adequately identify information attributes that explain information quality perceptions, and they reinforce the context or domain-specificity of even fundamental information quality attributes.

INTRODUCTION & BACKGROUND

Quality of information is a critical issue in today's interconnected society. Poor quality information has caused political controversy and high-profile disasters [16], clinical accidents [12,23], lost productivity [13,25], and failed enterprise [19,29] with estimated costs in the millions of US dollars [13]. Many studies of information quality employ models that posit or depend on theorized relationships between information attributes and overall information quality. For example, the FASB [15] models information usefulness as hierarchically dependent on attributes such as Relevance, Reliability, and Timeliness (Figure 1). Wang and Strong's factor model [29], widely used in management information systems (MIS), shows information quality as dependent on dimensions such as Accessibility and attributes such as Interpretability, Accuracy, Completeness and Timeliness (Figure 2). Many information quality models (and studies) derive from or depend on one or both of these models, presenting a conceptual hierarchy of elements (dimensions or attributes) leading to IQ. However, few such overall models have had the simultaneous relationships in their structure validated. The FASB model is explanatory and, although Wang and Strong's measurement model was developed through factor analysis, the model structure was not validated. Similarly, although Lee, et al. [22] assessed perceptions of information actually used, they were unable to assess the structural relationships of the model.

Software tools for covariance-based structural equation modeling (SEM) abound (e.g. LISREL [20] and AMOS [1]), and the factor analytic development of the latent constructs of Wang and Strong's popular model makes this an alluring choice for validating IQ model structure. Chae and Kim [6], for example, used LISREL to empirically test attributes of IQ in a user satisfaction model for mobile internet services. Although they reported robust fit indices, there are several reasons to suggest that such application of covariance-based SEM for IQ modeling is inappropriate.

First, as noted by Lee, et al.[1] [22], IQ elements (the dimensions and their attributes) are not necessarily independent. Second, covariance-based SEM assumes multivariate normality of the data, which may not hold for Likert-scale questionnaire data. Most importantly, however, IQ models tend to be formative or molar (e.g., Figures 1 – 3). That is, IQ is modeled as a latent construct that is *formed from* or *caused by* other latent constructs (Figure 4). A theoretical assumption for covariance-based SEM is that the covariance between constructs is explained by the causal influence of an underlying, higher-order latent construct [10]. This assumes that IQ *causes* such things as Accessibility, Interpretability, Relevance, Reliability, and Timeliness, rather than the impression of IQ being a result of them. To apply covariance-based SEM, a change in a lower-order latent construct should be associated with change(s) in the others, since all are causally dependent to some degree on the higher-order latent construct of IQ. This is not the case. Applying covariance-based SEM for path analysis under the conditions outlined above can result in inappropriate or inadmissible solutions [10, 22], such as negative variances.

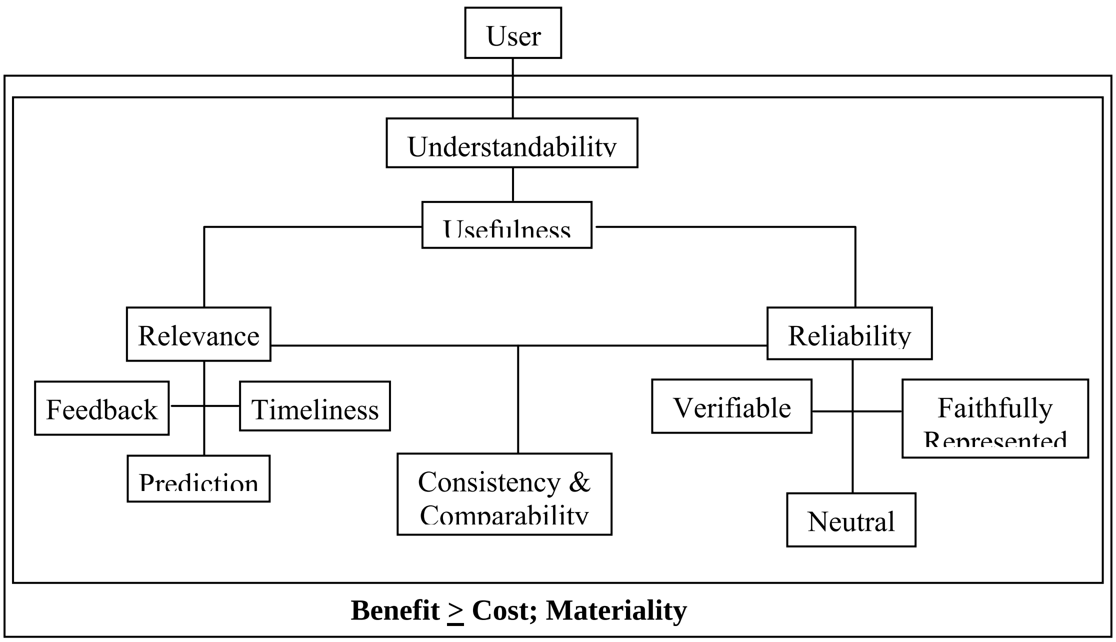

Figure 1. FASB [15] Model of accounting information usefulness. The two primary qualities leading to usefulness are *relevance* and *reliability*. Relevant information provides feedback or predictive value and is received in time to affect decision making (*timeliness*). Reliable information has sufficient evidence to be *verifiable*, is *neutral* (does not favor a particular outcome), and *faithfully represents* that which it purports to (is accurate, or in conformity with the thing of interest). Information can have decision usefulness yet lack *understandability*, as the FASB expected economically relevant accounting information to require a certain degree of specialized knowledge or training. Judgments of information usefulness are bounded by two constraints in the model: *materiality* (relative importance of the information) and *cost versus benefit* (information that cost more than the benefit derived from it would not be considered useful)

[1] Lee, et al. [22] used the Wang and Strong [29] IQ model.

