Journal of
Applied Sciences

Discovering Dependencies among Data Quality Dimensions: A Validation of Instrument

Payam Hassany Shariat Panahy, Fatimah Sidi, Lilly Suriani Affendey, Marzanah A. Jabar,
Hamidah Ibrahim and Aida Mustapha
Department of Computer Science, Faculty of Computer Science and Information Technology,
Universiti Putra Malaysia, Serdang, Selangor 43400, Malaysia

Abstract: Improving data quality is a basic step for all companies and organizations as it leads to increase opportunity to achieve top services. The aim of this study was to validate and adapt the four major data quality dimensions' instruments in different information systems. The four important quality dimensions which were used in this study were; accuracy, completeness, consistency and timeliness. The questionnaire was developed, validated and used for collecting data on the different information system's users. A set of questionnaire was conducted to 50 respondents who using different information systems. Inferential statistics and descriptive analysis were employed to measure and validate the factor contributing to quality improvement process. This study has been compared with related parts of previous studies; and showed that the instrument is valid to measure quality dimensions and improvement process. The content validity, reliability and factor analysis were applied on 24 items to compute the results. The results showed that the instrument is considered to be reliable and validate. The results also suggest that the instrument can be used as a basic foundation to implicate data quality for organizations manager to design improvement process.

Key words: Information quality, data quality, data quality dimensions, framework, empirical research, construct validity, factor analysis, reliability

INTRODUCTION

In the last decades, assessment and improvement of data and information quality issues have been grown quickly (Lee *et al.*, 2002). Moreover, many organizations have invested to gather, store, retrieve, elaborate huge amounts of data (Borek *et al.*, 2011; Carey *et al.*, 2006). There are some problems and difficulties in the nature of the data from the technical (data integration) to the nontechnical (lack of consistent strategy in organization). Assessing and improving the quality of data will improve business process and leads to make smart decisions (Madnick *et al.*, 2009). Missing value, duplicate and inaccurate data are examples of data quality problems that exist in the real world (Scannapieco *et al.*, 2005). Based on the researchers and experts' understanding; data quality has various definition in different field and period. According to quality management, data quality is "appropriate for use" or "to meet user needs" (Alizamini *et al.*, 2010).

Based on previous studies, data quality dimensions have dependency on each other (Barone *et al.*, 2010). Improving dependency structure among a set of data quality dimensions is referred to dependency discovery.

Logical interdependence analysis (Gackowski *et al.*, 2004), tradeoff analysis (Madnick *et al.*, 2009; Scannapieco *et al.*, 2005; Alizamini *et al.*, 2010; Gackowski *et al.*, 2004) and data dependency analysis are examples of researches to discover dependency structure among quality dimensions (Barone *et al.*, 2010). The effectiveness of information systems is related to the quality of information resources for solving problems and making decisions (Barone *et al.*, 2010). However, for improving activities and making correct decision, organizations need enough knowledge about correctness and relationship among data quality dimensions. Dependency model between quality dimensions can be divided to three main categories as: Prefect dependency, partial dependency and independency (DeAmicis *et al.*, 2006). However, there are some deficiencies on one dimension that can effect on another dimension. Relationship between quality dimensions is basically in terms of trade-off. The examples of trade-off are between accuracy and timeliness and completeness and consistency. The first trade off refers if information becomes better over time, it has more accuracy with negatively effect. Another trade-off between completeness and consistency dimensions indicates if

Corresponding Author: Fatimah Sidi, Department of Computer Science,
Faculty of Computer Science and Information Technology,
Universiti Putra Malaysia, Serdang, Selangor 43400, Malaysia Tel: +60389471739

data is more complete, its consistency is less. In another word if data is more complete it has more lack of consistency (Han and Venkatasubramanian, 2003; Fisher *et al.*, 2006; DeAmicis *et al.*, 2006). As a result, discovering data quality dependency is a basic step for making decision and developing activity (Lee *et al.*, 2002).

