

A DISCREPANCY-BASED MEASUREMENT APPROACH  
FOR DATA INTEGRATION

Gregory E. Truman

Stern #IS-95-24

A DISCREPANCY-BASED MEASUREMENT APPROACH  
FOR DATA INTEGRATION

Gregory E. Truman  
Department of Information Systems  
New York University  
Leonard N. Stern School of Business  
44 West 4th Street  
New York, NY 10012-1126

(212) 998-0800

GREGORY E. TRUMAN is a Visiting Assistant Professor of Information Systems at the Stern School of Business at New York University, where he earned his Ph.D. in 1994. His current research focus is on interorganizational information systems, systems integration and organization performance impacts. Dr. Truman has published articles in Management Information Systems Quarterly and Communications of the ACM, and is a member of INFORMS, ACM and the Association for Information Systems (AIS).

Working Paper Series  
Stern #IS-95-24

# **A Discrepancy-Based Measurement Approach for Data Integration**

## **Abstract**

This paper presents a measurement instrument for diagnosing the level of data integration at the organization and organization subunit levels. Incorporating consideration for specific theoretical properties related to data integration into its design, the measurement instrument relies on a discrepancy (or gap) measure to adequately account for these properties. Data from forty-eight organizations of the Group Insurance industry are used to test the reliability and validity of the measurement instrument. The results suggest that the discrepancy measure, based on the *difference* between respondents' ideal (i.e. normative) and actual estimates of data integration, appears to be a valid indicator of data integration. However the efficacy of using a discrepancy measure, over the simpler use of its parts, must be considered in light of practical and theoretical considerations.

## 1. Introduction

A perspective for managing information technology (IT) as an “*infrastructure*” or a “*platform*” has recently gained recognition [Weill 1993, Keen 1991]. Typically based on the premise of creating or enhancing organizations’ strategic advantages or opportunities, treatment of IT as infrastructure is frequently alluded to *as part* of a new and different managerial perspective—one that transforms senior management’s view of IT from an exclusive operational or back-office role to a supplementary strategic role intricately intertwined with the firm’s overall strategic direction [Henderson and Venkatraman 1993, Keen 1993, Luftman et al, 1993]. The infrastructure, hereafter referred to as the organizational computing infrastructure (OCI), may be defined as inclusive of several elements including computer hardware, systems software, application software, communications hardware and software, and data. However it is not handling of these elements in singular fashion that constitutes an organizational computing infrastructure, but rather managing the elements as a whole with the objective of promoting organizational coordination through integration.

As one means to adopt a pro-OCI management posture, Keen (1993) discusses the linking of business and IT imperatives. He suggests that when a (strategic) business imperative prescribes an IT imperative, it is critical that management view IT “...as a business priority, not a technical support function.” Keen (1993) continues:

“IT imperatives do not in general point to specific applications or systems but highlight where there is a need for shared information or communication resources. These are shared corporate infrastructures, which are here termed the I/T platform. ...Where the imperatives point to the I/T platform crossing functional, divisional, and political boundaries, senior managers must ensure that those boundaries do not become barriers. This often depends on big rules.” (p. 19)

The “big rules” are described as policy statements which clarify and guide cross-functional coordination in order to assure successful attainment of (strategic) business imperatives.

Business networking, Keen’s term to refer to the management perspective of

information technology as an organizational computing infrastructure, insists on a cross-functional triad of systems, *data stores* and communications. This cross-functional triad requires substantial effort and investment so that, evolving over time, the three parts collectively inhere greater integrative capabilities.

Granted through sustained management, an OCI is argued to position a firm to obtain comparatively greater efficiency and effectiveness benefits.<sup>1</sup> For example, greater IT integration is believed critical for the effective use of interorganizational systems--systems that support business processes that cut across organizational boundaries [Konsynski 1993, Hart and Estrin 1991, Rockart and Short 1990]. Konsynski (1993) states:

“The internal structure of the organization is not immune to impacts associated with these (interorganizational) linkages. An interesting emerging phenomenon is the recognition that the entire transaction set of an organization is potentially subject to EDI. Therefore, they need to establish and design a new architecture (i.e., infrastructure).” (p. 139)

In light of the evolving view of IT as an organizational computing infrastructure or platform, and the increasing preponderance of interorganizational systems, the importance of IT integration is elevated. Though the notion of IT integration applies broadly, encompassing computing hardware, systems software, application software, communications hardware and software, and data, this paper focuses exclusively on data.<sup>2</sup> More specifically, the objective of this paper is to offer a measurement instrument for assessing or diagnosing data integration at the organization and organization subunit levels. Incorporating consideration for specific theoretical properties related to data integration into its design, the measurement instrument relies on discrepancy (or gap) scores to adequately account for these properties.

---

<sup>1</sup> Drawing an analogy between IT infrastructure investments by companies to public infrastructure investments by government agencies, Weill (1993) proposed that such investments may enhance an organization's strategic positioning.

<sup>2</sup> As an essential part of the more inclusive notion of 'Strategic Data Planning' [Goodhue et al 1992a, Lederer and Sethi 1991], data integration assessment is critical to these, often massive and risky, endeavors.

## **2. The Nature of Data Integration**

Beginning with a brief statement of assumptions regarding organizations, this section proceeds to discuss the nature of data integration as an organizational coordination device while highlighting the efficacy of discrepancy measures as central to its effective measurement.

### **2.1 Organization as Hierarchy**

Though the constitutive features of an organization and its environment (i.e. where one organization ends and the next begins) are fluid and subject to debate, it is assumed that an organization as a self-contained entity can be defined. In addition, it is assumed that any organization can be conceptualized to reflect the common bureaucratic feature of hierarchy as defined in Max Weber's (1946) classical writings on bureaucracy.

This hierarchical conceptualization reveals the organization as an inverted pyramid-like structure of several levels, with each successive level holding multiple, and (typically) an increasing number of, subunits. For purposes of subsequent discussion, an organization is assumed to have multiple (i.e., two or more) levels, and multiple subunits at each level except the highest level--Level 0, which consists of one unit and represents the most inclusive level corresponding to the organization as a whole.