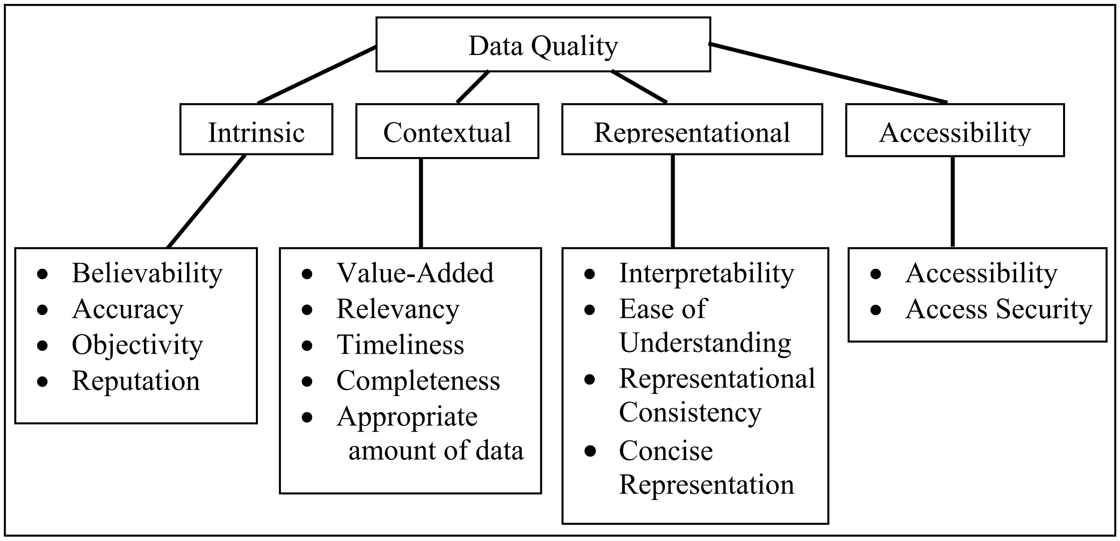

Figure 2. Wang and Strong's [29] model of data quality. Derived from factor analysis of consumer ratings of the importance of various terms to information quality, the resulting factors were then grouped in two sorting tasks by subjects. Consumers' perceptions of data quality are thus modeled as consisting of dimensions of *intrinsic*, *contextual*, *representational*, and *accessibility quality*.

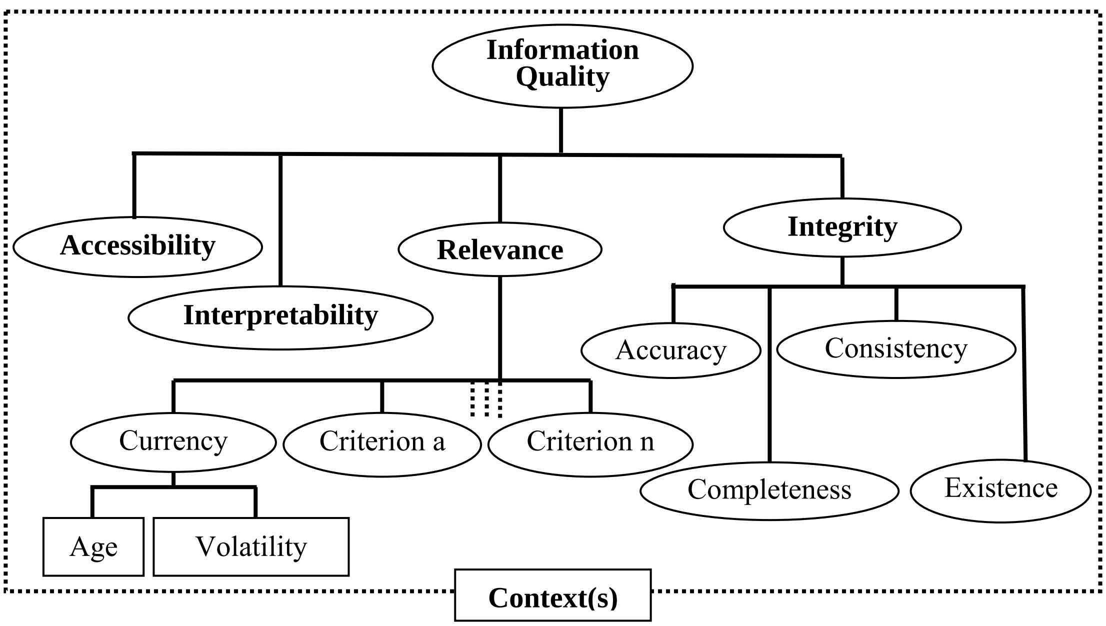

Figure 3. Modified Model of Information Quality [2, 3]. The main attributes of information quality in the proposed model can be summarized by the mnemonic of AIRI – *Accessibility*, *Interpretability*, *Relevance* and *Integrity*. Additional *Relevance* criteria other than *Currency* (e.g. "outpatients information only") are assumed to vary with the context and are not specifically tested here. *Integrity* has four elements: *Accuracy*, *Completeness*, *Consistency*, and *Existence* (which refers to information non-fictitiousness and non-redundancy). *Context* is a general boundary between information and the recipient, through which additional information and meaning can be derived. Context is the superset containing the criteria applied to each of the information quality attributes and elements during evaluation (e.g. a requirement that the information be in English to be interpretable)

RATIONALE AND PURPOSE

To deal with this, one might address the structure of the IQ model in question and its constructs, attempting to eliminate interdependencies between dimensions (e.g. see [2] or [3]), or gather huge volumes of data, as in Chae and Kim's study, to deal with distributional difficulties. Unfortunately, this does not resolve the fundamental discrepancy between the theoretical basis for covariance-based SEM and the formative, molar nature of IQ modeling. Even where inadmissible solutions do not occur, covariance-based SEM does not provide a predictive model and a large (potentially infinite) number of scores can be fitted to the model parameters. Given the mismatch between the theoretical assumptions underlying the method and IQ models, this is not comforting.