Based on the literature review, there are many data quality dimension that can be assessed and improved (Sidi *et al.*, 2012). In this work, we focus on the four important data quality dimensions which can be measured by standards applied to a database directly (Wang and Strong, 1996; Kahn *et al.*, 2002). The four most important data quality dimensions are: Timeliness, accuracy, consistency and completeness (O'Donoghue *et al.*, 2012). The generally known definitions of these dimensions in the field of data quality are as follows:

- **Timeliness:** To extent which age of data is appropriated for the task at hand (Wang and Strong, 1996)
- **Accuracy:** To extent, which data is correct, reliable and certified (Wang and Strong, 1996)
- **Completeness:** To extent to which data are of sufficient breadth, depth and scope for the task at hand (Wang and Strong, 1996)
- **Consistency:** To extent which data is presented in the same format and compatible with previous data (Wang and Strong, 1996)

Nevertheless, the progress of some researches is related to use of a standard framework and identifying its validity. Also, the results of one study cannot be generalized without repeating research in different sample and population.

The main objective of this study was, to empirically validate ACCTI (accuracy, completeness, consistency and timeliness, improvement process) framework. The results will provide a set of reliable and valid construct for assessing data quality dimensions which can be used to evaluate relationship among quality dimensions. Generally, applying different statistical methods provides evidence that the proposed framework can be used confidently in the future researches.

PROPOSED DATA QUALITY DIMENSIONS' FRAMEWORK

Based on the literature, information provider and consumer produce various qualities on the information system. However, quality of information will be judged by information consumers to check if it will be fit for users or not (Katerattanakul and Siau, 1999). Information quality
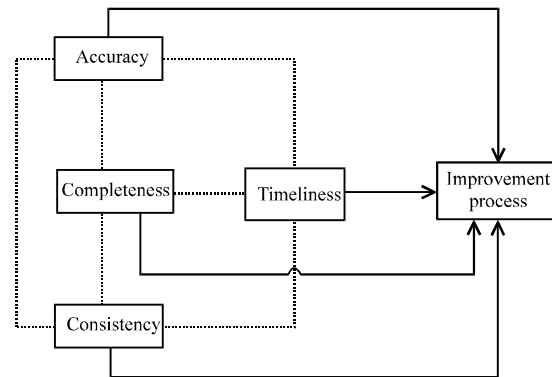

Fig. 1: ACCTI framework

should be accessed via framework by consumers who utilize the information based on their information quality needs (Katerattanakul and Siau, 1999; Strong *et al.*, 1997).

Moreover, as discussed earlier, data quality has different dimensions and; finding relationship among them is necessary for improving process quality in different application and domains (Sidi *et al.*, 2012). Finding effective dependencies and relationship can be useful and comprehensive for decision making process (Barone *et al.*, 2010). In fact, dependency discovery is consist of measuring attributes of quality dimensions and recovering dependency structure among them for extracting knowledge (Sadeghi and Clayton, 2002; Strong *et al.*, 1997). Selection of appropriate dimensions and analysis correlation among them is one of the ways that led to have high quality process (Katerattanakul and Siau, 1999).

The proposed framework shows dependency structure among following data quality dimensions; accuracy, completeness, consistency, timeliness and their relationship with data quality improvement process. The hypothesis is accuracy, completeness, consistency dimension are related to timeliness. Another hypothesis is all four quality dimension are related to improvement process. Figure 1 shows preliminary framework for four data quality dimensions.

In conclusion, the proposed of this framework is to discover dependent structure for evaluating data quality improvement via questionnaire and validate it in different information system.

METHOD, EMPIRICAL VALIDATION

Survey research has been used widely in the field of information system and there are number of validate quantitative instruments for the researcher in this field (Kock and Verville, 2012). Typically, questionnaire is used

for data collections in specific topics with the goal of generalize the result of the sample to a population (Creswell, 2009).

The objective of this research is to validate the propose framework in different information. Qualitative approach was used in this research with the aim of establishing meaning of variable from the view of participants.