### **2.2 IT as Coordination**

Lawrence and Lorsch's (1967) theoretical framework on organizations and environments suggests that out of varying environmental demands rise the efficacy of differentiating organizational subunits along certain dimensions, causing need for integrating their interdependent activities. Supported through empirical analyses, their argument indicates that appropriate integrative devices for coordinating the activities of differentiated organizational subunits will lead to improved performance. Though their use of the term 'integration' was more comprehensive in meaning, IT is one instrument to

furnish integrative mechanisms.<sup>3</sup> Therefore IT as an integrative device may, according to theoretical rationale, provide the benefit of enhanced coordination.

The *application* of IT does not grant uniform integrative support however; rather, there are varying discretionary technological arrangements regarding IT that may mitigate the integration effects afforded by it and lead to varied performance levels across organizations. Though most computers used in business are general purpose, IT is inherently varied across types (e.g., mainframes, midrange, micros), and across hardware and software vendors (e.g., Apple, IBM, Hewlett-Packard, Dell, etc.). Moreover, there are near-infinite design options in the *development* of databases and application systems (e.g., centralized versus decentralized computing architecture, relational versus hierarchical data architecture) which also manifest varied capacities to integrate IT. These and many other discretionary technological arrangements afford varying integrative capacities which, in turn, may influence the coordinative features theoretically presumed to extend from integration, and underscore incentives for measuring and diagnosing data integration.

### 2.3 IT as Coordination and Constraint

Goodhue, et al (1992b) provide convincing theoretically-based arguments that suggest high data integration may concurrently promote coordination of interdependent organizational subunits and constrain their flexibility or local autonomy. Consequently, they argue that the benefits of high data integration should be carefully weighed against

---

<sup>3</sup> Though IT is not specifically mentioned as an integrative device, it is assumed that Lawrence and Lorsch's use of the term 'paper systems' is intended to include computerized information systems. In a latter chapter entitled "Implications for Practical Affairs", they discuss control systems, payment systems, manpower selection, placement and promotion systems as specific management practices or options for attaining appropriate levels of integration in response to varying differentiation across subunits. And with regards to control systems, they write: "The degree of uncertainty of information could also be considered in control system design. Are the time interval and the detail of reporting adjusted for variations in certainty? The computer's great and growing capability for processing information makes such a flexibly designed control system an eminently practical choice." (p.226) Had the pervasiveness of computerized information systems in organizations predated their work of 1967, it is conceivable they might have explicitly recognized computerized information systems as an integrative device.

the costs when executing system design decisions. To paraphrase their rational<sup>4</sup> analysis of the benefits and costs associated with data integration, Goodhue et al (1992b) effectively argue that the “goodness” associated with increasing levels of data integration is not monotonically increasing. In order to measure internal systems integration therefore, it is necessary to develop a measure which is sensitive to this nonmonotonic property.

Wybo (1992) recognized that a “desirable” or “optimal” level of data integration between two organizational subunits’ application systems is contingent on the subunits’ perspective regarding their perceived level of interdependence. This contingency creates asymmetric perspectives regarding optimal data integration levels. Perceiving greater interdependence, Subunit 1 may desire that 80% of its data be integrated with Subunit 2’s data. Subunit 2, on the other hand, may perceive less interdependence with Subunit 1 and desire that only 30% of its data be integrated. Assuming the perceptions on behalf of each subunit reflect optimal data integration levels, these percentages, accepted as nominal or absolute figures, would reflect dramatically different data integration levels depending on whose perspective is assumed. A data integration measure should inhere a degree of relativity therefore, sensitizing the measurement instrument to this asymmetric property of data integration.

## 2.4 The Discrepancy Measure Concept

Discrepancy measures capture respondents’ judgment regarding the desired (or normative) *and* actual degree, amount or level of some phenomenon. The *discrepancy*, or difference, between the desired and actual scores are then typically used for subsequent empirical analysis. The utility of discrepancy measures extends from the more general *discrepancy concept*, introduced by Alutto and Belasco (1972) for measuring organization members’ participation in organization decision-making. The discrepancy between desired and actual assessments accommodates situations where no *universally* desired amount or level of some phenomenon can be adequately conveyed through some

---

<sup>4</sup> Goodhue, et al (1992) present a political analysis of data integration as well.



absolute value. With regards to participation in organization decision-making, Alutto and Belasco (1972) state:

Researchers too often maintain that the desire for increased participation is equally and widely distributed throughout an organization. Other evidence indicates that organizational populations are far from homogeneous in attitudes, sentiments, and expectations concerning a wide range of organizational issues. It is more reasonable to assume that not all segments of the population are equally desirous of additional participation in organizational life. If this is correct, then the crucial variable is the discrepancy between current (actual) and desired (ideal) rates of participation rather than a system member's absolute rate of participation. (p. 118)

The discrepancy concept has been applied for measuring other organizational phenomena. For example, discrepancy measures have been used extensively in the measurement of customer satisfaction [Bolton and Drew 1991, Churchill and Suprenant 1982]. Based on the so-called disconfirmation paradigm, this research is based on the premise that the discrepancy, or gap, between a customer's perceived (actual) and expected service performance intervenes to affect customer satisfaction. Parasuraman, et al (1985), (1988) used discrepancy measures<sup>5</sup> for development of their SERVQUAL concept which purports to measure service quality. Through extension of the SERVQUAL concept, discrepancy measures have been recently used to measure end-users' satisfaction with the IS function [Pitt et al 1995, Kettinger and Lee 1994]. Doll and Torkzadeh (1989) also used the discrepancy concept for measuring end-user involvement in systems development.

## **2.5 Data Integration Measures as Discrepancy Measures**

As noted above, there are costs as well as benefits associated with increasing data integration and organizational subunits are not homogeneous in their perspective regarding desirable data integration levels. As this paper's central contention, it is argued

---

<sup>5</sup> Parasuraman, et al referred to the discrepancy measures as 'gap measures'.

that discrepancy measures offer a reliable and valid alternative for measuring data integration through an accounting of both the nonmonotonic and asymmetric properties.