To address the lack of IQ model structure validation, this study instead empirically validated a general model of information quality [2, 3] using partial least squares (PLS), a *component-based SEM* method considered robust under the conditions outlined above. PLS is also referred to as 'soft modeling' due to its lack of 'hard' distributional assumptions regarding observed variables or their residuals [14, 24]. Since it applies ordinary least squares (OLS) and multivariate regression to maximize the explained variance of all dependent variables in the model, and arrives at determinate solutions for the proposed latent variables in a model, this makes PLS a useful technique where predictive solutions are desired [8]. However, PLS is also appropriate in situations where distributional assumptions cannot be met, where smaller sample sizes are involved, where formative indicator variables combine to form a latent variable rather than reflecting and covarying with it, and where (as with IQ models) latent variables combine in molar fashion to form higher order latent variables [7, 8, 9, 14, 24, 30].

To explain the variance in the dependent model variables, PLS iterates between two methods of estimating the proposed latent variables – the so-called *outside* and *inside* approximations [8, 14]. The 'outside' approximation estimates latent variables (LVs) from available values of the indicator variables. Reflective indicators are used from neighboring LVs to produce proxy values for each LV. Formative indicators are regressed on their respective LVs and the weights used to produce an estimate for the LV. Given these, the 'inside' approximation estimates the values of latent variables from their adjacent LVs according to the relationships in the model structure. PLS iteratively applies these solution approaches with the goal of maximizing the explained variance in the dependent latent and observed variables. Once the explained variance fails to decrease more than a specified amount from one iteration to the next (usually about $1/1000^{th}$), PLS is considered to have converged on a solution and OLS regression is used to determine the loadings, path coefficients, mean scores and location parameters for the latent variables and their indicators. PLS estimates LV values directly from indicator variables, so there are no difficulties with indeterminate LVs [8].

Sample sizes required for PLS can be considerably smaller than for covariance-based SEM. Suggested rules of thumb for the required number of cases are: a ratio of 5 cases per LV [14]; or, a minimum of ten times the larger of either a) the largest number of formative paths into a latent variable, or b) the greatest number of independent constructs influencing a dependent variable [8]. So, for example, in the IQ model tested here (Figures 3 and 4), there were four formative or molar paths from Accessibility, Interpretability, Relevance, and Integrity to IQ (or, alternately, four formative paths from Accuracy, Completeness, Consistency and Existence to the aggregate Integrity). Thus the largest number of formative paths or independent constructs influencing a dependent variable was four, and the total number of LVs was 9 (including Integrity). Using Falk and Miller's heuristic [14], the recommended minimum number of cases for analysis would have been 45. Using Chin's [8] heuristic, the recommended minimum would have been 40. Both estimates are considerably less than the 200+ that would be called for with covariance-based SEM.

Finally, PLS has been shown to produce robust results when sample data has a skewed distribution, shows multicollinearity, and even when indicator or latent variables are misspecified or missing [5]. Even omission of a manifest variable has been shown to have little effect on model estimates, both for inner and outer model structures. Similarly, skewed distributions or multicollinearity showed "no dramatic effects on fitted values of the manifest and latent variables" [5, pg S906]. Table 1 summarizes key differences between covariance-based and PLS SEM approaches.

The model evaluated was designed to address conceptual flaws in existing prominent models in the AIS and MIS literature, and to bridge IQ modeling in the two areas. Validation of this model (Figure