Utilized methodology for this research was based on common set of standards measuring for latent variables. The methods to conduct and identify quality dimensions dependency was as follow:

- Construct framework among quality dimensions based on literature review for assessing dependency
- To use questionnaire for mentioned dimensions, based on the last standard questionnaire
- To self-develop questionnaire for improvement process based on literature
- To implement a survey for gathering data to test the framework
- To use empirical methods for analyzing and validating the framework

Based on the proposed research framework, the close ended questionnaire was used and developed. Questions for data quality dimension, were adapted from the latest standard questionnaire (Lee *et al.*, 2002) and the questionnaire for the quality improvement was gained based on the comprehensive literature. The concept of questions for improvement process was related to quality improvement process, user satisfaction and evaluation criteria of individual information systems. All questions for each dependent and independent variable was measured based on the 9 point Likert scale.

The measure for this study was based on MIT research for developing AIMQ methodology (Lee *et al.*, 2002). The Likert scale used to assess each item was rating from 1 to 9, where 1 represented "not at all" and 9 represented "completely". Although, previous study used 11 scales for four quality dimension variables, due to 11 scales were so confusing for the target respondents, the 1-9 scale was used by confirming with professionals. For the latest variable, which was improvement process, questionnaire was self-developing based on the comprehensive literature review by testing same variable scale 1-9.

The respondent selected randomly by cross sectional methods in the period of 2 weeks. They were users of various information systems in different places and organization. Simple Random Sampling (SRS) was used to select respondent because, besides of its advantages;

free of classification error, requires minimum knowledge information of the population and effortless to interpret data collected (http://en.wikipedia.org/wiki/Simple_random_sample), all predictable statistical techniques can be applied by it (Creswell, 2009). The target respondent were; undergraduate student, postgraduate student, academician, manager and faculty staff. The respondents were likely to possess knowledge about data or information about in information systems area.

There are many ideas for sufficient sample size in Structural Estimating Modeling (SEM). Some researchers recommended; the adequate sample size should be more than 200 (Kline, 2010) but some others argued that minimum size of the sample should be fifth times of the number of variables or each variable should have at least 10 respondent (Hatcher, 1994). So, in this study cause there are five variables 50 respondents or more should be adequate. However, based on >200 samples rules, if there is restriction on population size less sample size does not have any problem (Kline, 2010). From 70 questionnaires, 55 were returned, 3 of them did not understand concept of questions and gave same mark to all questions and 2 questionnaires were deleted based on our predefined rules, as follows:

- The questionnaire will be deleted if the respondents answer less than half of the question
- The mean of answers for each item will be used to estimate missing value

The questionnaire was sent via email or manually and from the 55 returned, responsible rate was 78.57% for 50 usable questionnaires. Approximately, 60% of respondent were female and 40% were male, around 48% of the respondents' age were between 20-29 whom most of them were "postgraduate" student and around 54% mentioned that they chose "Student Information System" as a considering information system.

An empirical method with using quantitative and descriptive evidence is extensively used in any data quality research to study of real life context with finalizing observable evidence (Lee *et al.*, 2002; Wang and Madnick, 1989).

The questionnaire was developed, validated and used for collecting data on the different information systems.

ACCTI INSTRUMENT DEVELOPMENT

The ACCTI instrument was developed based on the standard methods for questionnaire and testing (Fynes *et al.*, 2005). The constructs was based on the extent literature and adapted to fit by running a pilot

study. The proposed framework provides a preliminary assessment of the item's reliability and validating; four major data quality dimensions and improvement process attributes. In this study, four data quality dimension; (accuracy, completeness, consistency, timeliness) and improvement process variables were measured and validated by gathering answers from the questionnaire's respondents. Validating this instrument is one step to improve process quality in information systems.

The questionnaire with 29 questions of the ACCTI instrument was printed as a seven-page booklet; first page for introducing researchers, organization, summary and explanation about data, data quality, data quality dimensions, data quality improvement and information systems of data quality; second page, for the demographic question, five pages of questions for each variable which was divided to five sections and a five blank line for respondents' suggestions. The questionnaire was given randomly to the target sample via email or manually. The questionnaire is as follows:

Questionnaire; All items are measured on a 1 to 9 scale where 0 refers to "Not at All " and 9 refers to "Completely", Items which labels with "(R)" are reverse coded