### **3. Methodology**

This section presents the design and development of the discrepancy measures employed for assessing data integration in the sample industry.

#### **3.1 A Data Integration Definition**

Wybo (1992) provides the following definition of data integration:

“The concept of data integration is defined as the proportion of the data items used by a subunit’s major information systems that are subject to standardized data definitions that are also used by another specified subunit in the organization”.

This definition captures two critical dimensions along which data integration may increase. First, Wybo (1992) suggests that enlarging the proportion of a given application system’s data fields subject to standard data definitions reflects an increase in data integration.<sup>6</sup> This would be enacted basically through the *creation of additional* data definitions for a system’s data, and may be characterized as an increase in *within-system* integration. Second, increased data integration occurs through increasing the number of application systems that subscribe to a given set of standard data definitions. This would be enacted basically through the *application of extant* data definitions on a system’s data, and may be characterized as an increase in *between-system* integration. This within-system versus between-system conceptual distinction is applied to the application systems at the organizational subunit level-of-analysis, incorporating *within-subunit* and *between-subunit* measures into the survey instrument design.

#### **3.2 The Group Business Unit**

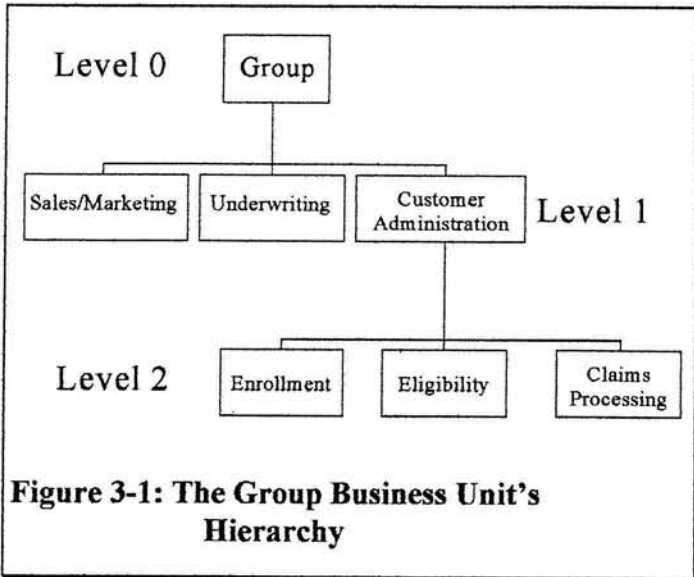
The Group Insurance Industry provides primarily medical, life, disability and

---

<sup>6</sup> Though alternative mechanisms for enacting data integration are available, such as schema integration, source tagging and federated databases as reviewed by Wybo (1992), it is through imposition of data standards that the most effective and reliable form of data integration is granted.

dental insurance services. A typical hierarchical breakdown of an insurance company's Group Business Unit follows a functional organization and is presented in Figure 3-1: The Group Business Unit's Hierarchy.<sup>7</sup> This hierarchical breakdown is adopted from an educational resource<sup>8</sup> of the Life Office Management Association (LOMA)<sup>9</sup>. Treating the entire Group Business Unit as level 0, the figure shows two additional levels of organizational *subunits*. The subunits of level 1 include Sales/Marketing, Underwriting, and Customer Administration; the subunits of level 2 include Enrollment, Eligibility, and Claims Processing.

Though no assertion is made as to the *absolute* degree of interdependency among these organizational subunits, it is argued that the three subunits at each level exist in a state of functional interdependency to some degree. Since subunit interdependency



requires coordinative devices as coping mechanisms [Galbraith 1977], some data integration as manifesting coordination mechanisms is expected. Additionally, because higher states of interdependency exist among organizational subunits at lower levels in the organization hierarchy

[Pfeffer and Salancik 1978], it is argued that more extensive coordination mechanisms at lower levels are needed. Consequently, it is logical to assume that higher levels of data

<sup>7</sup> This model is not intended as fully exhaustive of all Group functional units; rather, it is limited to those subunits incorporated into the survey instrument.

<sup>8</sup> This educational resource is used for administering LOMA's professional certificate educational programs to the Group Insurance Industry.

<sup>9</sup> LOMA is an independent association created and supported through collective efforts of North American insurance companies. They provide extensive education, training, data collection, data analyses and reporting services to their membership.

integration exist at lower hierarchical levels. In this study, we expect that data integration for Level 2 subunits is greater than for Level 1 subunits. Additionally, as a corollary, we expect that within-subunit data integration is greater than between-subunit data integration for each respective level.

### 3.3 The Relative Data Integration Measure

In this study, the ideal<sup>10</sup> and actual amount of data integration among the application systems *within* each subunit are measured yielding three pairs of within-subunit measures for Level 1 and 2. In addition for Level 1 and 2, the ideal and actual amounts of data integration *between* the application systems of each pairwise combination of subunits are measured yielding three pairs of between-subunit measures. Respondents were asked to indicate their subjective assessment in terms of percentage, rating the ideal and actual amount of data integration within and between subunits' application systems on a scale of 0 to 100. This procedure provides for six pairs of measures for data integration at each level--three within-subunit and three between-subunit measures.

Computed from each pair measures, the discrepancy measure is used to formulate a relative data integration (RDI) measure where RDI is defined as:

$$\text{Relative Data Integration (RDI)} = 100 - |\text{Ideal} - \text{Actual}|$$

High values of RDI represent greater data integration, and low values represent lesser data integration. Complete convergence of the ideal and actual measures, i.e., when the actual level of data integration matches exactly the ideal level of data integration, results in a RDI value of '100'. Moreover to the extent the ideal and actual measures diverge, the RDI measure will reflect data integration levels less than '100' regardless of whether the actual measure is above or below the ideal measure.