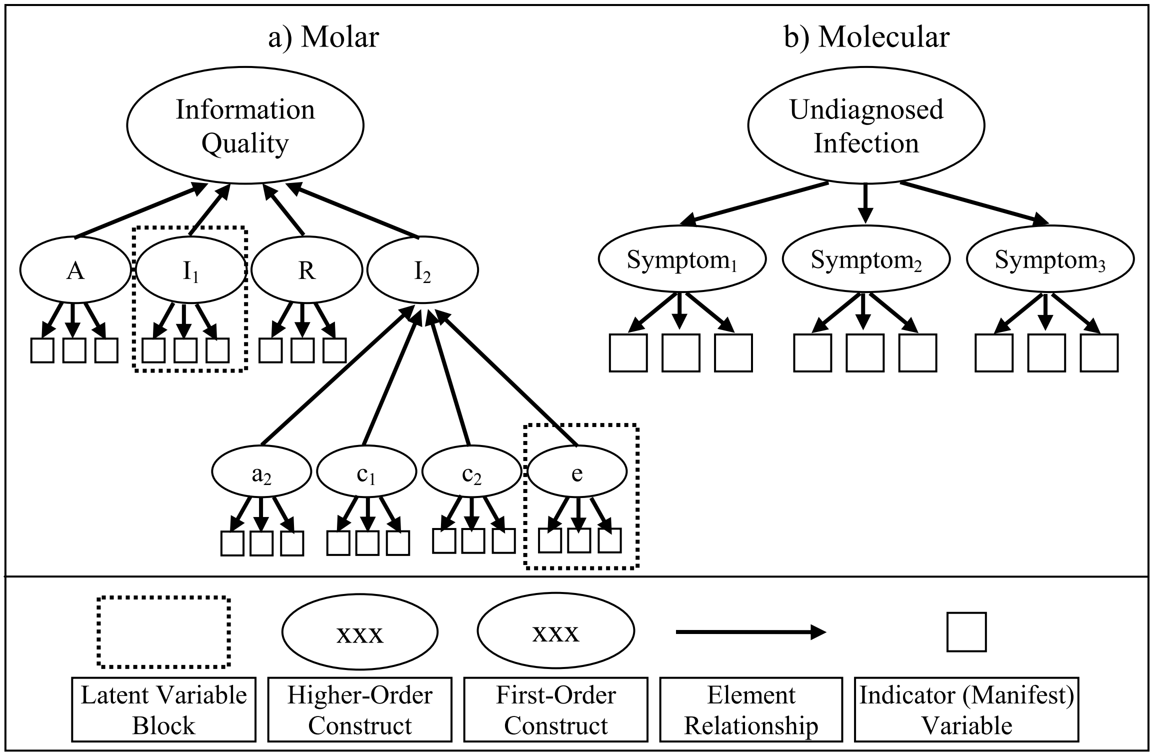

Figure 4. Molar (formative) and molecular (reflective) models. IQ is formed by lower-order constructs that are also formative, or are 'blocks' reflected in manifest variables. An undiagnosed infection could be modeled as reflected in symptom constructs that are reflected in manifest variables. For brevity, not all details of the IQ model tested here are depicted.

3) presented the possibility of a predictive model for comparative assessments of IQ within or across domains. For more detailed discussion of the model see [2, 3].

METHODS

Partial least squares analysis of the data set was performed using PLS Graph 3.0, Build 1060 [10]. Details about its operation can be found in the PLS Graph help file, online slides from a conference tutorial by Chin [9], and in an introductory tutorial by Buche [4]. In PLS modeling, loadings on each indicator are estimates of the first principle component loadings [14]; squared loadings (*communalities*) represent the amount of variance an indicator shares with other items through their common LV. Higher loadings represent greater shared variance among block indicators and therefore greater convergent validity of items for a LV. Loadings that equal or exceed 0.55 indicate at least 30% of the variance for an indicator is shared through the LV with other block indicators [14]. *Composite reliability* (ρ_c) [8] consists of the square sum of the item loadings (λ_i) of a block divided by that plus the sum of the block item variances ($var(\varepsilon_i)$), and gives an estimate of the reliability of LV block indicators:

$$\rho_c = (\sum\lambda_i)^2/[(\sum\lambda_i)^2 + (\sum_I var(\varepsilon_i))].$$

Higher composite reliability indicates more item variance for the block is shared through the LV. *Average variance explained* (AVE) [17] estimates the amount of variance captured by a LV from its indicator variables relative to the variance due to measurement error, and provides a measure of convergent validity:

$$AVE = (\sum\lambda_i^2/[\sum\lambda_i^2 + \sum_I var(\varepsilon_i)]).$$

For discriminant validity, Chin [8] suggests the AVE for a latent variable be greather than the largest squared correlation among LVs in the model. Thus more variance is explained within the construct and its block of indicators than between the construct and some other, purportedly different, indicator block. Falk and Miller [14] suggest covariance of less than 0.20 between residuals of items from different LVs indicates satisfactorily discriminant constructs.

ASPECT	COVARIANCE	LEAST SQUARES
Example Software Tools	LISREL, EQS, AMOS	LVPLS, PLS-GUI, PLS-Graph
Applications	Model building – exploratory Model testing - confirmatory	Exploratory or confirmatory prediction
Explanation Method	Latent Variable Model Fit	Latent Variable Prediction
Estimation Method	Maximum likelihood estimation	Ordinary least squares regression, and multivariate regression
Dependence on Theory	Higher	Lower
Sample Size Required	Much Larger ($\geq$ 200+ cases)	Smaller (10 times the larger of formative paths into a construct or indicators for a single construct)
Distributional Requirements	Multivariate normal	None Assumed
Indicator-to-Latent; Latent-to-Latent Relationships	Reflective	Reflective, Formative, or both
Indices of Success	Minimized residual covariance matrix	Maximized dependent variance explained
Latent variable values	Indeterminate	Determinate
Evaluation	Fit indices	Indicator loadings, path coefficients, and R^2 values
Misinterpretation of reflective nature of indicators, or violations of sample size or distributional assumptions	Improper and indeterminate solutions possible (e.g. 'Heywood' cases)	NA

Table 1. Structural Modeling Approach Differences. Summary of key differences between covariance-based and component-based or partial least squares approach to structural modeling.

The variance explained in target endogenous variables of interest in a model is a principal measure of the model [8, 14]. Where the goal is the ability of the model to predict these variables, their average R^2 is the criterion. Although no hard and fast rule exists regarding the acceptable minimum total variance explained for an endogenous variable, Falk and Miller [14] recommend at least 10%. They also suggest each individual LV predictor should account for at least approximately 1.5% of the variance in its endogenous target. The product of the weight and loading values assigned by PLS gives an estimate of this measure.

Where model validation is of interest, strength of the theorized model paths, as indicators of the model fit with data, are also of importance. T-statistics for paths and loadings were produced by bootstrap resampling, with replacement, of the original data set. Significant t-statistics indicated a path or loading value that was sufficiently stable to warrant its contribution to the model.

DATA SAMPLE

Study participation was solicited by direct contact (cold-calling) of the business manager or individual(s) responsible for claims billing at healthcare providers (hospitals, and single and group practices of physical therapy, chiropractic or clinical medicine). Sample data was gathered by an anonymous, confidential Web-based survey that consisted of general statements about information quality and the attributes and elements in the model, carefully drafted to be conceptually consistent with the pool of preexisting questions and theoretical constructs from the IQ literature (Table 2 shows statements from the section regarding Interpretability). In each set the first statement was directed toward perceived importance of the model construct and the remaining statements toward perceived presence or absence of it. Each statement was accompanied by a seven-point Likert scale, the anchored meanings of which were provided by illustrated example at the beginning of the survey (Figure 5). Survey items were analyzed by Kruskal-Wallis test [26, 27] for differences across self-reported demographic groups (Table 3). No meaningful patterns or trends in significant differences were evident. Summary demographics for the final sample of 198 surveys are provided in Tables 4 and 5.