Section A: General Information (GI)
Please select the most appropriate answer that describes your statues
A1: Please select your gender:
• Male
• Female
A2: Which category below includes your age?
• 20-29
• 30-39
• 40-49
• 50-59
• 60-older
A3: Which of the following categories describe your status?
• Undergraduate students
• Postgraduate student (Master or PHD)
• Academician (researcher/Lecturer)
• Faculty Staff
• Manager(Dean of Faculty/Head of Department)
A4: Please select one information system that you are considering to answer this questionnaire
• Student's information system
• Human Resource
• Financial System
• Learning Management System (LMS)
• Medical System
• Library System
• Others
  "This information" refers to the information system that you are involved such as student information, human resource information employee data, financial, etc
Section B: Accuracy
Q1: This information is correct
Q2: This information is incorrect (R)
Q3: This information is accurate
Q4: This information is reliable

Section C: Completeness
Q1: This information includes all necessary values
Q2: This information is incomplete(R)
Q3: This information is complete
Q4: This information is sufficiently complete for our needs
Q5: This information covers the need for our task
Q6: This information has sufficient breadth and depth for our task
Section D: Consistency
Q1: This information is consistently presented in the same format
Q2: This information is not presented consistently
Q3: This information is presented consistently (R)
Q4: This information is represented in a consistent format
Section E: Timeliness
Q1: This information is sufficiently current for our work
Q2: This information is not sufficiently timely (R)
Q3: This information is not sufficiently current for our work (R)
Q4: This information is sufficiently timely
Q5: This information is sufficiently up-to-date for our work
Section F: Improvement process
Q1: The process and people involved in the process should have opportunities for improvement
Q2: Quality improvement in the process is obvious
Q3: This information or data should be modified or redesigned for quality improvement
Q4: This information will provides value to our organization
Q5: This information or data has resulted in overall quality improvement for consultations

Content validity and reliability: The first process in the construct validity is establishing content validity. The six information system was selected as a domain. For data quality improvement, from 8 questions which were developed for these part 2 questions were deleted based on 5 IS expert's suggestions. Totally, the feedback obtained from 5 experts suggested us that the items were representative of the construct and the measures are suitable for construct validity.

Second process is to test reliability. The test of reliability will be specifying the amounts of random error that exist in the measurement (Nunally and Bernstein, 1994). The scales were used to discover construct's relationship; if the amount of error is increased the correctness of the result will be decrease (O'Leary-Kelly and Vokurka, 1998). Reliability was assessed by reporting Cronbach alpha reliability coefficient. An instrument is considered to be reliable if Cronbach alpha is at least 0.7. Rule of thumb for Cronbach alpha is as follow; more than 0.9 is considered as "Excellent", 0.8 is "Good", more than 0.7 is "Acceptable", more than 0.6 is Questionable, more than 0.5 is "poor" and <0.5 is considered "Unacceptable" (Gliem and Gliem, 2003). In fact, the numbers of the items will have effect on the value of alpha. So in this study, totally Cronbach alpha of 0.923 was showed "Excellent" result with a good internal consistency of the item for each variable. Furthermore, the value of Cronbach alpha for each variable calculated separately and compared with

previous research (Lee *et al.*, 2002). Cause of developing questionnaire for improvement process there wasn't any result to compare with it.

Table 1 shows computed Cronbach alpha of this study and previous study. The results shows alpha for the accuracy was 0.86, which indicate the items from a scale has reasonable internal consistency (excellent result) and it was almost same with 0.91 compare with previous study. Similarly, the alpha 0.88 for both completeness and consistency and 0.84 for timeliness indicate good internal consistency and they were almost same with previous study which was 0.87, 0.83, 0.88 but the 0.69 alpha for the improvement process scales indicated minimally adequate reliability (questionable result). The further result of reliability analyzes in detail is shown in Table 2 which indicate; total Cronbach alpha for 24 items was 0.929 which shows all the items have relatively high internal consistency. Moreover, according to "Corrected item-total correlation" because most correlations are moderately high to high, items will make a good component of a summated rating scale. So in total, obtaining results confirmed the items have relatively high internal consistency.

However, it should be remember that dimensionality of the items cannot be estimated by just having high value of Cronbach alpha and to determine the dimensionality of the items, factor analysis should be used (Gliem and Gliem, 2003) as it was the next process.