Use of this RDI measure, through application of the more general discrepancy concept, grants two advantages in terms of accounting for data integration's

---

<sup>10</sup> The normative assessment has been variously referred to as 'desired', 'desirable', 'expected', and, in this study, as 'ideal'.

nonmonotonic and asymmetric properties. First, the RDI measure will accommodate the nonmonotonic property because its value decreases as the ideal and actual measures diverge, regardless of the direction of divergence. For example, when the actual amount of data integration is 10% higher or lower than the ideal amount, the RDI measure indicates 90%. Second, the asymmetric property is accounted for because the *ideal* measure, which implies or “anchors” a nominal or absolute value of data integration, is allowed to vary or “float” across subunits.

## 4. Results

The mean difference t-test results for significant differences in data integration between Levels and between Within-subunit versus Between-subunit are reported first. Next, validity testing procedures are presented.

### 4.1 Mean Difference t-Test Results

Primary data were gathered from 48 Group Insurance' organizations in the United States and Canada which represents a 73% response rate. A sample of the survey instrument is shown in Appendix A.

Level 1	Within				Between		
	Ideal	Actual	RDI		Ideal	Actual	RDI
Sales/Marketing	87 (14) <sup>1</sup> n=31	48 (36)	61 (31)	Sales/Marketing & Underwriting	83 (13) n=34	36 (33)	53 (29)
Underwriting	92 (8) n=34	57 (31)	65 (29)	Sales/Marketing & Customer Admin	70 (29) n=43	35 (34)	65 (28)
Customer Admin	90 (12) n=48	65 (27)	74 (24)	Underwriting & Customer Admin	84 (19) n=46	53 (31)	66 (26)
Level 1-Within Overall	89 (11) n=48	57 (26)	68 (22)	Level 1-Between Overall	79 (14) n=46	42 (28)	61 (23)
Level 2	Within				Between		
	Ideal	Actual	RDI		Ideal	Actual	RDI
Enrollment	93 (10) n=47	64 (31)	71 (30)	Enrollment & Eligibility	93 (16) n=47	66 (29)	71 (29)
Eligibility	95 (9) n=48	69 (28)	73 (27)	Enrollment & Claims Processing	89 (22) n=47	61 (34)	70 (31)
Claims Processing	92 (13) n=48	70 (31)	77 (28)	Eligibility & Claims Processing	96 (9) n=48	70 (31)	73 (29)
Level 2-Within Overall	93 (10) n=48	68 (28)	74 (25)	Level 2-Between Overall	93 (13) n=48	66 (28)	71 (27)

1 Scale: 0-100

**Table 4-1: Mean and Standard Deviation Results**

In some cases, missing data occur due to an inability to ascertain or unwillingness to share the information.

The mean and standard deviation of the Ideal, Actual and RDI measures for the subunit and subunit pairs at each level are shown in Table 4-1. Additionally, ‘Overall’ figures are shown revealing the average across the respective subunits (for Within measures) and subunit pairs (for Between measures) for each level.

	Ideal	Actual	RDI
<b>Level 1 versus Level 2</b>			
Within Overall	(89 vs 93)**	(57 vs 68)**	(68 vs 74)*
Between Overall	(79 vs 93)**	(42 vs 66)**	(61 vs 71)*
<b>Within versus Between</b>			
Level 1 Overall	(89 vs 79)**	(57 vs 42)**	(68 vs 61)
Level 2 Overall	(93 vs 93)	(68 vs 66)	(74 vs 71)
** p<.01; * p<.05			

**Table 4-2: Mean Difference t-Test Results**

As expected, t-tests indicate that the Level 1 versus Level 2 mean difference for the Overall Ideal, Actual and RDI means are significantly higher for Level 2 at  $p<.01$ ,  $p<.01$  and  $p<.05$  respectively. This result holds for both the Within-subunit and Between-subunit measures. (Refer to Table 4-2.)

With regards to mean difference t-tests for Within-subunit versus Between-subunit Overall means by Level, expectations are supported at Level 1 only with significantly higher Ideal and Actual Within-subunit means at  $p<.01$ . The t-test for the RDI Within-subunit measure is not significant. There are no significant differences between the three Within-subunit versus Between-subunit means at Level 2.

One notable result is in regards to the Ideal measures’ standard deviation. As can be observed, there exists comparatively less variance in the Ideal measures versus the Actual measures. In many cases, the Ideal measures’ standard deviation is less than half of that of its respective Actual measures’ standard deviation

## 4.2 Nomological Validity

Beginning with introduction of several criterion variables for measuring data

integration, this section reports the nomological validity test results based on correlation analyses.

#### 4.2.1 The Criterion Variables

The nomological validity of the RDI variable is based on assessment of convergence with four other variables--referred to as criterion variables.<sup>11</sup> Their selection based on existing literature, these criterion variables are assumed to be suggestive of high data integration levels.<sup>12</sup> A seven-point Likert scale is used for each criterion variable to gauge the level of data integration for the three Level 1 functional areas,<sup>13</sup> resulting in a total of twelve criterion variables.

Using Cronbach's alpha for reliability testing, the reliability coefficient is .82, .77 and .80 across the four criterion variables for Sales/Marketing, Underwriting and Customer Administration respectively. All reliability coefficients are above the recommended .80 threshold except for the Underwriting functional area.

A factor analysis of the twelve criterion variables using varimax rotation showed three factors emerging. As shown in Table 4-3, Criterion Variables 3 and 4 load on factor one except for the Sales/Marketing Criterion Variable 4 which loads on factors one and two. Criterion variables 1 and 2 load on factors two and three.<sup>14</sup> A second factor analysis, constraining the factor solution to two factors using varimax rotation, showed Criterion Variables 3 and 4 loading on factor one, Criterion Variables 1 and 2 loading on factor two, and no criterion variable loading onto more than one factor. These results suggest that data integration may consist of *at least* two dimensions. For sake of parsimony, one criterion variable of each dimension emerging out of the two-factor

---

<sup>11</sup> Section 7 Appendix A contains examples of the criterion variable in questions 3.2-3.5.