How understandable and meaningful is the information that you receive from claims payers? That is, how interpretable is it to you?
It is important that information be understandable
Information is easy to understand
Information always makes sense
Information is routinely meaningless or unintelligible
Interpreting information is seldom easy
Making sense of information is a struggle

Table 2. Survey Statements Regarding Interpretability. The set of statements was preceded by the listed framing questions, and each statement was followed by a 7-point Likert scale that had been previously anchored (see Figure 5). Italics (used here only) highlight reverse-coded statements.[2] In all cases the first statement addressed the importance of the construct to enable comparison of the perceived importance of constructs ratings with perceptions of attributes of actual information received.

[2] Survey items for the model attributes and demographics can be found at http://www.bsad.uvm.edu/files/iciq2004/boveeSurvey.mht , http://www.bsad.uvm.edu/files/iciq2004/boveeSurvey.htm , and http://www.bsad.uvm.edu/files/iciq2004/boveeSurvey.pdf

1	2	3	4	5	6	7
Strongly Agree	Agree	Somewhat Agree	Neutral	Somewhat Disagree	Disagree	Strongly Disagree

Figure 5. Seven-point Likert Scale Anchors. These phrases anchored the scale points as shown for all Likert scale responses throughout the survey. This figure preceded each major section of information quality statements to be rated. Answer coding for negative statements was reversed prior to analyses.

Survey Demographic Groups	
Participant	Sex
	Processing Role
	Educational Level
Place of Employment	Firm Size
	Provider Type
	HIPAA Requirements Implementation Status
	Internet Speed
	Location by State

Table 3. Survey Demographic Groups. Survey data was examined by Kruskal-Wallis test [26, 27] for significant differences within the following dimensions. No meaningful trends or patterns were found.

Demographic		**Mean (s.d.)**
Years coding experience		11.66 (8.34)
Age (yr)		44.9 (10.03)
		%
Gender	• Female • Male	89.8% 10.2%
Education	• High School or Equivalent • Technical School Certification • Junior College or Associates • Bachelors • Masters • Doctorate or MD	28.3% 7.5% 21.4% 23.5% 9.6% 9.6%
Role	• Enter/transmit claims • Problem resolution • Multiple aspects • Manage people only • Healthcare only • Healthcare and claims	6.5% 3.8% 56.5% 18.3% 4.3% 10.8%
Job Title	• Clerical • Managerial • Senior Administration • Owner • Healthcare Provider	20.0% 55.0% 5.0% 5.0% 15.0%

Table 4. Participant Demographics. Distribution of the final sample by respondent coding experience, age, gender, education, claims processing role, and job title.

Demographic		Mean (s.d.)
	% Electronic Claims	76.31 (22.18)
	% Rejected Claims	9.64 (11.52)
	% Electronic Claims	**%**
Healthcare Provider Type	• Chiropractic • MD/DO • Physical Therapy	16.3% 67.9% 15.8%
Organization Type	• Single Practice • Group Practice • Clinic • Hospital • Other	40.4% 37.4% 4.5% 13.6% 4.0%
Number of Providers	• 1 • 2 – 5 • 6 – 10 • 11 – 25 • More Than 25	33.0% 36.0% 11.7% 5.6% 13.7%

Table 5. Facility Demographics. Details of the distribution of the final survey sample by facility attributes are summarized.

RESULTS

As illustrated in Table 2 the first statement in each group required participants to rate the importance of the model attribute to information quality. Participants rated all attributes of the model as very important to the quality of patient claims information they received (Mean: 1.26, Range: 1.16-1.36; A rating of 1 indicated participants very strongly agreed that the attribute was important to the quality of patient claims information received).

Overall results of the outer (measurement model) suggested satisfactory reliability, convergent and discriminant validity of the theorized latent constructs. Mean (s.d.) loading of items in the individual LV blocks across all indicator variables was 0.779 (0.089). Composite reliabilities for the LVs ranged from 0.781 (Currency) to 0.941 (Accessibility), with a mean (s.d) of 0.867 (0.044). AVE for all LVs in the model ranged from 0.504 (Integrity) to 0.842 (Accessibility), with a mean (s.d) of 0.650 (0.094). Composite reliability and AVE results are summarized in Table 6.

No correlations between indicator residuals from within a block and residuals of indicators from outside the block exceeded 0.200. Table 10 summarizes the squared latent variable correlations for the model applied to consumer data. In all cases, the AVE for each latent variable was larger than the largest squared correlation between that LV and all others[3], and the mean AVE for all LVs was considerably larger than the mean squared correlation between LVs. Thus the final outer model was deemed to have satisfactory convergent and discriminant validity.

Results of the model fit with data from the context of health care claims receipt are summarized in Figure 6. The variance in overall Information Quality explained in the data by the model was over 59%, and the variance in Relevance explained by Currency was over 35% (Table 7). Both values were above the recommended 10% minimum and highly significant by the F test [18]($p < 0.0001$).

[3] Integrity is not represented in this table since it is an aggregate formed entirely from other LVs with no indicator variables of its own.