Exploratory factor analysis: There are several multivariate statistical techniques that it is used to analyze theoretical model. One of these techniques is factor analysis that attempts to classify any underling factors which have responsibility for the group of independent variable (Walker, 1999). The aim of factor analysis is to reduce the number of the items and shows which variable explain or determine relationship (Leech *et al.*, 2007). Moreover, exploratory factor analysis is determining empirically

Table 1: Comparing Cronbach's alpha with previous research

Data quality dimensions	Cronbach's alpha of this study	Cronbach's alpha of previous study*
Accuracy (free of error)	0.86	0.91
Completeness	0.88	0.87
Consistency (consistent representation)	0.88	0.83
Timeliness	0.84	0.88
Improvement process	0.68	-

*Sidi *et al.* (2012)

Table 2: Summary items analysis from SPSS

Scale statistics	No. of Items		Mean		Variance		SD
	24		156.0400		572.039		23.91734
Summary item statistics	Mean	Min	Max	Range	Max/Min	Variance	No. of Item
Item means	6.502	5.520	7.180	1.660	1.301	0.177	24
Item variances	2.747	1.806	4.194	2.389	2.323	0.499	24
Inter-item correlations	0.344	-0.159	0.855	1.014	-5.385	0.038	24
Reliability statistics	Cronbach's alpha		Cronbach's alpha based on standardized items				No. of Items
	0.923		0.926				24
Item-total statistic	Scale mean if item deleted	Scale variance if item deleted	Corrected item-total correlation	Squared multiple correlation			Cronbach's alpha if item deleted
Accuracy Q1	148.8600	517.511	0.740	0.874			0.917
Accuracy Q2	148.9400	520.302	0.609	0.880			0.919
Accuracy Q3	149.1000	516.296	0.790	0.879			0.916
Accuracy Q4	149.4200	516.167	0.608	0.841			0.919
Completeness Q1	149.5200	524.949	0.735	0.844			0.918
Completeness Q2	149.8200	523.987	0.515	0.875			0.921
Completeness Q3	150.0400	515.753	0.597	0.869			0.919
Completeness Q4	149.3200	539.896	0.475	0.820			0.921
Completeness Q5	149.4200	538.085	0.472	0.826			0.921
Completeness Q6	149.9000	518.133	0.674	0.849			0.918
Consistency Q1	149.3000	523.071	0.639	0.872			0.919
Consistency Q2	149.6800	521.855	0.492	0.789			0.922
Consistency Q3	149.6400	514.847	0.677	0.875			0.918
Consistency Q4	149.6600	513.698	0.782	0.950			0.916
Timeliness Q1	149.3600	524.807	0.695	0.801			0.918
Timeliness Q2	150.2200	524.012	0.517	0.576			0.921
Timeliness Q3	150.5200	527.438	0.458	0.707			0.922
Timeliness Q4	150.0200	525.326	0.620	0.704			0.919
Timeliness Q5	149.8000	522.980	0.660	0.706			0.918
Improvement process 1	148.9400	549.119	0.297	0.623			0.924
Improvement process 2	149.5800	547.024	0.308	0.682			0.942
Improvement process 3	149.5000	556.296	0.141	0.522			0.928
Improvement process 4	149.3200	526.589	0.616	0.750			0.919
Improvement process 5	149.0400	542.651	0.431	0.761			0.922

Table 3: Total variance explained

Components of each variable		Initial eigenvalues	
		Total	Variance (%)
Accuracy	1	2.915	72.873
	2	0.662	16.538
	3	0.290	7.260
	4	0.133	3.329
Completeness	1	3.890	64.836
	2	0.975	16.243
	3	0.451	7.520
	4	0.262	4.358
	5	0.242	4.028
	6	0.181	3.015
Consistency	1	3.043	76.086
	2	0.629	15.732
	3	0.193	4.826
	4	0.134	3.357
Timeliness	1	2.942	58.842
	2	0.800	15.994
	3	0.540	10.796
	4	0.416	8.317
	5	0.303	6.052
Improvement process	1	2.265	45.299
	2	0.953	19.068
	3	0.797	15.946
	4	0.619	12.386
	5	0.395	7.301