<sup>12</sup> Criterion Variables 3 and 4 are reverse-coded to control for common response bias. Linearly transformed variables to reverse the direction of Variables 3 and 4 are used in subsequent data analysis and in reporting results, providing expected positive correlations among all criterion variables.

<sup>13</sup> The criterion variables were not collected for the Level 2 functional areas. Convergent validity for the Ideal, Actual and RDI variables of the Enrollment, Eligibility and Claims Processing functional areas is tested against the Customer Administration criterion variables.

<sup>14</sup> The conventional .50 threshold is used as the loading cutoff.

solution is selected for further analyses.

<b>Three-factor Solution</b>	<b>Factor 1</b>	<b>Factor 2</b>	<b>Factor 3</b>
Underwriting-CV3	.91		
Customer Admin-CV3	.87		
Sales/Marketing-CV3	.83		
Underwriting-CV4	.82		
Customer Admin-CV4	.73		
<b>Two-factor Solution</b>			
Sales/Marketing-CV1		.89	
Sales/Marketing-CV2		.89	
Underwriting-CV2		.78	
Underwriting-CV1		.68	
Sales/Marketing-CV4	.53	.66	
<b>Two-factor Solution</b>			
Customer Admin-CV2			.85
Customer Admin-CV1			.83
<b>Two-factor Solution</b>	<b>Factor 1</b>	<b>Factor 2</b>	
Underwriting-CV3	.91		
Customer Admin-CV3	.86		
Sales/Marketing-CV3	.85		
Underwriting-CV4	.84		
Customer Admin-CV4	.73		
Sales/Marketing-CV4	.58		
<b>Two-factor Solution</b>			
Sales/Marketing-CV2		.89	
Sales/Marketing-CV1		.85	
Underwriting-CV2		.84	
Customer Admin-CV2		.82	
Underwriting-CV1		.81	
Customer Admin-CV1		.79	

**Table 4-3: Factor Analyses Results (Varimax Rotation)**

#### 4.2.2 Correlation Results

We assess the nomological validity of the RDI measures using correlation analyses between the Overall RDI variable and Criterion Variables 2 and 4. Correlation analyses are also conducted for Overall Ideal and Actual variables in order to comparatively evaluate the superiority of the RDI variable. All correlation results are shown in Table 4-4.



		CV2			CV4		
		Ideal	Actual	RDI	Ideal	Actual	RDI
Level 1	Within	-1	36**	46**	19	35**	37**
	Between	30*	46**	47**	10	36**	36**
Level 2	Within	7	58**	65**	6	40**	45**
	Between	2	49**	53**	12	41**	40**

\*\*p<.01; \*p<.05

**Table 4-4: Results of Correlation Analyses**

The Overall Actual and RDI measures correlate significantly with Criterion Variables 2 and 4 at  $p<.01$ . Moreover, the correlations between the RDI measures and the criterion variables is larger in most cases. The Ideal measures do not correlate significantly with the criterion variables, except for the Level 1 Between Ideal measure with Criterion Variable 2 at  $p<.05$ .

## 5. Discussion

That results showing significantly higher data integration at Level 2 than at Level 1 is unsurprising, and confirms a central proposition presented by Goodhue et al (1992b). They proposed that data integration should increase as the associated benefits, which are theoretically argued to vary directly with the level of interdependence between subunits, increase. Since interdependence is greater at lower levels of the organization, the associated benefits should be greater at lower levels as well. Though the associated benefits are not measured here, data integration is significantly higher at Level 2 where higher interdependence among subunits should theoretically be found.

When compared to those of the Actual measures, the lower standard deviations on the Ideal measures indicate that respondents are more uniform in their subjective judgments regarding ideal data integration than they are regarding actual data integration. Though both measures are based on the respondents' perceptions, these results suggest that, even though the organizations of this sample are comparatively uniform in estimating the ideal amount of data integration, they fall short and vary in their ability to establish the desired amount of data integration. Further research investigating possible antecedents to varying amounts of actual data integration would be a useful managerial

contribution, and attests to the diagnostic value of a reliable and valid measurement instrument.

The RDI measures correlate significantly with the criterion variables in all cases. Though these results demonstrate convergent validity, the Actual measures also correlate significantly with the criterion variables. A similar pattern appearing in other empirical studies [Cronin and Taylor 1992, Babakus and Boller 1991, Parasuraman, et al 1991, Carman 1990], these results challenge the methodological value of the RDI discrepancy measures. Since the Actual measures seem to dominate the RDI, the utility of the Ideal measures is questionable. However there are both practical and theoretical considerations in judging the methodological merits of using both the Actual and Ideal measures, i.e., the discrepancy measures, against using the Actual measures alone.

From a practical standpoint the discrepancy measures provide substantive diagnostic value over using the Actual measures alone. In addition, the correlations are generally higher between the RDI and criterion variables than between the Actual and criterion variables suggesting that the RDI is a stronger predictor. However the *differences* between these dependent sets of correlations are not significant [Steiger 1980], indicating that the RDI measures are not significantly better than the Actual measures for assessing data integration.

From a theoretical perspective the asymmetric property of data integration may be adequately captured through using discrepancy measures, though the methodology as employed does not allow empirical confirmation of the existence of asymmetry between organizational subunits for this sample of organizations. For any two organizational subunits, two sets of Ideal and Actual measures, one from the perspective of each subunit, are necessary to ascertain the presence of asymmetry. The data collection process was constrained by practical limitations regarding the availability of respondents in nearly all organizations.

The nonmonotonic property of data integration was adequately captured through the employed methodology, but situations where the actual amount exceeded the ideal amount were rare. Though this could be the result of response bias [Wall and Payne

1973], it *may be* that the organizations of this sample have not attained ideal amounts of data integration..

To argue for reliance on Actual measures alone, one may consider the greater parsimony in methodology granted through removal of the Ideal measures in survey design. Moreover, the RDI variable, as a difference measure, is less reliable than the Actual variable since it is based on two error terms [Cohen and Cohen 1983].