Construct	Reliability	AVE
IQ	0.864	0.615
Accessibility	0.941	0.842
Interpretability	0.913	0.723
Relevance	0.889	0.667
Currency	0.781	0.551
Integrity	0.929	0.504
Accuracy	0.872	0.630
Completeness	0.845	0.646
Consistency	0.877	0.704
Existence	0.827	0.614

Table 6. Outer Model Results – Reliability and Convergent Validity. Composite reliability and Average Variance Explained for all latent constructs are shown.

All of the LVs theorized to contribute to prediction of Information Quality explained more than the recommended minimum of 1.5% variance (Table 8). However, Accessibility (4.0%) and Relevance (6.4%) both explained the lowest percentages of variance in IQ and had non-significant path values, suggesting they did not contribute meaningfully to the prediction of variance in Information Quality in the model (Table 9). Interpretability (17.3%) and Integrity (31.5%) explained significant amounts of variance and had significantly stable path values. The path value from Currency to Relevance (0.597) was highly significantly stable, and Currency explained over 35% of the variance in Relevance. Path values for all LVs theorized to form Integrity (Table 9) were highly significant ($p < 0.0005$), suggesting that they were stable and interacted as modeled.

DISCUSSION

The general model of information quality developed by Bovee, et al. [2] and Bovee [3] (Figure 3) and tested here had *Accessibility*, *Interpretability*, *Relevance*, and *Integrity* as attributes of overall Information Quality, *Currency* as a sub-attribute of *Relevance*, and *Accuracy*, *Completeness*, *Consistency*, and *Existence* as sub-attributes of *Integrity*. Seven out of nine of the model attributes were shown to have significantly stable paths (Tables 8-10), to contribute variance to higher-order constructs as predicted, and to explain or significantly predict variance in their target constructs. Based on the structural analysis by PLS, the results validated both a majority of the developed model structure and its ability to significantly predict perceptions of overall Information Quality. This provided support for the theoretical and conceptual foundations of the model, and the theorized formative nature of the relationships between the attributes and overall Information Quality.

Overall, study participants considered all elements evaluated to be very important to overall Information Quality.[4] Yet Accessibility and Relevance of information received explained little of the variance in perceived overall Information Quality and had non-significant paths to IQ. This finding is very surprising because both Accessibility and Relevance are prominent attributes in many other information quality studies. Relevance, for example, occupies a significant place in the FASB [15] hierarchy of information usefulness. Accessibility is listed by Wang et al. [28] and Bovee, et al. [2] as a logically necessary prerequisite for all other information quality attributes.

[4] While no statements regarding model construct *importance* were stated in the negative or directed toward the criticality of construct absence to information quality, each statement set regarding presence of a construct in information received did contain positively and negatively worded statements.

When considering the quality of information received, it seems unlikely participants may have ignored whether the information was accessible or relevant. They may have assumed that such attributes were given byproducts of pre-existing coding and transmission standards. If so, however, this assumption could be expected to impact other attributes as well, such as Interpretability, and it did not.

As information consumers, they perceived Accessibility and Relevance as pertinent attributes. As consumers of information they should have been able to judge these attributes and they still did not contribute significantly to the prediction of perceived Information Quality. It is unclear whether the attributes of Accessibility and Relevance simply do not determine as much of the perception of information quality as conventional wisdom and theory indicate, whether the context of health care claims process presents an environment in which the tested theoretical model of information quality does not wholly apply.

These results are important – they cast doubt on the empirical contribution of two otherwise well-accepted information attributes as determinants of individuals' perceptions of overall Information Quality. Even if such findings are limited to health care claims processing, or the health care information-processing field, this represents a large number of individuals and organizations. If evaluations and implementations of information quality in these or other similar domains that rely on assessed perceptions of Accessibility and Relevance are misguided, the impact is widespread.

LIMITATIONS

Due to the inherent difficulties in gaining access to patient data, the study did not examine actual information quality. Instead the results are based on the perceptions of the participants. With the implementation of HIPAA guidelines further safeguarding patient information privacy, overcoming this limitation may not be feasible. Since the study relied solely on survey data to empirically validate the information quality model developed, there is the potential for mono-method bias. While this was offset by pilot studies used to judge whether testing of information quality in the health care provider claims processing domain was warranted, the study was also restricted to this context. Finally, though the number of subjects in the sample was more than adequate for the method of analysis, it represented a small fraction of the total number of healthcare providers in the United States. All conditions limited the ability to generalize from the results.

CONCLUSIONS

Nonetheless, using modeling methods consistent with and robust to the theoretical assumptions inherent in IQ models, the study empirically validated a predictive model of information quality designed to combine elements from dominant preexisting models in AIS and MIS and simultaneously address theoretical and conceptual problems in them. Since assumptions underlying the major information quality models preclude the use of SEM tools such as LISREL and AMOS, IQ researchers should consider PLS for model path analysis or predictive IQ modeling instead. The validated model may enable comparative studies of information quality between the two domains and, as a general model, in additional domains.

The results also cast doubt on the practice of rating the importance of information quality attributes as a means of selecting attributes that explain variance in individuals' perceptions of information quality. In addition, the results suggested that Accessibility and Relevance, two attributes of information that occur frequently in the literature and that play a prominent role in many of the major information quality models, may not contribute to the perception of information quality in the health care claims processing domain. Analysis of the phenomenon of information quality in this domain is important, since the potential cumulative impact of poor quality claims information on administrative costs is enormous.

Construct	R^2	F-Statistic	Sig.
IQ	.592	70.01	<.0001
Relevance	.352	106.47	<.0001

Table 7. Target Construct Variance Explained. Variance explained in target endogenous variables and the associated significances is listed.