Extraction method: Principal component analysis

Table 4: Principal component analysis

Component	Questions	Factor loading
Accuracy	Accuracy Q1	0.922
	Accuracy Q2	0.978
	Accuracy Q3	0.917
	Accuracy Q4	0.674
Completeness	Completeness Q1	0.886
	Completeness Q3	0.801
	Completeness Q6	0.799
	Completeness Q4	0.794
	Completeness Q2	0.790
Consistency	Consistency Q3	0.932
	Consistency Q1	0.916
	Consistency Q4	0.912
	Consistency Q2	0.711
Timeliness	Timeliness Q1	0.841
	Timeliness Q5	0.795
	Timeliness Q4	0.761
	Timeliness Q2	0.735
	Timeliness Q3	0.695
Improvement process	Improvement process 5	0.842
	Improvement process 4	0.707
	Improvement process 3	0.674
	Improvement process 2	0.551
	Improvement process 1	0.547

whether participant's responses to the each part's question are more similar to other parts (Leech *et al.*, 2007).

Based on the assumption of exploratory factor analysis that data distribution should be normal (Leech *et al.*, 2007), all the skewness of our data was checked. All skewness was between -2 to 2 that showed all items were distributed normally.

In addition, to transform a group of observed variable into an another group of variable "Principal Component Analysis" method was applied (Kim and Mueller, 1978). It was conducted with choosing scores greater than 0.4 in the factor matrix (it makes factor matrix be more readable) and with Varimax rotation techniques to extract factors with Eigenvalues more than 1, which shows the number of required factors. Eigenvalue<1 indicated that variables cannot contribute with average amount of explaining the variance and solution cannot be found for redundancy problem (Leech *et al.*, 2007). Table 3 shows total variance explained; that indicates how the variance is divided among the items for each factor. The eigenvalues for the first factor of each variable were: accuracy 2.915, completeness 3.890, consistency 3.043, timeliness 2.942 and improvement process 2.265 that were greater than the next factors of related variable. Also, it shows the first factor of each variable accounted total variance of each variable as follow: accuracy 72.783%, completeness 64.836, consistency 76.086%, timeliness 58.842% and improvement process 45.299%.

Hence, it can be seen all 5 components of each variable have "total initial eigenvalues" more than 1 and it can be concluded; all items can be matched with their related variable and be grouped on one factor.

Factor loading value is the correlation between a variable and a factor, where only single factor is involved. If variable has higher loading it means that variable is closely at associated with the factor. The items with low loading less than.2 do not have high loading on the same factor (Leech *et al.*, 2007). As it can be seen in Table 4 all components were loaded in the range between 0.547 and 0.932 (the loading was suppressed at 0.4) and factor loading for most components were more than.7 which is considered significant. So, based on these results structure of the construct for the proposed framework was confirmed.

The result of Kaiser-Meyer-Olkin (KMO) and Bartlett's test which applied to measure sampling adequacy are shown in Table 5.

Researchers believe that if KMO be >0.7 there are sufficient items for each factor which is above of acceptable level of 0.5 that indicates the sample size is inadequate. So, based on KMO; enough items were predicted by each factor. Also, the Bartlett's test was significant because it was less than 5 which indicates that all variables are highly correlated enough to provide reasonable basis for factor analysis (Walker, 1999).

Furthermore, a graphical representation of incremental variance for each factor in the model which is called Scree plot showed all the number of factors should be remained (Leech *et al.*, 2007).

Table 5: KMO and Bartlett's test of the variables

| | | Bartlett's test of sphericity | | | |
	KMO	Approx. chi- square	df	Sig.	Determinant
Accuracy	0.751	121.579	6	0.000	0.015
Completeness	0.810	181.643	15	0.000	0.020
Consistency	0.772	140.639	6	0.000	0.050
Timeliness	0.789	855.272	10	0.000	0.160
Improvement process	0.688	43.875	10	0.000	0.389

Several assumptions were tested. All variable determinants were more than 0.0001 so a factor analytic solution could be obtained; (KMO) indicated that adequate items were predicted by each factor. Also, the reasonable basis for factor analysis was test by Bartlett test that showed there is high correlation between variables.