## **6. Conclusions**

The RDI measure based on the difference between ideal and actual measures appears to be a valid indicator of data integration across organizational subunits at different organizational levels. However the efficacy of using RDI or a discrepancy measure, over the simpler use of the difference score's parts--namely the Actual measure, is questionable. A test for a significant difference between dependent correlations showed no significance difference. The simpler use of the difference score's parts may be more efficacious, albeit lessening the diagnostic value of the survey instrument and compromising the ability to reflect the asymmetric and nonmonotonic properties of data integration.

## 7. Appendix A

### Section 3: Internal Application Systems Data

This section asks for information regarding the *internal* application systems used in the Sales/Marketing, Underwriting and Customer Administration functional areas within Group Insurance (Life, Medical, LTD, ASO) operations. The focus is on assessing the general level of integration among these internal application systems and the satisfaction with these systems' output information. If any question does not apply to your company, then indicate so with 'n/a'. If you cannot determine the exact answer to a question, then you may provide an estimate. In the event estimation is not possible, then indicate '?'.

**Part I.** The following questions address characteristics of the internal application systems within each respective functional area. Please indicate the answer that best describes the situation at your company.

3.1 A company may manage its data processing (DP) resource in four ways: (1) through their *own DP* department with predominately *customized* software development; (2) through their *own DP* department with predominately *packaged* software development; (3) through their *own DP* department with a *hybrid* of customized and packaged software development; and (4) through *outsourcing*—using the data processing services of a service bureau or facilities management company. Indicate according to *approximate* percentage breakdown how your DP resource is managed in each respective functional area.

Internal Applications Only (Note: Rows should add to 100%.)	(1) Own DP / Customized	(2) Own DP / Packaged	(3) Own DP / Hybrid	(4) Outsourcing	Total
Sales/Marketing	%	%	%	%	100 %
Underwriting	%	%	%	%	100 %
Customer Administration	%	%	%	%	100 %

Use the following scale to answer questions 3.2-3.5.

1	2	3	4	5	6	7
Strongly Disagree	Moderately Disagree	Slightly Disagree	Neither Agree nor Disagree	Slightly Agree	Moderately Agree	Strongly Agree

#### Criterion Variable 1

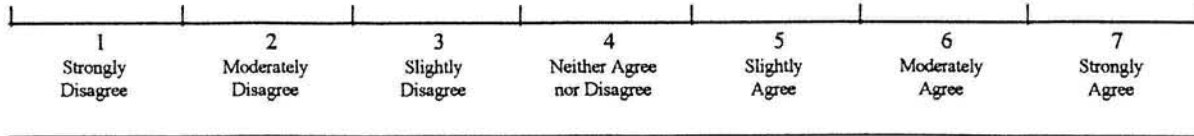
3.2 Data must be rekeyed as it used and reused by different employees of (functional area).

Sales/Marketing	1	2	3	4	5	6	7
Underwriting	1	2	3	4	5	6	7
Customer Administration	1	2	3	4	5	6	7

#### Criterion Variable 2

3.3 The *same* data is often inconsistent in terms of format and meaning across *different* internal application systems of (*functional area*).

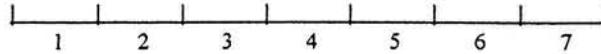
Sales/Marketing	1	2	3	4	5	6	7
Underwriting	1	2	3	4	5	6	7
Customer Administration	1	2	3	4	5	6	7



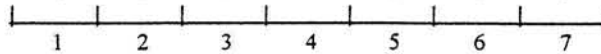
**Criterion Variable 3**

3.4 For at least *some* data fields, there are data standards imposed and enforced across *most* of the (*functional area's*) internal application systems.

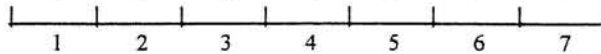
Sales/Marketing



Underwriting



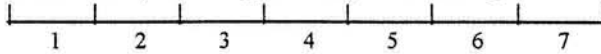
Customer Administration



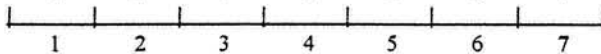
**Criterion Variable 4**

3.5 A high percentage of all data fields residing in the (*functional area's*) internal application systems are assigned data standards.

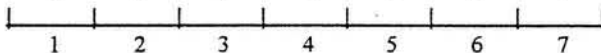
Sales/Marketing



Underwriting

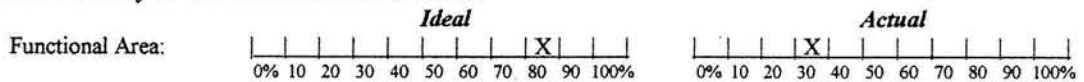


Customer Administration

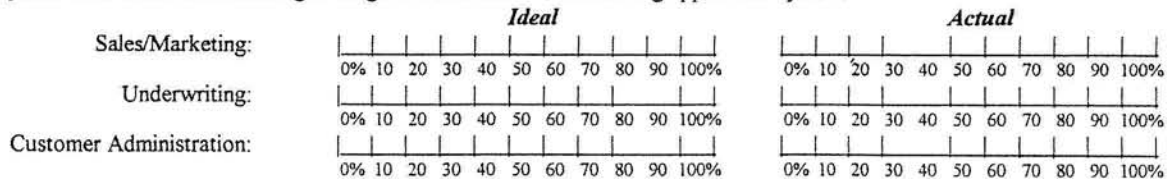


**Part II.** Integration of internal application systems is generally defined as the ability to share data among them. For example, if two internal application systems are integrated, then it can be said the data is shared between them in an organized and efficient manner. In most instances, the level of integration desired is greater than what actually occurs. However, a 100% integration level may not be ideal either as there are tradeoffs involved when integrating internal application systems' data (e.g., an increased level of data integration may result in less flexibility to changing business requirements and needs). The next two questions ask you to indicate both the *ideal* and *actual* data integration levels of the internal application systems *within* each respective functional area (3.6) and *between* functional areas (3.7). Use the following scale to answer questions 3.6-3.9.