LV	Dependent LV	Path	Correlation	Approximate Variance Explained
Accessibility	IQ	0.073	0.545	4.0%
Interpretability	IQ	0.256	0.675	17.3%
Relevance	IQ	0.106	0.602	6.4%
Currency	Relevance	0.597	0.593	35.4%
Integrity	IQ	0.431	0.732	31.5%
Accuracy	Integrity	0.354	0.692	24.5%
Completeness	Integrity	0.254	0.656	16.7%
Consistency	Integrity	0.292	0.662	19.3%
Existence	Integrity	0.228	0.537	12.2%

Table 8. Inner Model LV and Path Values. Path and correlation values between latent variables, and approximate target construct variance explained by each LV are listed.

Construct	Target LV	Path	T-Statistic	Sig.
Accessibility	IQ	0.073	1.336	NS[5]
Interpretability	IQ	0.256	3.106	<.005
Relevance	IQ	0.106	1.472	NS[6]
Currency	Relevance	0.597	12.726	<.0005
Integrity	IQ	0.431	5.804	<.0005
Accuracy	Integrity	0.354	23.870	<.0005
Completeness	Integrity	0.254	22.411	<.0005
Consistency	Integrity	0.292	23.763	<.0005
Existence	Integrity	0.228	15.446	<.0005

Table 9. Model Path Stability. Construct-to-target paths and their t-statistics are provided. Paths with significant t-statistics, generated by bootstrapping with resampling, are stable enough to warrant their contribution to the prediction of variance in the model.

[5] Significant at $p < 0.10$.
[6] Significant at $p < 0.10$.

AVE	Construct	Squared Correlation							
		IQ	Accessibility	Interpretability	Relevance	Currency	Accuracy	Completeness	Consistency
0.842	Accessibility	0.456	–	–	–	–	–	–	–
0.723	Interpretability	0.477	0.477	–	–	–	–	–	–
0.667	Relevance	0.297	0.335	0.410	–	–	–	–	–
0.551	Currency	0.364	0.434	0.415	0.215	–	–	–	–
0.630	Accuracy	0.258	0.308	0.408	0.187	0.352	–	–	–
0.642	Completeness	0.421	0.354	0.570	0.298	0.307	0.303	–	–
0.705	Consistency	0.312	0.319	0.429	0.225	0.255	0.289	0.475	–
0.614	Existence	0.438	0.413	0.529	0.303	0.469	0.341	0.552	0.430
0.672 (0.087)	**AVERAGE (s.d.)**	0.373 (0.096)							

Table 10. Discriminant Validity. Average Variance Explained (AVE) per latent variable and squared correlations (r^2) between all latent variables are provided. AVE for each latent variable exceeds the maximum r^2 with any other latent variable, suggesting good discriminant validity.

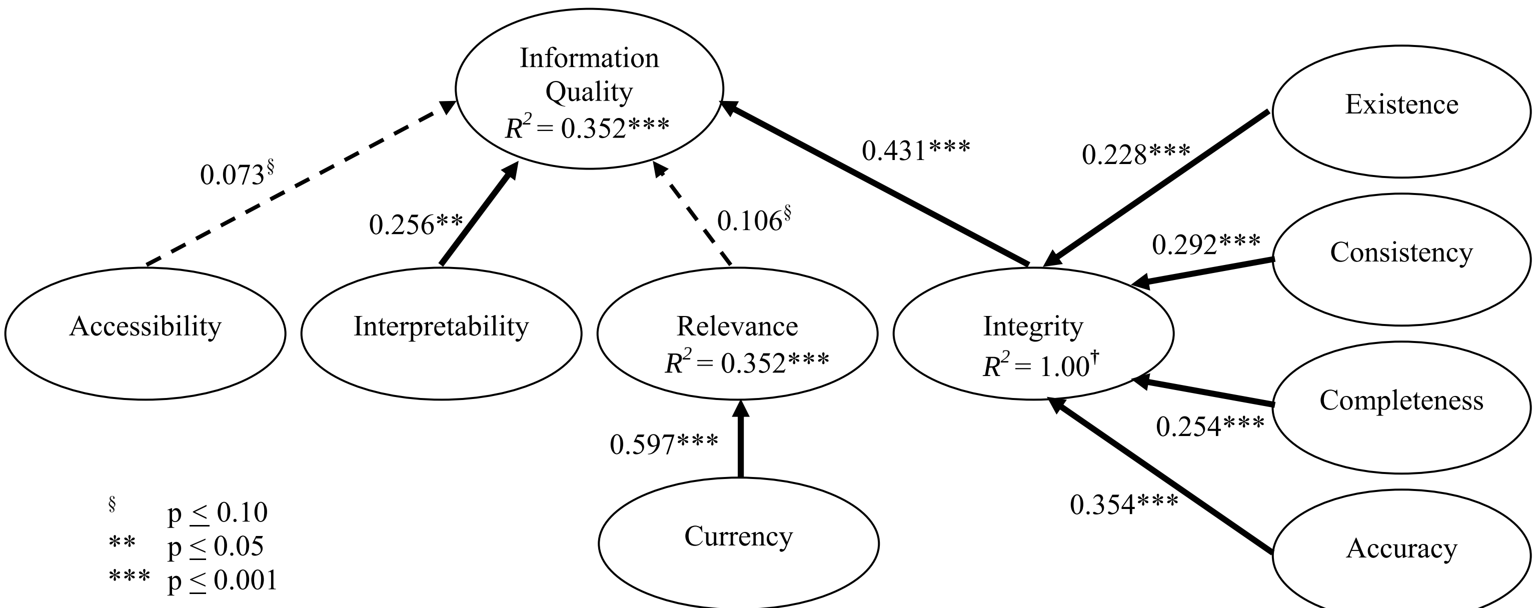

Figure 6. Path Diagram. The figure shows path values, endogenous variable variance explained, and significance levels based on data fitted to the proposed model. Integrity is an aggregate latent construct created from Accuracy, Completeness, Consistency, and Existence, with all its variance explained†**.**

REFERENCES

[1] Arbuckle, J.L. and W. Wothke (1999) *Amos 4.0 User's Guide*. SmallWaters Corporation, Chicago, IL, USA.