CONCLUSION AND FUTURE IMPLICATION

This study initiates that the ACCTI instrument is reliable and valid for the evaluation of the data quality dimensions and improvement process. The developing four major data quality dimension's framework is important in the field of data quality and for improving quality of data. In this study, a valid and reliable instrument for data quality dimensions was developed. The proposed framework was justified through the empirical methods with the aim of being useful for measuring, assessing and analyzing four major quality dimensions. The construct validity was evaluated based on data that gathered from the questionnaire. Systematically, the content validity, reliability and factor analysis were applied on 24 items. Based on the results, ACCTI instrument factors-Completeness, Accuracy, Timeliness, Consistency and quality improvement-were considered to have construct validity. The values of the Cronbach's alpha, correlation coefficients and composite reliability were indicated, ACCTI framework is reliable and valid. Factor loading showed that all items were clustered on the same suggested variable. Thus, the overall results indicated that our objective was accomplished and these factors could be confidently adopted for use to discover path relationship in the quality dimensions. Dependency discovery seems an appropriate approach to analyze the data quality dimensions. It can be a new way to study data quality based on their dimensions as well as to analyze the path among dimensions to assess and improve the quality of data in information systems. In this pilot study, based on dependency discovery structure, the framework was validated.

This framework can be used effectively in any information system in organizations and industrials. Several researches can be conducted by the direction of this framework. A possible future application can be the enhancing path analysis via Structural Estimating Modeling (SEM) to find detail relationship between quality dimensions with the objective of improving quality improvement process.

In the future, research can be extended to apply discovery dependency on data quality dimensions on the population. In such a way, the aim will be enhancing existing framework and approaches for improvement process in data quality in information system.

ACKNOWLEDGMENT

REFERENCES

Alizamini, F.G., M.M. Pedram, M. Alishahi and K. Badie, 2010. Data quality improvement using fuzzy association rules. Proceedings of the International Conference on Electronics and Information Engineering, Volume 1, August 1-3, 2010, Kyoto, Japan, pp: 468-472.

Barone, D., F. Stella and C. Batini, 2010. Dependency discovery in data quality. Proceedings of the 22nd International Conference on Advanced Information Systems Engineering, June 7-9, 2010, Springer, Hammamet, Tnisia, pp: 53-67.

Borek, A., P. Woodall and M. Oberhofer, 2011. A classification of data quality assessment methods. Proceeding of the 16th International Conference on Information Quality, November 18-20, 2011, Adelaide City West End, Australia, pp: 189-203.

Carey, M.J., S. Ceri, P. Bernstein, U. Dayal, C. Faloutsos, J.C. Freytag and J. Widom, 2006. Data-Centric Systems and Applications. Springer, Berlin Heidelberg, Germany.

Creswell, J.W., 2009. Research Design: Qualitative, Quantitative and Mixed Methods Approaches. 2nd Edn., Sage Publications Inc., London, UK., ISBN-13: 978-1412965576, Pages: 296.

DeAmicis, F., D. Barone and C. Batini, 2006. An analytical framework to analyze dependencies among data quality dimensions. Proceedings of the 11th International Conference on Information Quality, November 10-12, 2006, Cambridge, USA., pp: 369-383.

Fisher, C., E. Lauria, S. Chengalur-Smith and R. Wang, 2006. Introduction to Information Quality. M.I.T. Information Quality Program, USA., ISBN-13: 9780977759903, Pages: 206.

Fynes, B., C. Voss and S. de Burca, 2005. The impact of supply chain relationship quality on quality performance. Int. J. Prod. Econ., 96: 339-354.

Gackowski, Z., Q.U.D. Imensions, A.P.U. Ocused and V.I.E.W. On, 2004. Logical interdependence of some attributes of data/information quality. Proceedings of the 9th International Conference on Information Quality, November 5-7, 2004, Cambridge, USA., pp: 126-140.

Gliem, J.A. and R.R. Gliem, 2003. Calculating, interpreting and reporting cronbach's alpha reliability coefficient for Likert-type scales. Proceedings of the Midwest Research-to-Practice Conference in Adult, Continuing and Community Education, October 8-10, 2003, USA., pp: 82-88.