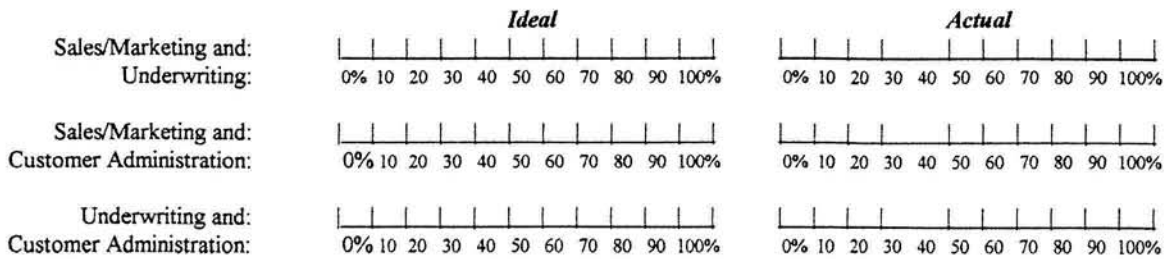
**Example:** Assume *ideally* you would like 80% of the data from your functional area's internal application systems integrated, however currently only about 30% *actually* is. You would mark the answer as:



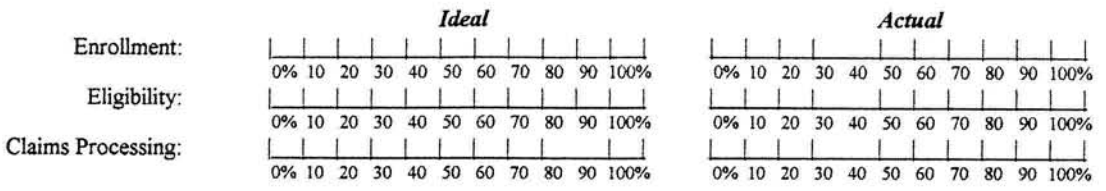
3.6 Indicate both the *ideal* and *actual* level of data integration among the internal application systems *within* each of the respective functional areas (i.e., *intrafunctional*). For example, assess the general degree to which a given internal application system used within Underwriting is integrated with *other* Underwriting application systems.



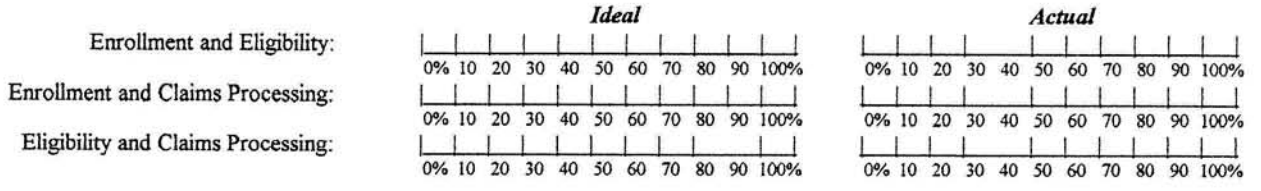
3.7 Indicate both the *ideal* and *actual* level of data integration *between* the following functional areas' internal application systems (i.e., *interfunctional* or *crossfunctional*). For example, assess the general degree to which the internal application systems used by Underwriting are integrated with the application systems of Sales/Marketing.



3.8 Within the Customer Administration functional area, a further breakdown by function might identify *Enrollment*, *Eligibility*, and *Claims Processing* as *subfunctions* (i.e., divisions, departments) within the Customer Administration area. Indicate both the *ideal* and *actual* level of data integration *within* the set of the internal application systems that support each subfunction's operations independent from the other two.



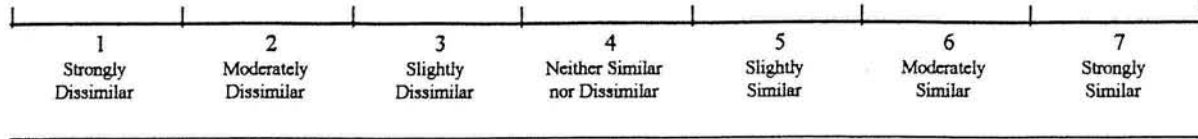
3.9 Indicate both the *ideal* and *actual* level of data integration *between* the following Customer Administration subfunctional areas' internal application systems. For example, assess the general degree to which the internal application systems used by Enrollment are integrated with the application systems of Eligibility.



3.10 Insurance companies may use different platforms (i.e., hardware and systems software) to support their data processing needs. Platforms across computer sizes differ, i.e., a mainframe computer is a different platform from a mini computer. And even within the same size of computers, the platform may vary. For example an IBM mainframe is a different platform from a Hewlett-Packard mainframe. And even among IBM mainframes the platform may vary. For example, a 4300-series IBM mainframe is a different platform from IBM's 3090-series. Indicate the number of *different* platforms currently used in each respective functional area. (Note: Two IBM 4381s should be counted as *one* platform.) If more than seven, then circle '>7'.

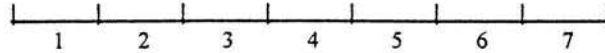
Sales /Marketing	1	2	3	4	5	6	7	>7	(circle one)
Underwriting	1	2	3	4	5	6	7	>7	(circle one)
Customer Administration	1	2	3	4	5	6	7	>7	(circle one)

Use the following scale to answer questions 3.11 and 3.13.

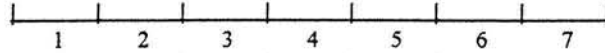


3.11 Using the definition of platform described above, indicate the general degree of similarity or dissimilarity of the platforms *between* the following pairs of functional areas.

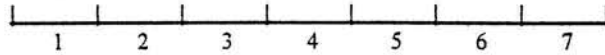
Sales/Marketing and Underwriting:



Sales/Marketing and Customer Administration:



Underwriting and Customer Administration :



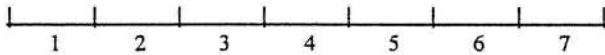
3.12 Using the definition of platform described above, indicate the number of *different* platforms used in each respective subfunction of Customer Administration. If more than seven, then circle '>7'.