[2] Bovee, M. W., Srivastava, R. P., and B. Mak (2003). "A Conceptual Framework and Belief-Function Approach to Assessing Overall Information Quality". *International Journal of Intelligent Systems* 18(1):51-74.

[3] Bovee, M. W. (2004) Information Quality: A Conceptual Framework and Empirical Validation, Ph.D. Dissertation, School of Business, University of Kansas, June.

[4] Buche, Mari. (2003) *IT Professional Work Identity: Construct and Outcomes*, Ph.D. Dissertation, School of Business, University of Kansas, July.

[5] Cassell, C. M., Hackl, P. and A. H. Westlund (2000). On Measurement of Intangible Assets: A Study of Robustness of Partial Least Squares. *Total Quality Management*, 11(7):S897-S907.

[6] Chae, M. and J. Kim (2001). "Information Quality for Mobile Internet Services: A Theoretical Model with Empirical Validation". *Proceedings of the 22nd International Conference on Information Systems*, New Orleans, Louisiana, USA.

[7] Chin, W.W. (1998a). Issues and Opinion on Structural Equation Modeling. *MIS Quarterly* 22(1):vii-xvi.

[8] Chin, W.W. (1998b) The Partial Least Squares Approach to Structural Equation Modeling, pp. 295-336 in *Modern Methods for Business Research*, G.A. Marcoulides (ed.), Erlbaum Associates, Mahwah, NJ, USA.

[9] Chin, W.W. (2000). Partial Least Squares For Researchers: An overview and presentation of recent advances using the PLS approach, *Proceedings of the 21st International Conference on Information Systems*, Brisbane, Queensland, Australia. http://disc-nt.cba.uh.edu/chin/icis2000plstalk.pdf

[10] Chin, W.W., & T. Frye (2003). PLS-Graph version 3.0. Software program licensed and distributed by Soft Modeling, Inc. Installed using Build 1060 with 200 variable limit.

[11] Chin, W. W., & P. R. Newsted (1995). The Importance of Specification in Causal Modeling: The Case of End-user Computing Satisfaction. *Information Systems Research*, 6(1):73-81.

[12] CNN.com (2001). DOD: Soldiers may get tainted blood. http://www.cnn.com/2001/HEALTH/11/07/military.blood.ap/index.html, Last accessed November 7, 2001 12:04 PM EST.

[13] English, L. P. (1999) *Improving Data Warehouse and Business Information Quality: Methods for Reducing Costs and Increasing Profits*. John Wiley and Sons, Inc, New York, New York, USA.

[14] Falk, Frank R. and Nancy B. Miller (1992). *A Primer For Soft Modeling*, The University of Akron Press, Akron, Ohio, USA.

[15] Financial Accounting Standards Board (1993). "Statement of Financial Accounting Concepts No. 2, Qualitative Characteristics of Accounting Information", *Original Pronouncements, Accounting Standards as of June 1, 1993, Volume II (AICPA Pronouncements, FASB Interpretations, FASB Concepts Statements, FASB Technical Bulletins),* Financial Accounting Standards Board, CN, USA.

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

[17] Fornell, C., & D. F. Larker (1981). "Structural Equation Models with Unobservable Variable and Measurement Error: Algebra and Statistics." *Journal of Marketing Research*, 18(February):382-388.

[18] Harris, Richard J. (2001) *A Primer of Multivariate Statistics, 3rd Ed.*, Lawrence Erlbaum Associates, Mahwah, New Jersey.

[19] Huang, K-T., Lee, Y.W and R. Y. Wang (1999). *Quality Information and Knowledge*, Prentice Hall, Upper Saddle River New Jersey, USA.

[20] Jöreskog, K.G. and D. Sörbom (1996). *LISREL 8: User's Reference Guide*, Scientific Software International, USA.

[21] Kahn, BK, Strong, DM, and R. Y. Wang (2002). Information Quality Benchmarks: Product and Service Performance. *Communications of the ACM* 45(4):184-192.

[22] Lee, Y.W., Strong, D.M., Kahn, B., R. Y. Wang (2002). AIMQ: A Methodology for Information Quality Assessment. *Information & Management* 40(2):133-146.

[23] Leveson, N. G. and C. S. Turner (1993). "An Investigation of the Therac-25 accidents". *Computer* 26(7):18-41.

[24] Lohmöller, J. (1989). *Latent Variable Path Modeling With Partial Least Squares*. Physica-Verlag, Heidelberg, Germany.

[25] Redman, Thomas C. (1996). *Data Quality For The Information Age,* Artech House, Boston, USA.

[26] Siegel, S. and N. J. Castellan, Jr. (1988). *Nonparametric Statistics For the Behavioral Sciences, 2nd Ed*. McGraw Hill, Inc., New York, New York, USA.

[27] SPSS for Windows (2002). Release 11.5.0 (6 Sep 2002), SPSS, Inc. 1989-2002.

[28] Wang, R. Y., Reddy, M. P. and H. B. Kon (1995). Toward quality data: An attribute-based approach. *Decision Support Systems*, 13(3/4):349-372.

[29] Wang, R.Y. and D.M. Strong (1996). "Beyond Accuracy, What Data Quality Means to Data Consumers." *Journal of Management Information Systems* 12(4):5-34.

[30] Wixom, B. H. & H. J. Watson (2001). An Empirical Investigation of the Factors Affecting Data Warehousing Success. *MIS Quarterly* 25(1):1-41.