Han, Q. and N. Venkatasubramanian, 2003. Addressing timeliness/accuracy/cost tradeoffs in information collection for dynamic environments. Proceedings of the 24th IEEE Real-Time Systems Symposium, December 3-5, 2003, Cancun, Mexico, pp: 108-117.

Hatcher, L., 1994. A Step-By-Step Approach to Using the Sas System for Factor Analysis and Structural Equation Modeling. SAS Institute, USA., ISBN: 9781555446437, Pages: 588.

Kahn, B.K., D.M. Strong and R.Y. Wang, 2002. Information quality benchmarks: Product and service performance. Commun. ACM., 45: 184-192.

Katerattanakul, P. and K. Siau, 1999. Measuring information quality of web sites: development of an instrument. Proceedings of the 20th International Conference on Information Systems, December 12-15, 1999, USA., pp: 279-285.

Kim, J.O. and C.W. Mueller, 1978. Factor analysis: Statistical Methods and Practical Issues. 8th Edn., Sage Publications Inc., California, USA., Pages: 88.

Kline, R.B., 2010. Principles and Practice of Structural Equation Modeling. 3rd Edn., The Guilford Press, New York, USA., ISBN: 9781606238769, Pages: 427.

Kock, N. and J. Verville, 2012. Exploring free questionnaire data with anchor variables: An illustration based on a study of it in healthcare. Int. J. Healthcare Inform. Syst. Inform., 7: 46-63.

Lee, Y.W., D.M. Strong, B.K. Kahn and R.Y. Wang, 2002. AIMQ: A methodology for information quality assessment. Inform. Manage., 40: 133-146.

Leech, N.L., K.C. Barrett and G.A. Morgan, 2007. SPSS for Intermediate Statics: Use and Interpretation. 3rd Edn., Routledge Academic, New Jersey, USA., ISBN-13: 978-0805862676, Pages: 58.

Madnick, S.E., R.Y. Wang, Y.W. Lee and H. Zhu, 2009. Overview and framework for data and information quality research. J. Data Inform. Qual., Vol. 1.

Nunally, J.C. and I.H. Bernstein, 1994. Psychometric Theory. 3rd Edn., McGraw-Hill, New York, USA., ISBN-13: 9780070478497, Pages: 752.

O'Donoghue, J., T. O'Kane, J. Gallagher, G. Courtney and A. Aftab *et al.*, 2012. Modified early warning scorecard: The role of data/information quality within the decision making process. Electron. J. Inform. Syst. Eval., 14: 100-109.

O'Leary-Kelly, S.W. and J.R. Vokurka, 1998. The empirical assessment of construct validity. J. Oper. Manage., 16: 387-405.

Sadeghi, A. and R. Clayton, 2002. The quality vs. timeliness tradeoffs in the BLS ES-202 administrative statistics. Federal Committee on Statistical Methodology, pp: 1-7. https://www1.oecd.org/std/financialstatistics/2640138.pdf

Scannapieco, M., P. Missier and C. Batini, 2005. Data quality at a glance. Datenbank-Spektrum, 14: 6-14.

Sidi, F., P.H. Shariat Panahy, L.S. Affendey, M.A. Jabar, H. Ibrahim and A. Mustapha, 2012. Data quality: A survey of data quality dimensions. Proceedings of the IEEE International Conference on Information Retrieval and Knowledge Management, March 13-15, 2012, Kuala Lumpur, Malaysia, pp: 300-304.

Strong, D.M., Y.W. Lee and R.Y. Wang, 1997. Data quality in context. Commun. ACM, 40: 103-110.

Walker, J.T., 1999. Statistics in Criminology and Criminal Justice: Analysis and Interpretation. Aspen Publisher Inc., USA., pp: 237-240.

Wang, J.R. and S.E. Madnick, 1989. The inter-database instance identification problem in integrating autonomous systems. Proceedings of the 5th International Conference on Data Engineering, February 6-10, 1989, Los Angeles, CA., USA., pp: 46-55.

Wang, R.Y. and D.M. Strong, 1996. Beyond accuracy: What data quality means to data consumers. J. Manage. Info. Syst., 12: 5-34.