Enrollment	1	2	3	4	5	6	7	>7	(circle one)
Eligibility	1	2	3	4	5	6	7	>7	(circle one)
Claims Processing	1	2	3	4	5	6	7	>7	(circle one)

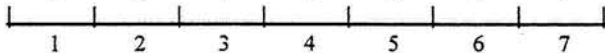
Use the scale at the top of the page to answer question 3.13.

3.13 Using the definition of platform described above, indicate the degree of similarity or dissimilarity of the platforms *between* the following pairs of subfunctions of the Customer Administration area.

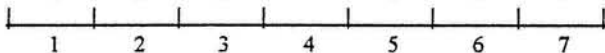
Enrollment and Eligibility:



Enrollment and Claims Processing:



Eligibility and Claims Processing:



## 8. References

- Alutto JA and Belasco JA. A Typology for Participation in Organizational Decision Making. *Administrative Science Quarterly*, 17:1, 1972, p.117-125.
- Babakus E and Boller GW. An Empirical Assessment of the SERVQUAL Scale. *Journal of Business Research*, Vol. 24, 1992, p.253-268.
- Bolton RN and Drew JH. A Longitudinal Analysis of the Impact of Service Changes on Customer Attitudes. *Journal of Marketing*, January 1991, p.1-9.
- Carman JM. Consumer Perceptions of Service Quality: An Assessment of the SERVQUAL Dimensions. *Journal of Retailing*, 1990, 66:1, p.33-55.
- Churchill GA and Suprenant C. *Journal of Marketing Research*, Vol. XIX, November 1982, p.491-504.
- Cohen J and Cohen P. Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences, 2nd Ed., Lawrence Erlbaum Associates, Hillsdale, New Jersey. 1983.
- Cronin JJ and Taylor SA. Measuring Service Quality: A Reexamination and Extension. *Journal of Marketing*, 56:3, 1992, p. 55-68.
- Doll WJ and Torkzadeh G. A Discrepancy Model of End-User Computing Involvement. *Management Science*, 35:10, October 1989, p.1151-1171.
- Galbraith J. Organization Design. Addison-Wesley, Reading, Massachusetts, 1977.
- Goodhue DL, Kirsch LJ, Quillard JA, and Wybo MD. Strategic Data Planning: Lessons from the Field. *MIS Quarterly*, 16:1, March 1992a, p.11-34.
- Goodhue DL, Wybo MD, and Kirsch LJ. The Impact of Data Integration on the Costs and Benefits of Information Systems. *MIS Quarterly*, 16:3, September 1992b, p.293-310.
- Hart P and Estrin D. Inter-Organization Networks, Computer Integration, and Shifts in Interdependence: The Case of the Semiconductor Industry. *ACM Transactions on Information Systems*, 9:4, October 1991, p.370-398.
- Henderson JC and Venkatraman N. Strategic Alignment: Leveraging Information Technology for Transforming Organizations. *IBM Systems Journal*, 32:1, 1993.
- Keen Peter G.W. Information Technology and the Management Difference: A Fusion Map. *IBM Systems Journal*, 32:1, 1993.



- Keen Peter G.W. Shaping the Future: Business Design Through Information Technology. Harvard Business School Press, Boston, Massachusetts, 1991.
- Kettinger WJ and Lee CC. Perceived Service Quality and User Satisfaction with the Information Services Function, 25:5/6, September-December 1994, p.737.
- Konsynski B. Strategic Control in the Extended Enterprise. IBM Systems Journal, 32:1, 1993.
- Lawrence PR and Lorsch JW. Organization and Environment: Managing Differentiation and Integration. Harvard Business School Press, Boston, Massachusetts, 1967.
- Lederer AL and Sethi V. Critical Dimensions of Strategic Information Systems Planning. Decision Sciences, Volume 22:1, Winter 1991, p.104-119.
- Luftman JN, Lewis PR, and Oldach SH. Transforming the Enterprise: The Alignment of Business and Information Technology Strategies. IBM Systems Journal, 32:1, 1993.
- Parasuraman A, Berry LL, Zeithaml VA. Refinement and Reassessment of the SERVQUAL Scale. Journal of Retailing, Vol. 67:4, Winter 1991, p.420-450.
- Parasuraman A, Zeithaml VA, Berry LL. SERVQUAL: A Multiple-Item Scale for Measuring Consumer Perceptions of Service Quality. Journal of Retailing, Vol. 64:1, Spring 1988, p.12-40.
- Parasuraman A, Zeithaml VA, Berry LL. A Conceptual Model of Service Quality and Its Implications for Future Research. Journal of Marketing, Vol. 49, Fall 1985, p.41-50.
- Pfeffer J and Salancik GR. The External Control of Organizations: A Resource Dependence Perspective. Harper & Row, Publishers, New York, New York, 1978.
- Pitt LF, Watson RT, and Kavan CB. Service Quality: A Measure of Information Systems Effectiveness. MIS Quarterly, 19:2, June 1995, p.173-185.
- Rockart J and Short J. Information Technology in the 1990s: Managing Organizational Independence. Sloan Management Review, 30:2, Winter 1989.
- Steiger JH. Tests for Comparing Elements of a Correlation Matrix. Psychological Bulletin, Vol 87, 1980, p.245-251.
- Wall TD and Payne R. Are Deficiency Scores Deficient? Journal of Applied Psychology, 58, 1973, p.322-326.

- Weber M. From Max Weber: Essays in Sociology, edited and translated by H.H. Gerth and C. Wright Mills, Oxford University Press, Inc. 1946.
- Weill P. The Role and Value of IT Infrastructure: Some Empirical Observations, in Strategic Information Technology Management: Perspectives on Organizational Growth and Competitive Advantage, Banker RD, Kauffman RJ and Mahmood MA Editors, Idea Group Publishing, Harrisburg, Pennsylvania, 1993.
- Wybo MD. Factors Affecting the Use of Data Integration in Manufacturing Firms. Unpublished Doctoral Dissertation, University of Minnesota, Minneapolis, Minnesota, 1992.